The role of the RNA 5' end cap in the regulation of gene expression

by

ONOFRIO ZANIN

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

School of Biosciences
College of Life and Environmental Sciences
University of Birmingham
December 2022

Abstract

The 5' RNA capping is an essential process conserved in the entire eukaryotic kingdom. It consists in the co-transcriptional addition of a 7-methylguanosine (m^7G) residue to the 5' end of the products of the RNA polymerase II. The m^7G cap mediates several processes such as the RNA maturation and nuclear export as well as the mRNA translation. The main function of the m^7G cap is thought to be limited to the protection of the RNA from the exonucleolytic degradation and to the interaction with cap binding proteins (CBPs) directly involved in many RNA processing pathways.

The work presented here aims to provide a global description of the role of the m^7G cap by investigating the events affecting the expression of cap defective RNAs in the yeast *Saccharomyces cerevisiae* (*S. cerevisiae*). The introduction of capping defects via the fast depletion of the capping enzymes, revealed complex dynamics related to the readjustments in the levels of non-capped mRNAs which resulted determined by their basal expression levels. The cap deficiency induced the preferential degradation of highly expressed RNAs along with the accumulation of lowly expressed transcripts. The effect was recapitulated upon the disruption of the 5'-3' exonuclease Xrn1 suggesting a correlation between the two events. Finally, contradicting the cap quality control paradigm, the transcription of uncapped pre-mRNAs is not prematurely terminated in capping enzymes depleted cells.

The evidence reported indicate a critical role of the m^7G cap in the Xrn1-mediated sensing and regulation of the cellular mRNA levels expanding the complexity in the function of this RNA modification and introducing a further layer of control in the regulation of gene expression.

Declaration

I hereby declare that this thesis work entitled "The role of the RNA 5' end cap in the regulation of gene expression" is my own original work and that it has not been submitted, in whole or in part, in any previous application for a degree.

Part of the experimental work was completed with the direct help of Dr. Pawel Grzechnik.

Most of the results have been submitted as preprint and deposited on bioRxiv -*Ceg1 depletion reveals mechanisms governing degradation of non-capped RNAs*

Onofrio Zanin, Daniel Hebenstreit, Pawel Grzechnik

doi: https://doi.org/10.1101/2022.09.24.509330

(https://www.biorxiv.org/content/10.1101/2022.09.24.509330v2)

Acknowledgments

List of figures

List of tables

List of schemes

Abbreviations list

4tU	4-thiouracil
6-AU	6-azauracil
AA	Anchor away
Ado-Hcy	S-adenosyl-L-homocysteine
Ado-Met	S-Adenosyl-L-methionine
ATP	Adenosine triphosphate
bp	Base pair
CBC	Cap binding complex
CBP	Cap binding protein
ChIP	Chromatin Immunoprecipitation
co-A	Coenzyme A
CTD	C-terminal domain
CUT	Cryptic unstable transcript
Dc	Decreased
DE	Differentially expressed
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSE	Downstream sequence element
EC	Elongation complex
eCUT	Extended cryptic unstable transcript (CUT)
eRNA	Enhancer RNA
GO	Gene ontology

GTF	General transcription factor
GTP	Guanosine triphosphate
HDV	Hepatitis delta virus
Ic	Increased
IRES	Internal ribosome entry site
kb	kilo base
lncRNA	Long noncoding RNA
m^{6}Am	N6, 2′-O-dimethyladenosine
m^{7}G	7-Methylguanosine
mRNA	Messenger RNA
NAD	Nicotinamide adenine dinucleotide
NFR	Nucleosome-free region
nt	Nucleotide
PAS	Polyadenylation signal
PCA	Principal component analysis
PCR	Polymerase chain reaction
PEP	Plastid-encoded RNA polymerase
PIC	Preinitiation complex
Pol I, II, III	RNA polymerase I, II, III
PPH	Pyrophosphohydrolase
pre-RNA	Precursor RNA
PROMPT	Promoter upstream transcript
PTM	Post translational modification
PTT	Premature transcription termination

qPCR	Quantitative PCR
Rap	Rapamycin
RBP	RNA binding protein
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
rRNA	Ribosomal RNA
RT-qPCR	Reverse transcription qPCR
SD	Synthetic defined
siRNA	Small interfering RNA
sno-lncRNA	snoRNA-associated long non-coding RNA
snoRNA	Small nucleolar RNA
snoRNP	Small nucleolar ribonucleoprotein
snRNA	Small nuclear RNA
SUT	Stable unannotated transcript
TES	Transcription end site
TMG	Trimethyl guanosine
TPM	Transcript per million
TR	Transcription rate
tRNA	Transfer RNA
USE	Upstream sequence element
UTR	Untranslated region
WT	Wild type
YPD	Yeast extract, peptone, dextrose

Table of Contents

1. Introduction

1.1 Eukaryotic gene expression

The genetic information encoded in the genome of living organisms is in the shape of DNA sequences that must be converted into molecules capable of solving the many cellular functions to sustain life. Gene expression involves the set of processes active in the conversion of genetic information into functional products such as RNA and proteins (Fig. 1.1). As a complex, multi-step process, gene expression involves a wide range of RNAs and proteins to occur. Importantly, gene expression is regulated at each of its steps either temporarily or spatially. This permits a dynamic adaptation aimed to ensure survival and increases the complexity of the system. The regulation of gene expression allows for sustaining cellular homeostasis, controlling the response to external stimuli and regulates differentiation and development.

Two main synthetic processes, in combination with many degradation systems, are responsible for maintaining the physiological concentration of RNAs and proteins in the cell. Transcription is the process which involves the synthesis of the RNA molecule from a DNA template. Many functional RNA classes are known nowadays (Brosius & Raabe, 2016; Eddy, 2001). They can arise from annotated transcriptional units or from what is considered intergenic regions. These latter have been named non-coding RNAs (ncRNAs) as they are not converted into proteins. A long list of RNAs belong to this group and their functional relevance is coming to light with the development of the technology. Among these, starting from the better characterised, there are transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) but also small nuclear and small nucleolar RNAs (respectively snRNAs and snoRNAs). Many are transient and less characterised; a few examples are cryptic unstable transcripts (CUTs) and stable unannotated transcripts (SUTs) whose functions are yet to be clarified. Many more ncRNAs are involved in essential

cellular processes, they are not listed here but have been described and deeply investigated (Cech & Steitz, 2014; Fu, 2014).

On the other end, the class of the messenger RNAs (mRNAs) do not exert any direct function but they are employed as a transitory and mobile piece of information needed for protein synthesis. Although mRNA class manages the production of the entire proteome, it constitutes a minority of the global cellular RNA, about 3-5% (Warner, 1999).

The second synthetic step required to guarantee gene expression is translation of the mRNA molecule into protein. In this process, interestingly, the conversion from the 4 nucleotides code into the 21 aminoacidic code of the proteins occurs giving name to translation. As for any other process involved in gene expression, the translation can be regulated at different levels to modulate the concentration of either RNAs or proteins (Gebauer & Hentze, 2004).
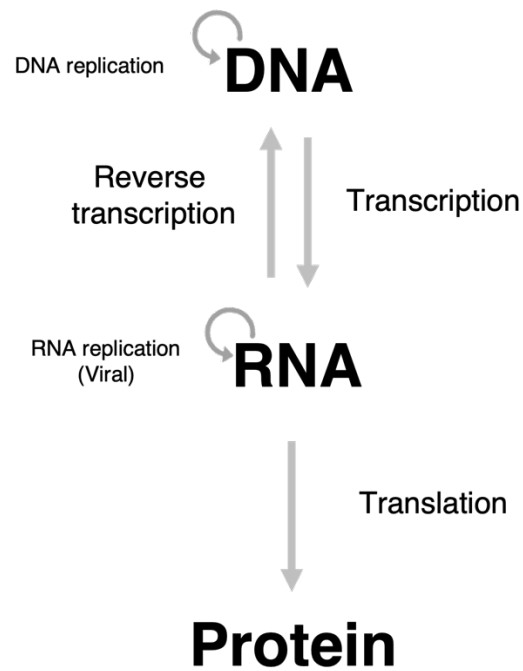
Figure 1.1 - The progression of the genetic information.

The genetic information encoded into the DNA (an auto replicative molecule) is converted in functional molecules of RNA with a process indicated as transcription. The RNA can serve as intermediate from DNA to the protein synthesis. The synthesis of a RNA into polypeptide is called translation.

1.1.1 Transcription

Transcription is the first step of gene expression and aims to the synthesis of an RNA molecule on the base of the template DNA. The synthetic reaction is universally conserved and is accomplished by an essential group of enzymes named DNA-dependent RNA polymerases (Sonntag & Darai, 1995). Eukaryotes possess three nuclear RNA polymerases (Pol) named respectively Pol I, II and III and the mitochondrial Pol (mtPol in humans and Rpo41 in yeast) (Ramachandran et al., 2016) or the plastid-encoded RNA polymerase (PEP) in the chloroplasts (Börner et al., 2015). Plants have more nuclear Pol (Pol IV and V) specialised in the synthesis of small interfering RNAs (siRNAs) and epigenetic modifications regulating gene expression (S. K. Archer, Shirokikh, & Preiss, 2014; J. Luo & Hall, 2007; Wierzbicki, Haag, & Pikaard, 2008; Zhou & Law, 2015).

The three main nuclear RNA polymerases (Pol I, II and III) show a high homology as they are multi-subunits enzymes sharing the same 10 core elements (5 of which shared with the prokaryotic and archaeal Pol). Pol I and II introduce some peripheral subunits that brings the count of the elements of the holoenzymes to 14 and 17 subunits respectively. Pol II is instead composed of 12 subunits (Khatter, Vorländer, & Müller, 2017). The structural differences lead to the fact that each eukaryotic RNA polymerase catalyses the synthesis of a different subset of genes. Pol I is indeed responsible for the synthesis of rRNAs. Pol III synthesise tRNAs and the small rRNA 5S. Pol II instead, transcribes protein-coding genes (mRNAs), snRNAs and snoRNA, and other ncRNAs (Carter & Drouin, 2009).

1.1.2 RNA Polymerase II (Pol II)-mediated transcription

The transcriptional cycle includes all the events that bring the Pol II from the start to the completion of the synthesis of an RNA molecule, from the binding to the DNA template to the release of the RNA synthesised. It is accomplished in three main stages: initiation, elongation and termination.

Transcription initiation starts with the binding of the general transcriptional factors (GTFs) such as TFIID, B, F, E and H (also indicated as initiation factors) to the DNA template that in turn permit the binding of the Pol to the specific genetic locus with the formation of the DNA-Pol-GTFs or pre-initiation complex (PIC). In the presence of adenosine triphosphate (ATP), the PIC induces the melting of the two strands of the DNA bound and the formation of the transcriptional bubble. One of the two DNA strands will be incorporated into the Pol II holoenzyme and will serve as template for the RNA synthesis (Sainsbury, Bernecky, & Cramer, 2015). Once the PIC is assembled, an intermediate phase called promoter clearance anticipate the release of the PIC into productive elongation. During the promoter clearance, Pol II performs several rounds of so-called abortive initiation/elongation synthesising and releasing short RNAs (McClure, 1985) remaining in the promoter region with a PIC partially disassembled.

The dissociation of initiation factors and the binding of elongation factors will convert the PIC into the elongation complex (EC). At this stage the Pol II C-terminal domain (CTD) is mainly hypophosphorylated and the kinase activity of the Kin28 (CDK7 in humans) subunit of TFIIH induces the phosphorylation of the CTD serine 5 residues. The release of Pol II from the initiation stage, is dictated by the phosphorylation of the serine 2 CTD residues mediated by positive transcriptional factors like Bur1-Bur2 and Ctk1 in yeast and P-TEFb in humans.

The EC is committed to transcription elongation which is the synthesis of the RNA molecule adding one nucleotide at a time to the 3' end of the upstream nucleotide introduced. During the elongation phase the concertised action of protein kinases and protein phosphatases induce the change in the CTD phosphorylation status accumulating phosphorylated serine 2 and dephosphorylating serine 5 residues (Sims, Belotserkovskaya, & Reinberg, 2004).

Finally, at the end of the transcriptional unit, the nascent RNA is released, the Pol II is displaced from the DNA template and recycled to start a new round of transcription. Different pathways have been described for the Pol II transcription termination of coding or non-coding RNAs. The most common termination

pathway of the protein-coding genes can be indicated as poly-A signal (PAS)-dependent pathway (Rodríguez-Molina, West, & Passmore, 2023). In yeast, the Nab3-Nrd1-Sen1 (NNS)-dependent termination mechanism has been observed for the termination of a subgroup of non-coding Pol II transcripts such as small nuclear and small nucleolar RNAs, respectively snRNAs and snoRNAs (Kuehner, Pearson, & Moore, 2011; Rodríguez-Molina, West, & Passmore, 2023). The transcription termination of non-coding RNAs differs in humans and depends on the multi-subunit Integrator complex (Baillat et al., 2005; Fianu et al., 2021).

The PAS-dependent pathway of transcription termination implies the recognition of cis-acting elements on the nascent RNA. The first described is the conserved consensus AAUAAA in mammals (PAS), more degenerated in yeast (Kyburz et al., 2003), essential to guide the transcription termination (Birse et al., 1998). Other signals are required in the nascent RNA to induce an efficient transcription termination, the downstream sequence element (DSE) U/GU-rich element (Gil & Proudfoot, 1984, 1987), which appears downstream of the PAS and the upstream sequence element (USE), upstream of the PAS (Moreira, Wollerton, Monks, & Proudfoot, 1995).

A large multi-complex protein structure recruited on the Pol II CTD, binds the nascent RNA on the PAS and co-transcriptionally cleaves the pre-mRNA. The protein complex cleavage and polyadenylation specificity factor (CPSF called cleavage and polyadenylation factor or CPF in yeast) binds the PAS and interacts with the cleavage factor I and II (CFI and CFII, typical of mammalian cells) and the cleavage stimulation factor (CstF called CFI-A in yeast) which interacts directly with the DSE (de Vries et al., 2000; Takagaki & Manley, 1997; Zhao, Hyman, & Moore, 1999). The RNA cleaved is polyadenylated at the 3' end by the poly-A polymerase enzyme PAP (Christofori & Keller, 1989; Ryner, Takagaki, & Manley, 1989).

After the cleavage and polyadenylation of the pre-mRNA, the Pol II progresses transcribing downstream of the transcriptional unit generating and exposing a 5' monophosphate end nascent RNA which is co-

transcriptionally degraded by the Rat1 (Xrn2 in mammals) exonucleolytic activity (M. Kim et al., 2004; West, Gromak, & Proudfoot, 2004). The Rat1/Xrn2 co-transcriptional degradation, combined with Pol II allosterically induced break downstream of the transcriptional unit, leads to the collision responsible for the displacement of the paused Pol II from the DNA and the complete degradation of the downstream nascent RNA according to the unified "*sitting-duck torpedo*" model (Cortazar et al., 2019; Eaton et al., 2020).

The alternative transcription termination pathway was described in the yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) for the processing of ncRNAs such as snRNAs, snoRNAs and CUTs and is mainly mediated by the nuclear protein Sen1 (Senataxin in human). Alike the PAS-dependent termination, the exposure of specific signals induces the recruitment and consequent activity of termination factors. In the Sen1-dependent termination the set of proteins that senses and activate the termination involves the proteins Nrd1, Nab3 and Sen1 (also known as NNS complex). Nrd1 and Nab3 are RNA binding proteins which bind respectively the GUAA and UCUU motif on the nascent RNA. Sen1 exerts a helicase activity able to affect the RNA:DNA heteroduplex thought to be the main force conferring the high stability of the EC. Sen1 is therefore thought to be responsible for the disruption of the EC (Han, Libri, & Porrua, 2017) at the end of ncRNA genes by inducing the separation of the 8 nucleotides long RNA:DNA pair hosted in the active site of transcribing Pol II (Arndt & Reines, 2015).

The dynamics of the Integrator-mediated termination of human ncRNAs are yet to be fully clarified but the structure of the complex has shown an endonuclease subunit responsible for the cleavage of different RNAs positioned next to the Pol II exit channel (Fianu et al., 2021).

The general structure of the transcriptional cycle is common for the different RNA polymerases, nevertheless the events occurring at each stage are characterised by some differences related to either the cis-regulatory elements of the DNA template, the TFs involved and the structural features of the transcribing Pol. The nuclear RNA polymerases recognise distinct promoter classes, Pol I and III

promoters have a restricted variety of elements, Pol II, instead, can start transcription from widely structured promoters. Another characteristic differentiating Pol II from Pol I and II is that Pol II has a more modular structure where a multiplicity of TFs can interact according to the transcriptional levels needed and the class of gene transcribed (Barba-Aliaga, Alepuz, & Pérez-Ortín, 2021). Pol III transcription is terminated with a mechanism that resemble the prokaryotic termination more than the termination mechanisms observed for the Pol I and II (Arimbasseri, Rijal, & Maraia, 2013).

Pol II exhibits structural peculiarities that distinguish it from the other nuclear RNA polymerases. The CTD of the Pol II-subunit Rpb1 is an unstructured repetition of the heptapeptide YSPTSPS. The number of the repetitions found in the Pol II CTD increases following the evolution starting from 26 in *S. cerevisiae* to 52 in humans. The CTD undergoes a wide and dynamic combination of post-translational modifications (PTMs) each of which is a precise hallmark of any step of the transcriptional cycle such as initiation, elongation, and termination (Davidson, Muniz, & West, 2014; Shah et al., 2018; Suh et al., 2016). Also, specific PTMs are now confirmed to coordinate co-transcriptional RNA modification events expanding the complexity and the function of the so called "*CTD code*" that is the term used to indicate the fine modulation of the CTD-PTMs adjusted to accomplish different functions (Bharati et al., 2016; Corden, 2016; Egloff & Murphy, 2008; Noe Gonzalez, Blears, & Svejstrup, 2021; Nojima et al., 2018; Shah et al., 2018). The CTD code acts modulating the Pol II interactome via the remodelling of PTMs combinations displayed on the CTD creating therefore the optimal microenvironment for any specific pathway related to the transcription (Harlen & Churchman, 2017; Harlen et al., 2016; Jeronimo, Collin, & Robert, 2016).

Most of the RNAs are generated as precursors (preRNAs) undergoing several maturation steps that subject the preRNA to some downstream modifications such as the addition of residues, trimming and cleavage (Chernyakov et al., 2008; Geerlings, Vos, & Raué, 2000; Kufel & Grzechnik, 2019; Proudfoot, Furger, & Dye, 2002). Pol II transcripts undergo three main co-transcriptional modification (Fig. 1.2): i) The addition of a 7-methylguanosine (m^7G) cap at the 5' end (capping), ii) for the pre-mRNAs there is the removal of

the introns and the joining of the exons (splicing) and iii) the cleavage and polyadenylation (mainly for protein-coding genes) of the 3'end. Some Pol II transcripts can be subjected to different 3' end modifications like the poly-uridylation of snRNA U6 (Munoz-Tello et al., 2015) or the TRAMP-mediated oligo-adenylation (Schmid & Jensen, 2018).
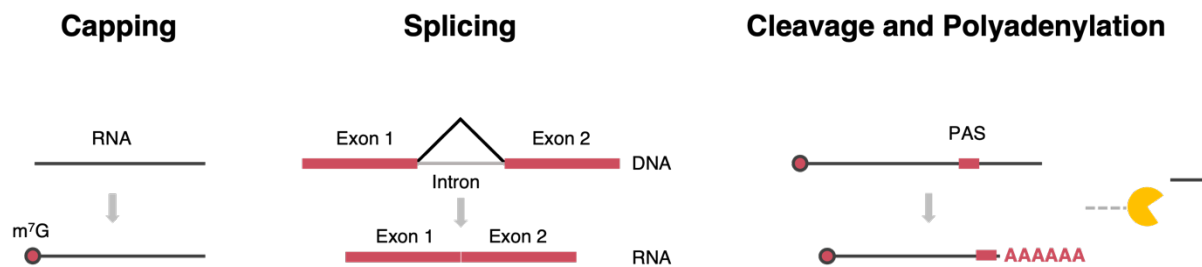
Figure 1.2 - Transcripts synthesised by the RNA polymerase II (Pol II) are subjected to three main co-transcriptional modification.

The first modification to appear on the nascent pre-mRNA is the addition of the 5'end m^{7}G cap. The eukaryotic genes are characterised by the succession of coding regions (Exons) interrupted by noncoding stretches of DNA or Introns. The exons are fused together and the introns are removed from the nascent RNA with a process named splicing aimed at the synthesis of a functional protein. To limit the progression of the Pol II into the downstream gene and to complete the maturation of the nascent transcript, at the end of the transcriptional unit, the transcription is terminated. The poly-A signal (PAS) induces, together with a protein multi-subunit complex, the cleavage of the nascent RNA and the 3' end polyadenylation of the pre-mRNA. The Pol II is displaced from the DNA and is recycled for a new transcriptional cycle.

1.1.3 5' RNA capping

The RNA capping is the first co-transcriptional modification involving all the products of the Pol II-dependent transcription. RNA capping consists of the binding of a 7-methylguanosine residue (m^7G) to the free 5' end of the nascent transcript. The steps of cap synthesis and its mechanism are evolutionary conserved from yeast to human (Ramanathan, Robb, & Chan, 2016; Yue et al., 1997). Three enzymatic reactions are required to ensure functional capping (Fig. 1.3A). Initially, an RNA thriphosphatase (TPase) removes the gamma (γ) phosphate from the first nucleotide of the nascent RNA generating a diphosphate 5'-end RNA with the release of an inorganic phosphate group. Next, the guanylyltransferase (GTase) reacts with a guanosine-triphosphate (GTP) to form a stable intermediate enzyme-guanosine-monophosphate (GTase-GMP). The GTase-GMP transfers the GMP moiety to the first ribonucleotide of the diphosphate 5'-end RNA generating a peculiar 5'-5' phosphodiester bond with the production of a guanosine 5'-5' capped RNA (GpppRNA) and the release of the free GTase. The methyltransferase (MTase) completes the synthesis of the minimal cap structure (Fig. 1.3B) with the transfer of the methyl residue from the S-adenosyl- L- methionine (AdoMet) to the N7 (the nitrogen in position 7 of the purine ring) of the guanosine to obtain the m^7G capped RNA (m7GpppRNA) releasing S-adenosyl-L-homocysteine (AdoHcy).

A
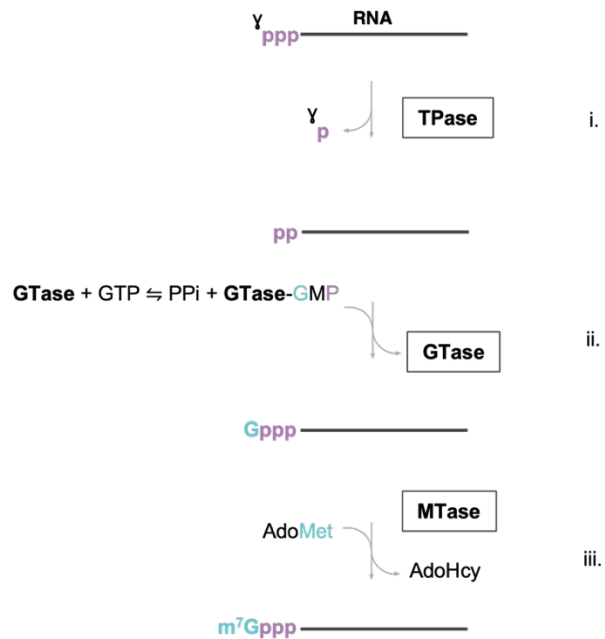
B
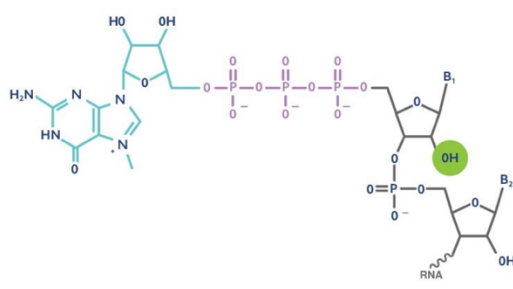
Figure 1.3 - The m⁷G cap synthesis and structure.

A) The schematic of the m⁷G cap synthesis. The RNA triphosphatase (TPase) removes the gamma (γ) phosphate group from the first ribonucleotide of the nascent RNA releasing an inorganic phosphate, p (i). The guanylyl transferase (GTase) reacts with a GTP molecule and forms a stable GTase-GMP complex releasing a pyrophosphate group, PPi. The GTase-GMP complex, transfer the GMP moiety to the 5' diphosphate RNA generating a guanosine capped RNA and releasing a free GTase enzyme (ii). Finally, the methyltransferase (MTase) transfers a methyl group from S-adenosyl-L methionine (AdoMet) to the guanosine cap with the release of S-adenosyl -L-homocysteine (AdoHcy) and the m⁷G capped RNA (iii). B) The molecular structure of the m⁷G cap. The 7-methylguanosine (green) is attached by a 5'-5' phosphodiester bond (purple) the first nucleotide of the nascent RNA (dark grey). B_n is any base found in the RNA. Image adapted from genengnews.com.

In *S. cerevisiae*, the enzymes responsible for pre-mRNA capping are encoded by three different genes, respectively *CET1* for the RNA triphosphatase, *CEG1* for the guanylyltransferase and *ABD1* for the methyltransferase.

Although the enzymatic activities needed for the synthesis of the m^7G cap remain conserved among the eukaryotes, some differences lie in the structural organization and protein-protein interactions of the capping enzymes between species. For instance, in metazoans the same polypeptide (mRNA-cap in *Drosophila melanogaster*; RNGTT in mammals) holds both the RNA triphosphatase and the guanylyltransferase activity. Else, the *S. cerevisiae* capping enzymes Cet1 and Ceg1 work in a strict association in a '*homodimeric quaternary structure*' as separate polypeptides (Hausmann, Pei, & Shuman, 2003). Such interaction results necessary for the binding of Ceg1 to Pol II and for the consequent capping activity. Contrarily, in *Schizosaccharomyces pombe*, the RNA triphosphatase Pct1 does not require the interaction with the guanylyltransferase Pce1 to be functional (Machineries et al., 2002).

1.1.3.1 Further cap modifications

The m^7G residue linked to the 5' end of the RNA forms what is called cap 0 structure, the minimal cap structure required to interact with the cap binding proteins (CBPs). Further modifications are introduced in the first nucleotide introduced by the transcribing Pol II (nucleotide +1). Specific methyltransferases (2'O methyltransferase, 2'O MTase), the hMTr1 and hMTr2 in human cells, catalyse the methylation of the 2' O in the ribose of the nucleotide +1 generating the so-called cap 1 structure (Bélanger et al., 2010) and nucleotide +2 with the formation of the cap 2 structure (Werner et al., 2011). A direct role for the cap 1 structure has been demonstrated in the identification of self RNA blocking the type I interferon (IFN) response to virus infections pathway. Nevertheless, some viruses have developed a capping machinery capable of generating a cap1 structure and evade the cellular surveillance (Ramanathan et al., 2016).

Another modification related to the cap structure and the adjacent nucleotide, concerns the methylation of the N6 and O 2' position of the nucleotide +1 when it is an adenosine (m^6Am). The modification, described at the very early stages of research on cap (Shatkin, 1976), increases RNA stability inhibiting DCP2-mediated decapping (Mauer et al., 2017).

Some sn- and snoRNAs display a 2,2,7-trimethylated guanosine (TMG) cap, sometimes indicated as m3G. The hypermethylation is mediated by the trimethylguanosine synthase 1 (Tgs1) in the yeast nucleus (Cougot et al., 2004; Mouaikel et al., 2002). In mammalian cells, TGS1 catalyses the synthesis of TMG cap in the cytoplasm which allows the return in the nucleus for the snRNAs whereas snoRNAs remain localised and modified in the nucleus (Warminski et al., 2017) (Fig. 1.4).
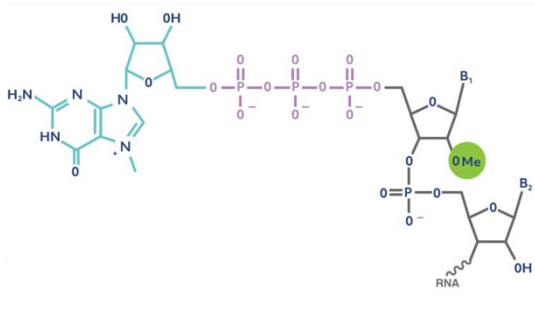
Cap 1   Cap 2
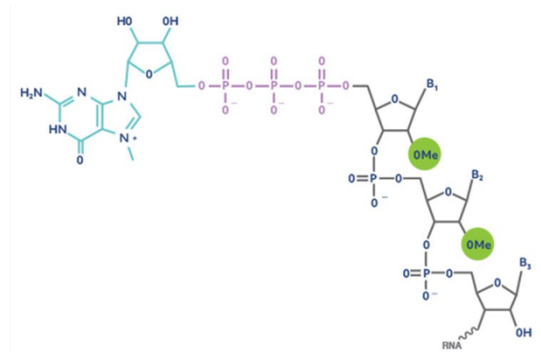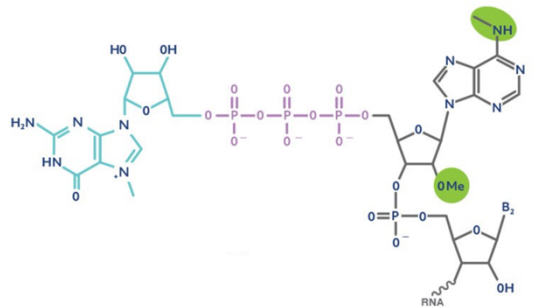
m⁶Am Cap   TMG
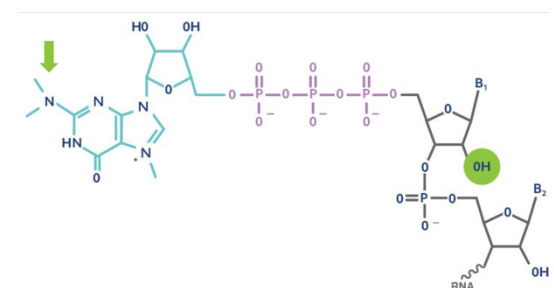
Figure 1.4 - Various cap modification and cap structures.

The molecular structure of cap 1 (a methyl group is attached to the ribose 2' O position of the nucleotide +1 of the RNA chain, green circle), cap2 (a methyl group is attached to the ribose 2' O position of the nucleotide +2 of the RNA chain, green circle), m⁶Am (when the nucleotide +1 is an adenosine, it can be methylated at the position 6 of the purine ring and at the 2' O of the ribose, green circles) and 2,2,7-trimethylated guanosine (TMG) cap (the 7-methilgianosine can be further methylated at the N in position 2 of the purine ring, green arrow). Image adapted from genengnews.com.

1.1.3.2 Interplay between the m^{7}G cap and Pol II

The RNA cap, the capping enzymes and the early phases of Pol II transcription are coupled by a complex feedback loop that leads to the assembly of the transcriptional machinery and ensures an efficient and regulated gene expression. First, capping enzymes require specific signals to achieve their interaction with the PIC and exert their function. Signals embedded in the DNA template can affect the RNA processing (Proudfoot et al., 2002), but a pivotal role is exerted by the Pol II subunit Rpb1 and specifically its CTD. Phosphorylation of the CTD serine 5 and 7, alongside the action of specific elongation factors such as Spt5 are essential for the efficient binding of capping enzymes to the Pol II at the promoter level (Bharati et al., 2016; Mandal et al., 2004; Noe Gonzalez et al., 2018; Schneider et al., 2010).

In *S. cerevisiae*, the capping enzymes contact the PIC as a Cet1-Ceg1 complex. The interaction involves the serine 5 phosphorylated CTD and the body of Pol II. It is accomplished at the very early phases of transcription when the nascent RNA is about 17 nt long (Martinez-Rucobo et al., 2015), they have been also shown to recruit certain transcriptional factors to the PIC (Sen et al., 2017). The cap is joined as soon as the nascent RNA is about 20-30 nucleotides (nt) long (Rasmussen & Lis, 1993; Aaron J. Shatkin & Manley, 2000). The engagement of Pol II into the elongation phase induces the increase in the CTD serine 2 phosphorylation, this permits to finalise the m^{7}G cap synthesis inducing the release of the capping enzymes from the PIC. About 110 nt downstream of the transcription start site (TSS) indeed, the Cet1-Ceg1 complex is released and Abd1 binds the EC to methylate the cap (Lidschreiber, Leike, & Cramer, 2013; Schroeder et al., 2000).

The cap structure and the capping enzymes keep their reciprocal interaction once the Pol II has escaped the promoter and is actively transcribing. In *S. cerevisiae*, the methyltransferase Abd1, binds the transcriptional machinery after the Cet1-Ceg1 complex but still in the very initial phases of transcription and maintains the interaction with Pol II until the 3' end of the gene (Lidschreiber, Leike, & Cramer, 2013).

The m[7]G cap is a prerogative of the Pol II products and marks the RNA for specific pathways of maturation and processing, it can be either directly or indirectly correlated to several aspects of the RNA metabolism. Even if capping is a peculiar modification concerning only the less representative species of the cellular RNA (about 5% of total), it results essential for cell viability, all the modifications impairing proper RNA capping tend to affect cell survival (Schwer & Shuman, 1994; Shibagaki et al., 1992).

1.1.3.3 The m[7]G cap functions

The most direct function the cap structure exerts is the shielding of the RNA molecule against the 5'-3' exonucleolytic degradation. Cap deficiency or cap defects have been correlated to a reduced mRNA half-life and loss of its functionality alongside an increased RNA degradation rate (Schwer, Mao, & Shuman, 1998). Uncapped RNA molecules have been shown to be degraded in the cytoplasm in structures called P-bodies (Liu & Kiledjian, 2006). Most of the research describing the m[7]G cap role on the RNA stability has been conducted via single gene analysis and often *in vitro* (Chang et al., 2012; Jiao et al., 2013; Schwer et al., 1998; Xiang et al., 2009) thus limited to a strict number of model genes. The genome wide investigation of the changes on gene expression in cells defective of the m[7]G cap synthesis is missing at the moment that this thesis is being compiled, therefore further aspects concerning this RNA modification might be missing and worth of a deeper investigation.

The enzymatic cap removal (decapping) is one of the starting points of RNA degradation involved in RNA turn-over and regulation of gene expression (Mugridge, Coller, & Gross, 2018). Intriguingly, decapping can positively induce the maturation of certain RNA species; the Rnt1 decapping activity has been shown to mediate the maturation of the box C/D snoRNAs by affecting their 3' end processing, cellular localization and overall, their function in the ribosomal RNA (rRNA) metabolism (Grzechnik et al., 2018).

This shed further light on the diverse functions of RNA cap and the many pathways related to it, not yet fully dissected but capable of affecting the global RNA metabolism and cellular homeostasis.

1.1.3.4 Cap analogues in eukaryotic cells and capping in bacteria

The RNA capping has been observed universally in the eukaryotic kingdom but recently, the existence of capping-like structures has been described in prokaryotes. First, using mass spectrometry methods 5' nicotinamide adenine dinucleotide (NAD)-linked short RNAs were detected in bacteria (Chen et al., 2009) in parallel with the description of small RNA molecules carrying the coenzyme A (co-A)-derivate linked at the 5' end (Kowtoniuk et al., 2009).

Genome wide analysis have identified some regulatory RNAs to be NAD capped besides a role for the NAD cap in bacteria was observed *in vitro* (Fig. 1.5). Similar to the role of the eukaryotic cap, the NAD residue attached at the 5' end of the prokaryotic RNAs increased the RNA stability inhibiting the endonucleolytic and exonucleolytic degradation which in turn can be induced after NAD decapping (Cahová et al., 2015). Both, NAD+ and the co-A derivates (as 3' dephospho co-A) are inserted into the nascent RNA as noncanonical first nucleotides incorporated by the bacterial RNA polymerase (J. G. Bird et al., 2016). Further research on the subject, has shown the incorporation of nucleoside tetraphosphate (Np4) as capping-like structure under certain conditions where the concentration of the Np4 precursors was increased (Luciano, Levenson-Palmer, & Belasco, 2019).

The cap-like structures have been observed on a fraction of bacterial RNAs, the incorporation of the noncanonical nucleotides has been shown to be promoter dependent upon certain stress conditions and affect the capped RNA stability. Intriguingly, the bacterial RNA polymerase seems to be the only enzyme responsible for the incorporation of the noncanonical nucleotide during transcription and therefore the main component of the bacteria capping complex (Jagodnik & Gourse, 2020; Luciano & Belasco, 2020).

A fraction of 5' NAD-RNAs have been purified from the *S. cerevisiae* transcriptome as nuclear and mitochondrial mRNAs which seems to incorporate the modification co-transcriptionally. It was also hypothesised the existence of a competition between ATP and NAD+ (at the adenosine transcription initiation sites) to be incorporated as first nucleotide by the Pol II (as well as the mtPol/Rpo41) regulated by the promoter sequence or additional TFs.
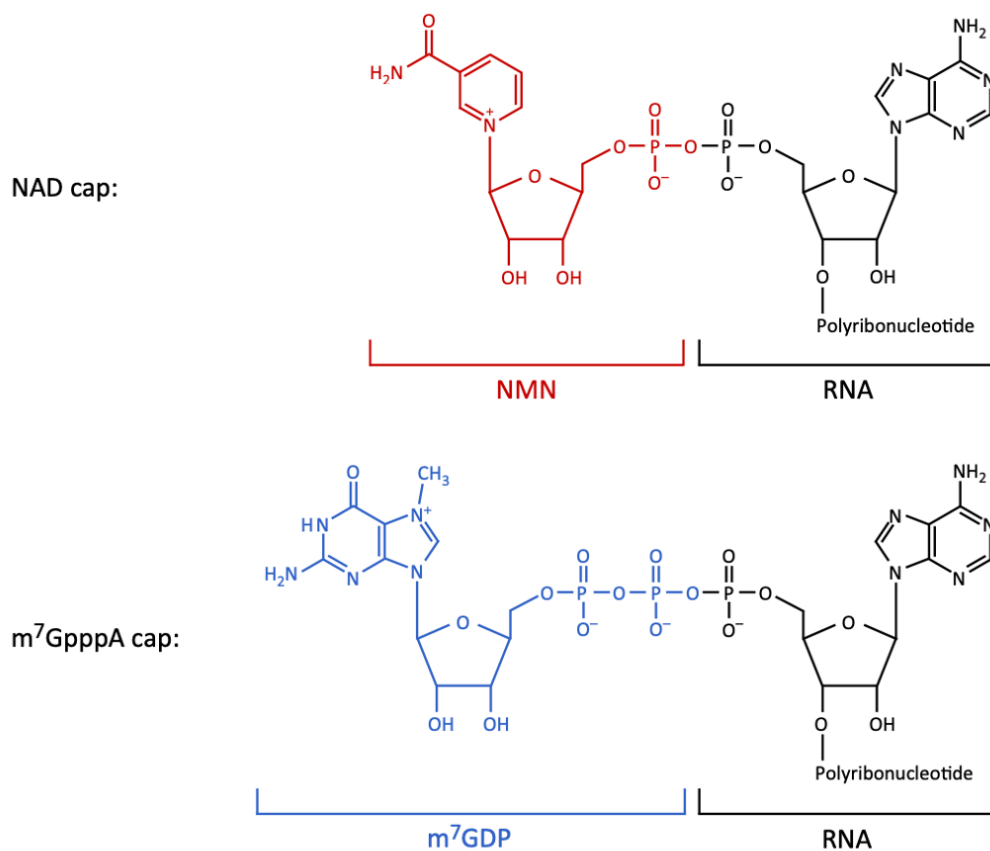
Figure 1.5 - NAD capped RNAs are found in eukaryotes but also in bacterial cells.

The molecular structures of the nicotinamide adenine dinucleotide (NAD) capped RNA. The nicotinamide mononucleotide portion (NMN) is in red (top) compared to the m7G cap. The 7-methyl guanosine biphosphate (m7GDP) is in blue (bottom). Adapted from (Luciano & Belasco, 2015).

Pol II can incorporate other alternative adenosine derivates (e.g., FAD+, co-A, NADH) which can represent alternative cap structures. Nevertheless the possible biological function for these 5' modifications is yet to be clarified (Walters et al., 2017).

5' NAD-RNAs have been observed also in mammalian cells where the NAD+ addition directly participates to the regulation of gene expression. It commits the RNA for degradation via enzymatic NAD removal or 'deNADding' mediated by DOX family protein otherwise responsible for the removal of aberrant m^7G cap (Jiao et al., 2017, 2010).

The observation of intronic ncRNA harbouring a 5' NAD cap suggested that the introduction of such modification (NADding) can be either co- or post transcriptional implying, therefore, more than one mechanism for its synthesis. Moreover, the addition of NAD cap is not equally distributed over all the adenosine transcription start sites indicating the controlled addition of the modification rather than the stochastic competition with ATP (Jiao et al., 2017). An intriguing suggestion for a possible biological meaning of the accelerated decay of NAD-RNAs in human cells has been proposed by Megerditch Kiledjian. The author refers to the resistance at the exonucleolytic degradation of the snoRNAs and snoRNA-associated long non-coding RNAs (sno-lncRNAs) the latter composed of a snoRNAs at each end (5' and 3') of a central lncRNA. As NAD-snoRNAs have been identified, this modification could promote the recruitment of DXO and induce the degradation of such stable RNA species. Another possible physiological significance of the NAD cap-mediated degradation would reside in the clearance of intermediates of degradation from the canonical m^7G cap decapping and exonucleolytic decay of the mRNAs. The author suggests that the incompletely degraded mRNAs could be NADded in the nucleus to recruit DXO and complete their degradation exploiting the exonucleolytic DXO activity to overcome the 5'-end decay block (Kiledjian, 2018).

The m^7G cap has been shown to be essential for the expression of exogenous RNAs originating for simpler genomes during viral infection. Cells have developed several mechanisms to distinguish the non-self,

invading RNA form the self, cellular RNA and in parallel, viruses have evolved mechanism and strategies to elude such control (Decroly et al., 2012). Mammalian viruses can protect the RNA 5' end by the covalent bond of a viral protein genome-linked (VPg)-like protein, usually used to promote viral RNA synthesis but that can be sufficient to bind the cellular synthetic apparatus and promote the viral RNA translation (Goodfellow et al., 2005). Another strategy adopted by several viruses is to generate a m^7G RNA cap either using the cellular or viral capping machinery, some viruses can use some short cellular capped RNAs and promote the viral RNA synthesis and processing with a mechanism called "cap snatching" (Decroly et al., 2012; Reich et al., 2014).

1.1.4 The cap binding complex

The m^7G cap establishes either nuclear or cytoplasmic interactions with the so-called cap binding proteins (CBPs) which mediate many of the cap functions.

In the nucleus, the m^7G cap is co-transcriptionally bound to the nuclear cap binding complex (CBC). The CBC is a protein complex consisting of two polypeptides, Cbp20 and Cbp80. The single subunits are unable to bind the m^7G cap, Cbp80 is required to stabilise Cbp20. The interaction Cbp80-Cbp20 induce conformational changes in Cbp20 increasing the specificity of the CBC for the cap. (Izaurralde et al., 1994). The direct interaction with the m^7G cap is mediated by the Cbp20 subunit but both subunits can bind the RNA (Castello et al., 2012; Mazza et al., 2001; Pabis et al., 2010).

CBC interacts with the cap and the transcriptional machinery since the very early phases of RNA synthesis and has been shown to participate to the activation or repression of the transcription (Lahudkar et al., 2011). Moreover, CBC mediates some of the RNA processing downstream events by targeting interactors involved in RNA splicing, transcriptional termination and RNA nuclear export.

Importantly the CBC contributes to the RNA stability. Decapping is an important step leading to the RNA degradation, the competition for the m^7G cap between the CBC and the decapping factors has been shown to hamper or delay the commitment of the RNA into degradation pathways (Gonatopoulos-Pournatzis & Cowling, 2014).

The nuclear CBC exerts important functions related to the gene expression, nevertheless it displays specie-specific functions. For instance, in *S. cerevisiae*, the CBC is only marginally involved in the RNA nuclear export (Izaurralde & Adam, 1998) and, more generally, unlike the m^7G cap, the CBC is not essential for the yeast cell survival (Fortes et al., 1999).

The nuclear CBC remains associated to the capped RNA during the nuclear export and maintains its interaction until the pioneer round of translation in the case of cap-dependent translated mRNAs. Subsequently, the Cbp20-Cbp80 is removed and the cap structure docks the eukaryotic translation initiation factor E (eIF4E) that, together with other translation initiation factors, drives the next round of translation (Isken & Maquat, 2008).

1.1.5 The effects of the RNA capping on transcription

Capping enzymes have been shown to affect the progression of transcription. In these terms, there are evidences of capping enzymes regulating the release of the Pol II from the promoter region leading it into productive elongation often, independently of their capping activity (Fujiwara et al., 2019; Guiguen et al., 2007; Kim et al., 2004; Schroeder et al., 2004; Sen et al., 2017).

Specifically, Ceg1 was reported to facilitate the transition of the Pol II into the productive elongation stage. The temperature-sensitive mutant *ceg1-63* displayed a transcription elongation defective phenotype at the permissive temperature expressed in combination with the failure of the Pol II to overcome synthetic pausing sites at the promoter level. These observations provide a basis to consider the RNA capping as a

checkpoint for the progression of the transcription from the initiation to the elongation step. The initial transcriptional arrest exerted by negative transcriptional factors like Spt4/5 complex (DSIF) or NELF (not found in yeast cells), could let time for the capping enzymes to bind the transcriptional machinery, and cap the nascent RNA. This in turn, recruits positive transcription factors like Ctk1 and Bur1-Bur2 (P-TEFb) with consequent release of the Pol II into the productive elongation. A mechanism developed to minimise the synthesis of cap defective RNAs (H. Kim et al., 2004; Lenasi, Peterlin, & Barboric, 2011).

Cet1, the RNA triphosphatase responsible for the m^7G cap synthesis in *S. cerevisiae,* has been shown to exert a direct influence on transcription. Cet1 promotes the release of Pol II from the promoter region independently of its capping activity (Lahudkar et al., 2014) and its N-terminal domain is responsible for the recruitment of the positive transcriptional factor FACT (Sen et al., 2017).

More observations link the RNA capping with the transcriptional elongation demonstrating the role of the capping enzyme Abd1 and the CBC in the recruitment of positive elongation factors (Bur1 and Ctk1) inducing the phosphorylation of the CTD serine 2, a condition required to promote the binding of transcriptional factor and permit the transition into Pol II elongation (Lidschreiber et al., 2013). The role of Abd1 in the promoter release of Pol II has been shown to be independent of the cap-methyltransferase activity besides, Abd1 stabilises the binding of Pol II to the promoter in a gene-specific fashion (Schroeder et al., 2004) differentiating the role of the capping enzymes in the synthesis of the m^7G cap and the progression of transcription.

1.1.6 The RNA cap quality control

Given the essential role it exerts on transcription and gene expression, a quality control mechanism of pre-mRNA capping has been proposed (Fig. 1.6). Decapping is the first reaction leading to mRNA degradation and turn-over. The main mRNA turn-over pathway involves the proteins Dcp2 (Dunckley &

Parker, 1999; Lykke-Andersen, 2002; Wang et al., 2002) and Nudt16 (Li, Song, & Kiledjian, 2011; Song, Li, & Kiledjian, 2010) responsible for mRNA decapping and the release of a 5' end monophosphate RNA that is the substrate of the quick degradation mediated by the 5' exonuclease Xrn1 (Decker & Parker, 1993; Hsu & Stevens, 1993). Cap-defective RNAs with 5' triphosphate ends as well as intermediates of the capping synthesis cannot be decapped and processed for degradation, therefore they represent a potential threat to the cell.

In the nucleus, the yeast ribonuclease Rai1 (Rat1 Interacting protein) plays a multivalent activity in this concern. It shows decapping endonuclease activity on RNAs with unmethylated cap and can also catalyse the pyrophosphohydrolysis (PPH) of transcripts missing the entire m^7G cap structure (Xiang et al., 2009). Both of Rai1-mediated reactions release unprotected 5' monophosphate RNA species that are the substrate of exoribonucleases like the nuclear Rat1 in yeast (Xrn2 in mammals) or the cytoplasmic Xrn1. The Rai1 homolog Dxo1 possesses decapping and 5'-3' exonucleolytic activity but cannot catalyse PPH reactions (Chang et al., 2012) contributing to the cap quality control mechanism with its dual function. In capping quality control, Rai1 is also needed to trigger the activity of Rat1 and the consequent co-transcriptional degradation of the RNA with aberrant or missing 5' cap *in vitro* (Jiao et al., 2010).

Co-transcriptional decapping of paused Pol II and Xrn2-mediated premature transcription termination (PTT) also known as Promoter-Proximal torpedo model, has been proposed to be a mechanism involved in the control of Pol II productive elongation and gene expression (Brannan et al., 2012). These observations have been taken in support of the mRNA capping quality control model.

The Rat1 co-transcriptional activity is largely involved in the transcription termination of protein-coding genes where the endonucleolytic cleavage of the nascent RNA generates a 5' RNA portion that is further processed and maturated in the final mRNA and a 3' portion with a 5' monophosphate end extended by the Pol II downstream transcription. Rat1 has been shown to processively degrade the unprotected nascent

RNA and displace the slowly transcribing Pol II from the DNA template downstream of the transcriptional unit (M. Kim et al., 2004; West et al., 2004).

Previous data demonstrated that Rat1 might act co-transcriptionally upstream of the transcription end site (TES) upon specific conditions. Rat1-mediated degradation was retained responsible for the PTT of the extra-long gene *FMP27* when Pol II has processivity defects inducing capping faults or when the capping is impaired in yeast Ceg1 temperature-sensitive mutants (*ceg1-63*). According to the authors, the free RNA 5' termini generated acts as an entry point for the Rat1 exonucleolytic activity that induce the 5'-3' co-transcriptional degradation of the nascent RNA. The kinetic competition between Rat1 and the Poll II for the nascent RNA causes the release of the Pol II from the DNA template upstream of the canonical TES, thus leading to PTT (Jimeno-González et al., 2010). The PTT observed when capping defects are in place on nascent transcripts might suggest that additional signals are needed for the premature release of the Pol II to occur. Rather than the only lack of m^7G cap, PTT observed upon the named conditions might be a function of either enzymatic decapping, Pol II elongation rate over the gene locus or could even require the loading of capping enzymes and their interactors onto the PIC. Importantly, the activation of the cap quality control is required to generate 5' monophosphate RNA species and permit the Rat1-mediated degradation.

Such a scenario is in line with the original torpedo model and the most recent '*sitting duck torpedo*' model proposed for the termination of protein-coding genes in which Rat1 still plays an essential role but in a much more complex circumstances (Cortazar et al., 2019; Eaton et al., 2020; Eaton & West, 2020). Therefore, also the cap quality control might require specific conditions to promptly arrest the synthesis of cap defective RNAs and might interest a specific subgroup of genes.

The m^{7}G cap quality control model
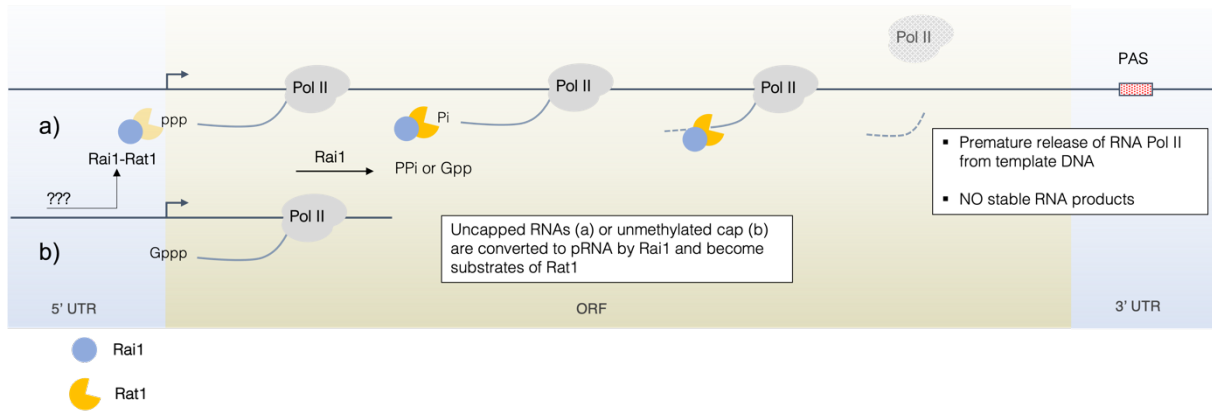
Figure 1.6 - The model proposed for the m^{7}G cap quality control mechanism.

The complex Rai1-Rat1 contacts the transcriptional machinery upon not yet clarified conditions (???). The activity of Rai1 converts cap defective pre-mRNAs in substrates of the Rat1 nuclear exonucleases. The co-transcriptional degradation Rat1-mediated should induce the decay of the nascent 5' monophosphate RNA and consequent premature transcription termination.

1.2 Cellular RNA degradation

The correct processing and maturation of the cellular RNAs is an essential requirement for cell viability. The production and the accumulation of aberrant RNA species is controlled at every stage of the RNA synthesis but also at the later stages of the RNA half -life, for instance at the translational and post-translational levels in the case of the mRNAs. The evolution has permitted the development of several mechanisms aimed at eliminating defective RNA molecules from the cellular pool which also deploy a defence against the invasion of exogenous nucleic acids. Nuclear and cytoplasmic ribonucleases (RNases) are the enzymes responsible for the degradation of the RNA. The cellular RNases are not specific and therefore the correct RNA processing is essential to protect and distinguish the functional transcripts, needed to exert their function, from the aberrant RNAs to remove (Houseley & Tollervey, 2009). Other RNA classes, different from mRNAs, establish strict interactions with RNA binding proteins (RBPs) to preserve their stability and avoid ribonucleolytic degradation (Moore, 2002; Schmid & Jensen, 2018).

1.2.1 Nuclear RNA degradation

In the nucleus, the RNA turn-over is mainly mediated by the exonucleolytic degradation accomplished by systems responsible for fundamental steps of the RNA maturation that can act, therefore, either co-transcriptionally or post-transcriptionally as soon as the RNA is released from the chromatin into the free nucleoplasmic region (Schmid & Jensen, 2018).

The main nuclear 5'-3' degradation is mediated by the exonuclease Rat1 (Xrn2 in mammals) involved in the transcription termination and 3' end processing of protein-coding genes.

Rat1 has been shown to co-transcriptionally degrade RNAs displaying a 5' monophosphate end. As discussed above, Rat1 has been proposed to take part in the quality control of the nuclear mRNAs (Jiao et al., 2010; M. Kim et al., 2004; West et al., 2004; Xiang et al., 2009).

Another fundamental player in the nuclear RNA degradation is the nuclear RNA Exosome (nExosome), a large and conserved multi-subunit complex exerting 3'-5' exonucleolytic and endonucleolytic RNase activity (Houseley, LaCava, & Tollervey, 2006; Kilchert, Wittmann, & Vasiljeva, 2016; Schmid & Jensen, 2018).

The canonical structure of the nExosome shows a 9-proteins channel through which the targeted RNA passes heading with its 3'end. At the end of the channel, the co-factor Dis3 exerts exo- and endonucleolytic activity (Staals et al., 2010). Rrp6 (EXOSC10 in humans) is another co-factor of the nExosome showing 3'-5' exonucleolytic activity. It is located on the top of the complex, and it is thought to be responsible for the degradation of structured RNAs that do not transit through the multi-proteins channel (Schmid & Jensen, 2018). The nExosome is responsible for the degradation of defective precursors of a variety of either coding or non-coding RNAs (Houseley et al., 2006).

As a common feature of the cellular RNA-decay factors, the nExosome partakes either in the maturation or the disruptive degradation of its targets. The nExosome mediates the trimming of the 3' ends of stable ncRNAs such as snRNAs and snoRNAs required for their maturation but also it is the master regulator of the degradation of the most unstable transcripts produced via the pervasive transcription of the genome (CUTs, PROMPTs, eRNAs).

The two activities are often strictly connected, indeed, the nExosome establishes a strict interaction with the Nrd1-Nab3-Sen1 (NNS) complex responsible for the transcription termination pathway alternative to the so-called polyadenylation signal (PAS)-dependent termination. Nrd1 of the NNS complex can interact and recruit Rrp6 of the nuclear exosome and the Trf4 subunit of the Trf4–Air2–Mtr4 polyadenylation (TRAMP) complex. In turn, the transcripts terminated via the NNS complex are rapidly subjected to the exonucleolytic activity of the nExosome which can lead to a trimming of the 3' end in the maturation of the ncRNAs like snoRNAs, or to the complete degradation of the RNA processed (Schmid & Jensen, 2018).

1.2.2 Cytoplasmic RNA degradation

In the cytoplasm several pathways for the RNA degradation are active and involve either the exonucleolytic (5'-3' and 3'-5') or the endonucleolytic cleavage and degradation of a variety of RNA species (Fig. 1.7). The cytoplasmic RNA degradation pathways contribute to the regulation of gene expression and constitute a surveillance system to control the abundance of aberrant or non-self (e.g., viral) RNA molecules and their potential toxicity.

Gene expression is regulated by a process of mRNA turnover aimed to control the transcript concentration according to the physiological needs of the cell. The turnover process of the mRNAs starts with the shortening of the poly-A (pA) tail mediated by different deadenylase protein complexes such as Ccr4-Pop2-Not complex (Daugeron, Mauxion, & Séraphin, 2001; Denis & Chen, 2003; Shimizu-Yoshida et al., 1999). In the Ccr4-Pop2-Not complex, the main deadenylation complex in yeast, Ccr4 and Pop2 are the nucleases responsible for the pA shortening, both are conserved in the eukaryotes (Dlakić, 2000; Moser et al., 1997). The presence of multiple nucleases in the same complex might increase the regulation feedback of the deadenylation process. Moreover, mutational studies on Pop2 suggested a supplemental role for this protein in the stabilization of the Ccr4-Not complex (Chen, Chiang, & Denis, 2002; Thore et al., 2003).

Another deadenylase complex in the cytoplasm is the PAB-specific ribonuclease 2-3 (Pan2-Pan3) complex which has been shown to work upon the inactivation of the main deadenylase complex Ccr4-Pop2-Not or to start the deadenylation process and then leave the main complex to overcome in the pA shortening. The Pan2-Pan3 complex is also involved in the gene-specific mRNA deadenylation (Brown & Sachs, 1998; Hammet, Pike, & Heierhorst, 2002; Tucker et al., 2001).

The Poly-A ribonuclease (PARN) is also involved in the pA shortening in mammals, its homologs have been described in different eukaryotes but are missing in *D. melanogaster* and *S. cerevisiae*. It belongs to the class of the RNaseD like Pop2 and Pan2 (Moser et al., 1997).
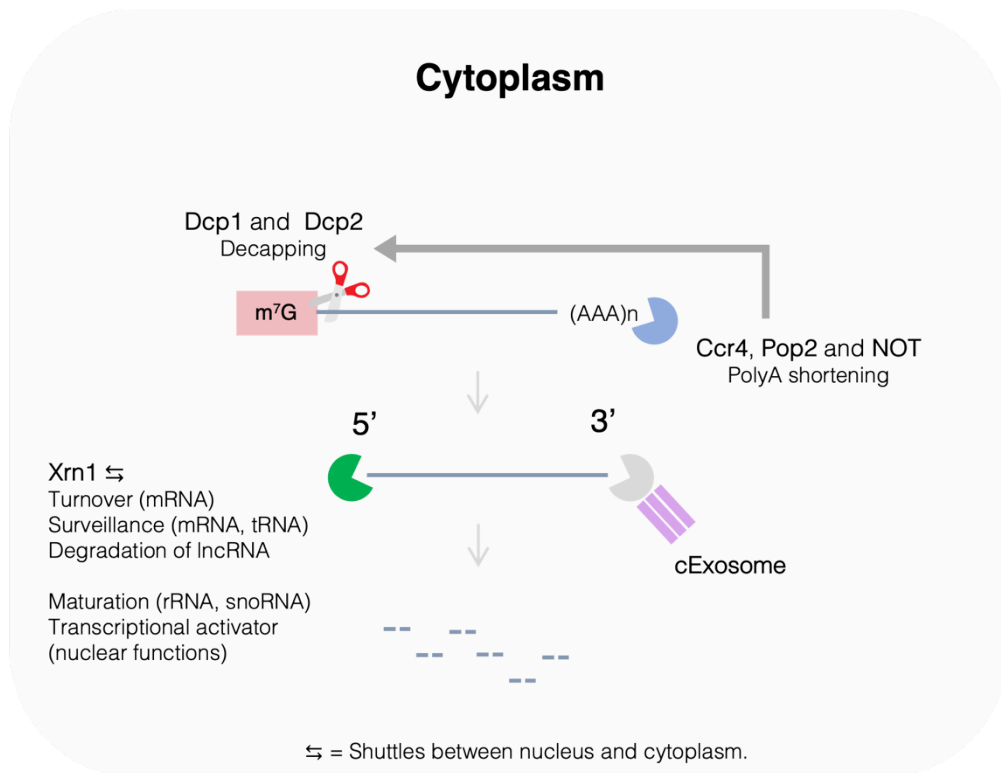
Figure 1.7 - The main cytoplasmic exonucleolytic degradation pathways taking part in the mRNA turnover.

Turnover regulates the cellular mRNA abundance, permits a fast response to environmental changes and maintains the cellular homeostasis. The mRNA degradation often starts with the polyA shortening which in turn, triggers the mRNA decapping. The polyA shortening and the decapping permit to the cytoplasmic exonucleases like the cytoplasmic exosome (cExosome) and Xrn1 to degrade the RNA.

1.2.2.1 3'-5' RNA degradation

The pA shortening can be followed by the 3'-5' mRNA degradation. The Exosome can also exert its functions in the cytoplasm (cExosome) where, alongside its role in the RNA quality control, is one of the main elements regulating the mRNA turnover. In the cytoplasm, the cExosome is differentiated from the nExosome as it establishes interactions with different cofactors (Houseley et al., 2006). The cExosome is the main 3'-5' degradation machinery involved in the clearance of aberrant transcripts and is responsible for the clearance of the products released from the translation-related surveillance pathways like nonsense-mediated decay (NMD), non-stop decay or no-go decay (Doma & Parker, 2006; Mitchell & Tollervey, 2003; van Hoof et al., 2002).

The pA tail shortening triggers the decapping mediated by the Dcp1-Dcp2 protein complex. The model proposed in mammals, involves the polyA-binding protein (PABP). According to the model, the long pA tail permits the interaction with many PABP units, sufficient to trigger the Pan2-Pan3 deadenylation activity and, at the same time inhibiting the Ccr4-Caf1 (Ccr4-Pop2 complex in yeast) and Dcp2 activity. The pA shortening, reduces the number of PABP units bound to the RNA and permits the activation of the Ccr4-Caf1 complex to complete the deadenylation and the activation of the decapping protein Dcp2 (Yamashita et al., 2005).

The Dcp1-Dcp2 holoenzyme is composed of two subunits, but the main decapping activity is exerted by Dcp2. The Dcp1 decapping activity is yet to be confirmed, its main role seems to be related with the stimulation of the Dcp2 decapping function (Dunckley & Parker, 1999; Dunckley, Tucker, & Parker, 2001; She et al., 2004).

Mammalian cells display a second decapping protein: Nudt16 (Song et al., 2010). Nud16 works in parallel with the Dcp1-Dcp2 complex to catalyse the decapping of different substrates. While Dcp2 is responsible for the decapping of the defective RNAs subjected to NMD, Nud16 exerts an overlapping function with

Dcp2 in the miRNA-mediated gene silencing and a gene-specific decapping activity on A-rich elements (ARE)-containing mRNA degradation has also been shown for Nud16 (Li et al., 2011).

1.2.2.2 5'-3' RNA degradation

The decapping reactions generate a free 5' end monophosphate RNA that is the substrate of the 5'-3' exonucleolytic degradation (Houseley & Tollervey, 2009; Parker & Song, 2004).

The main cytoplasmic 5'-3' exonuclease is Xrn1 (Hsu & Stevens, 1993). The best-known function of Xrn1 in the cytoplasm is related to the mRNA turnover.

Xrn1 is a multi-domain exonuclease conserved from yeast to human (Johnson & Kolodner, 1991; Stevens, 2001).

Its catalytic activity is highly specific for the 5' monophosphate RNAs. This characteristic, together with the high Xrn1 processivity ensure the degradation of RNAs subjected to the decapping and avoid the release of incompletely degraded intermediates like short RNAs (Pellegrini et al., 2008; Stevens, 2001).

Xrn1 has been shown to interact with the decapping complex Dcp1-Dcp2 suggesting a close connection between the processing and degradation steps. In this concern, cellular structures called P bodies have been described. In the P bodies RNAs released from the translation and the proteins involved in the mRNA turnover/surveillance, cluster (Eulalio, Behm-Ansmant, & Izaurralde, 2007).

1.2.3 Other functions of Xrn1

Xrn1 is also involved in the surveillance of tRNA modifications, specifically, Xrn1 as well as other components of the Xrn1 family like Rat1 (Xrn2 in mammals), have been shown to be part of the degradative apparatus clearing the cell from hypomodified tRNAs.

Interestingly, Xrn1 can shuttle between the cytoplasm and the nucleus where it exerts several functions in the RNA processing and gene expression. The Xrn1 nuclear import has been shown to be dependent on its degradative function as Xrn1 proteins able to bind uncapped RNAs and still exerting exonucleolytic activity are efficiently imported into the nucleus (Haimovich et al., 2013).

Xrn1 mediates the degradation of several ncRNAs and its depletion in *S. cerevisiae* (*xrn1Δ*) induces the accumulation of a specific class of antisense ncRNAs named Xrn1-sensitive unstable transcripts (XUTs). XUTs have been shown to be antisense ncRNAs of protein-coding genes related with the regulation of gene expression (Van Dijk et al., 2011).

Xrn1 has been described to be responsible for the degradation of defective 18S rRNA units that brings to the formation of elongation-defective ribosomes (Cole et al., 2009). Xrn1exerts more functions on the rRNAs, it is involved in the trimming of the 25S rRNA precursor to generate the mature 25S unit when the other 5'-3' exonuclease Rat1 is depleted (Chernyakov et al., 2008; Geerlings et al., 2000; Parker, 2012). Xrn1 has been shown to compensate the 5' end trimming activity of the nuclear Rat1 for a subset of polycistronic snoRNAs (Qu et al., 1999). Many observations have described the dual localization of Xrn1 to exert its function as 5' exonuclease in the degradation and processing of many RNA species.

1.2.4 Xrn1 as transcription regulator

The buffering of the cellular RNA levels is accomplished by the components responsible for RNA synthesis and degradation and the balance between the two processes. Intriguingly, constant feedbacks coming from the cellular decaysome, can directly regulate the transcription and affect the RNA synthesis. Previous studies have highlighted the role of Xrn1 as a transcriptional regulator. Xrn1 was shown to bind to chromatin and preferentially the promoter region of highly expressed genes (Haimovich et al., 2013). The disruption of the *XRN1* gene, has been shown to affect the synthesis rate of a certain group of mRNAs

while increasing the half-life of another group of mRNAs in a fashion indicated by the authors as "*compensatory manner*" (Fig. 1.8A). Genome wide analysis on the *xrn1Δ S. cerevisiae* strain showed a decrease in the transcriptional rate for a large proportion of genes with a stronger effect on highly expressed genes. The lack of Xrn1 was also reported to either affect the transcription initiation and elongation Pol II-mediated by preventing backtracking (Begley et al., 2019; Medina et al., 2014). Overall, Xrn1 can exert a double role as the main exonuclease in the cytoplasm and positive transcriptional factor for a specific group of genes, in the nucleus.

Independent results published in parallel, showed Xrn1 to be the main regulator of the cellular mRNA levels with a mechanism and dynamics in contrast with previous results. The data proposed from Sun et al., provided evidence of Xrn1 indirectly affecting the RNA synthesis, activating the transcription repressor Nrg1 in *S. cerevisiae*. According to the authors, in this case, the mRNA buffering was sustained globally by the compensation between degradation rate and synthesis of the mRNAs. According to the model, the molecular sensor of the cellular mRNA levels is Xrn1 itself. The concentration of *XRN1* mRNA and the resulting protein, when abundant, induce the transcriptional repressor Nrg1 with consequent downregulation of the expression levels. On the other hand, the RNA stability was preserved upon low Xrn1 concentration (Fig. 1.8B). This feedback loop is applicable to the entire mRNA set as low Xrn1 levels stabilises the global mRNA population maintaining the physiological balance (Sun et al., 2013).

Interestingly, a common necessary condition to observe the effects on the global mRNA population was the Xrn1 catalytic activity, the ability to bind the uncapped RNAs as well as the decapping process (Haimovich et al., 2013; Medina et al., 2014; Sun et al., 2013).

The participation of Xrn1 in the degradation and synthesis of the cellular mRNAs describe a continuous and regulated crosstalk between the transcriptional and degradative cellular system to maintain the mRNA homeostasis and respond to the environmental changes.
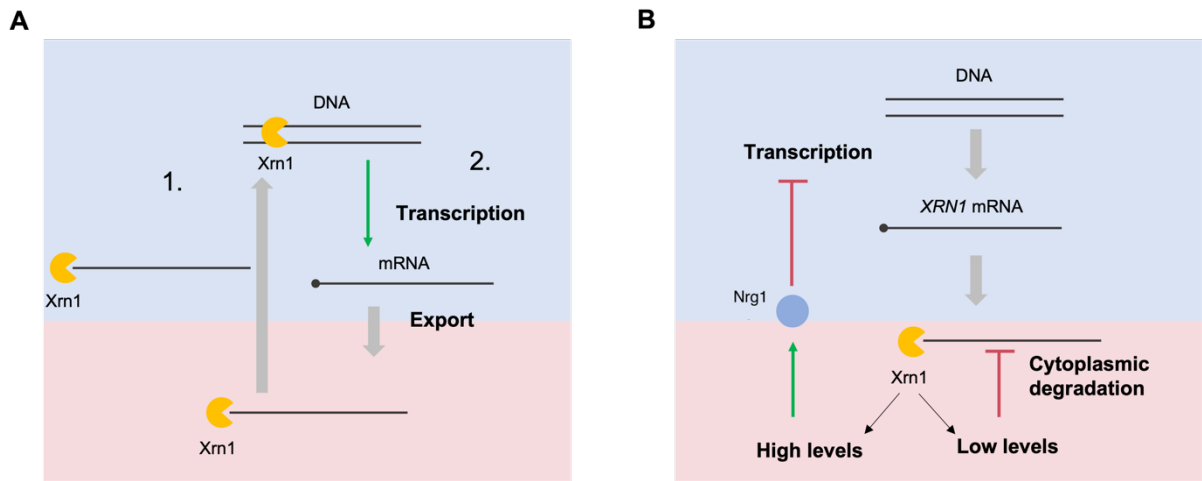
Figure 1.8 - Two models have been proposed to describe the role of XRN1 as a transcriptional regulator.

*A) XRN1 shuttles in the nucleus where it has been observed to interact with the chromatin at the promoter level of several genes (1). Here Xrn1 can promote transcription in a "compensatory manner" (2). **B)** The cytoplasmic levels of Xrn1 are responsible for the buffering of the cellular mRNA levels. High levels of Xrn1, activate the negative transcriptional factor Nrg1 which can inhibit the global cellular transcription. Low levels of Xrn1 increase the global mRNA half-life by reduced mRNA degradation. In both cases the interaction of Xrn1 with the decapped mRNA is essential. Figure based on (Haimovich et al., 2013; Sun et al., 2013).*

2.   Materials and Methods

2.1 Yeast media

Yeast strains were grown at 30 °C on YPD (1% w/v yeast extract, 2% w/v peptone, 2% w/v dextrose/glucose) media enriched in DMSO or rapamycin to the final concentration of 1 µg/mL where indicated. For 6-AU treatment, the yeast cells were grown in minimal media (yeast nitrogen base without amino acids) with synthetic dropout ura-. For solid media, agar was added to the final concentration of 2%.

2.2 RNA extraction

RNA was purified following the Phenol/Chloroform method. A single colony was inoculated in 50 mL of YPD and grown overnight at 30 °C with agitation. The pre-culture was diluted in YPD to the final concentration of $OD_{600} = 0.25$ and incubated at 30 °C with agitation to the final concentration $OD_{600} = 1$, pelleted in 50 mL falcon tubes by centrifugation for 2 mins at 200 x g at room temperature (RT). The pellet was frozen at -80 °C before further manipulation. Next, the pellet was resuspended in 1 mL of ice-cold RNase free water, moved to a 1.5 mL tube, centrifuged for 10 sec. at 4 °C max speed. Then the supernatant (water) was removed, and the pellet resuspended in 400 µL of filtered AE buffer (50 mM sodium acetate pH5.3; 10 mM EDTA), 40 µL of 10% w/v sodium dodecyl sulphate (SDS) and 400 µL of acid phenol (pH 4). The suspension was vortexed for 20 sec and incubated at 65 °C for 10 mins, then incubated for further 10 mins at -80 °C. The defrosted sample was next centrifuged for 5 mins, 16 000 x g at RT. 400 µL of 1:1 phenol:chloroform solution were added to the supernatant and vortexed for 30 sec. Then the sample was centrifuged for 10 mins, 16 000 x g at RT. The upper phase was transferred to a clean tube

and mixed with 400 µL of chloroform, vortexed and centrifuged for 5 mins, 16 000 x g at RT. The RNA was precipitated by adding to the upper phase 1 mL of 100% ethanol and 40 µL of 7.5 M ammonium acetate. Left at -80 °C for at least 2 hrs then centrifuged for 20 mins, 13 000 at 4 °C. The pellet was washed with 70% ethanol (in water) and resuspended in the adequate volume of RNase free water.

2.3 DNase treatment

20 U DNase I (Roche, Product No. 04716728001) were added to 5 µg RNA, 1x DNase Buffer (Roche) and 40 U RNase Inhibitor (RiboLock, EO0384) to the final volume of 50 µL with water. The sample was incubated at 37 °C for 30 mins then the RNA was re-precipitated following the Phenol/Chloroform method described above.

Reagent	Amount
RNA	5 µg
DNAse 10 U/µL	2 µL
Buffer 10X	5 µL
RNAse Inhibitor	1 µL
H_2O	To 50 µL

2.4 Reverse transcription (RT)

100 ng to 2 µg of RNA were resuspended in 10 µL of water and added with 2 µL 50 µM random hexamers (Invitrogen SO142), 1 µL of 10 µM dNTPs (Mix A). The mix was incubated at 65 °C for 10 minutes then added with 4 µL 5x RT Buffer, 2 µL DTT 0.1 M, 0.2 µL RNase inhibitor (RiboLock, EO0384) and 40 U SuperScript III-Reverse Transcriptase (Invitrogen 10368252) (Mix B). Following, the mix was incubated 10 mins at RT, 40 mins at 50 °C and 5 mins at 82 °C.

Mix A	Reagent	Amount
	H_2O	10 µL
	RNA (100 ng/µL)	1 µL
	Random hexamers 50 µM	1 µL
	dNTPs 10 µM	1 µL

Mix B	Reagent	Amount
	Buffer 5X	4 µL
	DTT 0.1 M	2 µL
	RNAse Inhibitor	0.2 µL
	Reverse Transcriptase	1 µL

2.5 Cap Immunoprecipitation

Total RNA from cells treated with DMSO or rapamycin was extracted and used for cap pull-down. 5 μg of total yeast RNA was spiked with 1 μg of total human RNA as immunoprecipitation efficiency control. Capped RNA species were immunoprecipitated using 5 μg of anti m^7G cap antibody (H-20).

Per each reaction, 50 μl of Dynabeads proteinG (Invitrogen 10765583) were washed three times in 1 mL of IPP150 buffer (150 mM NaCl/ 0.1% NP-40/ 10 mM Tris, pH 8.0) and three times in 1 mL of IPP500 buffer (500 mM NaCl; 0.1% NP-40; 10 mM Tris, pH 8.0).

After washes, the beads were resuspended in 50 μL of IPP500 buffer added with 5 μg of the anti m^7G cap H-20 monoclonal antibody (in parallel, a reaction without antibody was performed as negative control). The mix was rotated at 4 °C for 2 hrs. Next, the beads were washed three times with IPP150 and incubated upon rotation with IP reaction mix (5 μg of total yeast RNA, 1 μg of total spike RNA, 5 μL of 0.1M DTT, 50U of RNase inhibitor (RiboLock, EO0384) and IPP150 to a total volume of 200 μL at 4°C overnight. The beads were then washed five times in 0.5 mL of cold IPP150 buffer containing 2.5 mM DTT, resuspended in 200 μL of IPP150 containing 1mg/mL Proteinase K (Sigma 3115828001) and incubated at 37 °C for 30 min. The RNA from the supernatant, was extracted via phenol/chloroform precipitation and subjected to reverse transcription. The resulting cDNA was used to quantify the levels of capped RNA normalised against the *GAPDH* gene from spike RNA and to RNA levels pre immunoprecipitation as previously described (Jimeno-González et al., 2010).

2.6 Quantitative PCR (qPCR)

The cDNA obtained from the reverse transcription reaction of 100 ng RNA was first diluted 1:10 for Pol II genes and 1:40 for rRNA and used in the preparation of the qPCR mix using qPCRBIO SyGreen Mix

Lo-ROX (PCRBio PB20.11-05) following the manufacturer instructions. The reaction ix was composed of 7.5 µL of qPCRBIO SyGreen Mix Lo-ROX, 0.8 µL of each primer 10 µM, 0.9 µL of water and 5 µL of 1:10 or 1:40 cDNA.

Reagent	Amount
qPCRBIO SyGreen Mix Lo-ROX	7.5 µL
Forward Primer 10 µM	0.8 µL
Reverse Primer 10 µM	0.8 µL
H$_2$0	0.9 µL
cDNA (1:10 or 1:40 dilution)	5 µL

qPCR data were analysed with the ΔΔCt method (Schmittgen & Livak, 2008) normalised to 25S ribosomal RNA (rRNA) and control condition.

2.7 Growth assay in liquid media

To determine the growth rate of the yeast strain, a single colony was inoculated and grown overnight at 30 °C in 50 mL of YPD upon agitation. Then the cellular suspension was diluted in fresh YPD to OD_{600} = 0.2 in a final volume of 50 mL. The media was enriched of DMSO or rapamycin to a final concentration of 1 µg/mL and maintained at 30 °C with constant agitation. 1 mL of suspension was sampled at time 0 and at regular intervals of 15 mins to determine the cellular concentration through spectrophotometric methods. The value of OD_{600} was then plotted as function of time.

2.8 Chase experiment

2.8.1 FMP27 stability

A single colony was inoculated in 50 mL of YP-Galactose and growth overnight at 30 °C with agitation. Next, the suspension was diluted in 50 mL of fresh YP-Galactose to the final concentration of OD_{600}=0.25. The suspension was incubated at 30 °C under agitation to a final concentration of OD_{600}=0.8. 10 mL sample was harvested and processed for RNA purification, subsequently a solution of 20 % glucose was added to the media to the final concentration of 2%. Further samples were collected at 5, 10 and 30 minutes after the glucose addition. The samples were prepared for RNA extraction and subsequent RT-qPCR analysis.

2.8.2 6-azauracil treatment/ chase experiment

A single colony of the specific strain was inoculated in 50 mL of SD -ura liquid media with raffinose as main source of carbon and growth overnight at 30 °C with agitation. Next the suspension was diluted in 50 mL of fresh SD -ura to the final concentration of OD_{600}=0.25 and incubated at 30 °C with agitation to OD_{600}=0.8. Then 6-azauracil or ammonium hydroxide were added to the media to a final concentration of 50 µg/mL for 30 minutes at 30 °C with agitation. Subsequently, a 20% aqueous solution of galactose was added to the media to the final concentration of 2%. 10 mL of the suspension were collected a time 0, 5, 10 and 20 minutes and prepared for RNA purification.

2.8.3 6-azauracil treatment / ceg1-AA

A single colony of the specific strain was inoculated in 50 mL of SD -ura liquid media with glucose as main source of carbon and growth overnight at 30 °C with agitation. Next the suspension was diluted in

50 mL of fresh SD -ura to the final concentration of $OD_{600}=0.25$ and incubated at 30 °C with agitation to $OD_{600}=0.8$. Then 6-azauracil or ammonium hydroxide were added to the media to a final concentration of 50 μg/mL for 30 minutes at 30 °C with agitation. Subsequently DMSO or rapamycin to a final concentration of 1 μg/mL were added to the media. 10 mL of the suspension were collected a time 0, 45, 90 and 120 minutes and prepared for RNA purification.

2.9 6-azauracil (6-AU) treatment

Overnight culture was prepared inoculating a single colony on minimal media as described above. The next day the culture was diluted to $OD_{600} = 0.25$. The diluted culture was grown with agitation at 30 °C to $OD_{600} = 0.8$ and split in 4. Two cultures were added of 6-azauracil (6-AU) and the other two with ammonium hydroxide (NH_4OH, control) at the final concentration of 50 μg/mL for 30 minutes. Next, DMSO or rapamycin (Rap) at the final concentration of 1 μg/mL were added according to the following scheme: 6-AU+DMSO, 6-AU+Rap, NH_4OH+DMSO, NH_4OH+Rap. 10 mL of each suspension were collected a time 0, 45, 90 and 120 minutes after the addition of DMSO or Rap and prepared for RNA purification.

2.10 Ribosomal RNA (rRNA) depletion – RiboPOP method

The ribosomal RNA (rRNA) depletion was performed adapting the original method (Thompson, Kiourlappou, & Davis, 2020). The probes were designed to target the rRNA 18S and 25S of *S. cerevisiae* according to the authors code available on https://github.com/marykthompson/ribopop_probe_design. The probes generated were commercially synthesised and 3' biotinylated (IDT). Probes were re-hydrated at the final concentration of 100 μM with resuspension buffer (10 mM Tris, pH 8 and 0.1 mM EDTA)

and pulled together to the final concentration of 75 μM (75 pmol/μL) pool or 5 μM each. 200 μL of MyOne C1 Streptavidin beads (Invitrogen 65001) per reaction were prepared as follow: three washes with 1X B&W buffer (5 mM Tris-HCl, pH 7.5; 0.5 mM EDTA; 1 M NaCl) plus Tween-20 0.01%. Two washes with Dynabeads buffer A (0.1 M NaOH; 0.05 M NaCl), Two washes with Dynabeads buffer B (0.1 M NaCl) plus Tween-20 0.01%. Finally, the beads were washed twice with 1X Ribohyb buffer (1:10 dilution of the stock 10X Ribohyb buffer: 20X SSC; 0.1% Tween-20. 20X SSC: 3 M NaCl; 300 mM trisodium citrate; pH adjusted at 7 with HCl). The beads were resuspended in a 5X original volume of 1X Ribohyb buffer. In parallel the hybridization mix was prepared mixing 5 μg total RNA (DNased), 200 pmol probes mix, 1X final concentration of Ribohyb buffer, 1 μL RiboLock and water to 40 μL. The hybridization mix was incubated at 80 °C for 10 mins and then let cool down to 70 °. The hybridization mix was maintained at 70 °C for 5 mins and then incubated at RT for 10 mins. 60 μL of 1X Ribohyb buffer was added to the hybridization mix and the resulting 100 μL incubated with the beads removed of the resuspension solution. The reaction was then incubated at 50 °C for 30 mins with vigorous mix (100 rpm on Mixer HC, Starlab S8012-0000). The beads where then moved on the magnetic rack and the supernatant (containing the total RNA depleted of the rRNA fraction) was saved into a new tube. 100 μL of 1X Ribowash buffer (0.4X SSC; 0.01% Tween-20) were used to resuspend the beads. The beads were separated again taking them on the magnetic rack and the second supernatant added to the first saved volume. The final 200 μL of ribodepleted RNA were precipitated with 100% ethanol and 7.5 M ammonium acetate with the addition of glycogen. The solution was incubated at -80 °C overnight then spin for 30 mins at 4 °C. The pellet washed with 70% ethanol and resuspended in RNase free water.

2.11 RNA sequencing

Before library preparation, ribosomal RNA (rRNA) depletion was performed on 5 µg of total yeast RNA using 200 pmol of 3' biotinylated probes (IDT) designed in house following the RiboPOP method (Thompson, Kiourlappou, and Davis 2020). The libraries preparation was performed with Ultra II RNA Library Prep Kit for Illumina (NEB, E7770) and Multiplex Oligos for Illumina (NEB, E7335) according to manufacturer's instructions. The RNA sequencing was performed from the Genomics Facility at the University of Birmingham on Illumina NEXTseq apparatus.

The RNA-seq data are available in Gene Expression Omnibus (GEO) database with the accession number GSE213942 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE213942).

2.12 RNA-seq data analysis

The quality of the data was checked using FastQC v0.11.5 (Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

Reads were aligned to the yeast genome R64 (sacCer3) using HISAT2 v2.1.1 and sorted and indexed using SAMtools v1.15.1 (Wysoker et al., 2009). The read count per gene were calculated with LiBiNorm v.2.5 (N. Archer, Walsh, Shahrezaei, & Hebenstreit, 2016) with the following parameters count -u name-root -f --order=pos --minaqual=10 --mode=intersection-strict --idattr=gene_id --type=exon. Differential expression analysis was performed using DESeq2 package. GO analysis was generated using g:profiler (https://biit.cs.ut.ee/gprofiler).

2.13 4-Thiouracil (4tU) labelling and purification

Liquid yeast culture traded for 40 mins with rapamycin (1 µg/mL) or DMSO were added of 4tU 500 µM final concentration (Sigma-Aldrich, 440736) to final concentration 5 mM for 5 minutes and subsequently flash frozen in cold Ethanol (-80 °C). The suspension was pelleted RNA purified by Phenol/Chloroform extraction. 200 µg of total RNA from *S. cerevisiae* was spiked with *S. pombe* total RNA (4:1 ratio *S.cereviasiae* : *S. pombe*) treated with 4tU 500 µM final concentration (Sigma-Aldrich, 440736). The biotynilation was performed with 200ul EZ-link-HPDP-biotin (Thermo Fisher, 21341) for 30 minutes at 65 °C. The RNA was then incubated with 5M sodium chloride and isopropanol and incubated with Dynabeads M-270 Streptavidin (Thermo Fisher, 65305) for 1 hour at room temperature. Beads were washed five times with Binding & Washing buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA and 1 M NaCl) and 4tU labelled-RNA was eluted with 100 µM DTT, precipitated with 5M NaCl, isopropanol and glycogen, and resuspended DNase/RNase free water. Enrichment of nascent RNA was confirmed with qPCR using exon–intron primers of *ACT1* gene.

2.14 Yeast cell transformation

2.14.1 Long method

Cells were cultured to $OD_{600} = 1$ and pelleted by centrifugation for 5 mins at 2 400 x g at room temperature (RT). The pellet was resuspended in sterile 10 mM Tris-HCl pH 7.5 or water and centrifuged again as before. The pellet was resuspended in filtered LiT (10 mM Tris-HCl pH 7.5;100 mM Lithium acetate), 1 M DTT and incubated at RT for 40 mins with gentle shaking. After incubation the cells were pelleted again (as before) and resuspended in LiT and 1 M DTT. The suspension was added with LiT; and dsDNA (10 mg/mL); transforming DNA (0.1 to 1 µg or more) and incubated at RT for 10 mins. Next, PEG solution (1:1 PEG4000 : LiT) was added to the suspension and incubated at RT for 10 mins. DMSO was

then added, and the suspension incubated at 42 °C for 15 mins. After the incubation the suspension was pelleted and resuspended in 1 mL of YPD (1% w/v Bacto yeast extract, 2% w/v Bacto peptone, 2% w/v glucose/dextrose). Then the suspension was incubated at 30 °C for 1 hour, pelleted with 10 sec spin, max speed and inoculated on selective plates.

2.14.2 Quick yeast cells transformation

About 50 µL of fresh cells were directly scrubbed from the plate and washed twice with autoclaved water and pelleted. The cells were then resuspended in 100 µL of the PCR product of the transforming DNA. 240 µL of a solution of 50% w/v PEG-4000 in water were added to the cells followed by 8 µL of sterile 4 M lithium chloride. Next, 10 µL of 10 mg/mL carrier ssDNA were added to the mix. The suspension was incubated at 42 °C for 30 mins. After incubation, the suspension was spin at 2 400 x g for 1 min. The pellet was gently resuspended fresh YPD media and incubated at RT for 1 hour with rotation. The cells where then palleted and plated on the YPD plate according to the specific selection method (antibiotic or metabolic selection).

2.15 Growth spot assay

To evaluate the cell viability and growth phenotypes, yeast cells from a single colony were inoculated in 50 mL of YPM and incubated overnight at 30 °C with agitation. The suspension was diluted to OD_{600}=0.25 with fresh YPD and incubated at 30 °C with agitation to OD_{600} = 0.4. Serial dilutions 1:25 in YPD were performed and 5 µL of each dilution spotted onto YPD (2% w/v agar) plates added with DMSO or rapamycin to the final concentration of 1 µg/mL. The plates were grown at 30 °C in a static incubator for 24/48 hrs.

2.16 Chromatin immunoprecipitation (ChIP)

The specific yeast strain was grown over-night to OD600 = 1. Then 37% formaldehyde was added to the media to final concentration 1% v/v dropwise with gentle shaking for 15 mins. Next, 15 mL of 2.5 M glycine was added for further 5 mins keeping the shaking. The cells were pelleted by centrifugation at 2 400 x g for 3 mins at 4 °C. The pellet was resuspended in 40 mL of ice-cold 1x PBS and centrifuged at 2 400 x g for 5 mins at 4 °C, twice. The pellet obtained was then transferred and prepared for lysis. The pellet was resuspended in 1 mL of ice-cold FA1 (50 mM HEPES-KOH, pH7.5; 150 mM NaCl; 1 mM EDTA; 1% v/v triton-x 100; 0.1% v/v Sodium deoxycholate) added with proteinase inhibitor (Roche Mini-EDTA free) and about 400 μL zirconia beads. The cells were lysed in MagnaLyser (30 sec 7000 rpm, 3 cycles with 5 min incubation on ice in between). The lysate was drained out of the beads by centrifugation 200 x g for 1 min at 4 °C and SDS was added to a final concentration of 0.05% v/v. The lysate was sonicated as follow: 25 cycles 15 sec ON, 15 sec OFF on maximum power. Next, it was centrifuged for 20 mins at 15 000 x g at 4 °C. 200 μL of lysate was saved to check the fragmentation. The lysate was added with SDS to final concentration of 1% v/v and 0.05 mg of Proteinase K and incubated at 42 °C for 1 hour. Next it was moved to 65 °C for 4 hrs. The DNA de-crosslinked was precipitated by Phenol/Chloroform, the contaminant RNA removed by adding 2 μL of RNase A incubating for 1 hour at 37 °C and checked on agarose gel.

The rest of the lysate was centrifuged max speed for 5 mins at 4 °C. 150 μg of chromatin was used for immunoprecipitation. FA1 enriched with proteinase inhibitor was added to the DNA to a final volume of 700 μL. 20 μL of such mix was saved as input for the further ChIP analysis. 2 μg of the specific antibody was added to the sample and incubated on rotation over-night at 4 °C.

15 μL of protein A (Thermo Fisher, 10002D) and 15 μL of protein G-coated beads (Thermo Fisher, 10004D) were washed with 1 mL of FA1 and resuspended in 100 μL of FA1. The beads were added to the sample and incubated by rotation for 1 hour at 4 °C. After incubation the beads were washed 6 times

with FA1, 1 time with FA2 (50 mM HEPES-KOH, pH 7.5; 500 mM NaCl; 1 mM EDTA; 1% v/v Triton-X 100; 0.1% w/v Sodium deoxycholate), 1 time with FA3 (20 mM Tris pH 8.0; 250 mM LiCl; 0.5% v/v NP-40; 0.5% w/v Sodium deoxycholate; 1 mM EDTA) and TE buffer pH 8 (100 mM Tris-HCl; 10 mM EDTA). The beads were then resuspended in 195 µL of Elution buffer (50 mM Tris pH 7.5; 10 mM EDTA; 1% v/v SDS) and incubated for 10 mins at 65 °C and 0.05 mg of Proteinase K added. The Chromatin was then de-crosslinked incubating the sample at 42 °C for 1 hour and moved at 65 °C for 4 hrs. The DNA was finally purified using Quiagen cleanup (28104) kit following the manufacturer's instruction.

2.17 Yeast total DNA purification

To obtain the genomic DNA from yeast cells, the single colony was resuspended in 100 µL TE buffer (10 mM Tris, pH 8; 1 mM EDTA) and 1% v/v SDS. The suspension was vortexed for 2 mins then 300 µL of TE buffer and 400 µL of phenol (pH 8) were added. The suspension was then vortexed again for 2 mins and spin for 5 mins at max centrifuge speed. The upper aqueous phase separated and added to 400 µL of 1:1 phenol/chloroform mix, then vortexed for 30 sec and centrifuged for 10 minutes at top speed. The upper phase was taken and added to 400 µL of chloroform then vortexed for 30 secs and spin for 5 mins top speed. The DNA contained in the upper phase generated was precipitated with 100% ethanol and 3 M sodium acetate. The mix was left in ice for 30 mins and spin for 30 mins at 4 °C, top speed. The pellet was washed with 70% ethanol and resuspended in nuclease-free water.

2.18 Agarose gel electrophoresis

DNA samples were size fractionated through 1% w/v agarose-gel electrophoresis in TAE 1X (40 mM Tris base; 20 mM acetic acid; 1 mM EDTA) buffer. For the DNA detection SYBR Safe gel stain (Thermo Fisher, 10328162) was added to the melted gel to a concentration of 0.5 µg/mL. Before their loading the samples were re-suspended in TriTrack DNA Loading Dye (Thermo Scientific R1161). The run was conducted at 100 v/cm and 80 mA in TAE 1X running buffer. DNA was detected by UV rays trans illuminator machine, and their size estimated by 1 kb Plus DNA Ladder (Invitrogen 10787-026). The results were recorded by digital photography.

2.19 DNA gel purification

2.19.1 Commercial method

DNA bands were purified from the gel using the QIAquick Gel Extraction Kit (Quiagen 28704) according to the manufacturer instruction.

2.19.2 Freeze 'n Squeeze"

Alternatively, the DNA gel purification was performed following the "freeze 'n squeeze" method. A large (1 mL) filtered pipette tip was cut at its thin end to fit into a 1.5 mL Eppendorf tube with the filter on top. The gel portion containing the DNA band was introduced in the filter tip contained into the Eppendorf and frozen at -80 °C for 5 mins. Nest, the Eppendorf tube containing the frozen gel into the filter tip, was spin for 3 mins at RT, top centrifuge spin. The DNA solution was separated from the frozen gel matrix hold by the filter and used for downstream procedures.

2.20 Endpoint PCR

DNA amplification with different purposes was performed using the HiFi-App Polymerase (ARP041) in a reaction volume from 20 to 50 µL. The reaction mix was prepared mixing 5x HiFi reaction buffer to the final concentration 1X, forward and reverse primer 10 µM at the final concentration of 200 nM each, template DNA from 100 to 500 ng HiFi-App Polymerase (2U/µL) 1.25U - 5U per reaction and water to the final volume needed.

Reagent	Amount
5x HiFi reaction buffer	10 µL
Forward Primer 10 µM	1 µL
Reverse Primer 10 µM	1 µL
Template DNA	100 to 500 ng
HiFi-App Polymerase (2U/µL)	1.25U - 5U
H$_2$0	To 50 µL

2.21 DNA sequencing

The DNA samples were sequenced via Sanger method by Source Biosciences (www.sourcebioscience.com/home). Plasmid DNA was supplied at a concentration of 100 ng/µl in a volume of 5 µl per reaction and the sequencing primers at a concentration of 3.2 pmol/µl in a volume of 5 µl per reaction.

2.22 Gene targeting

The yeast genome was stably modified by either the introduction, substitution or deletion of DNA portions. To achieve such genetic modification a linear double stranded (ds)DNA sequence was introduced into the yeast cell via the transformation method described above. The endogenous gene targeting specificity was obtained introducing, via PCR, 50 nucleotides of homology region between the transforming DNA fragment and the endogenous genomic locus.

2.23 Restriction Enzyme RE digestion

DNA endonuclease digestion was performed using the specific digestion buffer suggested by the manufacturer's guidelines. The digestion was conducted with 100-500 ng of DNA and 5 U of the restriction enzyme in a final volume of 50 μL. The reaction was incubated for 2-4 hours at the temperature suggested by the manufacturer.

2.24 DNA ligation

DNA ligation was performed using the T4 ligase (New England Biolabs M0202S). Through this reaction two double stranded DNA ends are joined together by phosphodiester bond. The reaction was conducted with 20 ng of digested plasmid, 5-10-fold molar excess of insert fragment and 1X ligase buffer in a final volume of 20 μL. 1 U of T4 DNA ligase catalysed the reaction for 4-10 hours at room temperature.

2.25 Bacterial growth media

Competent *Escherichia coli* DH5α strain was employed for plasmid amplification. The colonies were maintained in agar-medium containing the specific antibiotic at 4 °C. For overnight liquid culture, Luria-Bertani (LB) medium was employed with the adding of the specific antibiotic at 100 μg/mL. Cells were incubated either in liquid or solid LB at 37 °C upon agitation or in a static incubator.

2.26 Bacterial transformation

Transformation was performed employing 1-100 ng of plasmid DNA. 50 μL of competent cells were thawed on ice then transferred in a tube containing the DNA and incubated on ice for 30 minutes. Then the heat shock was conducted in a heat block pre-heated at 42 °C for 45 sec. The cells were moved on ice for 2 minutes. 70 μL of pre-warmed liquid LB media were added to the cells and then incubated under soft shacking for 1 hour at 37 °C. from 20 to 50 μL of the suspension were plated on solid LB media with relative antibiotic and incubated over-night at 37 °C.

2.27 Hepatitis Delta Virus (HDV) Ribozyme dsDNA generation

For the generation of the dsDNA coding for the HDV ribozyme ready to be cloned into the receiving plasmid, two ssDNA oligo were ordered commercially (IDT). The two oligos showed each other complementary region at their 3' end and the restriction site for a different restriction enzyme at their 5' end (Scheme 1). To obtain a full-length dsDNA, the two oligos were rehydrated in nuclease-free water then 20 μg of each were put together (10 μL final volume) with SuperScript III-Reverse Transcriptase 5X buffer at the final concentration of 0.5X. The solution was incubated for 10 mins at 80 °C and slowly cooled down to 50 °C. The thermocycler heat block was pre-heated at 72 °C and subsequently the oligos

solution was added to 2U of HiFi-App Polymerase (2U/μL), 1X final concentration 5X Hi-Fi reaction

buffer and water up to 50 μL final volume. The mix prepared was incubated at 72 °C for 30 mins. The

constitution of the dsDNA was assessed on 2% agarose gel (Fig. 1).
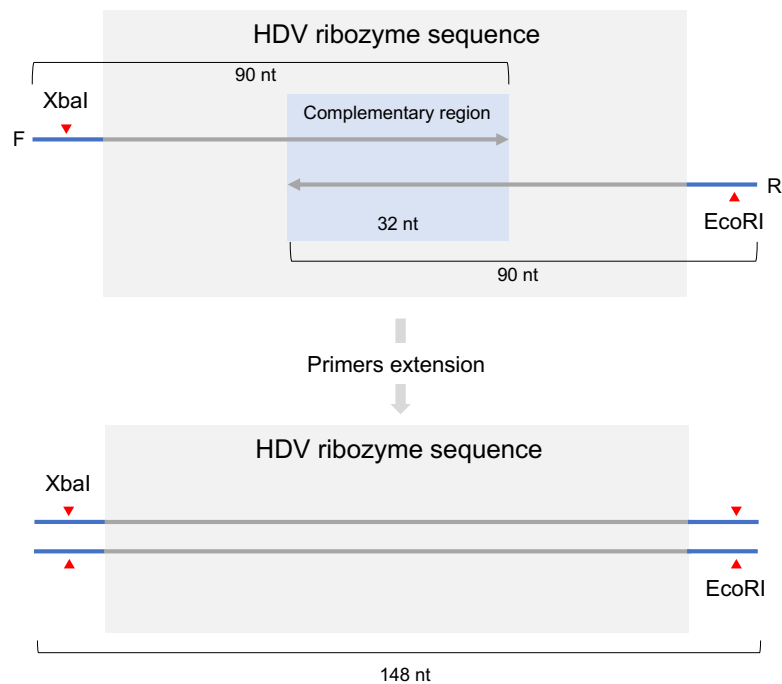
Scheme 1 - The principle behind the preparation of the dsDNA containing the Hepatitis Delta Virus (HDV) ribozyme sequence (grey portion of the arrows). The forward (F) and reverse (R) oligos were annealed by the complementary region (blue) and extended to obtain a full dsDNA. The constitution of the full dsDNA leaded to the formation of the restriction sites (red triangles) of XbaI and EcoRI used in the downstream step of gene cloning for the introduction of the HDV ribozyme sequence into the acceptor plasmid.
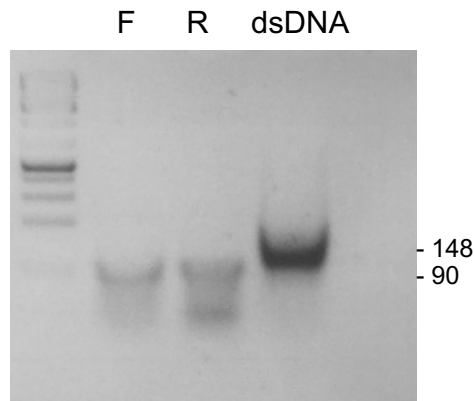
Figure 2.1 - The generation of the dsDNA was assayed an agarose gel.

The ssDNA forward (F) and reverse (R) oligos run on 2% agarose gel next to the dsDNA resulting from the primers extension reaction. From the left to the right: molecular ladder, forward primer (F), reverse primer (R) and double stranded DNA (dsDNA).

2.28 Site directed mutagenesis

To introduce the point mutation in the HDV ribozyme (RZ) necessary to generate the catalytical inactive isoform (C76U), the plasmid containing the wild type RZ was amplified using mutational oligos (PCR primers complementary to the RZ sequence on the plasmid except for a single nucleotide mismatch). The PCR resulted into the combination of wild type plasmids (RZ) and the resulting mutated plasmids (C76U) in the solution. The removal of the RZ plasmids (methylated from the bacterial amplification) was obtained via DpnI (methylation specific restriction enzyme) digestion. The PCR products (not methylated) C76U were not sensitive to DpnI endonucleolytic activity and therefore enriched in the solution after RE digestion (Scheme 2). The resulting plasmid was used to transform bacterial cells for amplification and

processing of the plasmid. The correct introduction of the point mutation was confirmed via Sanger sequencing and aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/ , Fig. 2).
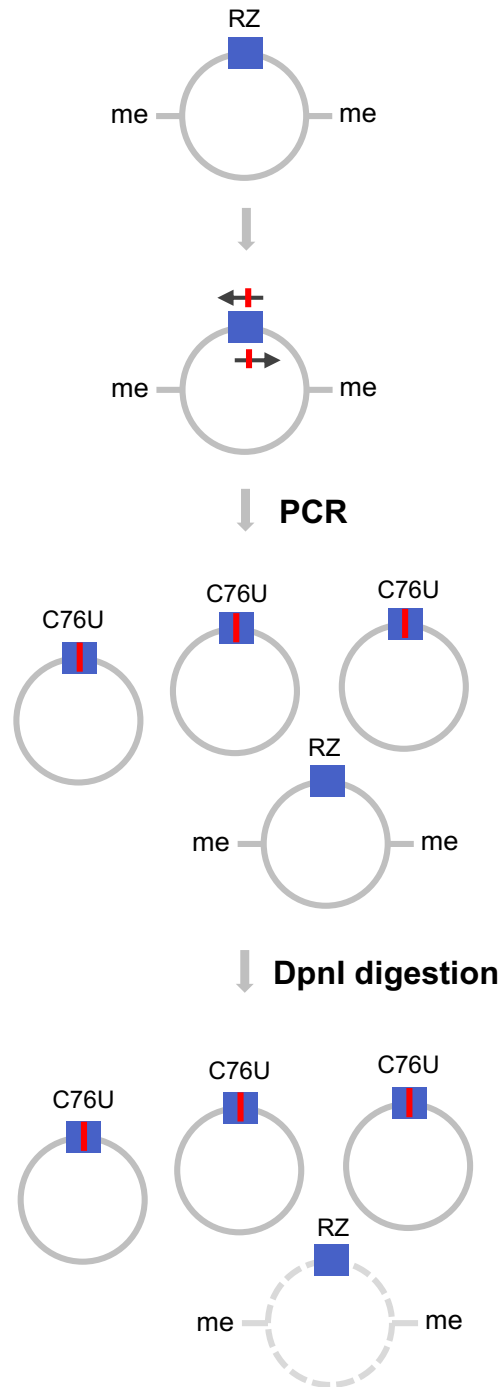
Scheme 2 - Site directed mutagenesis on RZ plasmid and mutants' selection. *The plasmid containing the wild type sequence of the HDV ribozyme (RZ) was amplified in bacteria cells (DH5α) and therefore*

methylated (me). RZ was subjected to PCR amplification using mutational oligos (black arrows)
containing the point mutation intended to be inserted (red bar) and generating the mutant C76U. The
remaining RZ plasmids were degraded into the PCR product mix RZ + C76U via DpnI digestion.

```
CLUSTAL O(1.2.4) multiple sequence alignment


hhRZ      AGGGCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGCCTGGGCAACATGCTTCGGCATGGC      60
C76U-1    AGGGCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGCCTGGGCAACATGCTTCGGCATGGT      60
C76U-2    AGGGCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGCCTGGGCAACATGCTTCGGCATGGT      60
C76U-3    AGGGCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGCCTGGGCAACATGCTTCGGCATGGT      60
          ************************************************************

hhRZ      GAATGGGACCAAAT      74
C76U-1    GAATGGGACCAAAT      74
C76U-2    GAATGGGACCAAAT      74
C76U-3    GAATGGGACCAAAT      74
          **************
```

Figure 2.2 - Clustal Omega check of the point mutation C76U.

The output from the Clustal Omega online tool for multiple alignment. The sequence obtained from Sanger sequencing of the RZ plasmid (hhRZ) was aligned with the sequence of the plasmid purified from three different mutant colonies (C76U -1, 2 or 3).

2.29 Plasmid DNA purification

For the purification of plasmid DNA from bacterial cells, the single colony was grown in 10 mL of LB media added of the specific antibiotic at 37 °C, overnight under agitation. The next day the cell suspension was processed with QIAprep Spin Miniprep Kit (Quiagen 27104) following the manufacturer's instructions.

Primers

Name	Sequence
ADH1_prom_F MP27_R	CTACATATACCGAGGTAAAGGGCTTTTCTAAGGATTCTTATGTTCTGATC- CATCGATGAATTCTCTGTCG
ADH1_prom_F MP27_F	GACATCTCAGAGACAAAGAACGGTAGCTATTACAAAGGACAGTAAAAGCA- CGTACGCTGCAGGTCGAC
25s RNA 1R (488)	TCCTCAGTCCCAGCTGGCAG
25s RNA 1F (489)	ACCGGGATTGCCTTAGTAACG
Xrn1D_kan F	AAAAATCAACACTTGTAACAACAGCAGCAACAAATATATATCAGTACGGTCGGATCCCCGGGTT AATTAA
Xrn1D_kan R	GATATACTATTAAAGTAACCTCGAATATACTTCGTTTTTAGTCGTATGTTGAATTCGAGCTCGTTT AAAC
Xrn1D test F	TTTTCAGCAGCTCTTCCGTTA
Kan-R	ATCGCGAGCCCATTTATACC
Xrn1D test-2 R	AGTGCAAAGCTACACCCTAC
CBP80-AA F	AATCCACTAGGAGAACAATTTCGAATTGGATTCAAGAAACAAAGGAAGTTCGGATCCCCGGGTT AATTAA
CBP80-AA R	TTGATTAAAAAATTAAAAAGCGGAGTGATAACGAATGTAGTCCATCCTCCGGAATTCGAGCTCGTT TAAAC
CBP80-AA test F	TGATAACGCTATTCCACACGA
CBP80-AA test R	AAAATTACGCTAGGTATGTAACATTGA
F adh1-xrRNA- test	CGTTCCCTTTCTTCCTTGTTTC
R adh1-xrRNA- test	CACCTTTTTGAGTTTCTGGGATAG
nel025c 3F	GCCACAATATAAATGGCCAAC
nel025c 3R	AGAGGCGTTAGCAAGACCAG
nel025c 2F	TGAAAGTGGTTAGAAGATTAATGTGA
nelo25c 2R	TTGGATCGCAGAGTTCTTACC
nel025c 1F	CAGTATCGTTCGTCCCGACT
nel025c 1R	TGCAGACAATTTGTTGTGCTT

nel025c pF	GATGACACCACTTGCCACAG
nel025c pR	CCTGTTCTGCCTCGTCAGTT
nel025c 1.5F	CAGAACAAGTTGTATCGAAATGA
nel025c 1.5R	CCAGGTCAAGCTACTAAAATGC
nel025c 4F	TTTCCTTATATCCACTATTGATTTCTC
nel025c 5F	CCTATAATATTCGCGGCAAAA
nel025c 5R	TGCTAAAAAGGTCATGAACGAA
nel025c 6F	AATAGGTGCCATCTGGACAC
nel025c 6R	CAATGGCATTTTAAGGCAAC
nel025c 4R	ATTTTTGAGCCCACATGCTT
t7-nel025c 2F	TAATACGACTCACTATAGGGAGATGAAAGTGGTTAGAAGATTAATGTGA
sut238 F	ATGAAGTTGCCGTTTGGTCT
sut238 R	GGCACATCACGATTGTAAAGG
sut253 F	CGAAAAGGGGTAGTTGCAGA
sut253 R	ATCACATCGCAAAGCAACAC
cut736 F	TGGCTGAAAATTGCAAAACA
cut736 R	CGTTAAAATTTCGTTGGAATCA
GAL1end F2	CTCGCGAAGAATTCACAAGAGACT
GAL1end R2	GGCAGTAAAGCTCGCTGTAGTCAT
GAL1end F1	CATATGGTTCCCGTTTGACCGGA
GAL1end R1	ATTGGCAAGGGCTTCTTTTACCT
adh1Term-kanMX6 F	AGCAACTCACAACTTATAATTCCTACACGGAGCCTGAAGAACTTCGTTAA GGCGCGCCACTTCTAAATAA
adh1Term-kanMX6 R	TTGATATGTATGATGCTACATTAAGAGATATGAAAAACTATAAATTTTTT GAATTCGAGCTCGTTTAAAC
RRP6 delta F	TAGACGAAATAGGAACAACAAACAGCTTATAAGCACCCAATAAGTGCGTTCGGATCCCCGGGTT AATTAA
RRP6 delta R	ATGAAAATTACCATAATTTATAAATAAAAAAAATACGCTTGTTTTACATAAGAATTCGAGCTCGTT TAAAC
RRP6 delta test F	GGCATCGGAAAATTTTTCAGTAATG
RRP6 delta test R	CTCCATGACACAGATATTCGATTAG

RAI1 delta F	TAATACTTCTGTAATATGGTGAAAGAATAGCGAAATATTAGACCAACATAGTGTATCCCACGGAT CCCCGGGTTAATTAA
RAI1 delta R	ATATTATTTAGATCCATACGTCGATGAGGATATGCGCAGGAAAGACATAAAGGAATATTGGAATT CGAGCTCGTTTAAAC
RAI1 delta test F	TAGGAACAGATGAGGTCTGGT
RAI1 delta test R	AAATCGGTTCGCCATGCATA
adh1 F 1	ACGGCCTTCCTTCCAGTTAC
adh1 R 1	AGGGAACGAGAACAATGACGA
adh1 F 2	CACACTGACTTGCACGCTTG
adh1 R 2	TTTTCACCCATGCCGACAAC
adh1 F 3	TCTGCTAACTTGATGGCCGG
adh1 R 3	CCTTACCTTCACCACCGTCA
adh1 F 4	TCAACCAAGTCGTCAAGTCCA
adh1 R 4	GCAAGGTAGACAAGCCGACA
adh1 F 5	AGGTCAGGTTGCTTTCTCAGG
adh1 R 5	GGGTGAAATGGGGAGCGATT
adh1 F 6	GCCCCTTACCTGAGCTTCAA
adh1 R 6	GCATGCACGTATACACTTGAGT
T7-ADH1 4R	TAATACGACTCACTATAGGGAGA-GCAAGGTAGACAAGCCGACA
rat1-AA F	AACAGCTACTCTCGGAATAACAAGCAAAGTCGGTATGACAATTCAAGAGCAAATAGGCGTCGGA TCCCCGGGTTAATTAA
rai1-AA R	TGTATGAAGATTTTATAAATTTGCGAAAACCTAAATTTACCATAAAATAAAATGCGCACGGAATT CGAGCTCGTTTAAAC
N6 R test	GCGTTGGCCGATTCATTAATGC
FMP27_pADH1_ test F	GAAATATACTTCCCATTGCATG
RZ fmp27 R	ACTGCAAATATAAGCCACTTGTACAGTAGAACATTAATCGGAAACATCAT ATTTGGTCCCATTCGCCATG
RZ crh1 F .2	ATAGCAAGAGAAAGGTTAAGATAATATAATATTCATAATAATTCAATACACGTACGCTGCAGGTC GAC

RZ crh1 R	GATAATAATGAAGAGGCACTGAGTACCGTTAGTAGGTCAAGCACTTTCAT ATTTGGTCCCATTCGCCATG
crh1 RZ test F	CAATGTTTCCGAATCGGGAATGTG
crh1 RZ test R	ACGAAGCAGTGCTAGAAGCT
RZ cut F	CCAAGCTGGCCGCTCTA
EcoRi-RibozC76-U R	CTCTTCTCCTCTAGAAGTTGTTCCGATGCAGAATTCATTTGGTCCCATTCACCATGCCGAAGCATG TTGCCCAGGCGGCGCCAGCGAGGA
XbaI-Riboz F	CTCCTATAACGGACTTCGACGCTGTTATCCTCTAGAAGGGCGGCATGGTCCCAGCCTCCTCGCTG GCGCCGCCTGGGCAACATGCTTCGG
EcoRI-Riboz R	CTCTTCTCCTCTAGAAGTTGTTCCGATGCAGAATTCATTTGGTCCCATTCGCCATGCCGAAGCATG TTGCCCAGGCGGCGCCAGCGAGGA
RZ C76-U F	GGCAACATGCTTCGGCATGGTGAATGGGACCAAATGAATTC
RZ C76-U R	GAATTCATTTGGTCCCATTCACCATGCCGAAGCATGTTGCC
RZ fmp extension F	TTAAGCCATTCAATGCGATATAAACTATAAATCCCTTTTAAAAGGGCCTAGACATCTCAGAGACA AAGAACGGTAG
RZ fmp extension R	TTATTCCTAAAAGCTTCCTTAGTAAAATTTTACAGCTCCATAAGAAAGTCACTGCAAATATAAGC CACTTGTACAG
RZ crh extension F	AACTTTCAATAATTGCCTACTCCTCTCCATATTCTACATTTCTACTAAATAGCAAGAGAAAGGTTA AGATAATATAATATTC
RZ crh extension R	CTGCAGTTGTACTGTCTGCAGTAGCAGTACTTTCGGCAGCCGCGAATGTAGATAATAATGAAGAG GCACTGAGTAC
BamHI-URA3 F	TAGCAGGTCG-GGATCC-ACAGCTTTTCAATTCAATTCATC
SacI-URA3 R	TGAGGTGTCA-GAGCTC-ACACCGCATAGGGTAATAAC
CRH1 F	GTGGTGTCGCAACATCAAGT
CRH1 R	GAGTTTGTCGAGCTGGCTTC
BIG1 F	GTTGCTTATTATGTGTGGAAGCTTTTG
BIG1 R	CACCTGGTCTACGTTACAATACTCC
HAC1 F	ACGACGCTTTTGTTGCTTCT
HAC1 R	TCAAATGAATTCAAACCTGACTG
Sc 18s rRNA F	TCTTGTGAAACTCCGTCGTG
Sc 18s rRNA R	GTACAAAGGGCAGGGACGTA

GAPDH F	GTCAGCCGCATCTTCTTTTG
GAPDH R	GCGCCCAATACGACCAAATC
DLD3 F 2	CCCATTGGATCTGCCTTCTA
DLD3 R 2	ATCTCACCGTTGGGTAGCAC
PMA1 F	CGGTTGGTGGTCTGAAAACT
PMA1 R	TGCATAGCAGCCATGAAGTC
S.p. MNH1x7 F	TGATGATCCAGAAGGGCTAAA
S.p. MNH1x7 R	CGGAGCTTAAAATTTAGGGAGA
PML39 p F	TTGATGATGAGATCAAACTACCAAA
PML39 p R	CCGAATGACTAGCGCTTGAA
PML39 1 F	TCGCTGGACAAGAATACCAA
PML39 1 R	GCCATTATGGGATTTCTTCTTG
PML39 2 F	ATGATGTCGCCGATTACACC
PML39 2 R	TCAACCTGATTTTCCCTCCA
PML39 3 F	TCGAGCATCCCTGAAGAAGT
PML39 3 R	GCCGATTAACTCCAGCAACA
PML39 4 F	TCGACAACCTCTTCAATTTCCT
PML39 4 R	AAGTTCGATGCCGAAATATG
PML39 5 F	GAATTCAACCGCAAGGTAGC
PML39 5 R	TTTCTAGCGAGAATGAAGAAAAGC
CTA1 F	CTCTACCGCGTTTTGGGTAA
CTA1 R	TAAACGCGCTGCTGTATTTG
STP2 F	GCGCTATCTCCACTTTCAGG
STP2 R	GTCCTCCTCGGTTTTGTCTG
V2-1	GATATATGGGCTGGAGCGGTCC
V2-2	GATAACCTTCGTTGCATTGACGGC
F CBP20-AA	TCAGACCAGGTTTCGATGAAGAAAGAGAAGATGATAACTACGTACCTCAG-CGGATCCCCGGGTTAATTAA
R CBP20-AA	ATATATATATATATATCTGTGTGTAGAATCTTTCTCAGATATAAATTG-GAATTCGAGCTCGTTTAAAC
F CBP20-AA-test	GCAATGGGTACTTTCAGACCAG

R CBP20-AA-test	TACATAACAGCATCATGCAAGCC
BET1_I1_F	TTCTTAAGTTAGAGTAGCGCTTTTTG
BET1_I1_R	CACGCTATAGTTAGTAAACGAAGTTCA
BET1_E2_F	AGGGGGAAACGCTTATCAAC
BET1_E2_R	CCCTAGCGCTGATGATACATT
FMP27p-A	TTCTCGTTCCCTTTCTTCCTTG
FMP27p-B	GGGATCCACTAGTTCTAGAGC
FMP27_0.1-A	GATTAATGTTCTACTGTACAAGTGG
FMP27_0.1-B	CTAGTTTAAACAGGTTAATCCATG
FMP27_0.5-A	CATATCAAGTCACTAAGATTTGATTCCT
FMP27_0.5-B	CCTACTTGAAGTCCATCCTTCAGAGG
FMP27_2-A	CATATCATCCACCCTAGGTGCTAGGTCGG
FMP27_2-B	GAGCTGACCAGACCTAACCATAGTAGCGTG
FMP27_4-A	AGATATTACTCGTTGTTCGTGCCCAG
FMP27_4-B	TCCCAAAACCCTAGTTTAACAGAAGG
FMP27_6-A	CGTACTGTTGAAATGGAACGAGGACGC
FMP27_6-B	ATCGCTTCCATACTCGTTGTATCATCAGTC
FMP27_8-A	GAGGGTCACAGATCTATTACTTGCCC
FMP27_8-B	GTTGTGAGTTGCTTCAGTGGTGAAGTG
FMP27_8.2-A	CCATGTTTCAGACCTTTATTGC
FMP27_8.2-B	TAATAACGCGGGGATCAG
FMP27_8.5-A	CCATTAAGTCAGATGACTGTTT
FMP27_8.5-B	TGCCATACACAATGTAAAGAGTA
PYK2 F	ACTACAGGGAACACGGCAAG
PYK2 R	AGACCTCGCCATTTCAACAC
ACT1 E2 F	GGTTGCTGCTTTGGTTATTG
ACT1 E2 R	CCCATACCGACCATGATACC

Yeast strain

Strain	Genotype	Source
HHY168	MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP	Euroscarf
ceg1-AA	MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP CEG1-FRB::HIS	This work
cet1-AA	MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP CET1-FRB::HIS	This work
abd1-AA	MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP ABD1-FRB::HIS	This work
cbp20-AA	MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP CBP20-FRB::HIS	This work
cbp80-AA	MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP CBP80-FRB::HIS	This work
ceg1-AA/xrn1Δ	MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP CEG1-FRB::HIS KAN::XRN1	This work
ceg1-AA/rrp6Δ	MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP CEG1-FRB::HIS KAN::RRP6	This work
ceg1-63	MATa ceg1Δ::HIS3 PGAL1-YLR454::URA3 ura3Δ [pRS315-ceg1-63]	(Jimeno-González et al., 2010)
ceg1-63/rat1-1	MATa rat1-1 ceg1Δ::HIS3 PGAL1-YLR454::URA3 ura3Δ [pRS315-ceg1-63]	(Jimeno-González et al., 2010)
fmp27-C76U	MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP CEG1-FRB::HIS URA3::RZc76u::YLR454	This work
fmp27-RZ	MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP CEG1-FRB::HIS URA3::RZ::YLR454	This work
fmp27-a	as ceg11-AA but ADH1::YLR454	This work
fmp27-short	as ceg1-AA but ADH1::YLR454[Δ1-5959]	This work
fmp27-t	as ceg1-AA but YLR454::ADH1t	This work
WT for Ceg1 mutants	MATa PGAL1-YLR454::URA3 ura3Δ	T. Jensen lab.

Ribo-Pop probes sequences

	Sequence	target	target_start	target_end	length	unique_id
1	GAGACAAGCATATGACTACTGGCAGGATCAACCAG	18S	4	38	35	18S_16338
2	TCTCAGGCTCCCTCTCCGGAATCGAACCCT	18S	374	403	30	18S_7643
3	TTGGAGCTGGAATTACCGCGGCTGCTGGCACCAGA	18S	562	596	35	18S_16896
4	ACTCCACCAACTAAGAACGGCCATGCACCACCACC	18S	1283	1317	35	18S_17617
5	GCAGGTTCACCTACGGAAACCTTGTTACGACT	18S	1800	1831	32	18S_12698
6	AGAGCTGCATTCCCAAACAACTCGACTCT	25S	274	302	29	25S_10716
7	AGACTCCTTGGTCCGTGTTTCAAGACGGGC	25S	660	689	30	25S_14581
8	GAGGGAAACTTCGGCAGGAACCAGCTACTAGATGG	25S	959	993	35	25S_32260
9	ACACTCCTTAGCGGATTCCGACTTCCATGGCCACC	25S	1265	1299	35	25S_32566
10	ACTAGAGGCTGTTCACCTTGGAGACCTGCTGCGG	25S	1900	1933	34	25S_29727
11	TGACATTCAGAGCACTGGGCAGAAATCACATTGC	25S	2276	2309	34	25S_30103
12	GCTAGATAGTAGATAGGGACAGTGGGAATCTC	25S	2419	2450	32	25S_23295
13	TAACAGATGTGCCGCCCCAGCCAAACTCCCCACCT	25S	2705	2739	35	25S_34006
14	ACCTGTCTCACGACGGTCTAAACCCAGCTCAC	25S	3047	3078	32	25S_23923
15	TCAGCAGATCGTAACAACAAGGCTACTCTACTGC	25S	3448	3481	34	25S_31275

3.   Results and Discussion

3.1 Application of the anchor away system (AA) to induce the rapid depletion of capping enzymes and cap binding complex proteins.

3.1.1 Introduction

The investigation of the function of genes as well as the discovery of new genes has been largely conducted employing a class of mutants named temperature sensitive mutants (*ts*). Such class of mutants are characterised by the substitution of one or few amino acids in the primary structure of a specific protein (missense mutation) which allows to maintain its function at the so-called permissive temperature (low temperature usually) and becomes inactive at the non-permissive temperature (high temperature), it is also possible to obtain a hypomorphic effect moving the *ts* cells at the semi-permissive temperature and therefore modulate the protein function.

In one of the first screening in *S. cerevisiae*, 400 *ts* mutants were reported to show defects in cell growth, morphology, protein and nucleic acids synthesis at non-permissive temperature (Hartwell, 1967). By the conditional activation of the mutant phenotype, *ts* mutants allow the investigation of the function of essential genes (genes whose function is required for cell viability) and the dissection of the primary effects of their silencing. The strict dependency of *ts* mutants on the temperature shift to express the mutant phenotype, poses some issues on the use of such mutations in certain conditions. The shift at non-permissive temperature can trigger the stress response mechanism causing physiological alterations that can mask the specific effects of the single protein inactivation. Moreover, the lack of structural studies makes difficult to predict the interactions of the *ts* variant with protein partners and interactors at the non-permissive temperature.

Methods involving the use of inducible promoters have been largely used to understand the gene function by the conditional appearance of specific phenotypes. One of the best characterised endogenous yeast promoters employed to control gene expression is the GAL promoter (P_{GAL}), the module controlling the transcription of the genes *GAL1*, *GAL7* and *GAL10* involved in the metabolism of the galactose (Maya et al., 2008). The expression of these genes depends on the main carbon source available in the media and on the Gal4 activator, a positive transcriptional factor that induce the gene expression in the presence of galactose. On the other hand, when glucose is available in the cell, the Gal80 repressor can exerts its function, leading to the transcriptional inhibition of the genes under the control of the P_{GAL}. The use of P_{GAL} owes its large success to the strong and quick activation of the gene under its control but the use of an energy source alternative to the glucose introduces metabolic adaptations that can represent a limiting factor for specific applications of the method (Ikushima, Zhao, & Boeke, 2015).

Other methods have been developed to induce or repress gene expression in *S. cerevisiae* engaging inducible promoters from bacteria. This is the case of the tetracycline regulator system. In this system, the synthesis of a transactivator (tTA) regulates the expression of the target gene downstream of the tetracycline operator (tetO) region in the presence of tetracycline, or most commonly its analogue doxycycline. The tetracycline/doxycycline presence can either induce (Tet-ON) or repress (Tet-OFF) the expression of the target gene according to the specific variant of the system (Garí et al., 1997).

Of most recent conception, the Auxin-Inducible Degron (AID) is the method that permits the rapid degradation of the target protein upon the treatment with the plant hormone auxin (Nishimura et al., 2009). The method has been developed from the naturally occurring auxin-mediated degradation of proteins displaying the degron sequence IAA17 during plant growth and development. Specifically, auxin mediates the binding of TIR1, part of the SCF-TIR1 E3 ubiquitin ligase complex to the IAA17 transcriptional repressor (Aid) epitope. Once bound to the target protein displaying the Aid, the SCF-TIR1 complex recruits the E2 enzyme that in turns polyubiquitylates the Aid transcriptional repressor. The

polyubiquitylation induce the proteasome-mediated degradation of the protein displaying the Aid epitope. The method, therefore, relies on the degradation of the target protein fused to the Aid tag. AID is applicable to non-plant cells expressing the fusion protein Aid-target and the plant-specific factor TIR1 (Natsume, Kiyomitsu, Saga, & Kanemaki, 2016).

The anchor-away technique (AA) is another relatively new tool developed in the yeast *S. cerevisiae* that relies on the rapamycin-induced mislocalization of the target protein from its physiological cellular compartment to another where it cannot exert its function. The conditional inactivation was developed on the mechanism of inhibition of the mammalian or mechanistic target of rapamycin (mTOR). mTOR is a protein displaying kinase activity that control a plethora of pathways involving cell growth and proliferation and it is also related to several diseases either metabolic, autoimmune or cancer (Zou et al., 2020). Rapamycin is a natural inhibitor of the mTOR activity and to exerts its inhibition interacts with the FK506-binding proteins family (FKBP) of which FKBP12 is a member. The interaction of the rapamycin with FKBP12 results in the formation of the FKBP12-rapamycin complex capable of binding the FKBP12-rapamycin binding domain (FRB domain) of mTOR (Fig. 3.1) hampering the binding of the mTOR substrates and therefore inhibiting the relative downstream pathways (Hausch et al., 2013).
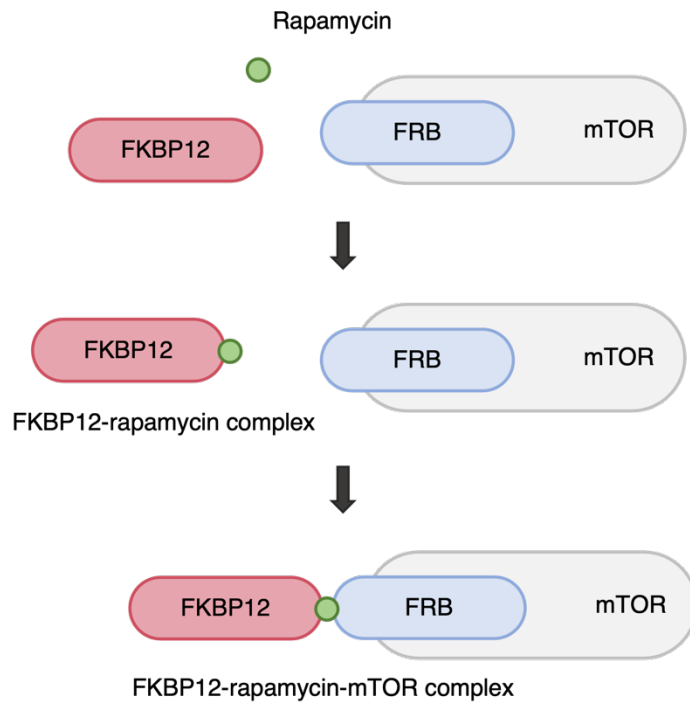
Figure 3.1 - The molecular mechanism for the inhibition of mTOR.

Rapamycin (green dot) can inhibit mTOR (grey) activities via its interaction with the protein FKBP12 (red). Rapamycin binds first to FKBP12, the FKBP12-rapamycin complex, in turn, can bind to the mTOR subunit FRB (blue) and prevent the binding of mTOR to its substrates.

In principles, the anchor away allows the rapamycin-mediated interaction of a target protein C-terminal fused to the FRB domain, with the anchor protein fused to the human FKBP12. A valid anchor protein for the depletion of nuclear protein targets is the ribosomal protein RPL13A that is abundant, and shuttles thought the nucleus to complete its maturation and eventually accumulate in the cytoplasm (Fig. 3.2). The fusion protein RPL13A-FKBP12 is currently the standard anchor in the application of the anchor away for the depletion of nuclear proteins (Ding et al., 2014; Haruki et al., 2008). To be applicable to *S. cerevisiae*, the anchor away requires other genetic modifications. To neutralise the toxicity of rapamycin the *tor1-1* mutant version of the gene *TOR1* must be introduced in the yeast genome. The depletion of the gene *FPR1* (*fpr1Δ*), the homolog of the human *FKBP12* is also necessary (Haruki et al., 2008). Therefore, the applicability of the anchor away system is subject to a series of modifications in the genomic background of the yeast cell. The anchor away permits the conditional and reversible depletion of several target proteins, in several organisms, within minutes (30-60) (Bosch, Pepperl, & Basler, 2020; Ding et al., 2014; Haruki et al., 2008).
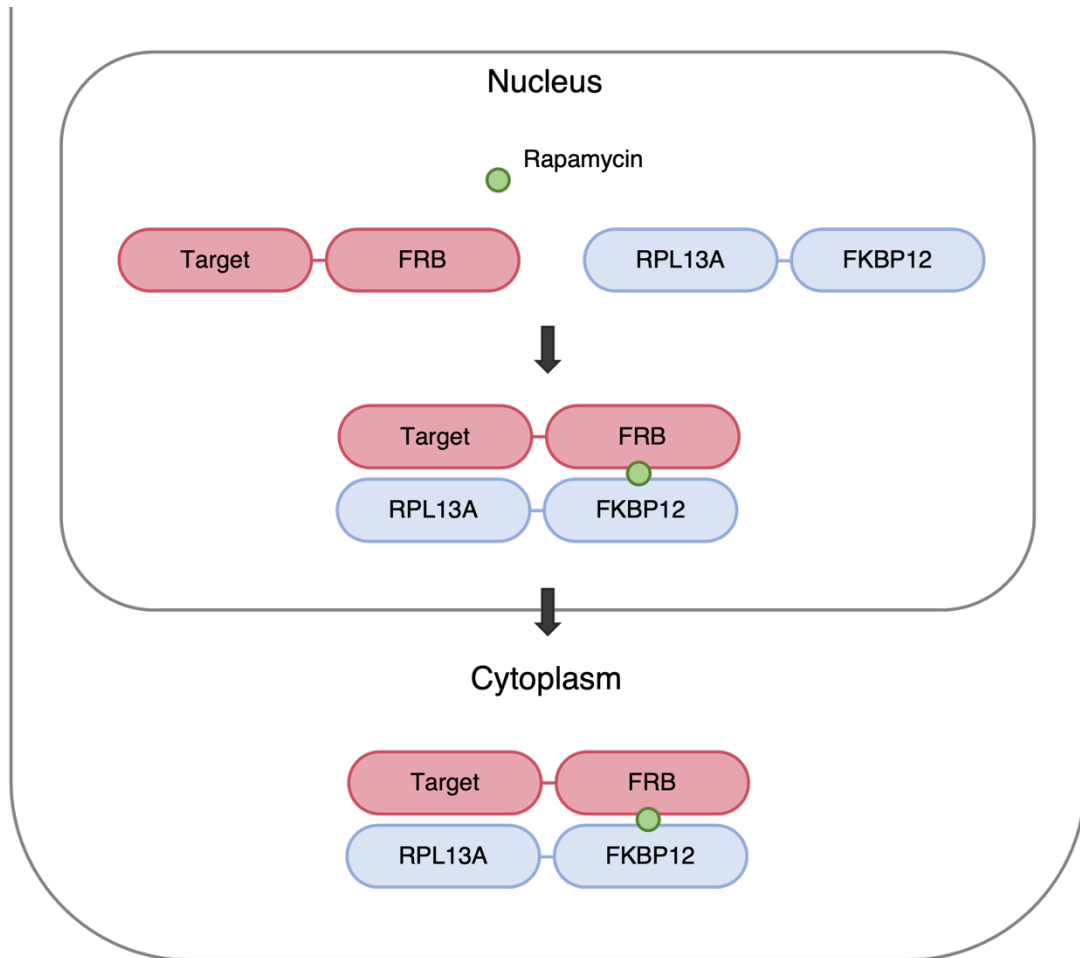
Figure 3.2 - The anchor away system for the depletion of a nuclear target protein.

Schematic cartoon showing the principles of the anchor away system. The target protein is fused with the FRB domain of mTOR (red) and the ribosomal protein RPL13A is fused to the rapamycin binding protein FKBP12 (blue). The two fusion proteins Target-FRB and RPL13A-FKBP12 can dimerise in the presence of rapamycin (green dot). The flux of the chimeric ribosomal protein from the nucleus towards the cytoplasm induces the conditional mislocalization and loss of function of the target fusion protein upon rapamycin treatment. Figure based on (Haruki et al., 2008).

Here I show the application of the anchor away system to induce the rapid depletion of the yeast capping enzymes Cet1, Ceg1 and Abd1, as well as the cap binding complex (CBC) proteins Cbp20 and Cbp80. The anchor away of the capping enzymes represents a valid method to induce the fast inactivation of the machinery responsible for the synthesis of the m^{7}G cap. This allows the investigation of the role of the RNA cap, the capping enzymes and the cap synthesis on the resulting uncapped RNA species and the effects on transcription. Moreover, the conditional depletion of the CBC proteins (Cbp20 and Cbp80) allows to differentiate the direct effect related to the lack of the capping enzymes in the nucleus from the effects related to the CBC defects.

3.1.2 The nuclear depletion of Ceg1 is sufficient to induce capping defective phenotype

In the yeast *S.cerevisiae*, the synthesis of the m^{7}G cap is accomplished in a conserved three-steps biochemical reaction catalysed by three different enzymes: the RNA triphosphatase Cet1, the guanylyltransferase Ceg1 and the methyl transferase Abd1. The deletion of each of the genes encoding for the capping enzymes is lethal confirming that each of the three genes is essential and highlighting on the lack of compensatory or redundant pathways available for the synthesis of the RNA cap (Schwer & Shuman, 1994; Shibagaki et al., 1992). For this reason, the conditional inactivation of their function is needed to investigate the effects of cap deficiency *in vivo*. Here I apply and characterize the anchor away system for the depletion of each of the capping enzymes. For the construction of each anchor away strain, I employed the yeast strain HHY168 (MATalpha tor1-1 fpr1::NAT RPL13AxFKB12::TRP) whose genetic background displays all the characteristics required for a functional nuclear protein depletion system. HHY168 harbours the *tor1-1* mutation and the deletion of the endogenous gene *FPR1* substituted by the fusion protein RPL13A-FKBP12. For this reason, HHY168 results rapamycin resistant. Compared to the parental strain W303 and the strain BY, HHY168 grows normally on YPD media or YPD media

enriched with DMSO (control condition for rapamycin treatment). On the other hand, HHY168 maintains

cell viability on media enriched with rapamycin where W303 and BY show a lethal phenotype   (Fig. 3.3)

and suitable for the introduction of the anchor away cassette in the locus of the designed target protein.
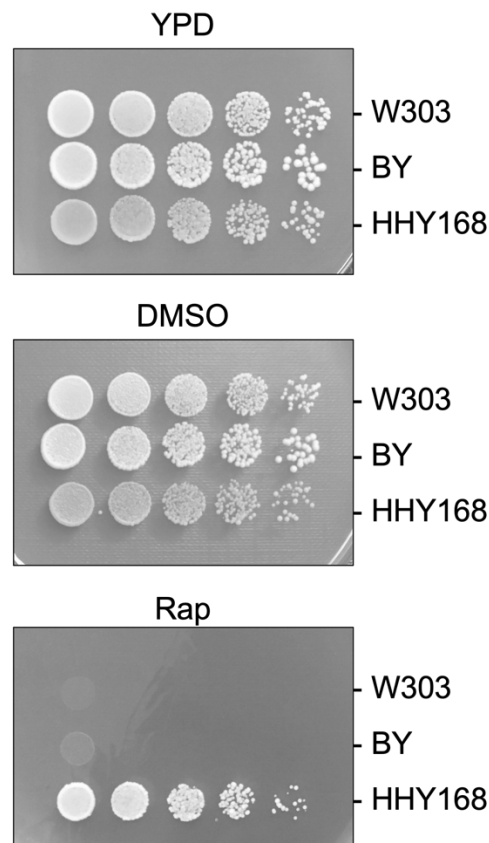
Figure 3.3 - The strain HHY168 is rapamycin resistant.

The spot growth assay of the yeast strains W303, BY and HHY168. Serial dilutions (1:1; 1:5; 1:25; 1:125; 1:625) from the liquid culture of each of the strains listed were spotted on media YPD or YPD enriched of DMSO (DMSO) or rapamycin (Rap) and grown at 30 °C. The strains W303 and BY show a lethal phenotype in the presence of rapamycin, HHY168 results rapamycin resistant.

I employed the anchor away system to induce the nuclear depletion of the three capping enzymes starting from the guanylyltransferase Ceg1. I introduced the commercially available anchor away cassette downstream of the open reading frame (ORF), upstream of the stop codon of the endogenous gene *CEG1* to generate the fusion protein Ceg1-FRB (Fig. 3.4A). The correct insertion of the cassette was verified by PCR on the genomic DNA of the resulting mutant *ceg1-AA* (Fig. 3.4B). To evaluate the effects of the genetic manipulation and the functionality of the Ceg1-AA system generated, the strain *ceg1-AA* was tested by growth spot assay and compared to the parental strain HHY168 on yeast extract peptone dextrose (YPD) media enriched of dimethyl sulfoxide (DMSO) as control or rapamycin at the final concentration of 1 μg/mL. The growth of the mutant *ceg1-AA* was not affected on media containing DMSO demonstrating that the introduction of the anchor away cassette did not affect the cellular metabolism in control conditions such to induce defects in the colony morphology, replication rate or cell viability. At the same time, I confirmed that DMSO had no effects on the growth of mutant (Fig. 3.4C). On the other hand, *ceg1-AA* showed a lethal phenotype on media containing rapamycin (Fig. 3.4C). This was consistent with the activation of the anchor away system with consequent nuclear depletion and loss of function of the capping enzyme Ceg1, essential for cell viability.
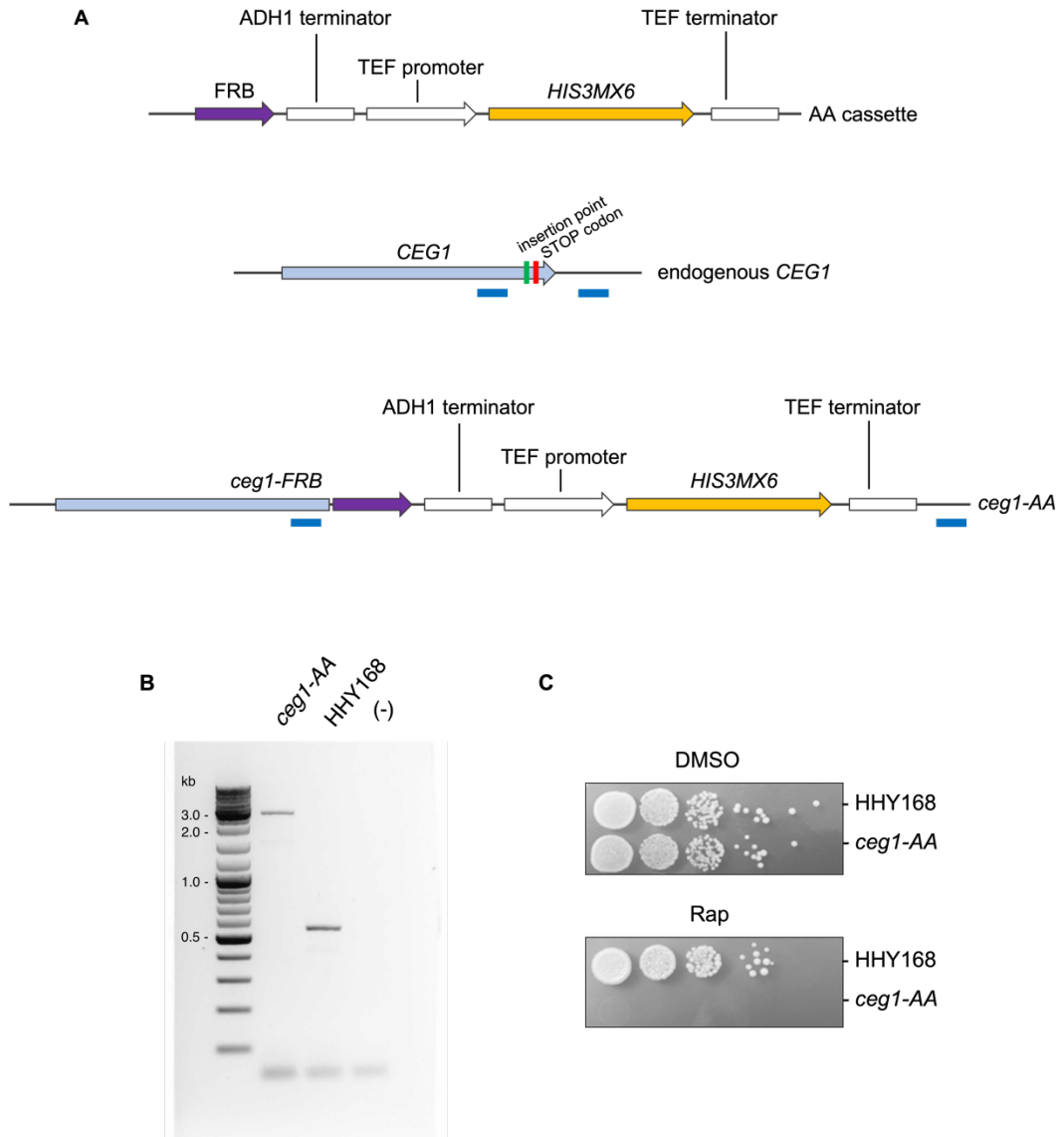
Figure 3.4 - The generation of the Ceg1-AA system.

A) Scheme showing the composition of the anchor away cassette on the top and the genetic locus for its insertion, in the middle. The anchor away cassette was introduced in the open reading frame (ORF) of CEG1 upstream of the stop codon (red bar) in the insertion point selected (green bar). At the bottom the resulting transgene ceg1-FRB. The blue horizontal bars under the genetic locus depicted, show the position of the PCR primers used to screen the correct introduction of the cassette. B) The PCR products run on 1% agarose gel showing the correct introduction of the anchor away cassette in the genomic locus targeted using the oligos located as in A. From the left to the right: molecular ladder, ceg1-AA mutant, parental strain HHY168 and PCR negative control (-). C) The spot growth assay of the strains HHY168 and ceg1-AA. Serial dilutions (1:1; 1:5; 1:25; 1:125; 1:625) of the liquid culture of each strain spotted on

To measure the time required for the Ceg1-AA system to generate a phenotype, I compared the growth of the mutant *ceg1-AA* on liquid media containing DMSO or rapamycin at the final concentration of 1 µg/mL (Fig 3.5A). Growth defects, expressed as the reduction of the exponential growth rate of *ceg1-AA*, appeared after 60 minutes of rapamycin treatment (Fig. 3.5B). This gap of 60 minutes represents the maximum time the cell can tolerate the nuclear depletion of Ceg1 remaining viable not displaying growth defects. It is reasonable to sustain that the actual Ceg1 depletion starts immediately after the cell gets in contact with the rapamycin and that the nucleus is emptied of Ceg1 quickly. At this stage, the molecular effects of the depletion can be appreciated with the minimal interference of secondary effects whose activation and accumulation reach the breakpoint after 60 minutes when the cell cannot sustain a physiological growth and proliferation.
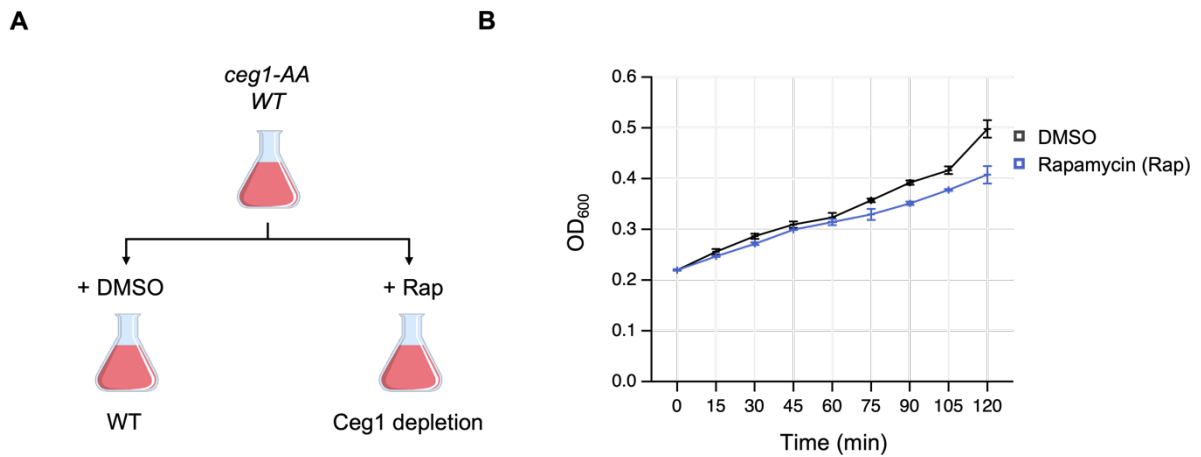
Figure 3.5 - The timing of the Ceg1-AA system.

A) Scheme of the experimental approach employed to assess the timing of the Ceg1-AA system. The liquid culture of the mutant ceg1-AA was grown overnight in YPD and then diluted to $OD_{600} = 0.2$. The culture was split in two and added respectively of DMSO (+DMSO) or rapamycin (+Rap) to the final concentration of 1 μg/mL. The cells were sampled at regular timepoints and the OD_{600} was assessed. B) The value of OD_{600} plotted as a function of time in minutes (min) for the strain ceg1-AA growing in DMSO (black) or rapamycin (Rap, blue).

Next, I investigated the molecular effects of the depletion of Ceg1 from the nucleus. First, I sought to confirm the capping defects induced by the Ceg1-AA system developed. I employed cap immunoprecipitation (IP) with the anti m^7G, m^3G cap antibody (H-20) to measure the abundance of capped mRNA levels in *ceg1-AA*. I observed that 45 minutes of rapamycin treatment compared to control (DMSO) were sufficient to reduce the capped RNA species of the 80% (Fig. 3.6). This confirmed that the Ceg1-AA system affects the m^7G cap synthesis and provides further details on the timing of the system as the cap-IP experiment defines 45 minutes of treatment as a period of time sufficient to induce strong capping defects before the rise of global secondary effects affecting the cellular growth.

Cap IP
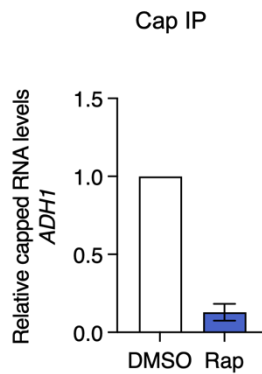
Figure 3.6 - Ceg1-AA affects the cap synthesis in 45 minutes of rapamycin treatment.

Plot showing the levels of capped ADH1 RNA after 45 min of rapamycin treatment (Rap) compared to control DMSO. Capped RNA species were immunoprecipitated (IP) with H-20 anti m⁷G cap antibody. RNA levels immunoprecipitated were quantified via RT-qPCR and normalised against the human GAPDH from spike RNA and pre-IP RNA levels. The error bar shows the standard deviation of three independent experiments.

The cap protects the RNA from 5' to 3' exonucleolytic degradation. The lack of cap has been shown to be correlated to increased RNA degradation and reduced half-life resulting therefore in lower cellular RNA levels (Furuichi, 2015; Schwer et al., 1998). To further characterise the molecular effects of the Ceg1-AA system generated, I addressed its impact on the RNA levels. I measured the RNA levels for three well characterised yeast genes, the highly expressed model genes *ADH1*, *PYK2* and *PMA1*. I quantified the RNA levels for these three genes with RT-qPCR at different timepoints, respectively 0, 45, 90 and 120 minutes of rapamycin treatment. Compared to control (DMSO) RNA levels progressively decrease over the time after rapamycin treatment for all the genes tested (Fig. 3.7). The first main drop was reported already between 0 and 45 minutes in line with the experiments on cell growth and cap-IP showed previously (Fig. 3.5B and 3.6). The sum of all the observations collected above, demonstrates that the nuclear depletion of the capping enzyme Ceg1 through the anchor away system is an efficient method to induce Ceg1 loss of function and consequent capping defects.
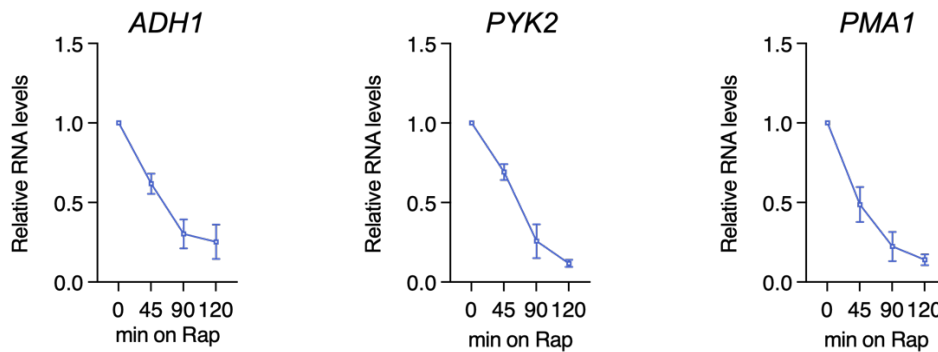
Figure 3.7 - Ceg1-AA induce the progressive decrease in the RNA levels of ADH1, PYK2 and PMA1.

Plot showing the RNA levels measured with RT-qPCR for ADH1, PYK2 and PMA1compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment. The error bars show the standard deviation of three independent experiments.

Along with the *ceg1-AA* strain, I also generated the mutant strains *cet1-AA* and *abd1-AA* with the aim of inducing the nuclear depletion of each of the capping enzymes found in *S. cerevisiae*. As described above, the anchor away cassette was stably introduced at the 3' end of the ORF of the genes *CET1* and *ABD1* (Fig. 3.8A-B) to obtain the chimeric proteins Cet1-FRB and Abd1-FRB respectively. Cell viability upon nuclear depletion of Cet1 and Abd1 was tested via growth spot assay. The strain *cet1-AA* displayed lethal phenotype in the presence of rapamycin compared to the control condition (DMSO) and the common parental strain HHY168 (Fig. 3.8C). This result confirms the essential role of Cet1 in the nucleus and the lack of enzymes excreting overlapping functions. On the contrary, the nuclear depletion of Abd1 had only minor effects on the cellular growth and maintained the viability of the *abd1-AA* strain on rapamycin (Fig. 3.8D). The cellular effects observed on cell viability caused by the depletion of Cet1 and Abd1 were confirmed molecularly via RT-qPCR on the RNA levels of *ADH1*. The conditional depletion of Cet1,

caused the progressive and continuous decrease of *ADH1* over the timepoint tested (Fig. 3.8E). On the other hand, *abd1-AA* did not show the decrease of *ADH1* over the time in the presence of rapamycin but rather showed a mild increase in the time window from 0 to 45 minutes to settle then in the interval 45 to 90 minutes and finally returned to the initial levels after 120 minutes of Abd1 nuclear depletion (Fig. 3.8F).

Figure 3.8 - Development and characterization of the anchor away system for Cet1 and Abd1.

A) *The PCR products run on 1% agarose gel showing the correct introduction of the anchor away cassette in the genomic locus targeted. From the left to the right: molecular ladder, the parental strain HHY168 (negative mutants) and three cet1-AA mutant replicates.* **B)** *The PCR products run on 1% agarose gel showing the correct introduction of the anchor away cassette in the genomic locus targeted. From the left to the right: molecular ladder, three replicates of the parental strain HHY168 (negative mutants) and the abd1-AA mutant.* **C)** *The spot growth assay of the strains HHY168 and cet1-AA. Serial dilutions (1:1; 1:5; 1:25; 1:125; 1:625) of the liquid culture of each strain spotted on YPD media enriched with DMSO or rapamycin (Rap). The strain cet1-AA is viable on DMSO but shows a lethal phenotype on Rap as*

*consequence of the activation of Cet1 nuclear depletion. **D)** The spot growth assay of the strains HHY168 and abd1-AA. Serial dilutions of the liquid culture of each strain spotted on YPD media enriched with DMSO or rapamycin (Rap). The strain abd1-AA is viable on DMSO and maintains its viability on Rap. **E)** Plot showing the RNA levels measured in cet1-AA with RT-qPCR for ADH1 compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment. The error bars show the standard deviation of three independent experiments. **F)** Plot showing the RNA levels measured in abd1-AA with RT-qPCR for ADH1 compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment. The error bars show the standard deviation of three independent experiments.*

3.1.3 The anchor-away (AA) of Cbp20 and Cbp80 permits to distinguish the effects of cap deficiency from CBC defects

An important aspect related to the various functions of the m^7G cap is that it is a necessary docking point for the binding of the so-called cap binding proteins (CBPs). CBPs like Cbp20 and Cbp80, the two components of the nuclear cap binding complex (CBC) contact the m^7G cap as soon as it gets to completion by the methylation of the N7 residue of the capping guanosine and remain associated until the RNA is exported into the cytoplasm. The CBC protects the cap from decapping and has also been shown to regulate RNA splicing, transcription termination and nuclear export (Gonatopoulos-Pournatzis & Cowling, 2014; Topisirovic et al., 2011). Once it is exported into the cytoplasm, the m^7G cap interacts with another CBP, the eukaryotic translation initiation factor 4E (eIF4E) that promotes the cap-dependent translation of the mRNA. The RNA cap and the CBC establish a strict and sustained interaction. Besides, the CBC mediates many functions of the cap including the increase of RNA stability. Therefore, it was important to isolate the effects observed upon the nuclear depletion of the capping enzymes Ceg1 and Cet1 from the consequent CBC deficiency. I generated two anchor away strains to induce the nuclear

depletion of each of the components of the nuclear CBC: Cbp20 and Cbp80 (Fig. 3.9A-B). Only Cbp20 takes direct contact with the m^7G cap, but the CBC works as a heterodimer, a feature conserved from yeast to human therefore, I generated the strains *cbp20-AA* and *cbp80-AA* to neatly dissect any CBC defect. Unlike what was observed for the capping enzymes Ceg1 and Cet1, the nuclear depletion of the CBC resulted in a viable phenotype (Fig. 3.9C) as both strains (*cbp20-AA* and *cbp80-AA*) showed only minor growth defects, visible only at the lowest concentrations (right side of the plate) compared to the parental strain HHY168. Moreover, the lack of the nuclear CBC had a minor effect on the RNA abundance as observed for the *ADH1* mRNA levels (Fig. 3.9D-E). RT-qPCR quantification of the *ADH1* mRNA revealed a mild decrease upon 90 minutes of rapamycin treatment without any further reduction up to 120 minutes. Altogether, the results shown here confirm that the CBC is not essential in *S. cerevisiae* and that, the effects induced by the depletion of the capping enzymes can be distinguished from those of the depletion of the nuclear CBC.
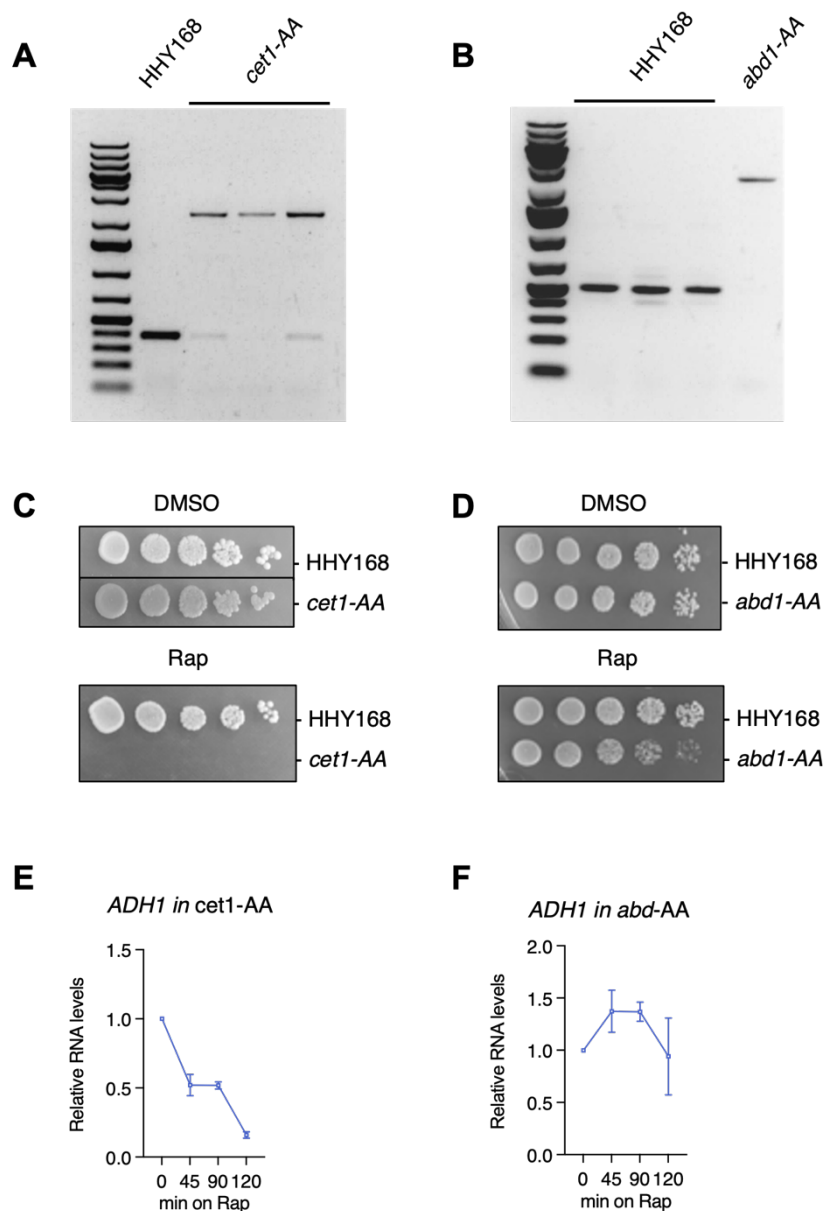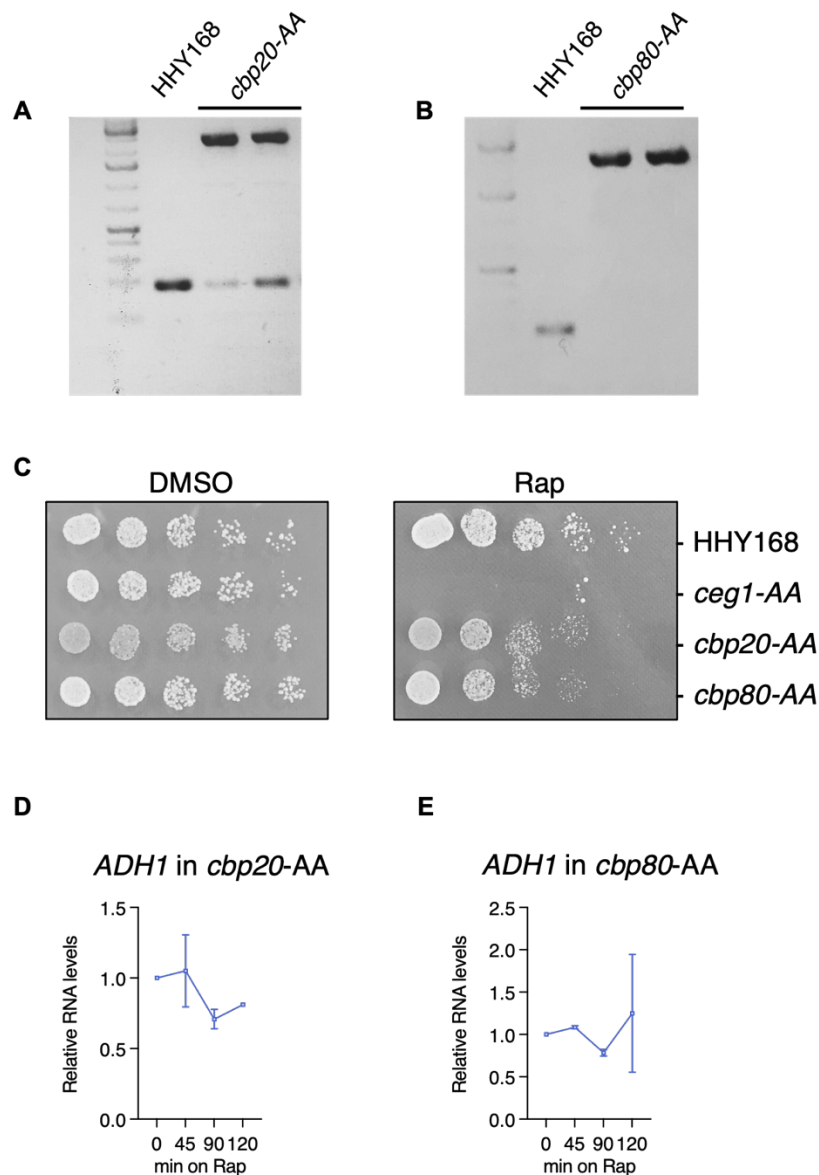
Figure 3.9 - Development and characterization of the anchor away system for Cbp20 and Cbp80.

*A) The PCR products run on 1% agarose gel showing the correct introduction of the anchor away cassette in the genomic locus targeted. From the left to the right: molecular ladder, the parental strain HHY168 (negative mutants) and two replicates of the cbp20-AA mutant. **B)** The PCR products run on 1% agarose gel showing the correct introduction of the anchor away cassette in the genomic locus targeted. From the left to the right: molecular ladder, the parental strain HHY168 (negative mutants) and two replicates of the cbp80-AA mutant. **C)** The spot growth assay of the strains HHY168, ceg1-AA, cbp20-AA and cbp80-AA. Serial dilutions of the liquid culture of each strain spotted on YPD media enriched with DMSO or rapamycin (Rap). Both strains cbp-20-AA and cbp80-AA are viable on DMSO and on Rap as consequence of the activation of CBC nuclear depletion. **D)** Plot showing the RNA levels measured in*

cbp20-AA with RT-qPCR for ADH1 compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment. The error bars show the standard deviation of three independent experiments.
***E)** Plot showing the RNA levels measured in cbp80-AA with RT-qPCR for ADH1 compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment. The error bars show the standard deviation of three independent experiments.*

3.1.4 Discussion

In this chapter I proposed the application of the anchor away system (Haruki et al., 2008) to rapidly deplete the capping enzymes from the nucleus to induce conditional cap defects and investigate the effects on the resulting uncapped RNA species and transcription.

To characterise the system, I started assessing the cellular effects induced by the nuclear depletion of the capping enzymes. The inactivation of Ceg1 achieved with the cellular relocation of the protein from the nucleus to the cytoplasm according to the anchor away system rationale, confirmed the essential role of the enzyme in the nucleus as the resulting phenotype was lethal in the presence of rapamycin. This showed that the loss of function of only one of the three capping enzymes active in *S. cerevisiae* is sufficient to induce global defects capable of affecting the cell viability (Fig. 3.4C and 3.5B). Next, I described the timing of the Ceg1-depletion system generated. The application of the anchor away aimed to avoid the detection of secondary effects caused by the prolonged inactivation of the capping enzyme. I demonstrated that 60 minutes of rapamycin treatment were sufficient to observe proliferation defects on the mutant *ceg1-AA* in liquid media (Fig. 3.5B). It is important to consider that this interval of time (60 minutes) is not the time required to initiate the depletion of Ceg1 from the nucleus. The depletion is more likely to start soon after the addition of the rapamycin to the media and reaches completion shortly. The molecular effects of the depletion, then, progress to the breaking point (60 minutes after the addition of rapamycin to the media) leading to global defects affecting cellular proliferation and viability. For this reason, a rapamycin treatment shorter than 60 minutes was required to obtain a pronounced molecular phenotype associated to the minimal onset of secondary effects. Importantly, I demonstrated the crucial role of Ceg1 in the synthesis of the m^7G cap and the lack of redundant pathways as I directly quantify 80% reduction in the abundance of capped RNA species upon 45 minutes of Ceg1 nuclear depletion (Fig. 3.6). This outcome validates the Ceg1-AA system to be appropriate to investigate capping defects as I could relate the cellular effects observed to the cap deficiency induced. I also observed a strong effect on the mRNA levels caused

by the activation of the Ceg1-AA system. *ADH1*, *PYK2* and *PMA1* mRNA levels decreased progressively starting from 45 minutes of rapamycin treatment (Fig. 3.7) defining a suitable duration for Ceg1 depletion that resulted sufficient to obtain an optimal cap-defective phenotype before the rising of the cellular decay observed at 60 minutes of rapamycin treatment. The loss of function of Ceg1 was shown to affect the abundance of the genes analysed that in turn could be caused by either the reduced RNA synthesis (Fujiwara et al., 2019; Lahudkar et al., 2014; Mandal et al., 2004; Sen et al., 2017) and/or increased RNA decay promoted by the lack of the m7G cap (Mugridge et al., 2018; Schwer et al., 1998; Xiang et al., 2009). This crucial difference was taken in account and will be discussed later in the results session.

I applied the anchor away system to the other two capping enzymes Cet1 and Abd1. In relation to the Cet1 depletion, the effects on cell viability and *ADH1* RNA levels were similar to the effects observed for the Ceg1 depletion confirming the essential role of Cet1 in the nucleus and in the synthesis of the m7G cap (compare Fig. 3.4C and 3.7 with Fig. 3.8C and E). Cet1 and Ceg1 work together in the cap synthesis assembled as two homodimers in a quaternary structure (Hausmann et al., 2003). Due to the strict interaction between Cet1 and Ceg1, together with the nature of the anchor away system that does not induce the degradation of the target protein but rather induce the relocation to a cellular compartment where the protein cannot exert its function, it is possible that by depleting one of the two interactors (Ceg1 or Cet1) from the nucleus, its partner would be co-mislocalised in the cytoplasm. It is possible, therefore, for the Ceg1-AA and Cet1-AA, to induce the cytoplasmic accumulation of the entire Cet1-Ceg1 nuclear complex. This effect is usually listed as a weakness of the anchor away system but in this case, it goes to our advantage as in this way, I obtain the complete lack of the central capping machinery and the possibility of inducing strong cap defects. As this effect, if present, would not have negatively affected the overall results, I did not investigate it further. Still, it remains an open question worth to be examined in future work.

ABD1 is an essential gene encoding for the methyltransferase Abd1 responsible for the methylation of the guanosine residue of the m^7G cap (Mao, Schwer, & Shuman, 1995). Abd1 is mainly nuclear but it was shown to accumulate in the cytoplasm in response to hypoxia (Dastidar et al., 2012). Possibly, the nuclear depletion of Abd1 had no effects on cell viability and RNA abundance (Fig. 3.8D and F). The cell viability observed upon depletion of Abd1 via anchor away system opens several considerations that have not yet been solved in the literature and exiled the focus of this work and for this reason, it opens the chance for more research on the subject. Nonetheless, hypotheses were formulated to explain such an outcome. The delocalization of Abd1 obtained via Abd1-AA could maintain the function of the methyltransferase in the cap synthesis delaying the methylation of the incomplete cap to the moment in which the mRNA was exported to the cytoplasm maintaining the stability and the functionality of the transcript not leading to any cap-deficient phenotype. The scenario where the methylation of the cap was abrogated upon Abd1-AA maintaining some unidentified cytoplasmic function of the capping enzyme related to its cytoplasmic accumulation reported during the hypoxic stress was also proposed. In this case, the lack of methylation in the cap structure would cause the failure of the CBC to bind the cap (Wong et al., 2007), a condition shown to be viable in *S.cerevisiae* (Fig. 3.9C and Fortes et al., 1999) but , as mRNA translation occurs with higher affinity for methylated cap (Cowling, 2010), a stronger phenotype was expected. The last hypothesis proposed involved the basic inapplicability of the anchor away system to Abd1 caused by incompatibility between the nature of the protein and the rationale of the method that might mask the effects of the inactivation of the enzyme.

A strict and sustained interaction is established between the m^7G cap and the complex Cbp20-Cbp80 or CBC (Aguilera, 2005). The CBC binds the methylated cap and remains associated to the nascent RNA during transcription, maturation and nuclear export. The CBC mediates many functions attributed to the cap and for this reason, capping defects directly impact on CBC binding and functionality (Gonatopoulos-Pournatzis & Cowling, 2014). I generated the mutants *cbp20-AA* and *cbp80-AA* to induce the conditional

depletion of the CBC from the nucleus. The phenotype observed in the anchor away of the CBC was viable, showing minor growth defects (Fig. 3.9C) confirming previous published results indicating the non-essential role of the CBC in yeast (Fortes et al., 1999). Consistently, only a mild and transient reduction in the abundance of *ADH1* mRNA was reported upon the nuclear depletion of each of the CBC proteins (Fig. 3.9D-E). Cbp20 and Cbp80 associate in a heterodimeric complex to be functionally associated to the cap therefore it is possible, similarly to the case proposed above for the Cet1-Ceg1 complex, that the mislocalization induced with the anchor away system could affect both partners simultaneously conferring a neat nuclear depletion of the entire CBC. The limited effects observed on cell viability and RNA stability for Cbp20-AA and Cbp80-AA, allowed me to separate the effects of the capping enzymes depletion and consequent cap defectiveness (reported here to be lethal, quick and strong) from CBC defects (compare Fig. 3.4C and t with 3.9C, D and E).

In the light of the outputs discussed here, I confirmed that the conditional depletion of the capping enzymes via the anchor away system is a suitable approach to investigate the molecular effects of capping defects on the resulting uncapped RNAs and transcription. This is the result of the anchor away capability of quickly inducing a neat and strong phenotype, avoiding the accumulation of secondary or unspecific effects.

3.2 Genome wide analysis of the mRNA levels in cap-defective cells

3.2.1 Introduction

Among the several functions the m^{7}G cap exerts, many are mediated by its either nuclear (the CBC) or cytoplasmic (eIF4E) interactors. This aspect connects the cap structure with the RNA processing like splicing and termination, nuclear export and translatability of the mRNAs (Gonatopoulos-Pournatzis & Cowling, 2014; Topisirovicet al., 2011). One of the most direct functions of the m^{7}G cap, on the other hand, resides on the protection of the capped RNA from the 5' to 3' exonucleolytic degradation (Furuichi, 2015; Gagliardi & Dziembowski, 2018). The presence of the cap at the 5' end of the RNA molecule has been shown to prevent the activity of the cellular exoribonucleases preserving the RNA stability. Capping defects, on the contrary, have been shown to increase RNA degradation leading to the reduced half-life of the cap-defective RNA (Galloway & Cowling, 2019; Schwer et al., 1998).

Although the m^{7}G cap structure as well as its synthesis and function have been largely investigated, to date a genome wide perspective on RNA stability is missing in cap-defective cells.

In this chapter I investigate the effects of the nuclear depletion of the capping enzyme Ceg1 on the transcriptome of *S. cerevisiae*.

To achieve a rapid loss of function of Ceg1 and obtain the consequent capping defects, I employed the anchor away system described and characterised in the previous chapter (Ceg1-AA). Another important aspect is that the application of the Ceg1-AA system induces the formation of the transcriptional machinery lacking the capping enzyme depleted. This also represents a new perspective on the topic as previous work investigating cap defects was mainly conducted on enzymatically inactive or temperature sensitive mutants potentially capable of interacting with the pre initiation complex and the elongating complex or subjecting the cells to a temperature shock to be inactivated at the non-permissive temperature.

Overall, I was here able to systematically analyse the abundance of cap-defective transcripts minimizing the stress sources like temperature shift or the use of alternative carbon sources to the glucose, reducing therefore the detection of secondary effects.

3.2.2 Application of the Ribo-Pop method for the depletion of the ribosomal RNA in *S. cerevisiae*

A necessary step in a variety of genome wide investigation of the RNA abundance is the depletion of contaminant transcripts that can affect the quality of the data. In our case, the depletion of the abundant and not capped RNA species was required. Specifically, ribosomal RNA (rRNA) are the most abundant RNA species transcribed in the model organism investigated here, *S. cerevisiae* (Moss & Stefanovsky, 2002; Reeder, 1998) , moreover, rRNAs are known to subtract coverage power in the sequencing step of the analysis, obscuring other classes of transcripts like the most lowly expressed mRNAs. Proprietary kits for rRNA removal are commercially available but they often suffer in the specificity of the sequence of the probes for the organism targeted and can also be prohibitive in terms of cost. I applied the Ribo-Pop method (Thompson, Kiourlappou, & Davis, 2020) for the depletion of rRNA species from the samples used in the library preparation for the RNA sequencing (RNA-seq). The method allows to generate the DNA sequences of the probes for the organism of interest following specific parameters of length and melting temperature (Tm), it is also possible to specify the target of the probes, in our case the rRNA 18S and 25S of *S. cerevisiae.* The probes are then biotinylated at the 3' end and used to hybridise the targets. The sample is then depleted of the hetero-duplexes probe/rRNA via precipitation using streptavidin coated beads. I generated 302 probes sequences targeting the 18S rRNA and selected 5 sequences to be optimal and 735 probes sequences targeting the 25S rRNA of which 10 selected for localization, Tm >70 °C and length between 29-35 nt (Fig. 3.10A-B). The probes are generated to evenly span over the target sequence
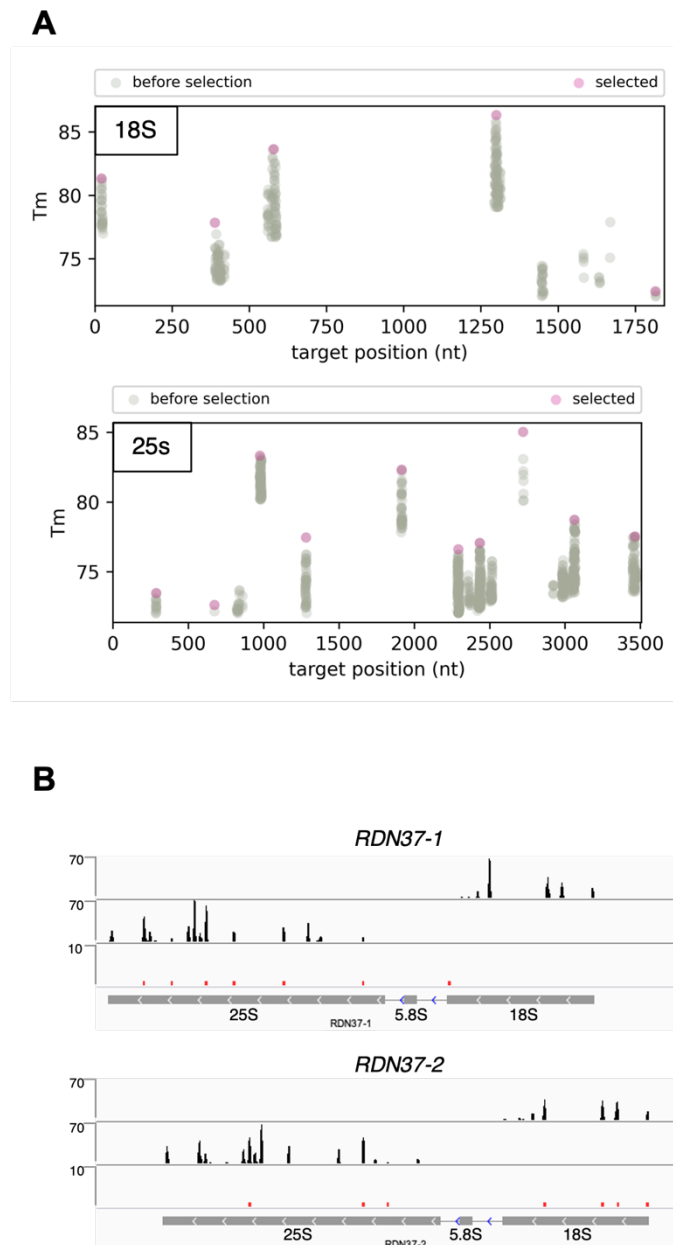
Figure 3.10 - The Ribo-Pop method for the rRNA depletion from S. cerevisiae.

*A)Plot of the probes generated for the 18S (top) and 25S (bottom) rRNA in S. cerevisiae. Each dot represents a probe generated, plotted according to target position (nt) and melting temperature (Tm) expressed in degrees Celsius (°C). The probes generated are in grey, the probes selected are in pink. **B)** The genome browser view of the probes generated with the Ribo-Pop method. In black all the probes generated and their localization on the ribosomal DNA sequences of the genes RND37-1 (top) and RDN37-2 (bottom). The red spots show the location and the number of the probes selected. Respectively 5 probes for the rRNA 18S and 10 for the 25S. Each of the genes transcribe a single polycistronic*

transcript of identical sequence that is subsequently processed to generate the rRNAs 25S, 18S and the small 5S.

Obtained the sequences, I used commercially synthesised probes biotinylated at their 3' end and tested them via RT-qPCR for their ability to deplete the rRNAs targeted from total yeast RNA samples. The quantification after incubation with streptavidin beads demonstrated that 200 pmols of the probes mix was sufficient to ablate the rRNA levels from 5 µg of total RNA compared to no probes control sample and input RNA (Fig. 3.11).
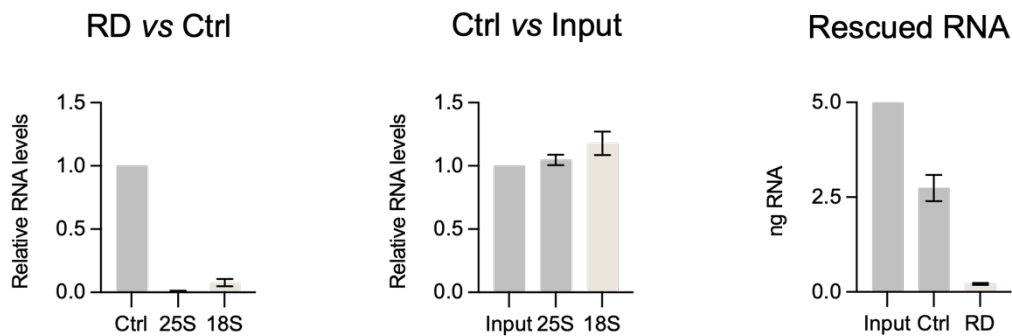
Figure 3.11 - The rRNA depletion power of the probes generated with the Ribo-pop method and commercially synthesised.

Plots showing the RNA levels measured for the rRNAs 18S and 25S depleted from the pool of 5 µg of total yeast RNA. On the left, the rRNA levels measured compared to the no probes control (Ctrl) after the incubation with streptavidin beads. In the centre, the rRNA levels for the rRNAs 18S and 25S in the no probes control (25S and 18S) after the incubation with streptavidin beads compared to the input (Input) composed of total RNA in the incubation mix and not incubated with streptavidin beads. On the right, the levels for the rRNAs 25S in the ribodepleted fraction (RD), no probes control (Ctrl) after incubation with streptavidin beads, compared to input (Input). The RT-qPCR samples were internally normalised to the non-target ACT1 gene mRNA. The error bars show the standard deviation of three independent experiments.

On the light of the results described here, I conclude that the Ribo-Pop was a suitable method to generate probes sequences capable of a strong depletion of rRNA moiety from total RNA samples and that such samples were therefore suitable for downstream library preparation for RNA-sequencing.

3.2.3 Ceg1 depletion has a different impact on mRNA abundance

The addition of the m^7G cap is essential for gene expression. The structure and the synthesis of the cap are conserved in all the eukaryotic kingdom from yeast to human, underlying its fundamental role for cell viability. Interestingly, cap analogues have recently been described in bacteria (Cahová et al., 2015; Chen et al., 2009; Jagodnik & Gourse, 2020; Kowtoniuk et al., 2009). Extensive literature is available on the structure of the m^7G cap, its synthesis and its functions. However, a global description of the effects of lack of the cap on the abundance of the mRNAs is missing at the moment this work is being produced. To fill this gap, I decided to induce the nuclear depletion of the capping enzyme Ceg1 in *S. cerevisiae* by activating the anchor away system for 45 minutes, then sequence the transcriptome and compare the global mRNA abundance to the control condition (DMSO).

The raw data generated were evaluated for their quality and reproducibility. I applied principal components analysis (PCA) to evaluate the reproducibility of the replicates of the samples sequenced (Fig. 3.12). The samples showed a strong difference between conditions as shown from the neat clustering of the replicates for the Ceg1 depletion and the relative control DMSO (PC1) and resulted sufficiently similar in the same group (PC2). This result anticipates in terms of intergroup variance, the differential accumulation of the mRNAs in Ceg1-depleted cells compared to the control.
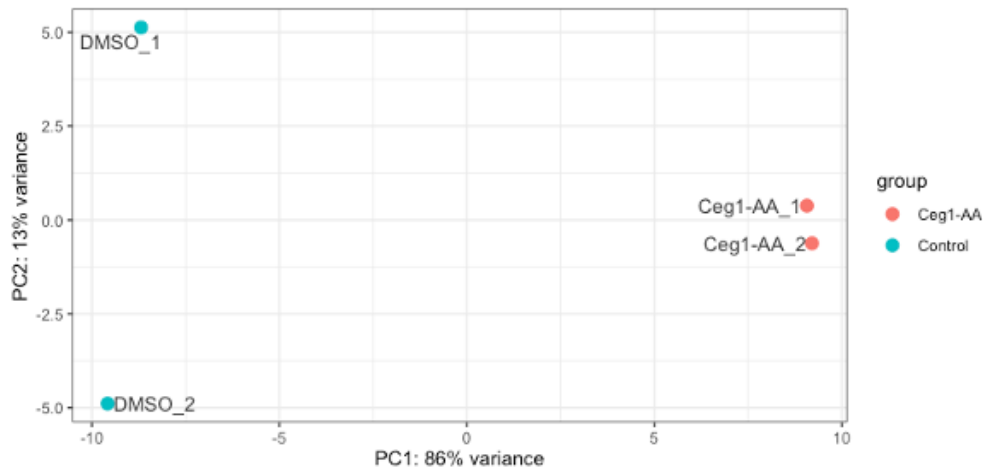
Figure 3.12 - The Principal Component Analysis (PCA) of ceg1-AA transcriptome sequenced.

The plot shows the PCA results showing the PC1 and PC2 from the samples sequenced from the ceg1-AA strain treated for 45 minutes with rapamycin Ceg1-AA_1 and 2 and the relative control DMSO_1 and 2.

Next, I measured the relative abundance of the mRNAs in Ceg1-depleted cells compared to the control condition. I selected the 6600 protein-coding genes from the yeast genome and analysed their differential expression (Fig. 3.13). I filtered differentially expressed genes (DE) displaying a log2 fold change >1 and < -1 with padj < 0.05. The filtering produced 1205 hits showing that 45 minutes of Ceg1 depletion was sufficient to at least double or half (log2 fold change ±1) the mRNA levels for the 18% of the protein-coding genes. Of the 1205 differentially expressed genes selected, 736 (61%) resulted reduced in their levels compared to the control (Dc). Unexpectedly, 469 (39%) increased their levels compared to control condition (Ic). The variation of the RNA abundance was on average -1.4 log2 fold chance for the genes showing decreased levels (Dc) and 1.5 log2 fold change for genes with increased levels (Ic) upon 45 minutes of Ceg1 depletion.

The role of the m^{7}G cap in the maintenance of the RNA stability by protecting the transcript from the exonucleolytic degradation is known and well established (Furuichi, 2015; Gagliardi & Dziembowski, 2018). An overall decrease of the RNA levels was expected upon Ceg1 depletion and the subsequent capping defects. The 736 genes whose levels resulted decreases compared to the control (Dc), represent the fraction of the uncapped mRNAs subjected to fast degradation. Interestingly, a non-negligible proportion of mRNAs (39% of the differentially expressed) resulted accumulated by the Ceg1 depletion (Ic). This interesting phenomenon has never been described before and therefore I investigated it further.
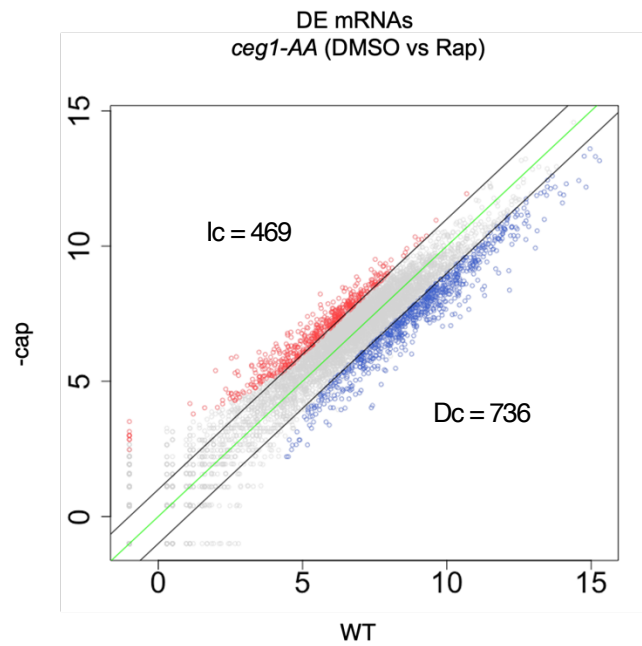
Figure 3.13 - The differential expression of mRNAs in Ceg1-depleted cells.

Plot showing the differential expression of protein-coding genes in cells depleted of Ceg1 for 45 minutes (-cap) compared to the control (WT). The results are the average of two replicates. Each dot represents one gene. The gene whose abundance was decreased with log2 fold change < -1 and padj < 0.05 upon Ceg1 depletion (Dc) are in blue. The genes whose abundance was increased with log2 fold change >1 and padj < 0.05 upon Ceg1 depletion (Ic) are in red. Genes labelled in grey were considered not differentially expressed. The green line shows the zero-change line, the black lines delimit the ±1 log2 fold change.

I followed the variation over time of the abundance of three genes selected from the fraction showing increased RNA levels in the RNA-seq dataset (Ic). I treated cells with rapamycin to trigger the nuclear depletion of Ceg1 and measured the levels of the RNA *PML39*, *STP2* and *CTA1* with RT-qPCR at 0, 45, 90 and 120 minutes of rapamycin treatment. All the genes investigated showed a progressive increase of their RNA levels over the timepoints. Specifically, for *PML39* and *STP2*, a peak in the RNA abundance was reached at 90 minutes of Ceg1 depletion with subsequent stabilization (Fig. 3.14). *CTA1* mRNAs showed instead a sustained increase up to 120 minutes of rapamycin treatment (Fig. 3.14) suggesting a possible difference in the accumulation dynamics for these transcripts.
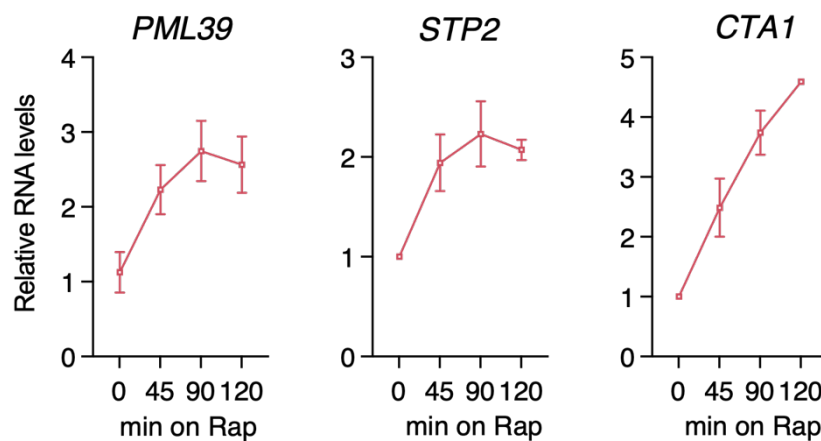
Figure 3.14 - The accumulation of PML39, STP2 and CTA1 upon Ceg1 depletion.

Plots showing the RNA levels measured via RT-qPCR for PML39, STP2 and CTA1 in ceg1-AA compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment. The error bars show the standard deviation of three independent experiments.

3.2.4 The basal expression levels correlate with the abundance of the mRNAs in Ceg1-depleted cells

The conditional depletion of the capping enzyme Ceg1 for 45 minutes is sufficient to induce the change in the RNA levels of a fraction of the protein-coding genes in *S. cerevisiae*. Specifically, in the *ceg1-AA* mutant upon Ceg1 depletion, 1205 (18%) of the 6600 protein-coding genes showed a log2 fold change >1 or < -1 compared to the control (DMSO).

As expected, most of the genes (736 of 1205, 61%) displayed decreased RNA levels (log2 fold change < -1) in line with previous observations and data from the literature showing that cap-defective mRNAs are subjected to increased exonucleolytic degradation and have an overall reduction of their half-life leading to the reduction of their cellular abundance (Furuichi, 2015; Gagliardi & Dziembowski, 2018; Galloway & Cowling, 2019; Schwer et al., 1998).

Unexpectedly, a consistent fraction of the DE mRNAs (469 of 1025, 39%) accumulated in Ceg1-depleted cells (log2 fold change >1) compared to the control (DMSO).

Due to the lack of data obtained from genome wide approaches investigating the fate of uncapped mRNAs, the accumulation observed here for some mRNA species (Ic), has not been described in the literature to date. To understand the differential expression observed for the uncapped mRNAs in Ceg1-depleted cells, I dissected the characteristics of the genes involved.

I observed an inverse correlation between the basal expression levels and the abundance of the uncapped mRNA in Ceg1-depleted cells (R^2=0.56). Data from the total ribo-depleted RNA-seq were interrogated to measure the basal expression level of the protein-coding genes in *ceg1-AA* cells under control conditions (DMSO). The value of the normalised reads counts (TPM) relative to each gene was used as a measure of the gene expression. The fraction of genes differentially expressed upon Ceg1 depletion, compared to the control (1205 genes), was divided in three sub-groups according to their expression levels. I defined, therefore, the group of highly expressed genes (High) from the 25% most expressed genes, 301 genes with a TPM range from 98.5 to 4593.4, the group of the 25% most lowly expressed genes (Low), 301 genes

with a TPM range spanning from 1.4 to 6.3 and finally I classed together the remaining 50% of the genes (Mid), 602 genes with a TPM range from 13.45 to 38.6 (Table 3.1). One gene (*DMO1*) scored 0 (zero) in the conversion to TPM for its expression levels and therefore was not considered here.

Table 3.1 - The differentially expressed protein-coding genes in ceg1-AA were classed according to their expression levels. *The table shows for each group: High, Mid and Low, the number of the genes grouped, the expression levels (TPM range), the average variation of expression upon Ceg1 depletion (Aver log2 FC) and the numner of decreased (Dc) or increased (Ic) genes.*

Group	Number of genes	TPM range	Aver log2 FC	Dc	Ic
High (Top 25%)	301	98.5 - 4593.4	-1.4	300	1
Mid (Middle 50%)	602	13.45 - 38.6	-0.6	426	176
Low (Bottom 25%)	301	1.4 - 6.29	1.5	10	291

1204*

*One gene scored TPM=0 in the conversion

The expression levels inversely correlated with the abundance of the mRNAs in Ceg1-depleted cells. The highly expressed genes, indeed, had decreased RNA levels upon Ceg1 depletion (Dc). 300 genes out of the 301 in the group High decreased upon Ceg1 depletion. On the other hand, 291 out of 301 lowly expressed genes were increased in Ceg1-depleted cells (Ic). Following the fashion indicated here, the genes from the group Mid had an overall intermediate dependency with the tendence to be decreased: 426 genes of the 602 total had negative log2 fold change in Ceg1-depleted cells (Fig. 3.15 and Table 3.1).
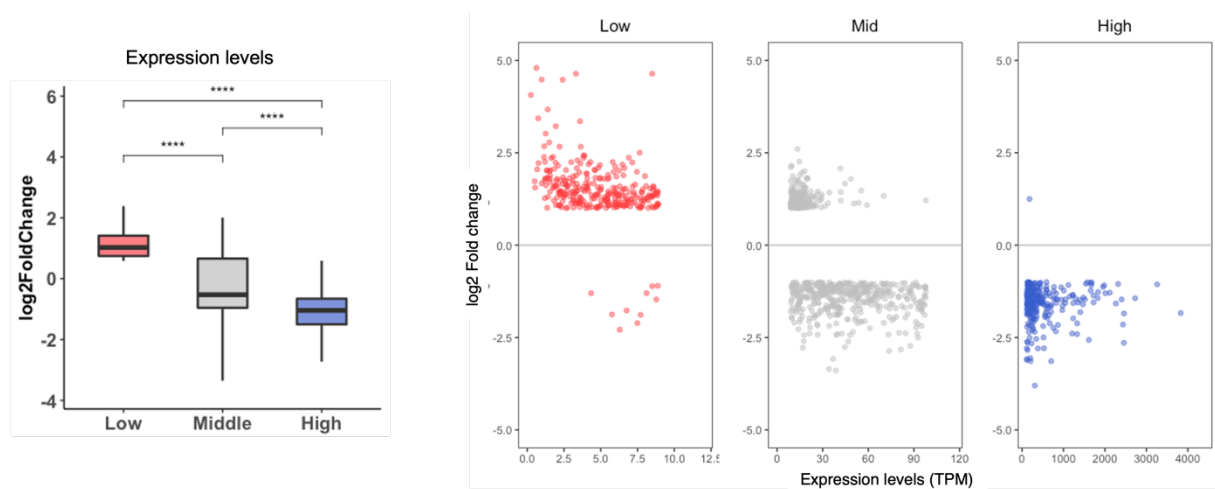
Figure 3.15 - The mRNAs levels in Ceg1-depleted cells have inverse correlation with their expression levels.

On the left, the box plot showing the log2 fold change distribution of the differentially expressed genes in ceg1-AA grouped according to their basal expression levels in High, Mid and Low. The asterisks (*) indicate the statistical significance calculated via ANOVA test (ns=p-value > 0.05; *=p-value ≤ 0.05; **=p-value ≤ 0.01; ***= p-value ≤ 0.001; ****= p-value ≤ 0.0001). On the right. The scatter plot showing the distribution at single gene levels of the differentially expressed genes in ceg1-AA according to their expression levels. Highly expressed genes (blue) show on average a decrease in Ceg1-depleted cells, on the contrary, lowly expressed genes (red) are increased upon Ceg1 depletion.

Of the three groups generated, the group High and Low showed the same average variation in terms of log2 fold change described previously for the total protein-coding genes regardless their expression levels, respectively -1.4 and 1.5 confirming the dependency of the mRNAs' abundance displayed upon Ceg1 depletion from the basal expression levels. Genes from the group Mid displayed on average a less negative variation in their abundance (average log2 fold change = -0.6) as the ratio between increased and decreased genes was less pronounced in this group compared to Low and High (Table 3.1).

I investigated other features related to the expression levels of the genes differentially expressed in Ceg1-depleted cells compared to the control (DMSO). I employed data available in the literature concerning the RNA polymerase II (Pol II) density and the transcription rate (TR) (Pelechano, Chavez, & Perez-Ortın, 2010). In line with our previous results, genes showing high Pol II density and TR, both characteristics related to highly expressed genes, were mainly decreased in Ceg1-depleted cells compared to the control (Fig. 3.16) confirming the dependency of the uncapped mRNA abundance from the basal expression levels in our system.

A

DE mRNAs in *ceg1-AA* (Rap)
RNA Pol II density (per kb)

B

DE mRNAs in *ceg1-AA* (Rap)
TR (mol/min)
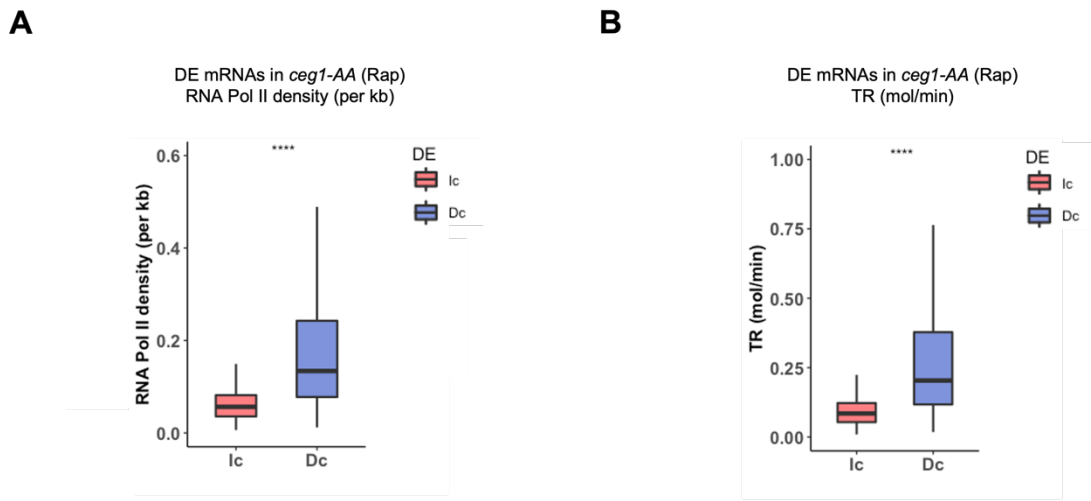
Figure 3.16 - The relation between the RNA polymerase II density and the transcription rate with the mRNA abundance in Ceg1-depleted cells.

A) Box plot showing the Pol II density of the differentially expressed genes (DE) in Ceg1-depleted cells. Decreased genes (Dc) have an overall higher Pol II density compared to the increased genes (Ic). The asterisks (*) indicate the statistical significance calculated via t-test (ns=p-value > 0.05; *=p-value ≤ 0.05; **=p-value ≤ 0.01; ***= p-value ≤ 0.001; ****= p-value ≤ 0.0001). B) The box plot showing the transcriptional rate (TR) of the differentially expressed genes in Ceg1-depleted cells. Decreased genes (Dc) have an overall higher TR compared to the increased genes (Ic). The asterisks (*) indicate the statistical significance calculated via t-test (ns=p-value > 0.05; *=p-value ≤ 0.05; **=p-value ≤ 0.01; ***= p-value ≤ 0.001; ****= p-value ≤ 0.0001).

Gene length is a structural characteristic described to be related to gene expression. Short genes, indeed have been shown to have higher expression levels (Urrutia & Hurst, 2003). I related our data about the accumulation of the mRNAs in Ceg1-depleted cells to the gene length. Genes encoding for the mRNAs that increased their levels in Ceg1-depleted cells, were on average longer than the genes encoding the mRNAs that were decreased, the average gene length was respectively 1430 for Ic and 1113 nt for Dc (Fig. 3.17A). This outcome supports our finding relating the expression levels to the mRNA abundance upon Ceg1 depletion. Longer genes, being less expressed accumulated in Ceg1-depleted cells.

Another feature investigated of the differentially expressed genes upon Ceg1 depletion was the mRNA stability in terms of RNA half-life. I compared data from the literature regarding the RNA half-life (Geisberg et al., 2014) to our data concerning the mRNA abundance in Ceg1-depleted cells. This time I found a direct correlation between RNA half-life and mRNA abundance upon Ceg1 depletion (Fig. 3.17B). Indeed, the mRNAs accumulating in Ceg1-depleted cells had on average a longer half-life compared to the genes whose abundance decreased (respectively 41.9 for Ic and 30.7 minutes for Dc).
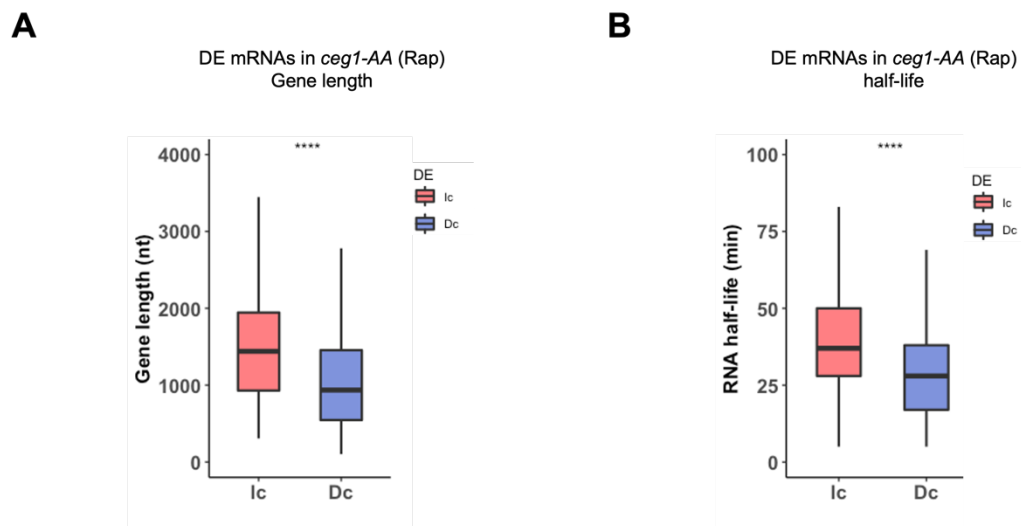
Figure 3.17 - The relation between the gene length and the half-life with the mRNA abundance in Ceg1-depleted cells.

A) Box plot showing the average gene length of the differentially expressed genes (DE) in Ceg1-depleted cells. Decreased genes (Dc) are shorter compared to the increased genes (Ic). The asterisks (*) indicate the statistical significance calculated via t-test (ns=p-value > 0.05; *=p-value ≤ 0.05; **=p-value ≤ 0.01; ***= p-value ≤ 0.001; ****= p-value ≤ 0.0001). B) The box plot showing the half-life of the differentially expressed genes in Ceg1-depleted cells. Decreased genes (Dc) have an overall shorter half-life compared to the increased genes (Ic). The asterisks (*) indicate the statistical significance calculated via t-test (ns=p-value > 0.05; *=p-value ≤ 0.05; **=p-value ≤ 0.01; ***= p-value ≤ 0.001; ****= p-value ≤ 0.0001).

Since I observed the interdependency between the basal expression levels and the log2 fold change on the uncapped mRNA in Ceg1-depleted cells, I tested whether affecting the synthesis on the highly expressed gene *ADH1* could affect the rate of its downregulation. I used 6-azauracil (6-AU) which was shown to affect Pol II processivity and elongation rate in *S. cerevisiae* (Mason & Struhl, 2005) to affect the RNA synthesis in Ceg1 depleted cells (Fig. 3.18A). First, I tested the efficiency of the 6-AU treatment on *ceg1-AA* cells before Ceg1 depletion therefore with functional capping enzymes and able to generate capped mRNAs. I grew cells in minimal synthetic media in the absence of uracil (-ura) and with raffinose as the main carbon source. Cells were then treated with 6-AU for 30 minutes or ammonium hydroxide (NH_4OH) as control. Then I added galactose to the media to induce the transcription of the inducible gene *GAL1* and followed its mRNA synthesis over time via RT-qPCR (Fig. 3.18B). I observed a slower accumulation of *GAL1* in 6-AU treated cells compared to the control. The test confirmed, therefore the effectiveness of the 6-AU treatment on the mRNA synthesis in our system. Next, I applied the 6-AU treatment to Ceg1-depleted cells (Rap) and the relative control (DMSO). I treated cells with 6-AU or NH_4OH for 30 minutes and then for each condition, cells were treated with rapamycin or DMSO (Fig. 3.18C). I monitored the *ADH1* mRNA levels over time (Fig. 3.18D) and observed a slower decrease in the uncapped mRNA levels in 6-AU-treated cells compared to the control. This data showed that the reduction in the synthesis had an impact on the degradation of the uncapped mRNA which resulted mildly stabilised by the 6-AU effects indicating a global re-adjustment of the cellular uncapped mRNA degradation dynamics.

A

Growth on RAFFINOSE

GAL1 transcription OFF

NH$_4$OH/6-AU Treatment, 30 min

Control/Slow Pol II

Growth on GALACTOSE

GAL1 transcription ON

Sampling

GAL1 mRNA quantification

B

GAL1

□ NH$_4$OH
□ 6-AU

Relative RNA levels

min on Galactose

C

SD -ura

NH$_4$OH/6-AU

DMSO/Rap

OD$_{600}$ = 1

30 min

0, 45, 90 120 min

D

ADH1

□ NH$_4$OH
□ 6-AU

Relative RNA levels

min on Rap
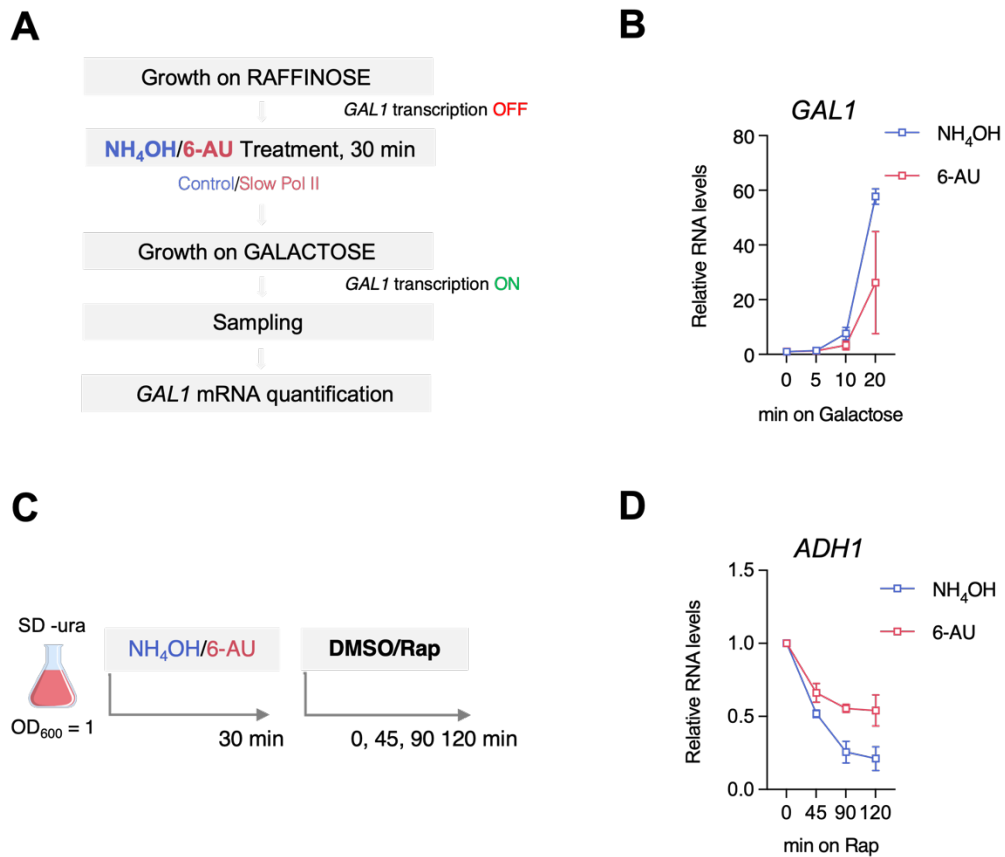
Figure 3.18 - The 6-azauracil (6-AU) treatment induce stabilization of the uncapped ADH1 mRNA in Ceg1-depleted cells.

A) The schematic of the 6-AU/ NH$_4$OH treatment performed to confirm the effects on Pol II processivity and elongation rate. B) Chase experiment for the GAL1 gene induced after 30 minutes 6-AU treatment. RT-qPCR levels shown over time after the induction of GAL1 expression. The error bars show the standard deviation of three independent experiments. C) The schematic of the 6-AU/ NH$_4$OH treatment followed by the DMSO/Rap treatment. D) RT-qPCR levels of the ADH1 mRNA in cells treated with 6-AU/ NH$_4$OH for 30 minutes followed by DMSO/Rap treatment. The error bars show the standard deviation of three independent experiments.

Although the m^{7}G cap is essential for the translation of most of the mRNAs, some of them can be translated via the so-called cap-independent translation (Iizuka et al., 1994). The rise of stress conditions induces the use of internal structures in the mRNAs employed as binding site for the ribosome subunits and trigger the translation of a specific subset of genes. Internal ribosome entry sites (IRES) have been described for viral and yeast RNAs but also in humans (Hellen & Sarnow, 2001; Peguero-Sanchez, Pardo-Lopez, & Merino, 2015). In yeast, the cap-independent translation has been shown to occur for genes displaying longer 5' untranslated region (UTR) from less structured IRES composed of A-rich stretches interacting with the polyA binding protein Pab1(Gilbert et al. 2007). To investigate the possibility in which the increased mRNAs observed in Ceg1-depleted cells (Ic) were stabilised by the sustained cap-independent translation, I started investigated the characteristics of the 5' UTRs in this subset of genes. I obtained the sequences of the 5' UTRs of the yeast genes from online repositories (https://www.yeastgenome.org) and calculated their length for the genes displaying a differential abundance upon Ceg1 depletion in the RNA-seq experiment. I did not find any correlation ($R^2 = 8$ x 10^{-6}) between the log2 fold change and the 5' UTR length in Ceg1-depleted cells (Fig. 3.19A) suggesting that this structural RNA feature is not responsible for the accumulation of the mRNAs in Ceg1-depleted cells. I also inspected the differentially expressed genes for the presence of IRES in their coding sequences. I employed data from the literature listing the genes predicted to harbour IRES (Peguero-Sanchez, Pardo-Lopez, & Merino, 2015). No specific overlapping was reported between IRES genes and the genes observed to accumulate in Ceg1-depleted cells (Fig. 3.19B). Only 12 of the 469 (2.5%) genes increased upon Ceg1 depletion were predicted to contain IRES and none of the genes increased in Xrn1 defective, Ceg1-depleted cells showed this feature. This outcome, together with the data relative to the 5' UTR length, suggested the independency between the accumulation reported for a group of genes in Ceg1-depleted cells and the cap-independent translation.
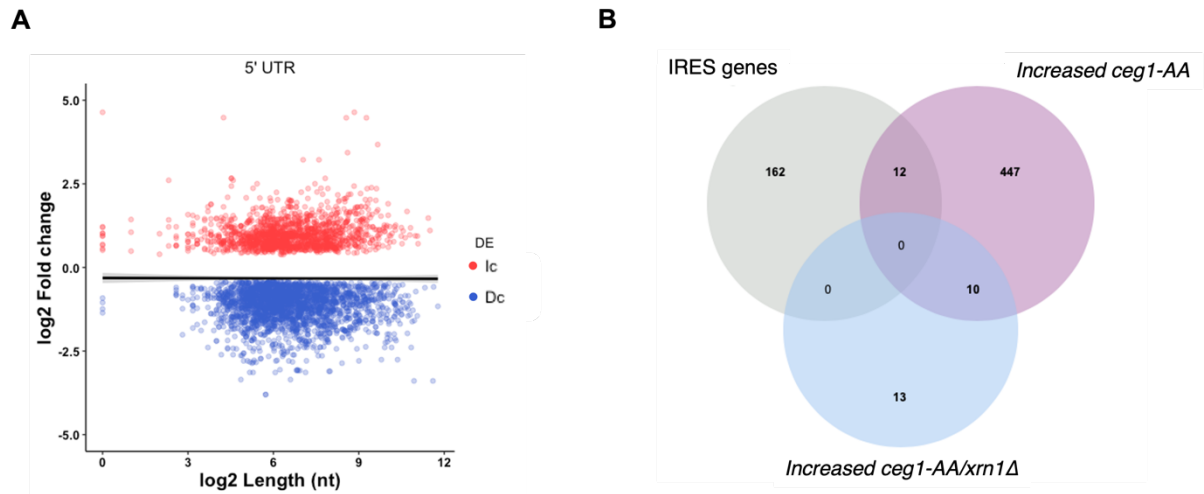
Figure 3.19 - The independency between the mRNA accumulation in Ceg1-depleted cells and the cap-independent translation.

A) The scatter plot showing the distribution of the length of the 5'UTRs (expressed as the log2 of the length in nucleotides) and the log2 fold change of the mRNAs differentially expressed in ceg1-AA cells upon Ceg1 depletion. In blue the decreased genes (Dc) and in red the increased genes (Ic). The linear relationship between the two parameters is expressed with the regression line (black). B) The Venn diagram showing the overlap between the genes listed to have internal ribosome entry site (IRES), the increased genes in ceg1-AA and the increased genes in ceg1-AA/xrn1Δ.

Finally, I investigated the functional characterization of the genes differentially expressed in Ceg1-depleted cells (DE) and I inspected their coding regions to possibly find specific motifs differentiating the Dc genes from the Ic genes upon Ceg1 depletion. Gene Ontology (GO) represents a large collection of biological information available for the community (http://geneontology.org). I performed GO enrichment analysis for the genes differentially expressed in Ceg1-depleted cells compared to the control (DMSO). Genes accumulating in Ceg1 depleted cells have the tendency to correlate at a certain degree with transcription and DNA binding (Fig. 3.20). Some genes are specifically related to Pol II transcription and are reported in Table 3.2. For genes displaying decreased levels upon Ceg1 depletion, on the other hand, I did not identify a specific enrichment pattern in the molecular function of the genes (Fig. 3.20).
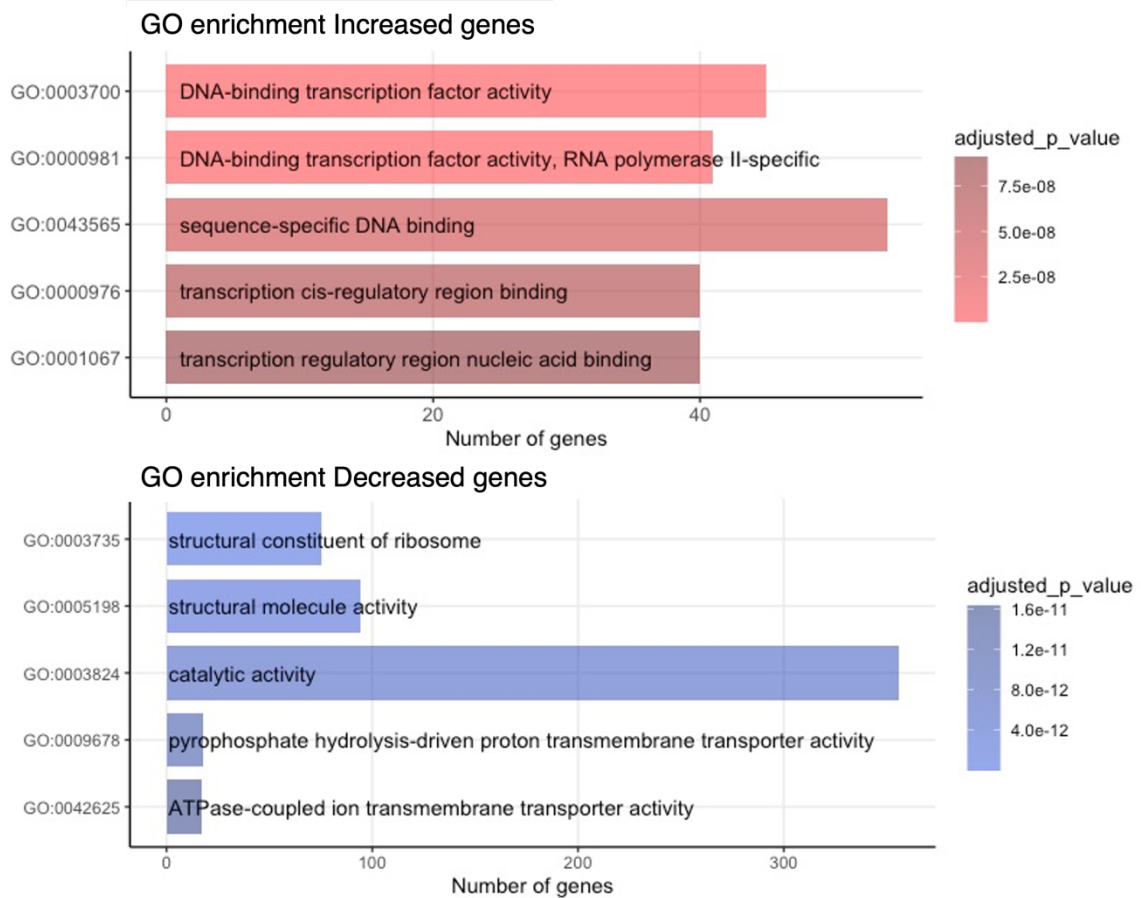
Figure 3.20 - GO enrichment analysis of the genes differentially expressed in Ceg1-depleted cells.

Top, the top 5 functional classes enriched in the subgroup of genes increased upon Ceg1 depletion in ceg1-AA strain. The bars are listed from hight to low confidence and coloured with increasing padj (adjusted_p_value) from light to dark red. Most of functional classes listed are related to DNA binding and Pol II-dependent transcription. Bottom, the top 5 functional classes enriched in the subgroup of genes decreased upon Ceg1 depletion in ceg1-AA strain. The bars are listed from hight to low confidence and coloured with increasing padj (adjusted_p_value) from light to dark blue. No specific functional pattern is described here.

Table 3.2 - DNA-binding transcription factor activity_ RNA polymerase_II-specific. *The table lists the gene ID, gene name and the functional description of the genes related to Pol II transcription increased in Ceg1-depleted cells.*

Gene ID	Name	Description
YGL162W	SUT1	Zn(II)2Cys6 family transcription factor; positively regulates sterol uptake genes under anaerobic conditions; involved in hypoxic gene expression; represses filamentation-inducing genes during vegetative growth; positively regulates mating with SUT2 by repressing expression of genes that act as mating inhibitors; repressed by STE12; relocalizes from the nucleus to the cytoplasm upon DNA replication stress; SUT1 has a paralog, SUT2, that arose from the whole genome duplication [Source:SGD;Acc:S000003130]
YPR065W	ROX1	Heme-dependent repressor of hypoxic genes; mediates aerobic transcriptional repression of hypoxia induced genes such as COX5b and CYC7; repressor function regulated through decreased promoter occupancy in response to oxidative stress; contains an HMG domain that is responsible for DNA bending activity; involved in the hyperosmotic stress resistance [Source:SGD;Acc:S000006269]
YGL096W	TOS8	Homeodomain-containing protein and putative transcription factor; found associated with chromatin; target of SBF transcription factor; induced during meiosis and under cell-damaging conditions; TOS8 has a paralog, CUP9, that arose from the whole genome duplication [Source:SGD;Acc:S000003064]
YMR136W	GAT2	Protein containing GATA family zinc finger motifs; similar to Gln3p and Dal80p; expression repressed by leucine [Source:SGD;Acc:S000004744]
YGR067C	YGR067C	Putative protein of unknown function; contains a zinc finger motif similar to that of Adr1p [Source:SGD;Acc:S000003299]
YDR259C	YAP6	Basic leucine zipper (bZIP) transcription factor; physically interacts with the Tup1-Cyc8 complex and recruits Tup1p to its targets; overexpression increases sodium and lithium tolerance; computational analysis suggests a role in regulation of expression of genes involved in carbohydrate metabolism; YAP6 has a paralog, CIN5, that arose from the whole genome duplication [Source:SGD;Acc:S000002667]
YBR240C	THI2	Transcriptional activator of thiamine biosynthetic genes; interacts with regulatory factor Thi3p to control expression of thiamine biosynthetic genes with respect to thiamine availability; acts together with Pdc2p to respond to thiaminediphosphate demand, possibly as related to carbon source availability; zinc finger protein of the Zn(II)2Cys6 type [Source:SGD;Acc:S000000444]

YKR034W	DAL80	Negative regulator of genes in multiple nitrogen degradation pathways; expression is regulated by nitrogen levels and by Gln3p; member of the GATA-binding family, forms homodimers and heterodimers with Gzf3p; DAL80 has a paralog, GZF3, that arose from the whole genome duplication [Source:SGD;Acc:S000001742]
YML027W	YOX1	Homeobox transcriptional repressor; binds to Mcm1p and to early cell cycle boxes (ECBs) in the promoters of cell cycle-regulated genes expressed in M/G1 phase; expression is cell cycle-regulated; phosphorylated by Cdc28p; relocalizes from nucleus to cytoplasm upon DNA replication stress; YOX1 has a paralog, YHP1, that arose from the whole genome duplication [Source:SGD;Acc:S000004489]
YGL071W	AFT1	Transcription factor involved in iron utilization and homeostasis; binds consensus site PyPuCACCCPu and activates transcription in response to changes in iron availability; in iron-replete conditions localization is regulated by Grx3p, Grx4p, and Fra2p, and promoter binding is negatively regulated via Grx3p-Grx4p binding; AFT1 has a paralog, AFT2, that arose from the whole genome duplication; relative distribution to the nucleus increases upon DNA replication stress [Source:SGD;Acc:S000003039]
YJL056C	ZAP1	Zinc-regulated transcription factor; binds to zinc-responsive promoters to induce transcription of certain genes in presence of zinc, represses other genes in low zinc; regulates its own transcription; contains seven zinc-finger domains [Source:SGD;Acc:S000003592]
YIL056W	VHR1	Transcriptional activator; required for the vitamin H-responsive element (VHRE) mediated induction of VHT1 (Vitamin H transporter) and BIO5 (biotin biosynthesis intermediate transporter) in response to low biotin concentrations; VHR1 has a paralog, VHR2, that arose from the whole genome duplication [Source:SGD;Acc:S000001318]
YOL108C	INO4	Transcription factor involved in phospholipid synthesis; required for derepression of inositol-choline-regulated genes involved in phospholipid synthesis; forms a complex, with Ino2p, that binds the inositol-choline-responsive element through a basic helix-loop-helix domain [Source:SGD;Acc:S000005468]
YPL038W	MET31	Zinc-finger DNA-binding transcription factor; targets strong transcriptional activator Met4p to promoters of sulfur metabolic genes; involved in transcriptional regulation of the methionine biosynthetic genes; feedforward loop controlling expression of MET32 and the lack of such a loop for MET31 may account for the differential actions of Met31p and Met32p; MET31 has a paralog, MET32, that arose from the whole genome duplication [Source:SGD;Acc:S000005959]
YOR380W	RDR1	Transcriptional repressor involved in regulating multidrug resistance; negatively regulates expression of the PDR5 gene; member of the Gal4p family of zinc cluster proteins [Source:SGD;Acc:S000005907]

YPL089C	RLM1	MADS-box transcription factor; component of the protein kinase C-mediated MAP kinase pathway involved in the maintenance of cell integrity; phosphorylated and activated by the MAP-kinase Slt2p; RLM1 has a paralog, SMP1, that arose from the whole genome duplication [Source:SGD;Acc:S000006010]
YER040W	GLN3	Transcriptional activator of genes regulated by nitrogen catabolite repression; localization and activity regulated by quality of nitrogen source and Ure2p [Source:SGD;Acc:S000000842]
YPL133C	RDS2	Transcription factor involved in regulating gluconeogenesis; also involved in the regulation of glyoxylate cycle genes; member of the zinc cluster family of proteins; confers resistance to ketoconazole [Source:SGD;Acc:S000006054]
YDR043C	NRG1	Transcriptional repressor; recruits the Cyc8p-Tup1p complex to promoters; mediates glucose repression and negatively regulates a variety of processes including filamentous growth and alkaline pH response; activated in stochastic pulses of nuclear localization in response to low glucose [Source:SGD;Acc:S000002450]
YCR018C	SRD1	Protein involved in the processing of pre-rRNA to mature rRNA; contains a C2/C2 zinc finger motif; srd1 mutation suppresses defects caused by the rrp1-1 mutation [Source:SGD;Acc:S000000611]
YBL066C	SEF1	Putative transcription factor; has homolog in Kluyveromyces lactis [Source:SGD;Acc:S000000162]
YML076C	WAR1	Homodimeric Zn2Cys6 zinc finger transcription factor; binds to a weak acid response element to induce transcription of PDR12 and FUN34, encoding an acid transporter and a putative ammonia transporter, respectively [Source:SGD;Acc:S000004541]
YOR162C	YRR1	Zn2-Cys6 zinc-finger transcription factor; activates genes involved in multidrug resistance; paralog of Yrm1p, acting on an overlapping set of target genes; YRR1 has a paralog, PDR8, that arose from the whole genome duplication [Source:SGD;Acc:S000005688]
YBR182C	SMP1	MADS-box transcription factor involved in osmotic stress response; SMP1 has a paralog, RLM1, that arose from the whole genome duplication [Source:SGD;Acc:S000000386]
YHR178W	STB5	Transcription factor; involved in regulating multidrug resistance and oxidative stress response; forms a heterodimer with Pdr1p; contains a Zn(II)2Cys6 zinc finger domain that interacts with a pleiotropic drug resistance element in vitro [Source:SGD;Acc:S000001221]
YJL089W	SIP4	C6 zinc cluster transcriptional activator; binds to the carbon source-responsive element (CSRE) of gluconeogenic genes; involved in the positive regulation of gluconeogenesis; regulated by Snf1p protein kinase; localized to the nucleus [Source:SGD;Acc:S000003625]
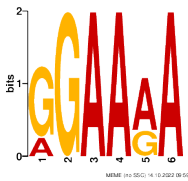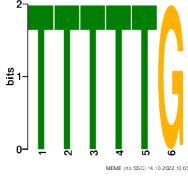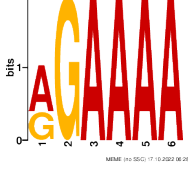
YLR098C	CHA4	DNA binding transcriptional activator; mediates serine/threonine activation of the catabolic L-serine (L-threonine) deaminase (CHA1); Zinc-finger protein with Zn[2]-Cys[6] fungal-type binuclear cluster domain [Source:SGD;Acc:S000004088]
YMR021C	MAC1	Copper-sensing transcription factor; involved in regulation of genes required for high affinity copper transport; required for regulation of yeast copper genes in response to DNA-damaging agents; undergoes changes in redox state in response to changing levels of copper or MMS [Source:SGD;Acc:S000004623]
YBL103C	RTG3	bHLH/Zip transcription factor for retrograde (RTG) and TOR pathways; forms a complex with another bHLH/Zip protein, Rtg1p, to activate the pathways; target of Hog1p [Source:SGD;Acc:S000000199]
YJL110C	GZF3	GATA zinc finger protein; negatively regulates nitrogen catabolic gene expression by competing with Gat1p for GATA site binding; function requires a repressive carbon source; dimerizes with Dal80p and binds to Tor1p; GZF3 has a paralog, DAL80, that arose from the whole genome duplication [Source:SGD;Acc:S000003646]
YLR014C	PPR1	Zinc finger transcription factor; contains a Zn(2)-Cys(6) binuclear cluster domain, positively regulates transcription of URA1, URA3, URA4, and URA10, which are involved in de novo pyrimidine biosynthesis, in response to pyrimidine starvation; activity may be modulated by interaction with Tup1p [Source:SGD;Acc:S000004004]
YOR337W	TEA1	Ty1 enhancer activator involved in Ty enhancer-mediated transcription; required for full levels of Ty enhancer-mediated transcription; C6 zinc cluster DNA-binding protein [Source:SGD;Acc:S000005864]
YPL248C	GAL4	DNA-binding transcription factor required for activating GAL genes; responds to galactose; repressed by Gal80p and activated by Gal3p [Source:SGD;Acc:S000006169]
YCR106W	RDS1	Putative zinc cluster transcription factor; involved in conferring resistance to cycloheximide [Source:SGD;Acc:S000000703]
YOR172W	YRM1	Zinc finger transcription factor involved in multidrug resistance; Zn(2)-Cys(6) zinc finger transcription factor; activates genes involved in multidrug resistance; paralog of Yrr1p, acting on an overlapping set of target genes [Source:SGD;Acc:S000005698]
YER184C	TOG1	Transcriptional activator of oleate genes; regulates genes involved in fatty acid utilization; zinc cluster protein; deletion confers sensitivity to Calcufluor white, and prevents growth on glycerol or lactate as sole carbon source [Source:SGD;Acc:S000000986]
YDR421W	ARO80	Zinc finger transcriptional activator of the Zn2Cys6 family; activates transcription of aromatic amino acid catabolic genes in the presence of aromatic amino acids [Source:SGD;Acc:S000002829]

YDL170W	UGA3	Transcriptional activator for GABA-dependent induction of GABA genes; binds to DNA elements found in the promoters of target genes and increases their expression in the presence of GABA (gamma-aminobutyrate); zinc finger transcription factor of the Zn(2)-Cys(6) binuclear cluster domain type; localized to the nucleus; examples of GABA genes include UGA1, UGA2, and UGA4 [Source:SGD;Acc:S000002329]
YER130C	COM2	Transcription factor that binds IME1 Upstream Activation Signal (UAS)ru; COM2 transcription is regulated by Haa1p, Sok2p and Zap1p transcriptional activators; may bind the IME1 promoter under all growth conditions to negatively regulate its transcription in the absence of a positive regulator that binds more effectively; repressor activity may depend on phosphorylation by PKA; C. albicans homolog (MNL1) plays a role in adaptation to stress [Source:SGD;Acc:S000000932]
YLR266C	PDR8	Transcription factor; targets include ATP-binding cassette (ABC) transporters, major facilitator superfamily transporters, and other genes involved in the pleiotropic drug resistance (PDR) phenomenon; PDR8 has a paralog, YRR1, that arose from the whole genome duplication [Source:SGD;Acc:S000004256]
YHR006W	STP2	Transcription factor; activated by proteolytic processing in response to signals from the SPS sensor system for external amino acids; activates transcription of amino acid permease genes; STP2 has a paralog, STP1, that arose from the whole genome duplication [Source:SGD;Acc:S000001048]

I investigated the coding sequence of the differentially expressed genes reported in the RNA-seq analysis searching for motifs differentiating the two groups of genes whose abundance was decreased (Dc) or increased (Ic) upon Ceg1 depletion.

Employing the MEME suite (https://meme-suite.org/meme/), I performed the analysis with the Differential Enrichment mode for the discovery of the motifs enriched in the sequences of the genes increased in Ceg1-depleted cells (Ic). I found a strong enrichment of the motif (G/A)GAA(A/G)A, E-value = 1.6 x 10^{-6} and TTTTTG, E-value = 2.0 x 10^{-9} (Table 3.3). I also analysed the RNA sequences of the Ic genes confirming the enrichment in the motif (A/G)GAAAA, E-value = 5.1 x 10^{-13}. The motif (G/A)GAA(A/G)A has not been reported in yeast but it represents the binding site of the transcriptional factors family Nuclear factor of activated T cells (NFAT) NFATc2, NFATc3 and NFATc5 (Badran et al., 2002; Yu et al., 2015). The genes reported to increase in their abundance upon Ceg1 depletion (Ic) are also enriched in the T-rich motif TTTTTG that has not been reported previously. Due to the strong end neat enrichment of the motifs listed above, it is possible that these stretches of sequences may represent a supplemental feature affecting the accumulation of the mRNAs in Ceg1-depleted cells and therefore this aspect is certainly worth of further investigation.

Table 3.3 - The motifs enriched in the coding DNA sequences and in the RNA sequences of the genes increased in Ceg1-depleted cells. *The statistical significance of the motif found are expressed in E-value. "The E-value of a motif is based on its log likelihood ratio, width, sites, the background letter frequencies (given in the command line summary), and the size of the training set. The E-value is an estimate of the expected number of motifs with the given log likelihood ratio (or higher), and with the same width and site count, that one would find in a similarly sized set of random sequences (sequences where each position is independent and letters are chosen according to the background letter frequencies). Source: https://meme-suite.org/meme/.*

Nucleic acid	Motif	Logo	E-value
DNA	G/AGAAA/GA		1.6×10^{-6}
DNA	TTTTTG		2.0×10^{-9}
RNA	A/GGAAAA		5.1×10^{-13}

3.2.5 *XRN1* deletion restores the levels of the differentially expressed uncapped mRNAs in Ceg1-depleted cells

Xrn1 is the cytoplasmic 5'-3' exoribonuclease responsible for the degradation of the mRNAs subjected to their physiological turnover (Nagarajan et al., 2013). The deletion of *XRN1*, gene encoding for the homonymous exonuclease, has been shown to accumulate non capped RNAs in temperature sensitive mutants of the capping enzyme Ceg1: *ceg1-3* and *ceg1-13* (Schwer et al., 1998).

To investigate the differential expression in our system upon Ceg1 depletion, I deleted the gene *XRN1* in the *ceg1-AA* strain generating the strain named *ceg1-AA/xrn1Δ*. The deletion was introduced to clarify the role of Xrn1 in the degradation of uncapped mRNAs in Ceg1-depleted cells. The deletion was obtained by exchanging the *XRN1* endogenous locus with the sequence encoding for the aminoglycoside phosphotransferase gene (*kanMX*) that confers resistance to the kanamycin and neomycin (Fig. 3.21A). The correct insertion of the transgene was tested by PCR (Fig. 3.21B) and the cell viability of the new mutant, checked via growth spot assay (Fig. 3.21C). The mutant *ceg1-AA/xrn1Δ* showed mild growth defects compared to the parental strains HHY168 and *ceg1-AA* in line with previous observations reported in the literature (Grousl et al., 2015; Larimer & Stevens, 1990). Importantly, the Ceg1 depletion remains lethal in the *ceg1-AA/xrn1Δ* strain (Fig. 3.21C) confirming the essential role of the m^{7}G cap for the RNA functionality.

A

kanMX

transformant DNA

XRN1

endogenous *XRN1*

kanMX

xrn1Δ

B

HHY168   *ceg1-AA*   *ceg1-AA/xrn1Δ*   (-)

C

DMSO

- HHY168
- *ceg1-AA*
- *ceg1-AA/xrn1Δ*

Rap

- HHY168
- *ceg1-AA*
- *ceg1-AA/xrn1Δ*
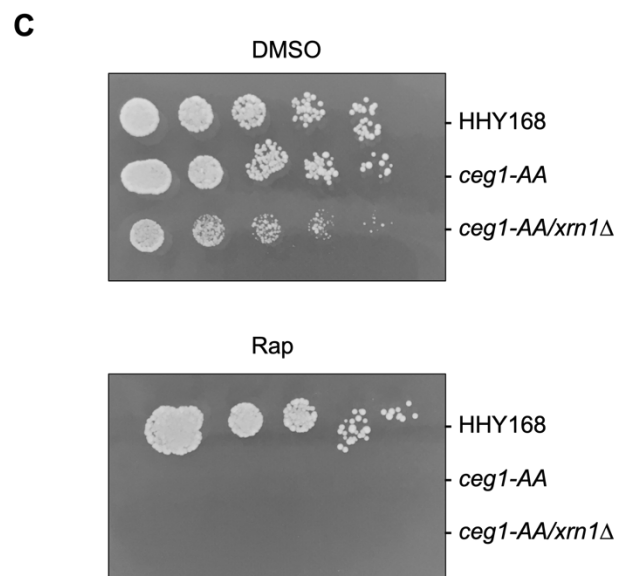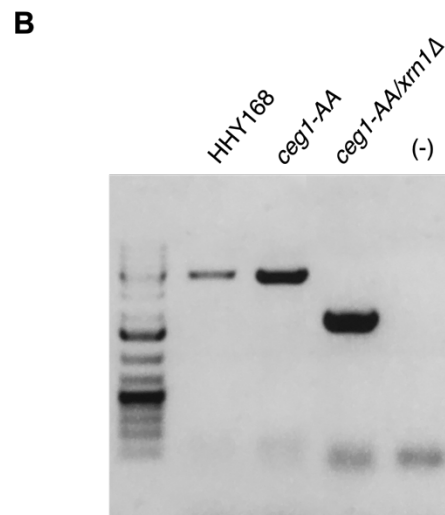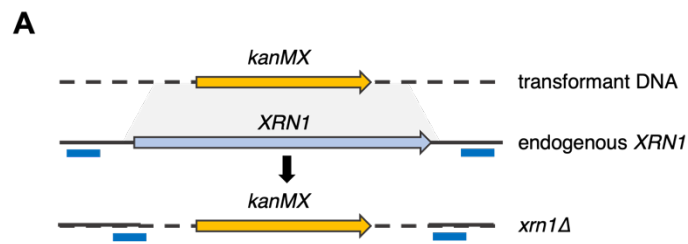
Figure 3.21 - The generation and characterization of the ceg1-AA/xrn1Δ mutant.

A) The scheme showing the gene targeting inducing the depletion of the XRN1 locus from the yeast genome and concomitant introduction of the kanMX selection marker gene. The blue horizontal bars

*under the genetic locus depicted, show the position of the PCR primers used to screen the correct introduction of the transgene. **B)** The PCR products run on 1% agarose gel showing the correct introduction of the transgene kanMX in the genomic locus targeted using the oligos located as in A. From the left to the right: molecular ladder, parental strain HHY168, the parental strain ceg1-AA, the mutant ceg1-AA/xrn1Δ and the PCR negative control (-). **C)** The spot growth assay of the strains HHY168, ceg1-AA and ceg1-AA/xrn1Δ. Serial dilutions (1:1; 1:5; 1:25; 1:125; 1:625) of the liquid culture of each strain spotted on YPD media enriched with DMSO or rapamycin (Rap). The strain ceg1-AA is viable on DMSO but shows a lethal phenotype on Rap as consequence of the activation of Ceg1 nuclear depletion. The strain ceg1-AA/xrn1Δ shows a lethal phenotype on Rap as consequence of Ceg1 depletion even in the absence of Xrn1-mediated degradation of uncapped RNAs, confirming the role of the m^7G cap for the mRNA functionality.*

Next, I sequenced the *ceg1-AA/xrn1Δ* upon 45 minutes of rapamycin treatment and the relative DMSO control. I assessed the complete knockout of the *XRN1* gene reporting the read counts normalised and expressed in transcript per million (TPM) of the resulting RNA products under control conditions (DMSO) in *ceg1-AA/xrn1Δ* compared to the strain *ceg1-AA* (Fig. 3.22). The total number of reads counted to *XRN1* was on average 713 in *ceg1-AA* and zero for both the replicates tested in the strain *ceg1-AA/xrn1Δ* confirming the deletion of the gene.
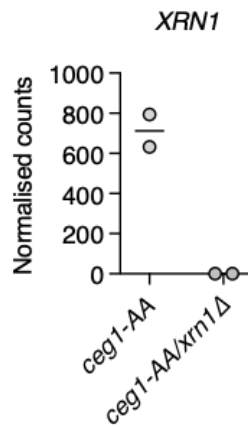
Figure 3.22 - The knockout of the XRN1 gene.

Plot showing the expression levels expressed as normalised counts (TPM) for XRN1 in ceg1-AA and ceg1-AA/xrn1Δ in control conditions (DMSO). The depletion of the single copy of the endogenous XRN1 is sufficient to completely abrogate its expression generating 0 (zero) reads in the RNA-seq experiment.

The genome wide analysis of the RNA abundance in response to Ceg1 depletion for the Xrn1-defective cells (*ceg1-AA/xrn1Δ)*, showed the recovery of the RNA levels of the genes displaying decreased levels (Dc) in the *ceg1-AA* strain upon rapamycin treatment (Fig. 3.23A). Only 18 protein-coding transcripts displayed the decreased abundance in *ceg1-AA/xrn1Δ (Dc)* compared to the 736 calculated for the *ceg1-AA* strain. Of the 18 genes decreased in *ceg1-AA/xrn1Δ*, 13 were consistently decreased in both *ceg1-AA* and *ceg1-AA/xrn1Δ* strains (Fig. 3.23B) with no specific functional profiling suggesting their degradation in Ceg1-depleted cells to be efficient even in the absence of Xrn1 activity and to be related to their structural features rather than to their function.

Unexpectedly, the inactivation of Xrn1 affected the accumulation of the genes showed to increase their abundance (Ic) in Ceg1-depleted *ceg1-AA* strain (Fig. 3.23A). Indeed, 469 protein-coding genes increased in *ceg1-AA* strain upon Ceg1 depletion compared to the 23 increased in the Ceg1-depleted *ceg1-AA/xrn1Δ* strain (Ic). Of these, 10 genes increased in both strains upon Ceg1 depletion (Fig. 3.23C).

A weak functional correlation was detected for this group as 2 genes (*ARG3* and *ARG4*) were involved in the arginine synthesis via ornithine pathway (respectively GO:0042450 and GO:0006591).

This result indicates that the accumulation of a fraction of protein-coding genes in *ceg1-AA* (Ic) can be related to the decrease of the other mRNAs (Dc).

A

DE mRNAs
ceg1-AA/xrn1Δ (DMSO vs Rap)

Ic = 23

Dc = 18

-cap -Xrn1

-Xrn1

B

Decreased (Dc) genes

ceg1-AA   *ceg1-AA/xrn1Δ*

723   13   5

RPA43
YBR08
9W
RNR1
TOS4
APT2
GAS3
POL30
GNP1
INA1
GDA1
HTA2
NOP1
PMI40

C

Increased (Ic) genes

ceg1-AA   *ceg1-AA/xrn1Δ*

459   10   13

YHR022C
YPR150W
ECM13
ARG3
YMR193C-A
SDP1
SIP4
MUM3
ARG4
YGL117W
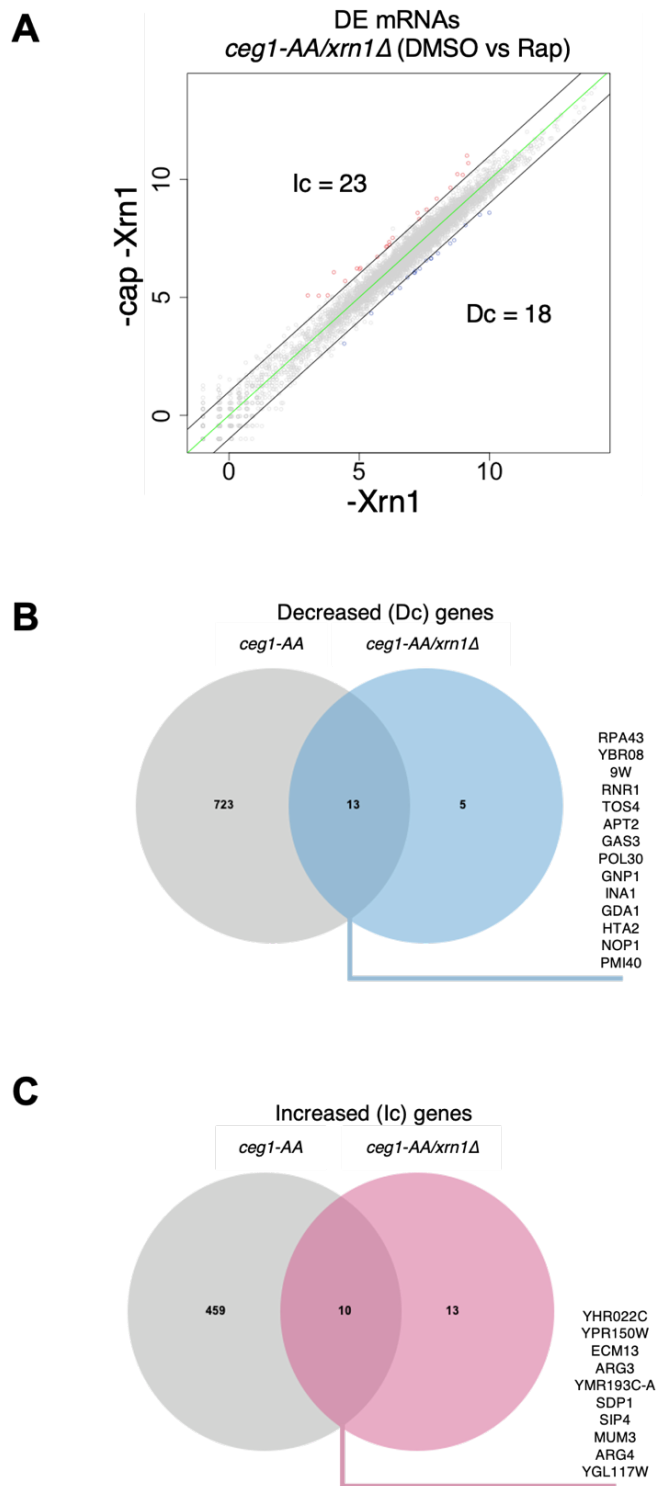
Figure 3.23 - The differential expression of mRNAs in ceg1-AA/xrn1Δ cells.

A) Plot showing the differential expression of protein-coding genes in ceg1-AA/xrn1Δ cells depleted of Ceg1 for 45 minutes (-cap -Xrn1) compared to the control (-Xrn1). The results are the average of two replicates. Each dot represents one gene. The gene whose abundance was decreased with log2 fold change < -1 and padj < 0.05 upon Ceg1 depletion (Dc) are in blue. The genes whose abundance was increased with log2 fold change >1 and padj < 0.05 upon Ceg1 depletion (Ic) are in red. Genes labelled in grey were considered not differentially expressed. The green line shows the zero-change line and the black lines the ±1 log2 fold change. The depletion of Xrn1 maintain the stability of uncapped mRNAs and abolishes the accumulation of the transcripts differentially expressed in ceg1-AA cells. B) Venn diagram showing the overlap between the genes decreased in ceg1-AA and ceg1-AA/xrn1Δ cells upon Ceg1 depletion. The overlapping genes are listed. C) Venn diagram showing the overlap between the genes increased in ceg1-AA and ceg1-AA/xrn1Δ cells upon Ceg1 depletion. The overlapping genes are listed.

To obtain a more direct comparison, I labelled the genes differentially expressed in Ceg1-depleted cells on the plot showing the differential expression in the *ceg1-AA/xrn1Δ* cells in the presence of active or depleted Ceg1 (Fig. 3.24). As expected, the differentially expressed genes reported for the Ceg1-depleted cells in the presence of the functional Xrn1 were distributed within the log2 fold change limit (±1) to be considered differentially expressed. Nevertheless, it is interesting to note how the genes showed the tendency to distribute according to their differential expression observed in Ceg1-depleted cells with the functional Xrn1. This result suggested that the introduction of the capping defects had a minimal impact on the abundance of the mRNAs in the presence of inactive Xrn1 but the tendency to the differential expression resulted in line with that of Ceg1-depleted cells.
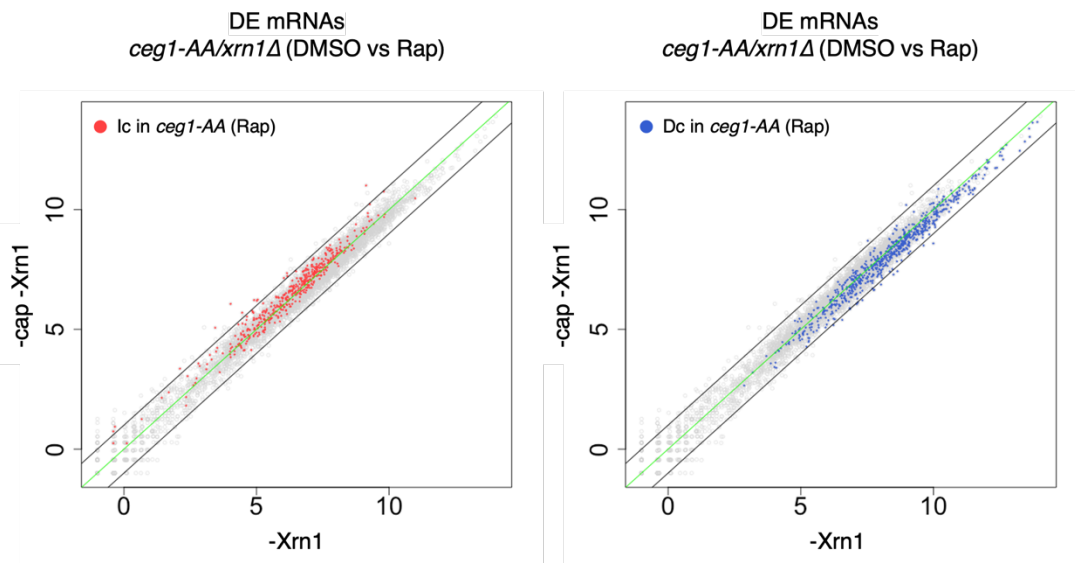
Figure 3.24 - Differentially expressed (DE) genes in Ceg1-depleted cells showed the same tendency in Ceg1-depleted ceg1-AA/xrn1Δ cells.

Genes differentially expressed (DE) in Ceg1-depleted cells with a functional Xrn1 are labelled on the plot showing the differential expression in xrn1Δ cells upon Ceg1 depletion (-cap -Xrn1) compared to control (-Xrn1). Increased (Ic) genes in Ceg1-depleted cells are labelled in red and accumulate on the positive side of the zero-change line (green). Decreased (Dc) genes in Ceg1-depleted cells are labelled in blue and accumulate on the negative side of the zero-change line. The results are the average of two replicates. Each dot represents one gene. The gene whose abundance was decreased with log2 fold change < -1 and padj < 0.05 upon Ceg1 depletion (Dc) are in blue. The genes whose abundance was increased with log2 fold change >1 and padj < 0.05 upon Ceg1 depletion (Ic) are in red. Genes labelled in grey were considered not differentially expressed. The green line shows the zero-change line, the black lines delimit the ±1 log2 fold change.

I also assayed the stabilization of the mRNAs in the Ceg1-depleted *ceg1-AA/xrn1Δ* strain over the time. I tested the RNA levels for three genes: *ADH1*, *PYK2* and *PMA1* whose RNA levels were shown to decrease progressively up to 120 minutes of rapamycin treatment in *ceg1-AA*. RT-qPCR analysis showed the stabilization at single gene level of all three genes investigated (Fig. 3.25). The uncapped mRNAs

remain stable up to 120 minutes of Ceg1 depletion displaying minor fluctuations in the RNA levels but not the tendency to drop.
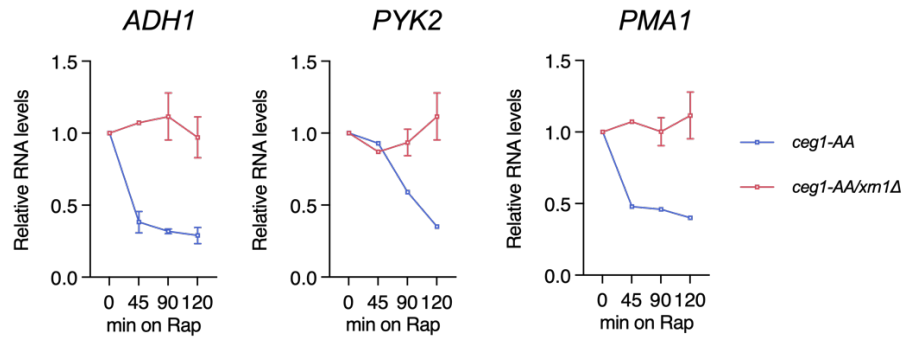
Figure 3.25 - The deletion of the gene XRN1 stabilises the RNA levels of ADH1, PYK2 and PMA1 in Ceg1-depleted cells.

Plots showing the RNA levels measured with RT-qPCR for ADH1, PYK2 and PMA1 compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). The RNA levels upon Ceg1 depletion in ceg1-AA and ceg1-AA/xrn1Δ are shown. The depletion of XRN1 abolishes the degradation of the uncapped mRNAs in Ceg1-depleted cells. Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment. The error bars show the standard deviation of three independent experiments.

Taken together, the results presented here indicate Xrn1 as the main exonuclease responsible for the control of the expression of the uncapped mRNAs generated in Ceg1-depleted cells. The inactivation of the Xrn1 enzymatic activity via gene knockout, rescued the RNA levels of the 98% of the mRNAs whose levels were decreased in Ceg1-depleted *ceg1-AA* cells (Dc). The Xrn1 inactivation abolished the accumulation of the 95% of the genes whose RNA levels increased upon Ceg1 depletion in the *ceg1-AA* strain (Ic). This observation links the expected and predominant increased degradation of the uncapped mRNAs to the accumulation of other uncapped mRNAs in Ceg1-depleted cells.

3.2.6 *XRN1* deletion recapitulates the mRNA differential expression dynamics of the Ceg1 depletion

To further investigate the role of Xrn1 in the regulation of the expression levels of the uncapped mRNAs, I started interrogating our data on the effect of the *XRN1* deletion on capped mRNAs, before the Ceg1 depletion, a condition I will refer to as *xrn1Δ* from now on.

I focused the attention on understanding how the Xrn1 inactivation in *ceg1-AA/xrn1Δ* (Rap) abolished the overexpression of the increased genes observed in the strain *ceg1-AA* upon Ceg1 depletion (Rap).

The global mRNA levels upon *XRN1* deletion were obtained comparing the gene expression in the strain *ceg1-AA* with the expression in the strain *ceg1-AA/xrn1Δ*, both in control condition (DSMO) therefore in the absence of capping defects. The analysis revealed that the lack of Xrn1was sufficient to induce the differential expression of 1140 protein-coding genes (Fig. 3.26) of which 543 upregulated (Ic, average log2 fold change 1.5) and 605 downregulated (Dc, average log2 fold change -1.4).
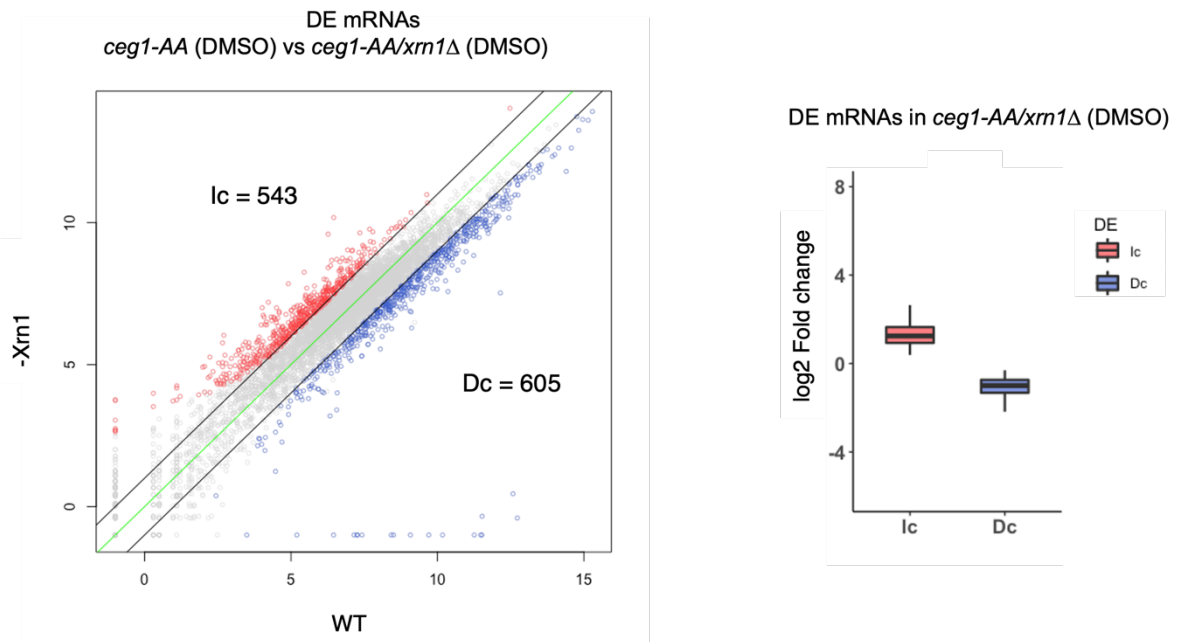
Figure 3.26 - The differential expression of mRNAs in xrn1Δ cells.

On the left. The plot showing the differential expression of protein-coding genes in xrn1Δ cells (-Xrn1) compared to the control (WT). The results are the average of two replicates. Each dot represents one gene. The gene whose abundance was decreased with log2 fold change < -1 and padj < 0.05 (Dc) are in blue. The genes whose abundance was increased with log2 fold change >1 and padj < 0. (Ic) are in red. Genes labelled in grey were considered not differentially expressed. The green line shows the zero-change line, the black lines delimit the ±1 log2 fold change. On the right. The log2 fold change of the differentially expressed genes (DE) in xrn1Δ cells. Decreased genes (Dc) showed log fold change -1.4, increased genes (Ic) have log2 fold change +1.5.

A deeper investigation on the genes differentially expressed in *xrn1Δ* cells brought to light some similarities between this group and the differentially expressed genes in Ceg1-depleted cells. Specifically, I focused our attention on the expression levels of the of the genes increased and decreased upon Xrn1 depletion. I compared the log2 fold change of the differentially expressed genes from *xrn1Δ* cells with their basal expression in the cells with the functional Xrn1. I observed that the genes showed to accumulate (Ic) had a lower basal expression level compared to the genes downregulated (Dc) upon Xrn1 inactivation

(Fig. 3.27A) The upregulated genes (Ic) were expressed at lower level (median log2 TPM = 2.69) compared with the downregulated fraction, Dc (median log2 TPM = 6.18).

I generated three groups of genes according to their basal expression, the 25% less expressed were fitted into the group named Low, the 25% highly expressed in the group High and the remaining 50% was grouped and named Mid. This grouping provided a clearer view of the changes in the mRNA abundance relatively to the basal gene expression (Fig. 3.27B). Indeed, I observed that of the 95% of genes belonging to the group Low (lowly expressed) were upregulated in *xrn1Δ* and, on the other hand the 99% of the genes from the group High were instead downregulated. The genes with intermediate basal expression levels were more equally distributed but still displaying a similar trend of dependency of log2 fold change from the basal expression levels (Fig. 3.27C and Table 3.4).
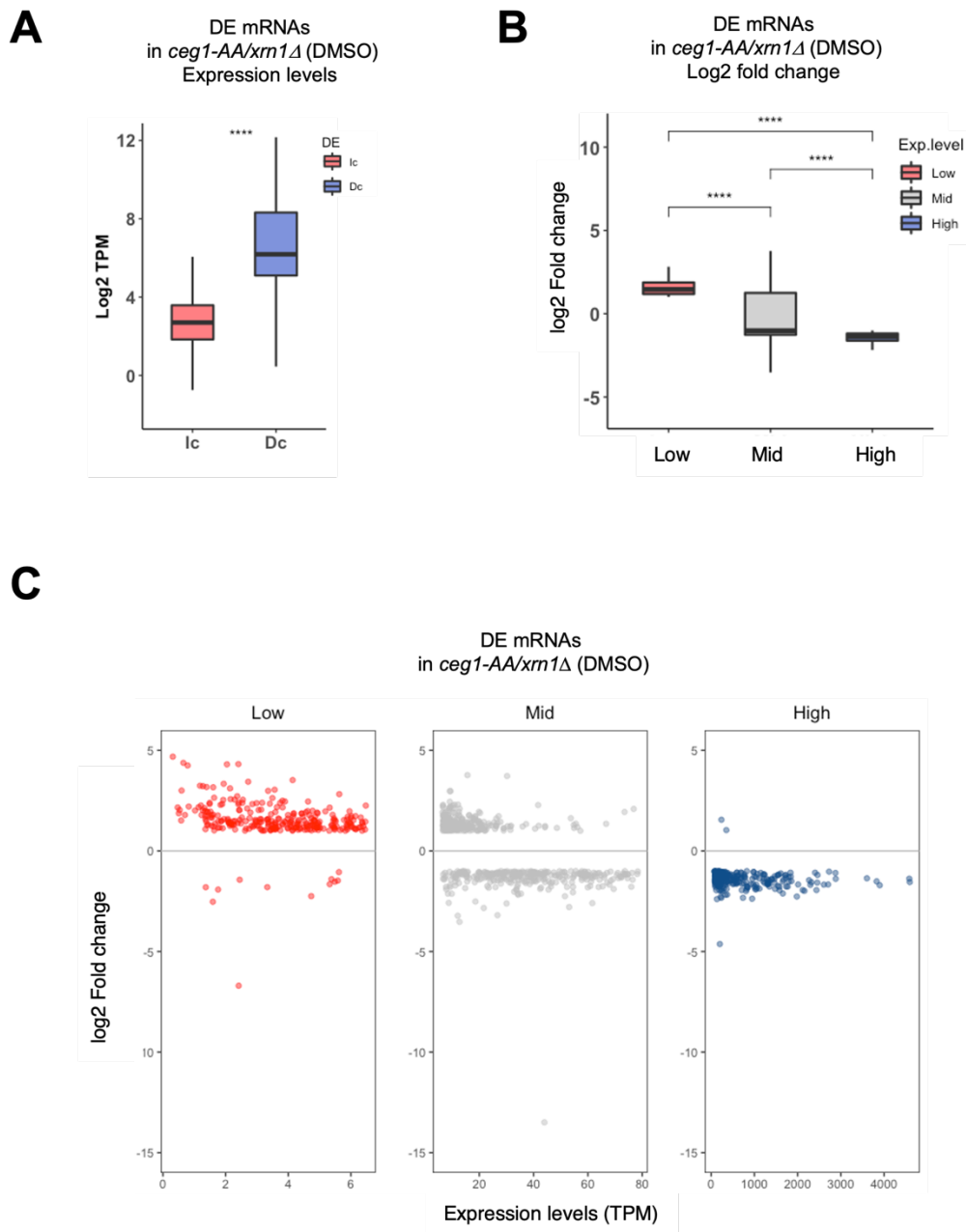
Figure 3.27 - The mRNAs levels in xrn1Δ cells have inverse correlation with their expression levels.

*A) The basal expression levels (expressed as log2 TPM) of the increased (Ic) and decreased (Dc) genes in xrn1Δ cells. Ic genes displayed overall, lower basal expression compared to Dc genes. The asterisks (*) indicate the statistical significance calculated via t-test (ns=p-value > 0.05; *=p-value ≤ 0.05; **=p-value ≤ 0.01; ***= p-value ≤ 0.001; ****= p-value ≤ 0.0001).* ***B)*** *The distribution of the differentially expressed genes (as log2 fold change) in ceg1-AA/xrn1Δ cells grouped according to their expression levels in High, Mid and Low. The asterisks (*) indicate the statistical significance calculated via ANOVA test (ns=p-value > 0.05; *=p-value ≤ 0.05; **=p-value ≤ 0.01; ***= p-value ≤ 0.001; ****= p-value ≤*

0.0001). ***C)*** *The distribution at single gene levels of the differentially expressed genes in ceg1-AA/xrn1Δ according to their expression levels. Highly expressed genes (blue) show on average a decrease in xrn1Δ cells, on the contrary, lowly expressed genes (red) are mainly increased.*

Table 3.4 - The differentially expressed protein-coding genes in xrn1Δ were classed according to their expression levels. *The table shows for each group: High, Mid and Low, the numner of increased (Ic) or decreased (Dc)genes.*

xrn1Δ	Increased (Ic)	Decreased (Dc)	Tot
High (Top25%)	2	280	282
Mid (Middle 50%)	268	296	564
Low (Bottom25%)	270	12	282

I also related the differential expression of the mRNA in *xrn1Δ* cells with their half-life (Fig. 3.28) using data available in the literature (Geisberg et al., 2014). Alike the genes differentially expressed upon Ceg1 depletion, the genes upregulated (Ic) in *xrn1Δ* cells had a longer half-life (median 30 min) compared to the downregulated genes (Dc, median 17 min).

DE mRNAs in *ceg1-AA/xrn1Δ* (DMSO)
half-life
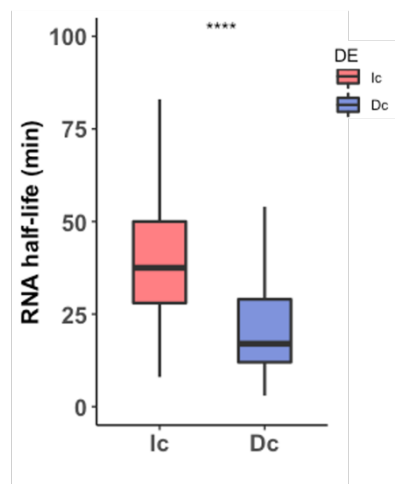
Figure 3.28 - The plot showing the half-life of the differentially expressed genes in xrn1Δ cells.

Decreased genes (Dc) have an overall shorter half-life compared to the increased genes (Ic). The asterisks (*) indicate the statistical significance calculated via t-test (ns=p-value > 0.05; *=p-value ≤ 0.05; **=p-value ≤ 0.01; ***= p-value ≤ 0.001; ****= p-value ≤ 0.0001).

Due to the similarity of these results to the effects reported for the Ceg1-depleted cells with the functional Xrn1(described in the previous paragraphs), I compared the genes differentially expressed in these two conditions (i. e., capping defective with functional Xrn1 and capped with *XRN1* deletion). The comparison (Fig. 3.29) revealed a wide but not total overlap between the genes differentially expressed upon capping defects (*ceg1-AA*) or Xrn1 inactivation (*xrn1Δ*). Indeed, of the 1205 genes differentially expressed in *ceg1-AA* and the 1148 genes differentially expressed in *xrn1Δ*, 463 (20%) were differentially expressed (DE) in both conditions. More specifically, 190 genes were commonly upregulated (Ic) and 264 downregulated (Dc) in both conditions (9 genes showed differential expression of opposite sign between the two conditions: *CWP1*, *MBF1*, *DHH1*, *ALD4*, *IDH2*, *AQY2*, *YLL053C*, *LCL1* and *YMR193C-A*).
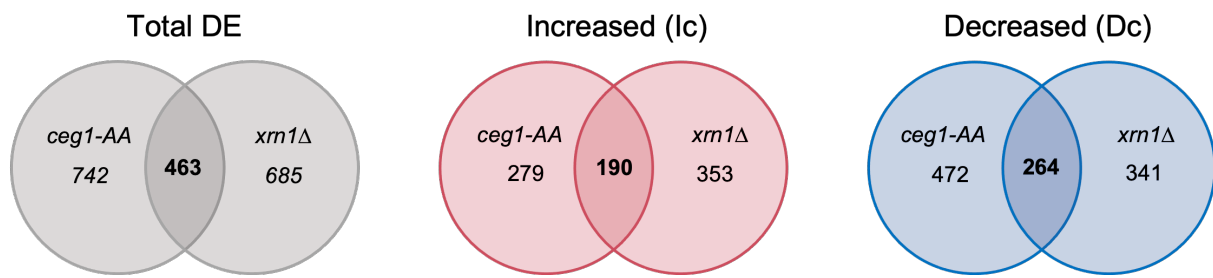
Figure 3.29 - Differentially expressed genes (DE) in ceg1-AA (Rap) and xrn1Δ (DMSO) cells.

Number of differentially expressed genes in ceg1-AA (Rap) and xrn1Δ (DMSO) are shown and compared in grey. The increased (Ic) genes are shown in red and the decreased (Dc) in blue.

This comparison took in account the genes classified as differentially expressed according to the parameters selected in both datasets (i.e., |log2 fold change| > 1). Therefore, to better understand the overlap between the effects of Ceg1 depletion and Xrn1 inactivation, I tested the expression levels of all differentially expressed genes in Ceg1-depleted cells, in *xrn1Δ* cells (Fig. 3.30A). I found that the differentially expressed genes from Ceg1-depleted cells were affected in *xrn1Δ* in a similar fashion. The overexpressed (Ic) genes in Ceg1-depleted cells maintained an average positive log2 fold change in *xrn1Δ* cells (average from 1.5 to 1.28). Downregulated (Dc) genes in Ceg1-depleted cells, similarly, maintained an overall negative decreased abundance in *xrn1Δ* cells (average log2 fold change from -1.4 to -1.01).

To further investigate the trend, I performed the reverse analysis and evaluated the abundance of differentially expressed genes from *xrn1Δ* cells in the Ceg1 depletion context (Fig. 3.30B). Both increased (Ic) and decreased (Dc) genes from *xrn1Δ* cells maintained their redistribution in the Ceg1-depleted cells (Ic from 1.5 in *xrn1Δ* to 1.2 in *ceg1-AA* and Dc from -1.4 in *xrn1Δ* to -1.1 in *ceg1-AA*).

These results suggested that either the lack of the m⁷G cap induced by the Ceg1 depletion and the Xrn1 inactivation obtained via gene knockout, might affect gene expression of the mRNAs through a common mechanism. This hypothesis is in line with the observation whereby the introduction of capping defect in

cells lacking the functional Xrn1 did not permit to appreciate the pre-existing differential expression resulting therefore in the identification of only 23 upregulated (Ic) and 18 downregulated (Dc) protein-coding genes (see Fig. 3.23A).
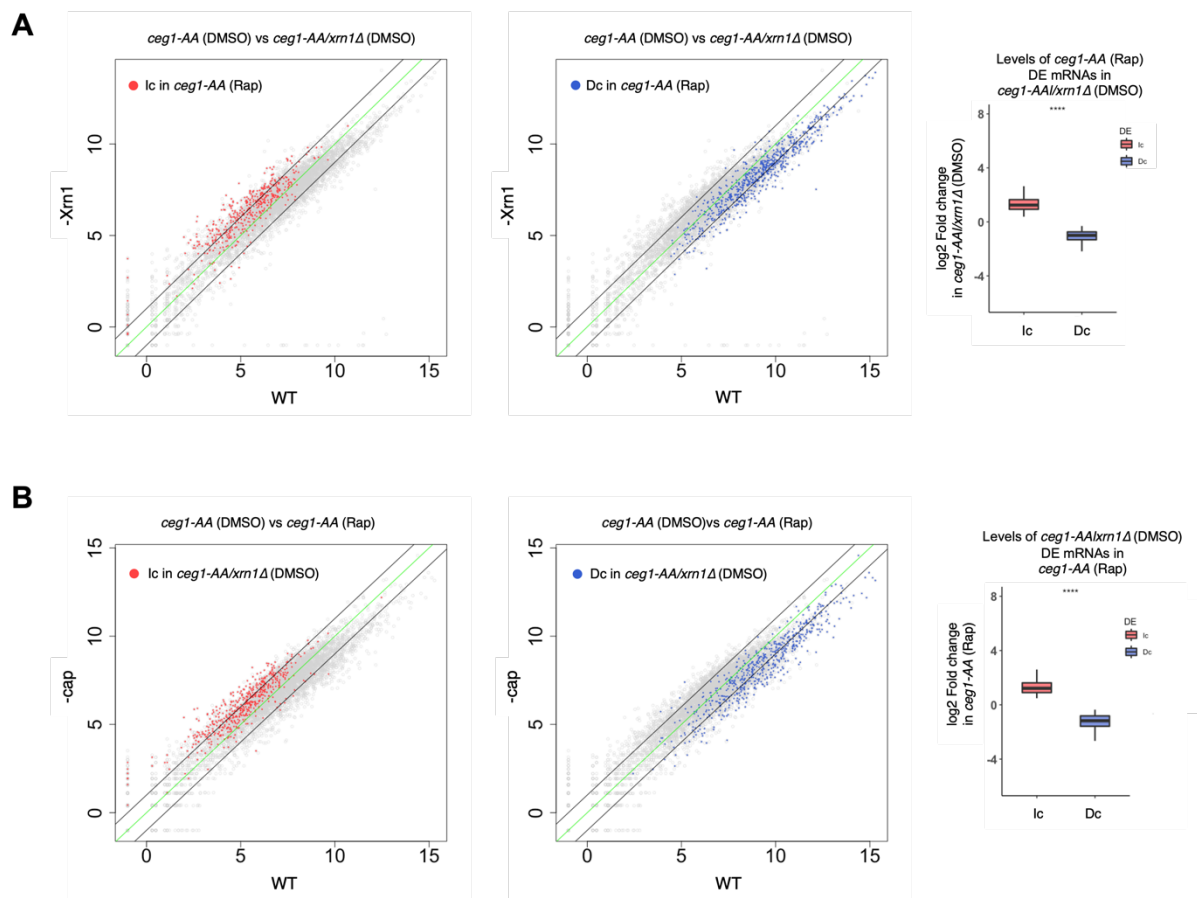
Figure 3.30 - The Ceg1 nuclear depletion and the Xrn1 knockout (xrn1Δ) similarly affect the mRNA expression.

A) *Differentially expressed genes (DE) in Ceg1-depleted cells with a functional Xrn1 are labelled on the plot showing the differential expression in xrn1Δ cells (-Xrn1) compared to control (WT). Increased (Ic) genes in Ceg1-depleted cells are labelled in red and accumulate on the positive side of the zero-change line (green). Decreased (Dc) genes in Ceg1-depleted cells are labelled in blue and accumulate on the negative side of the zero-change line. The results are the average of two replicates. Each dot represents one gene. The gene whose abundance was decreased with log2 fold change < -1 and padj < 0.05 upon Ceg1 depletion (Dc) are in blue. The genes whose abundance was increased with log2 fold change >1 and padj < 0.05 upon Ceg1 depletion (Ic) are in red. The black lines delimit the ±1 log2 fold change. On the right the value of log2 fold change of the DE genes in Ceg1-depleted cells, measured in xrn1Δ cells.*
B) *Differentially expressed genes (DE) in xrn1Δ cells are labelled on the plot showing the differential expression in Ceg1-depleted cells (-cap) compared to control (WT). Increased (Ic) genes in xrn1Δ cells are labelled in red and accumulate on the positive side of the zero-change line (green). Decreased (Dc) genes in xrn1Δ cells are labelled in blue and accumulate on the negative side of the zero-change line. The*

results are the average of two replicates. Each dot represents one gene. The gene whose abundance was decreased with log2 fold change < -1 and padj < 0.05 upon Xrn1 inactivation (Dc) are in blue. The genes whose abundance was increased with log2 fold change >1 and padj < 0.05 upon Xrn1 inactivation (Ic) are in red. The black lines delimit the ±1 log2 fold change. On the right the value of log2 fold change of the DE genes in xrn1Δ cells, measured in Ceg1-depleted cells.

3.2.7 Discussion

The lack of genome wide analysis on cap-defective cells has led to a limited view of the global effects of the m^7G cap on gene expression. In this chapter, I showed data expanding the perspective on the regulation of the uncapped mRNAs abundance upon Ceg1 depletion.

First, I generated a rRNA-depleted pool of RNA from yeast cells employing the Ribo-Pop method (Thompson et al., 2020) that allowed to obtain the sequences for 15 antisense probes targeting the rRNA 28S and 18S from *S. cerevisiae*. I demonstrated that the method induces a strong depletion of the rRNAs targeted (Fig. 3.11) in a reliable and cost-saving fashion. As indicated from the authors of the method (Thompson et al., 2020), the 5S and 5.8S small species of the cellular rRNAs were not targeted as they represent a negligible source of contamination in the sequencing protocols.

I generated the *ceg1-AA* strains (*ceg1-AA* and *ceg1-AA/xrn1Δ*) that resulted ideal to quickly depleting the nuclear protein and capping enzyme Ceg1 causing its loss of function. As a direct consequence, the activation of the anchor away abolishes the mRNA capping. I first observed that 45 minutes of Ceg1 depletion resulted in the differential expression of a subset of protein-coding genes (1205 of 6600, 18% of total protein-coding genes). The partial effect on the RNA abundance indicated that different genes could respond with different timing to the lack of the m^7G cap. The prolonged depletion of Ceg1 could have induced a stronger effect on the cellular transcriptome either in terms of the number of genes involved or in the amplitude of the differential expression. Nevertheless, the sustained depletion of the essential protein Ceg1 could have left time for the detection of misleading secondary effects. Previous observations on cap-defective mRNAs were conducted through a single-gene approach and all converged towards the concept by which RNAs whose capping was impaired are quickly degraded in the nucleus via a dedicated cap quality control pathway leading to a general decrease of the cap-deficient RNA levels (Jiao et al., 2013; Jimeno-González et al., 2010; Zhai & Xiang, 2014). The outcomes reported in this chapter add important details to this paradigm. Via nuclear depletion of the capping enzyme Ceg1, I observed the expected

decrease in the levels of a fraction of uncapped mRNAs in parallel with the unexpected accumulation of other cap-deficient mRNAs (Fig. 3.13). This indicated that the lack of the m^7G cap has a differential effect on the global mRNA distribution.

Further investigating this phenomenon, I observed a correlation between the basal expression levels of the protein-coding genes and their differential expression upon Ceg1 depletion (Fig. 3.15). Highly expressed genes displayed a decreased mRNA abundance, on the other hand, lowly expressed genes showed increased levels in Ceg1-depleted cells compared to the control. Other genes features were shown to be related to the uncapped mRNA expression upon Ceg1 depletion. A solid correlation was reported for Pol II occupancy and transcription rate (TR) confirming the role of the expression levels in the accumulation of lowly expressed mRNAs in our model. Low Pol II occupancy and low TR can be directly associated to low expression and our results indicated the mRNAs with lower Pol II occupancy and TR were upregulated upon Ceg1 depletion (Fig. 3.16A and B). The accumulation of uncapped lowly expressed mRNAs in Ceg1-depleted cells was also related to higher gene length and RNA half-life (Fig. 3.17A and B). Overall, all these observations point on the preferential degradation of highly expressed and uncapped RNAs. It is plausible the hypothesis by which the most highly transcribed mRNAs have a preferential access to the degradation machinery when uncapped and that the global degradation caused by the capping defects inhibits the degradation of the lowly expressed and cap-deficient mRNAs, possibly following saturation dynamics. The treatments aimed to reduce Pol II processivity and elongation rate (6-AU treatment) induced the mild stabilisation of the *ADH1* mRNA (Fig. 3.18D) suggesting the existence of a global threshold level in the regulation of the expression of uncapped mRNAs upon Ceg1 depletion.

The differential expression of the mRNAs in Ceg1-depleted cells did not correlate with the other features tested like 5' UTR length or the presence of IRES and therefore with the cap-independent translation process (Fig. 3.19A and B). These results suggested that the genes reported to accumulate are unlikely to

be stabilised by structural characteristics of the untranslated regions (UTRs) or shielded by the translation into proteins.

The functional characterization of the differentially expressed genes depicted a mild tendency of the increased mRNAs to correlate with the Pol II-dependent transcription process as being mainly represented by transcriptional factors or DNA binding proteins with different functions. A clear pattern was not recognised for the mRNAs decreased in Ceg1-depleted cells (Fig. 3.20 and Table 3.2).

Finally, two sequence motifs resulted enriched in the subgroup of the genes upregulated (Ic) in Ceg1-depleted cells. The motifs (G/A)GAA(A/G)A and TTTTTG were highly represented in the coding region of these genes at DNA level. The motif (A/G)GAAAA was enriched also in the analysis of the RNA sequences. This motif has been described to function as binding site for a class of human transcriptional factors belonging to the NFAT family. It could exert a function in *S. cerevisiae* regulating the expression levels or the stability of the uncapped mRNAs.

This observation was not further investigated and represents a hint for an interesting parallel work on mRNA expression and stability in yeast.

I also observed that the deletion of *XRN1* (*xrn1Δ*), the gene encoding for the main cytoplasmic 5'-3' exonuclease responsible for the degradation of the mRNAs, equally abolished the decrease and the accumulation of the uncapped mRNA generated in Ceg1-depleted cells (Fig. 3.23A and Fig. 3.25). The lack of Xrn1 function, has been shown to induce the stabilization of uncapped mRNAs in temperature sensitive Ceg1 mutants (Schwer et al., 1998) therefore the rescue of the downregulated genes in Ceg1-depleted cells was expected. On the other hand, the effects of the Xrn1 inactivation on the levels of the upregulated mRNAs observed in Ceg1-depleted cells was unexpected and therefore further investigated. The depletion of Xrn1, indeed, affected the levels of the increased fraction of mRNAs, abolishing their accumulation. To understand this phenomenon, I assessed the effects of the *xrn1Δ* mutation on the expression of capped mRNAs. I observed that the lack of Xrn1-exonucleolytic activity induced the

redistribution of the mRNA levels similar to that observed in Ceg1-depleted cells. I observed a symmetrical upregulation and downregulation of protein-coding gene expression recapitulating the correlation of the basal expression levels and mRNA stability described for the cap deficient cells. Interestingly, the genes differentially expressed upon the two conditions (i.e., Ceg1 depletion and *xrn1Δ*) showed a certain overlap (Fig. 3.29) that was further expanded by the global tendency of the genes differentially expressed on one condition to distribute similarly on the other (and vice versa). Genes upregulated in one condition were overall upregulated in the other, similarly for downregulated genes (Fig. 3.30A and B).

This output indicates that either the cap deficiency or the lack of the functional Xrn1, exert a common effect on the balance of the cellular mRNAs suggesting that these two conditions affect the same mechanism. This is also in line with our observation by which the induction of Ceg1 depletion in *xrn1Δ* cells did not result in the differential expression of the cellular mRNAs masked by a pre-existing unbalance in the gene-expression caused by the lack of the functional Xrn1.

In conclusion, the results presented here show that the stability of the uncapped mRNAs in Ceg1-depleted cells is not affected exclusively in a negative way but rather, global capping defects can induce the accumulation of certain mRNA species. The stability observed is related to the expression levels of the genes which would favour the highly expressed genes in the access to the cytoplasmic degradation machinery (Casolari et al., 2004). The uncapped mRNAs generated in the Ceg1-depleted cells seem to evade the nuclear capping quality control. They are transported into the cytoplasm and degraded via the general Xrn1-mediated degradation pathway, described for the functional mRNAs' turnover. Xrn1 has been shown to regulate the cellular mRNA sensing to control the physiological balance between RNA synthesis and degradation (Braun et al., 2012; Sun et al., 2013). The differential expression observed in the *xrn1Δ* strain can be the result of a combinatory effect of the lack of mRNA degradation and the dysfunctional sensing of the cellular mRNA levels. This will lead to the adjustments related to the rate of

the RNA synthesis favouring the accumulation of lowly expressed and the degradation of highly expressed genes. Due to the similarities in the protein-coding gene expression reported between Ceg1 and Xrn1-depleted cells, the m^7G cap can be involved in the Xrn1-mediated buffering of the cellular mRNA levels.

3.3 The mRNA levels of *FMP27* are stable upon Ceg1 depletion

3.3.1 Introduction

The results described above show that the Ceg1 depletion has different effects on the mRNA abundance. The nuclear depletion of the capping enzyme, indeed, induces the degradation of highly expressed genes and the accumulation of the genes expressed at a lower level. The ideal duration for the Ceg1 depletion has been set at 45 minutes. This time window was demonstrated to be sufficient to induce a strong cap-defective phenotype before the rising of any major cellular response observed as proliferation defects soon after 60 minutes of rapamycin treatment, required to activate the Ceg1-AA system. The genome wide analysis showed a fraction of the protein-coding genes to be differentially expressed in Ceg1-depleted cells for 45 minutes. I postulated that a longer depletion could have had a deeper effect on the transcriptome either in terms of number of genes involved or in the degree of variation of the differentially expressed genes. I investigated further the features of the genes whose mRNA levels were increased upon Ceg1 depletion to describe the phenomenon of their accumulation.

In this chapter I will investigate the model gene *FMP27* (also known as *YLR454W*) widely employed in the literature studying yeast transcription but also capping defects and co-transcriptional degradation. *FMP27* was stable upon Ceg1 depletion therefore I investigated its characteristics to dissect the phenomenon and further clarify the dynamics of the mRNA stability in Ceg1-depleted cells.

3.3.2 The mRNA levels of FMP27 are stable upon Ceg1 depletion

The model gene *FMP27* has been often investigated in studies related to the Pol II-dependent transcription in *S. cerevisiae*. *FMP27* spans for 8 kb representing an exceptionally long transcriptional unit in the yeast genome where the average length is about 1.4 kb (Fig. 3.31A). This characteristic offers the space to

investigate the co-transcriptional events besides, *FMP27* is not essential, it encodes for a putative polypeptide found in the analysis of the mitochondrial proteome (Reinders et al., 2006) and therefore it can be an ideal endogenous locus where to introduce modifications and tools for the transcriptional studies. Previous reports, showed that *FMP27* is co-transcriptionally degraded in yeast mutants with cap-defective phenotypes (Jimeno-González et al., 2010) therefore I treated our strain *ceg1-AA* with rapamycin for a time window of 120 minutes to induce the depletion of the capping enzyme Ceg1 and recorded the mRNA levels of *FMP27* at different timepoints. The analysis showed that the abundance of *FMP27* is not perturbed by the global lack of cap induced by Ceg1 depletion (Fig. 3.31B). This outcome is in contrast with the data from the literature where global capping defects were induced in the temperature sensitive mutants of Ceg1 (Jimeno-González et al., 2010).

To better understand the physiological decay of *FMP27* mRNA, I performed a chase experiment followed by RT-qPCR. I employed the yeast strain mutant where the expression of *FMP27* is under the control of the inducible *GAL1* promoter (*GAL1::FMP27*). I grew the mutant in the presence of galactose as the main source of carbon, a metabolic condition that induces the expression of *FMP27* then shut down its expression by the introduction of glucose in the media (Fig. 3.31C). The mRNA levels of *FMP27* were measured over time at different timepoints (Fig. 3.31D). I compared the mRNA levels of *FMP27* with the levels of another gene under the control of the *GAL1* promoter, the gene *GAL1* as a positive control to confirm the metabolic switch and the constitutively expressed *ADH1* as a negative control for the expression of the genes independent of the main carbon source. The results of the chase experiment indicate that *FMP27* is not a super-stable transcript. By shutting down its transcription, the decrease of the mRNA levels was already observed 5 minutes after the introduction of the glucose in the media, the reduction continued progressively up to 30 minutes time, when the sampling stopped. The same decrease was observed for the gene *GAL1* confirming the uptake of the glucose by the cells. The levels of *ADH1* remained stable, although only mildly perturbed, confirming that the decrease in the mRNA levels of

FMP27 and *GAL1* were caused by the transcriptional shut down induced by the metabolic switch towards the glucose catabolism. Overall, the chase experiment shows that *FMP27* is not a super-stable transcript and therefore its stability upon capping defects in Ceg1-depleted cells is not caused by the accumulation of the pre-existing transcripts but more likely by the impairment of its degradation.

A Gene length distribution in *S.cerevisiae*

B

C

D Chase experiment
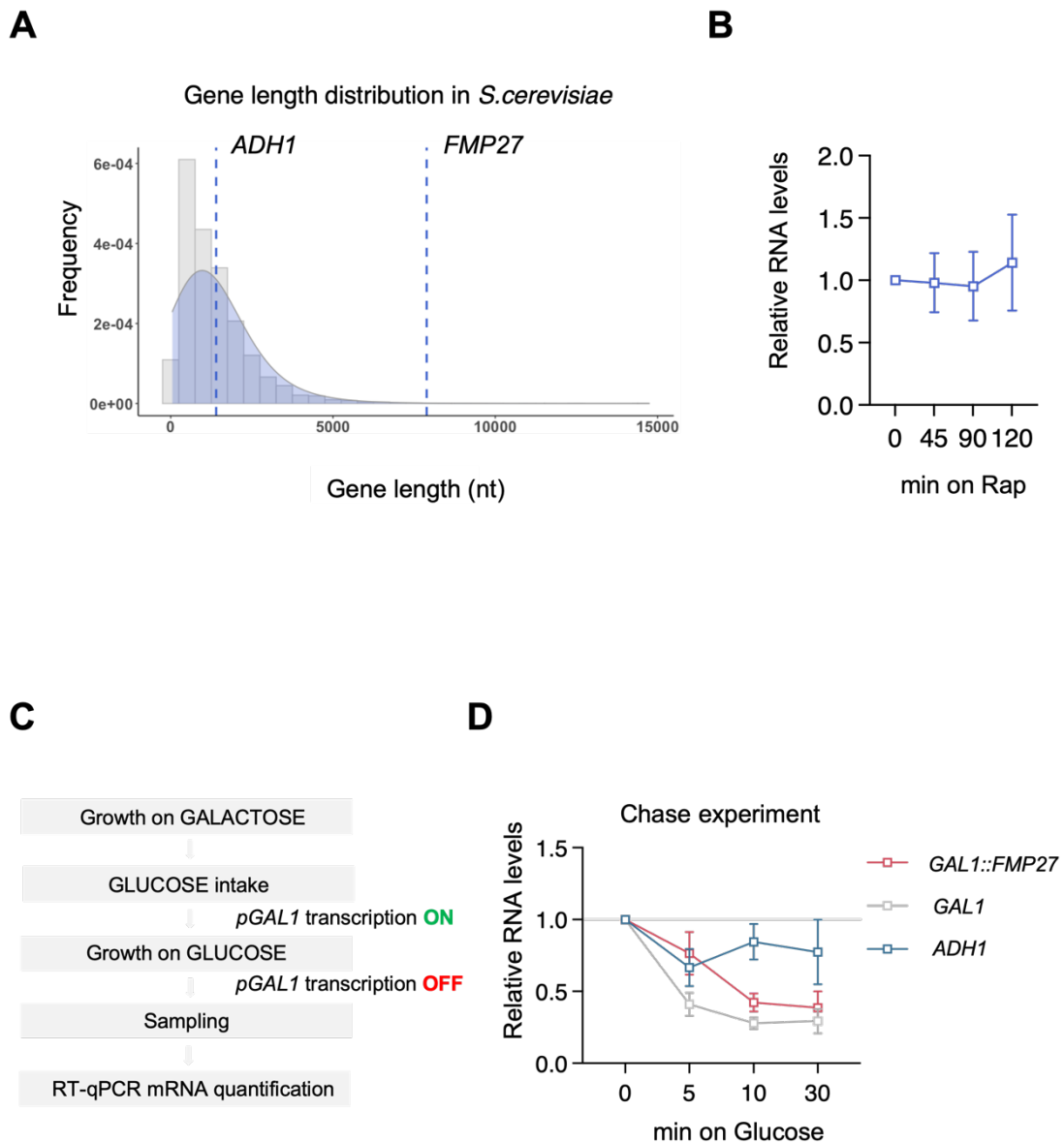
Figure 3.31 - FMP27 mRNA levels are stable upon Ceg1 depletion but it is not a super-stable transcript.

A) The plot showing the distribution of the gene length in Saccharomyces cerevisiae. The grey bars show the frequency of the recurrence for the same gene length. The area underneath the trend line is shown in semi-transparent blue. The dashed blue lines indicate the value of the gene length for ADH1 (1047 nt) and FMP27 (7887 nt) respectively. B) The plot showing the RNA levels measured via RT-qPCR for FMP27 in ceg1-AA compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment. The error bars show the standard deviation of three independent experiments. C) The schematic of the

chase experiment. **D)** *The plot showing the RNA levels measured via RT-qPCR for GAL1::FMP27, GAL1 and ADH1 after the addition of glucose to the media and consequent inhibition of the GAL1 promoter-mediated transcription. The RNA abundance was normalised against the ribosomal RNA (rRNA) 25S. The error bars show the standard deviation of three independent experiments.*

At this point I tried to recapitulate the degradation of *FMP27* in the temperature sensitive mutant *ceg1-63* (Fresco & Buratowski, 1996), I also investigated the mRNA decay in the double temperature sensitive mutant for ceg1 and for the nuclear exoribonuclease Rat1. In both of the mutants and in the wild type strain (WT), the expression of *FMP27* was under the control of the *GAL1* promoter (*ceg1-63/rat1-1,* Jimeno-González et al., 2010). The cells were grown in the presence of galactose to induce the expression of *FMP27* at the permissive temperature of 25 °C (time zero) and then moved to the non-permissive temperature of 37 °C. I performed RT-qPCR to measure the mRNA levels of *FMP27*, *ADH1* and *GAL1*, over time, up to 120 minutes compared to time zero (Fig. 3.32). The temperature inactivation of Ceg1 over time, and the consequent capping defects, induced the progressive decrease of the mRNA levels measured for *FMP27*, *ADH1* and *GAL1* suggesting that the lack of the m^7G cap in the mutant *ceg1-63* was sufficient to trigger the mRNA decay also for *FMP27* that was stable when the capping defects were induced via Ceg1 nuclear depletion. The mRNA decay for *FMP27*, *ADH1* and *GAL1* was also observed in the mutant *ceg1-63/rat1-1* at the non-permissive temperature in a very similar fashion as for the *ceg1-63* mutant (Fig. 3.32). This result confirmed the mRNA degradation resulted from the capping defects induced but poses an interesting point on the degradative pathways that the uncapped mRNAs undergo. Specifically, this outcome shows that Rat1 is only marginally or not responsible for the degradation of the cap-defective mRNAs in this system as its inactivation had no effect on the RNA abundance quantified.
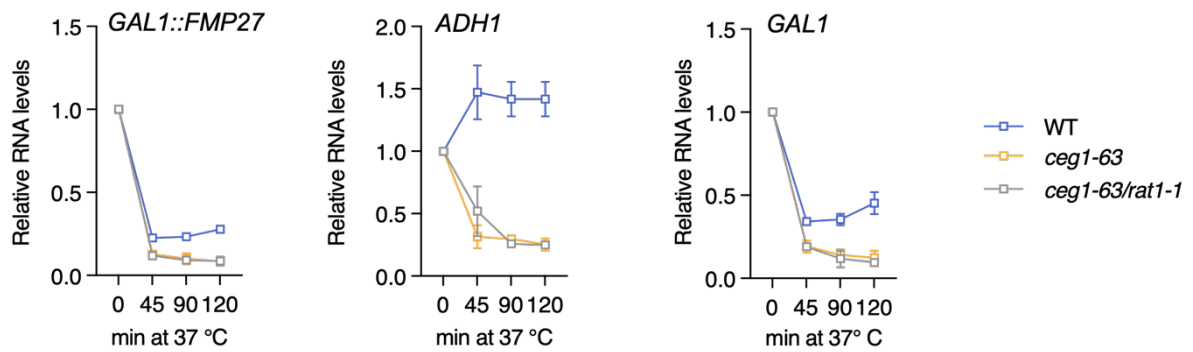
Figure 3.32 - The mRNA levels of GAL1::FMP27, ADH1 and GAL1 in Ceg1 and Rat1 temperature-sensitive mutants.

The plots showing the mRNA levels measured via RT-qPCR for GAL1::FMP27, ADH1 and GAL1 normalised against the ribosomal RNA (rRNA) 25S. Cells were grown at 25 °C (0 min) then moved to the non-permissive temperature (37 °C) and sampled at different timepoints. The temperature-sensitive mutant strains (ceg1-63 and ceg1-63/rat1-1) and wild type strain (WT) harbour the mutation GAL1::FMP27 where the endogenous FMP27 promoter was replaced with the inducible GAL1 promoter. The error bars show the standard deviation of three independent experiments.

Importantly, I observed the unexpected reduction of the *FMP27* and *GAL1* mRNAs levels over time, in the WT strain at the non-permissive temperature (Fig. 3.32) suggesting that in these mutant cells (WT but *GAL1::FMP27*) the abundance of these two mRNAs (*GAL1* and *GAL1::FMP27*) is affected regardless the capping defects. The mRNA decay observed was less severe but still consistent compared to the temperature sensitive mutants *ceg1-63* and *ceg1-63/rat1-1* (Fig. 3.32) suggesting a cooperative effect acting on the decrease of the mRNA levels in these mutants.

The decrease was not observed for *ADH1* in the same WT strain at the non-permissive temperature. Rather, a mild and sustained increase of its mRNA levels was observed (Fig. 3.32), possibly caused by a metabolic burst induced by the increased temperature (Postmus et al., 2008).

Rat1 Interacting Protein (Rai1) is the nuclear decapping protein which triggers the degradation Rat1-mediated of the transcript that fail a proper capping (Doamekpor et al., 2020). Rai1 has decapping and

pyrophosphohydrolase activity such as it can convert either RNAs with unmethylated cap and non-capped RNAs in 5' monophosphate RNAs that are the ideal substrate of Rat1 that in turn can co-transcriptionally degrade such uncapped RNAs exerting a cap quality control (Jiao et al., 2010). The formation of the complex Ra1-Rat1 is needed for the activity of both components (Xiang et al., 2009) therefore, to further investigate the role of Rat1 in the degradation of the uncapped RNAs in the Ceg1-depleted cells, I induced the deletion of the *RAI1* gene, which is a viable null mutation (Xue et al., 2000), to affect the binding of Rat1, whose null mutation is instead lethal (W. Luo, Johnson, & Bentley, 2006), to the transcriptional complex engaged in the synthesis of uncapped RNAs. The *RAI1* genetic locus was deleted by gene targeting with the introduction of the aminoglycoside phosphotransferase coding sequence (*kanMX*) used as selection marker for the transformant cells as it confers resistance to kanamycin and neomycin (Fig. 3.33A). The correct introduction of the transgene was tested via PCR (Fig. 3.33B). Next, the mRNA levels of *ADH1* and *FMP27* were measured after 90 minutes of rapamycin treatment to induce the Ceg1 depletion. A longer single timepoint was preferred to ensure a strong capping defective phenotype in the *ceg1-AA/rai1Δ* strain (Fig. 3.33C). I did not observe any difference on the uncapped mRNA levels for *ADH1* and *FMP27* between Ceg1-depleted cells expressing or not Rai1 (*ceg1-AA* and *ceg1-AA/rai1Δ* respectively). For both strains the rapamycin treatment induced the decrease of the *ADH1* levels leaving unperturbed the expression of *FMP27*. This setup aimed to disrupt the elements of the cap quality control system and test the effects on the uncapped mRNAs in Ceg1-depleted cells. The role of the Rai-Rat1 complex in the degradation of the uncapped mRNAs resulted minimal or null as its disruption in the cell did not affect the fate of the uncapped mRNAs whatever it was. This result, together with the results obtained with the temperature sensitive mutants (*ceg1-63/rat1-1*) indicate that the cap quality control mechanism was not active in the conditions investigated.
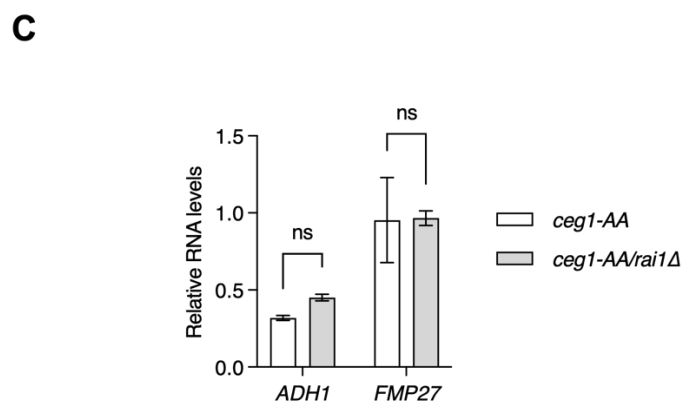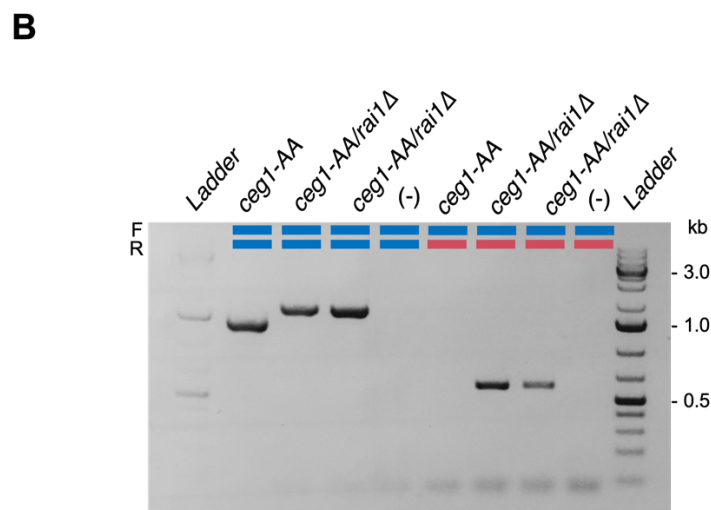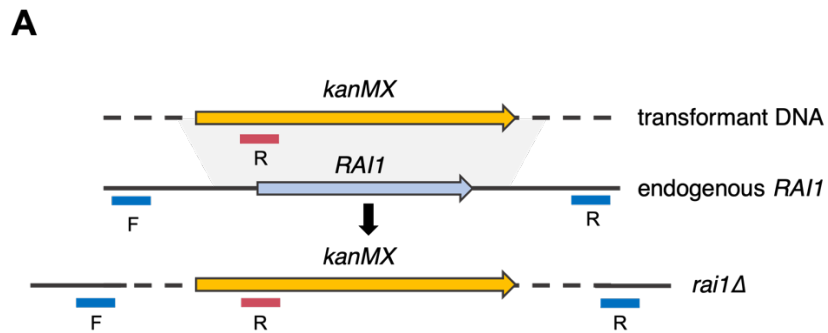
Figure 3.33 - The construction of the mutant ceg1-AA/rai1Δ and the effects of the rai1Δ mutation on the uncapped mRNA levels for ADH1 and FMP27.

A) The scheme showing the gene targeting inducing the deletion of the RAI1 locus from the yeast genome and concomitant introduction of the kanMX selection marker gene. The blue and the red horizontal bars

Overall, these results indicated that the temperature sensitive mutants used here are not a reliable tool to investigate the RNA decay of the uncapped *FMP27* as the shift at the non-permissive temperature caused some unexpected effects that have the potential to cause the misinterpretation of the data generated.

To affect the mRNA stability observed upon Ceg1 depletion for *FMP27*, I generated some mutants of the structure of the endogenous *FMP27* gene by introducing some modifications aimed to alter its expression. *FMP27* is less expressed than the model gene *ADH1*, about its 0.9% (Fig. 3.34A) and overall considered a low expressed gene. To investigate the rules governing its stability upon Ceg1 depletion, I generated a mutant in the *ceg1-AA* genetic background where *FMP27* was expressed under the control of the *ADH1* promoter (Fig. 3.34B). The correct introduction of the promoter was tested by PCR (Fig. 3.34C) and the strain *ceg1-AA/fmp27-a* (named *fmp27-a* from now on) was tested for its viability with growth spot assay (Fig. 3.23D). The strain *fmp27-a* had a viable phenotype, suggesting that the modification introduced did not alter the cellular physiology to the point to generate proliferative and growth defects. The phenotype on rapamycin remained lethal due to the activation of the anchor away for the capping enzyme Ceg1.
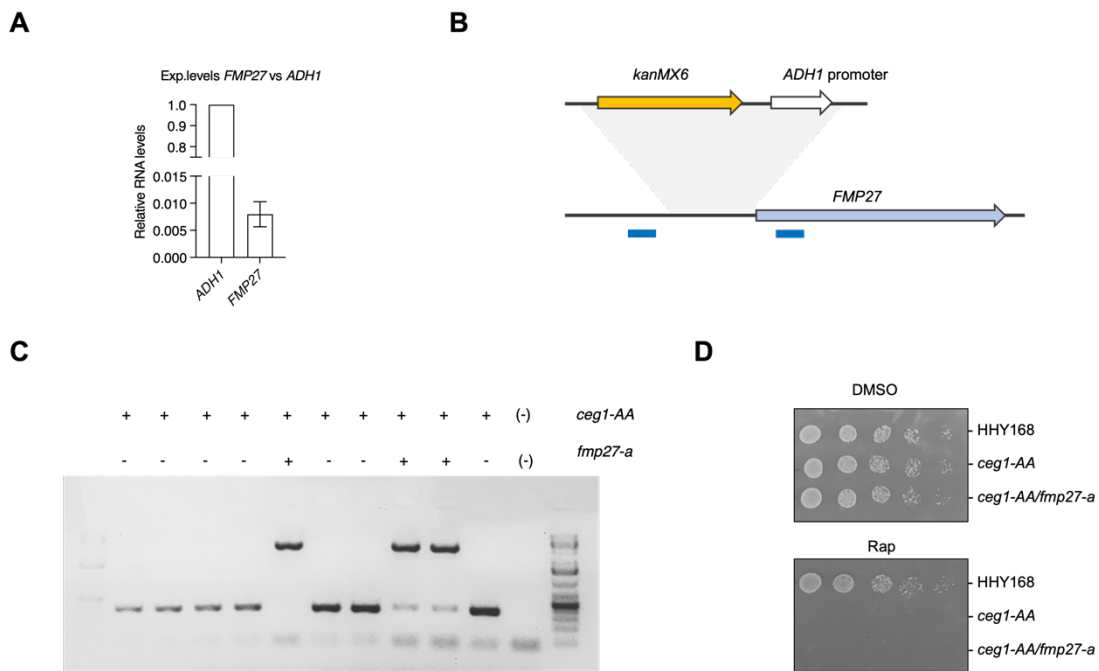
Figure 3.34 - The expression levels of FMP27 compared to ADH1 and the characterisation of the fmp27-a mutant.

*A) The expression levels of FMP27 compared to ADH1 in ceg1-AA cells grown in the presence of DMSO. RNA levels were normalised against the ribosomal RNA (rRNA) 25S and, for FMP27, normalised against the ADH1 levels arbitrarily set to 1. The error bars show the standard deviation of three independent experiments. **B)** The scheme showing the gene targeting for the introduction of the ADH1 promoter upstream of the FMP27 coding region. The blue horizontal bars under the genetic locus depicted, show the position of the PCR primers used to screen the correct introduction of the cassette. **C)** The PCR products run on 1% agarose gel showing the correct introduction of the cassette in the genomic locus targeted using the oligos located as in B. The PCR negative control is labelled with (-). **D)** The spot growth assay of the strains HHY168, ceg1-AA and ceg1-AA/fmp27-a. Serial dilutions of the liquid culture of each strain spotted on YPD media enriched with DMSO or rapamycin (Rap). The strain ceg1-AA is viable on DMSO but shows a lethal phenotype on Rap as consequence of the activation of Ceg1 nuclear depletion. The strain ceg1-AA/fmp27-a shows a lethal phenotype on Rap as consequence of Ceg1 depletion.*

I measured the levels of *FMP27* in the strain *fmp27-a* with RT-qPCR (Fig. 3.35A). Compared to the parental strain (*ceg1-AA*), the mRNA levels of *FMP27* resulting from the transcription controlled by the *ADH1* promoter (which I will refer to as *fmp27-a* for the name of the strain to distinguish this transcript from the transcript produced from the endogenous promoter), showed a six-fold increase. Nevertheless, the expression levels of *fmp27-a* were still not comparable to the levels of *ADH1* representing only the 4.5% of its expression (Fig. 3.35B). I induced the Ceg1 nuclear depletion via anchor away and measured the mRNA levels of *fmp27-a* over time (Fig. 3.35C). Compared to the control conditions, the levels of *fmp27-a* increased for the first 90 minutes of Ceg1 depletion to return to the basal levels after 120 minutes. Although affected, the mRNA levels of *fmp27-a* did not show a clear trend, therefore the introduction of a stronger promoter to drive the expression of *fmp27-a* was not sufficient to clearly alter its stability upon Ceg1 depletion.

A — Exp levels *FMP27* vs *fmp27-a*

B — Exp.levels *fmp27-a* vs *ADH1*
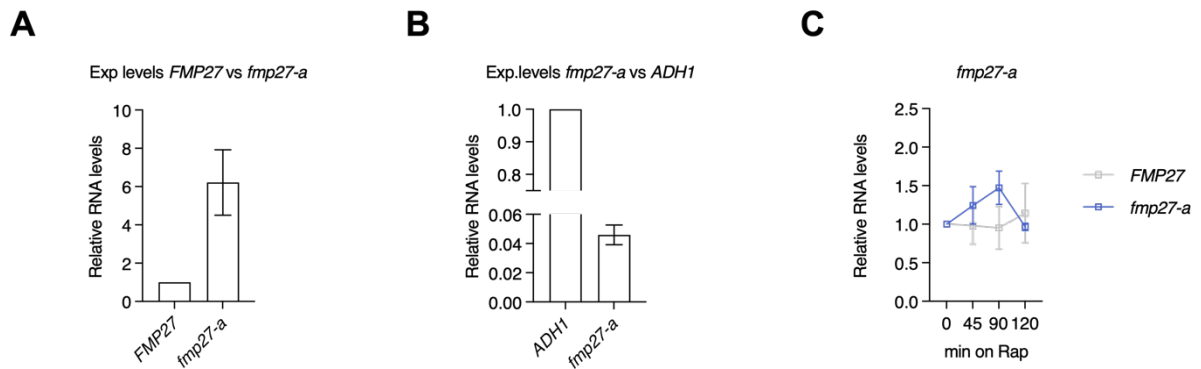
C — *fmp27-a*

Figure 3.35 - The mRNA levels of fmp27-a compared to FMP27, ADH1 and over time, upon Ceg1 depletion.

A) Plot showing the basal expression levels of fmp27-a compared to FMP27. RNA levels were normalised against the ribosomal RNA (rRNA) 25S and, for fmp27-a, normalised against the FMP27 levels arbitrarily set to 1. The error bars show the standard deviation of three independent experiments. B) Plot showing the basal expression levels of fmp27-a compared to ADH1. RNA levels were normalised against the ribosomal RNA (rRNA) 25S and, for fmp27-a, normalised against the ADH1 levels arbitrarily set to 1. The error bars show the standard deviation of three independent experiments. C) Plot showing the RNA levels measured in ceg1-AA with RT-qPCR for FMP27 and fmp27-a compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment. The error bars show the standard deviation of three independent experiments.

Next, I tested the relation between the length of *FMP27* and its stability upon Ceg1 depletion. I generated a shorter version of *FMP27* by inserting the *ADH1* promoter strategically to transcribe the last 2 kilobases of the gene and deleting the upstream 6 kilobases of coding sequence generating the mutant *fmp27-short* (Fig. 3.36A). The correct genetic manipulation was tested by PCR (Fig. 3.36B). The abundance of the mRNA resulting from the transcription of the mutant gene *fmp27-short* was assessed over time via RT-qPCR (Fig. 3.36C). Compared to the control condition, the mRNA levels of *fmp27-short*, showed an about two-fold progressive increase from 0 to 90 minutes of Ceg1 depletion to then re-adjust to the initial levels

after 120 minutes. This result indicated that the shortening of the *FMP27* gene length can transitionary affect its mRNA levels in Ceg1-depleted cells but do not reduce its stability.
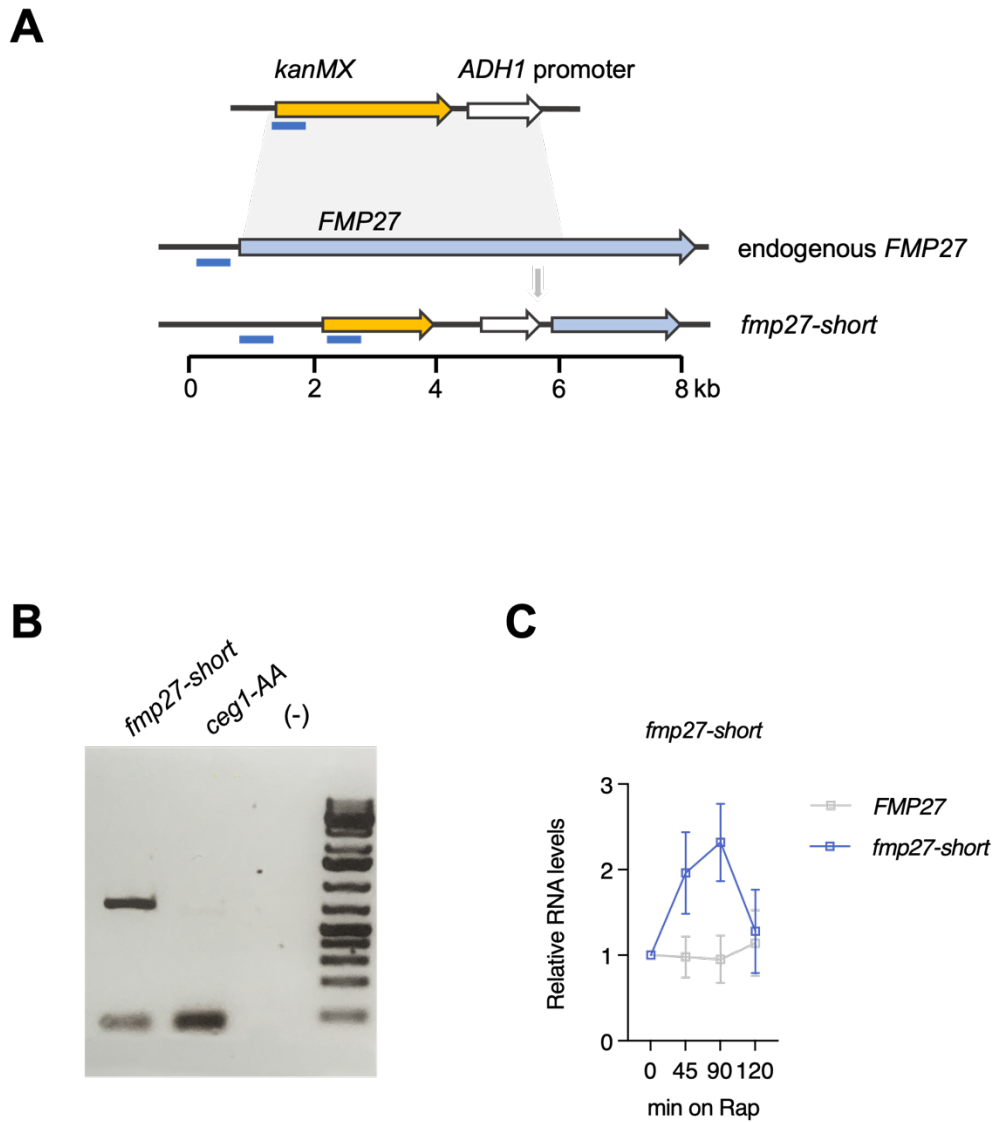
Figure 3.36 - The characterisation of the mutant fmp27-short and its expression levels upon Ceg1 depletion.

A) The scheme showing the gene targeting for the introduction of the ADH1 promoter in the coding region of FMP27 to obtain the mutant fmp27-short. The blue horizontal bars under the genetic locus depicted,

The last modification introduced in the endogenous *FMP27* involved the elements of the gene related to its transcription termination. Specifically, I introduced the strong terminator sequence of the model gene *ADH1* downstream of the coding region of the endogenous *FMP27* in the *ceg1-AA* strain background generating the mutant gene *fmp27-t* (Fig. 3.37A). The correct introduction of the terminator sequence was tested via PCR (Fig. 3.37B) and the levels of the resulting *fmp27-t* transcript measured via RT-qPCR over time upon Ceg1 depletion (Fig. 3.37C). The introduction of the *ADH1* terminator had a negligible effect on the mRNA levels of the *fmp27-t* transcript indicating that the termination of the wild type *FMP27* was similar in the efficiency to the termination induced by the *ADH1* terminator sequence and therefore no effects were appreciated on the stability of the transcript upon Ceg1 depletion.
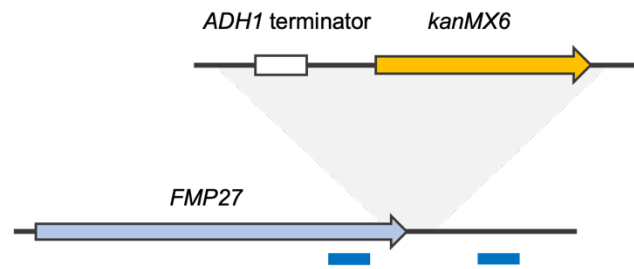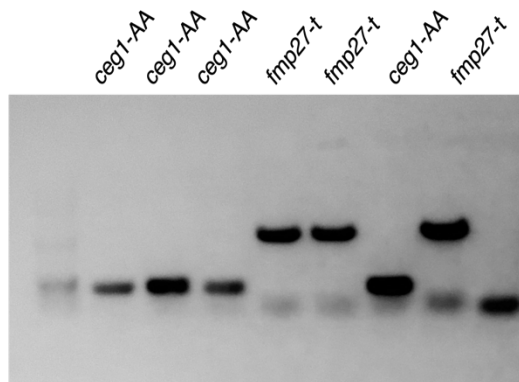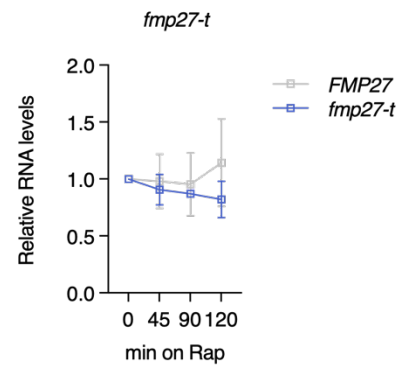
Figure 3.37 - The characterisation of the mutant fmp27-t and its expression levels upon Ceg1 depletion.

A) The scheme showing the gene targeting for the introduction of the ADH1 terminator downstream of the coding region of FMP27 to obtain the mutant fmp27-t. The blue horizontal bars under the genetic locus depicted, show the position of the PCR primers used to screen the correct introduction of the cassette. *B)* The PCR products run on 1% agarose gel showing the correct introduction of the cassette in the genomic locus targeted using the oligos located as in A. The PCR negative control is labelled with (-). *C)* Plot showing the RNA levels measured in ceg1-AA with RT-qPCR for FMP27 and fmp27-t compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment. The error bars show the standard deviation of three independent experiments.

The global degradation of the uncapped mRNAs consequent to the Ceg1 nuclear depletion induces the accumulation of the less expressed genes, possibly due to the limited access of this subgroup of RNAs to the cytoplasmic degradation system and concomitant defects in the cellular RNA buffering (see previous chapters). Due to the strict relation existing between the basal expression level and uncapped mRNA degradation/accumulation upon Ceg1 depletion, I asked whether the decrease in the *FMP27* transcription rate could affect either way its stability, generating an accumulation or a decreased abundance, upon Ceg1 depletion. Therefore, I assessed the *FMP27* mRNA abundance over time via RT-qPCR, affecting the Pol II processivity and elongation rate via 6-AU treatment (Fig. 3.38) as shown above (Fig. 3.18C). The levels of the uncapped *FMP27* remained overall stable over time upon either normal of reduced Pol II processivity and elongation rate suggesting, in this case, that the lower expression was not sufficient to induce the accumulation of *FMP27* like observed for the lower expressed genes upon Ceg1 depletion.
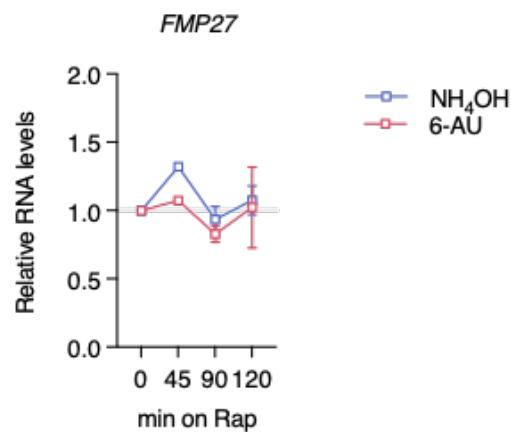
Figure 3.38 - The FMP27 stability upon Ceg1 depletion is maintained after 6-azauracil (6-AU) treatment.

RT-qPCR levels of the FMP27 mRNA in cells treated with 6-AU/NH₄OH for 30 minutes followed by DMSO/Rap treatment. The error bars show the standard deviation of three independent experiments.

To test whether the stability of the uncapped *FMP27* upon Ceg1 depletion was induced by this effect or from other uninvestigated features of the transcript, I generated the mutant *fmp27-ribozyme* (*fmp27-RZ*). The mutant was conceived to be capped and then soon de-capped via intramolecular self-cleavage in the early stages of its transcription. The resulting transcript was an uncapped mRNA with an unprotected 5'-OH end that is still a substrate of the cellular exonucleases (Doamekpor et al., 2020; Jinek, Coyle, & Doudna, 2014) and therefore suitable to investigate the degradation of *FMP27*.

The hepatitis delta virus (HDV) genome is a single-stranded RNA molecule. The HDV RNA contains some structured regions displaying catalytic activity that are named ribozymes. HDV ribozymes can induce their self-cleavage generating the scission in the RNA required in the replicative cycle of the virus (Wadkins et al., 1999). I introduced the sequence of the HDV ribozyme in the 5' UTR of the endogenous *FMP27* in the background of the *ceg1-AA* strain (*fmp27-RZ*) (Fig. 3.39A). The catalytic activity of the HDV ribozyme can be completely abolished by the point mutation C76-U (Bird et al., 2005; Perrotta, Shih, & Been, 1999), therefore I introduced the mutation in the wild type ribozyme sequence and generate the mutant strain *fmp27-C76U* as control. The introduction of the point mutation was tested via Sanger sequencing of the resulting plasmid (Fig. 3.39B). The correct introduction of the HDV ribozyme sequence in the 5' UTR of the endogenous *FMP27* was tested via PCR (Fig. 3.39C). The resulting transcript was therefore generated harbouring the m^{7}G cap, which was promptly cleaved out with a short portion of the 5' end of the nascent transcript as soon as the ribozyme sequence was transcribed and exposed to the nuclear environment. I measured the level of the self-cleaving *fmp27-RZ* transcript compared to the catalytically inactive *fmp27-C76U* via RT-qPCR (Fig. 3.39D). The levels of the uncapped transcript (*fmp27-RZ*) were reduced compared to the control (*fmp27-C76U*) of about the 64%. This result indicates that the degradation of the uncapped *FMP27* is active in the absence of the global degradation of the uncapped cellular mRNAs and might require the presence of the cap or the capping enzymes in the early phases of transcription. It can also be deduced that there are not intrinsic structural or functional features

preserving the stability of the uncapped *FMP27* in which case would have preserved its stability when uncapped via HDV ribozyme self-cleavage.
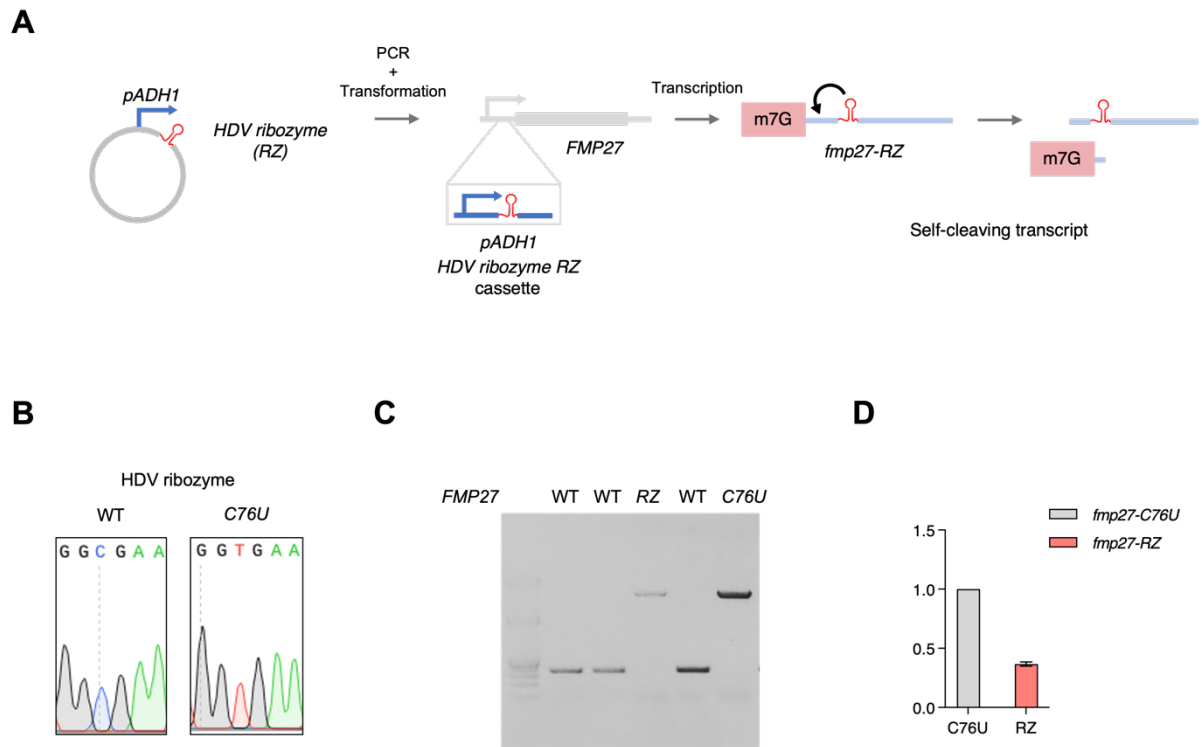
Figure 3.39 - The characterisation of fmp27-RZ and fmp27-C76U mutant and their expression levels.

A) The schematic showing the process of the generation of the fmp27-RZ mutant. The hepatitis delta virus (HDV) ribozyme sequence (RZ, red) was introduced in the 5' UTR of the endogenous gene FMP27. The resulting self-cleaving transcript was able to remove the m⁷G cap releasing a 5' uncapped product. B) The results from the Sanger sequencing performed on the plasmids used to generate the mutants fmp27-RZ and fmp27-C76U. The point mutation introduced in the HDV ribozyme sequence (WT) to generate the catalytically inactive variant C76U is shown. C) The PCR products run on 1% agarose gel showing the correct introduction of the RZ/C76U cassette in the genomic locus.

3.3.3 Discussion

This chapter pointed the attention on the stability displayed by the uncapped *FMP27* upon Ceg1 depletion and aimed to understand this phenomenon investigating the features of the gene potentially related to this observation. *FMP27* is a gene of 7887 bp, therefore largely above the average gene length observed in *S. cerevisiae* of 1.4 kb (Hurowitz and Brown, 2003). It encodes for a putative protein of unknown function observed in previous mitochondrial proteomic analysis. *FMP27* is not a highly expressed gene, it falls in the group Mid according to our arbitrary classification (Table 3.1) and therefore supposed to be more prone to the degradation upon Ceg1 depletion.

By shutting down its expression, I demonstrated that *FMP27* is not a super-stable transcript in wild type conditions (Fig. 3.31D) and therefore the stability displayed upon Ceg1 depletion is unlikely to be related to its capped pre-existing RNA copies.

I employed the temperature sensitive mutants of Ceg1 (*ceg1-63* and *ceg1-63/rat1-1*) with the aim of recapitulating previous results demonstrating the decay of *FMP27* via premature transcription termination (PTT) upon capping defects (Jimeno-González et al., 2010). Although I observed the progressive decrease of the *FMP27* mRNA levels over time at the non-permissive temperature of 37 °C in *ceg1-63* (Fig. 3.32), a certain decrease was observed in the wild type strain (WT) which did not show capping defects after the temperature shift. I demonstrated this point by the control measurement of the *ADH1* mRNA levels (Fig. 3.32) which did not display the same mRNA decrease. In our experiments, in fact, the temperature shift caused the increase of the *ADH1* mRNA levels in the WT. This result was possibly related to the burst in the energetic metabolism pathways, induced by the increased temperature, of which Adh1 is a fundamental enzyme (Postmus et al., 2008).

A crucial point raised from our experiments on the temperature sensitive mutants was related to the unexpected decrease in the mRNA levels of *FMP27* and *GAL1* in the WT strain when moved at the non-permissive temperature. The decrease in the WT was milder compared to the temperature sensitive

mutants but still consistent and sustained over time (Fig. 3.32). This effect could be caused by the response of the *GAL1* promoter, regulating the expression of *FMP27* and *GAL1*, to the temperature shift. Therefore, the mRNA decrease measured in the temperature sensitive mutants could be the effect of a combination of factors, some of which unrelated to the cap defects, affecting the stability or the expression of *FMP27* and *GAL1* in this context.

Another important outcome arose from the experiments employing the temperature sensitive mutants and involved the role of the nuclear exonuclease Rat1 in the degradation of the uncapped mRNAs. Rat1 is the yeast nuclear 5'-3' exonuclease widely shown to co-transcriptionally degrade the uncapped RNA in the transcriptional termination of the protein-coding genes (M. Kim et al., 2004; W. Luo et al., 2006). Exploiting the co-transcriptional exonucleolytic activity, Rat1 in combination with Rai1 takes part and regulates the cap quality control that ultimately causes the degradation of the mRNAs that either lack the entire m^7G or are modified with an unmethylated cap (Doamekpor et al., 2020; Jiao et al., 2010; Zhai & Xiang, 2014). The inactivation of the Rat1 catalytic activity in the temperature sensitive mutant *ceg1-63/rat1-1* did not rescue the levels of the uncapped *FMP27*, *GAL1* or *ADH1* mRNAs at the non-permissive temperature. Taken in account the effects generated by the temperature shift in WT stain on the capped *FMP27* and *GAL1*, the residual decrease in their mRNA levels observed in the *ceg1-63* and ceg*1-63/rat1-1* strains could be attributed to the capping defects induced at the non-permissive temperature. For the genes investigated (*FMP27*, *ADH1* and *GAL1*), the inactivation of Rat1 had no effect on the decrease in the mRNA abundance, following the decay pattern of the mutant with a functional exonuclease (*ceg1-63*). A milder degradation was expected in the mutant *ceg1-63/rat1-1* compared to *ceg1-63* in the scenario where Rat1 was the main or an important player in the degradation of the uncapped mRNAs. I confirmed these data investigating the mutant deleted of the Rat1-interactor Rai1. I stably deleted Rai1in the context of the *ceg1-AA* strain generating the mutant *ceg1-AA/rai1Δ*. The stable depletion of Rai1 was induced as the *rai1 null* mutation is viable contrarily to the *rat1 null* (Jiao et al.,

2010; W. Luo et al., 2006; Xiang et al., 2009). The aim was to induce the disruption of the Rai1-Rat1 complex and evaluate the concomitant effects on the degradation of the uncapped mRNAs in the Ceg1-depleted cells. Confirming the results obtained with the temperature sensitive mutants, the role of the elements of the cap quality control (Rai1 and Rat1) is only marginal also in the definition of the fate of the uncapped mRNAs upon Ceg1 depletion.

This result indicated that Rat1 has a minimal or null impact on the uncapped mRNAs levels suggesting that the capping quality control might require additional signals to occur (Dengl & Cramer, 2009). Nevertheless, an effect of the Rat1 inactivation on the *FMP27* mRNA levels at the sub-permissive temperature (34 °C) was previously demonstrated (Jimeno-González *et al.*, 2010). In this case the *GAL1::FMP27* expression could also be higher compared to the expression of the gene controlled by the endogenous promoter. The higher expression therefore could induce the degradation of *FMP27* mRNA due to the interdependency between the expression levels and the uncapped mRNA degradation discussed above.

The use of the temperature sensitive mutants at 34 °C, supports the hypothesis by which the capping enzymes might be needed even if catalytically inactivated (ceg1-63) to support the cap quality control as the Rat1-dependent degradation of *GAL1::FMP27* has been shown (Jimeno-González et al., 2010). In the case of the Ceg1-depleted cells instead, where the capping enzymes are depleted from the nucleus, the 5' tri-phosphate mRNAs generated might not be quicky degraded as not being the optimal substrate of Rat1 and Xrn1, but they could be exported in the cytoplasm and directed to other degradation pathways. The capping enzymes might be involved in the recruitment of the elements of the cap quality control (Rai1 and Rat1) on the transcriptional machinery. By direct consequence of the Ceg1 anchor away, the complex Rai1-Rat might not be recruited and therefore the cap quality control abolished allowing the nuclear export of the 5' triphosphate uncapped mRNAs.

Structural features were investigated for *FMP27* concerning the promoter, the gene length and the transcription termination (Fig. 3.35-3.37). Expressing *FMP27* under the control of the strong *ADH1* promoter, induced a six-fold increase of its mRNA levels that remained though consistently lower compared to *ADH1* possibly because of the relocation of the promoter in a different genomic locus. The resulting *fmp27-a* transcript remained stable upon Ceg1 depletion indicating that its stability was not promoter-dependent and that the increased transcription was not sufficient to rise its levels to the point to induce its degradation in the Ceg1-depleted cells following the dependence of the differential expression from the expression levels described in the previous chapters. The stability was maintained also for the mutant *fmp27-short* composed of the last 2 kb of the 3' end of the endogenous *FMP27*. Both, *fmp27-a* and *fmp27-short* displayed a transient increase in the mRNA levels up to 90 minutes of Ceg1 depletion to return to the initial levels after 120 minutes. This effect indicates that the stability of the two mutants was influenced by the Ceg1 depletion in an unexpected fashion but certainly not promoting their degradation. The introduction on the *ADH1* terminator sequence downstream of the coding sequence of *FMP27* (*fmp27-t*) had no effect on the mRNA levels of the resulting mutant suggesting that the transcription termination of *FMP27* was similar in terms of efficiency to the highly expressed *ADH1* whose mRNA levels progressively decrease upon Ceg1 depletion. This experiment allowed us to exclude the transcription termination efficiency as a determinant of the *FMP27* stability in Ceg1-depleted cells.

No difference was observed in the stability of the uncapped *FMP27* affecting its synthesis via 6-AU treatment (Fig. 3.38). This result indicated that the changes in the *FMP27* mRNA synthesis were not sufficient to alter the balance between synthesis and degradation responsible for the stability of this uncapped mRNA.

Finally, I demonstrated that the *FMP27* transcript can be degraded when not capped via the RNA self-cleavage mediated by the hepatitis delta virus (HDV) ribozyme (Fig. 3.39D). I introduced the HDV ribozyme in the 5' UTR of the endogenous *FMP27* generating the *fmp27-RZ* mutant capable of cleave

out the m^{7}G cap and a short portion of the 5' UTR generating an uncapped transcript synthesised in the presence of functional capping enzymes. This result clarified the dynamics regulating the stability or the degradation of such transcript when uncapped in different conditions.

I concluded that the stability of the uncapped *FMP27* in Ceg1-depleted cells was still a consequence of the global degradation of the uncapped mRNAs generated, which limited the access to the degradation machinery for the lowly expressed *FMP27*. In the previous chapters, transcripts with a low expression level have been shown to accumulate upon Ceg1 depletion. In the case of *FMP27*, the balance between the transcription rate and degradation might be maintained in Ceg1-depleted cells leading to a constant level of the mRNA measured. The use of the uncapped *fmp27-RZ* mutant, in combination with the structural mutants and the other experiments discussed above, confirmed that the stability shown by the uncapped *FMP27* in Ceg1-depleted cells is not mediated by structural or functional characteristics of the transcripts and reinforce the model suggested where the abundance of the uncapped mRNAs upon Ceg1 depletion is mainly dependent on the expression levels of the relative genes.

3.4 Genome wide analysis of the snoRNAs in Ceg1-depleted cells

3.4.1 Introduction

PoI II is responsible for the transcription of a variety of genes. Along with protein-coding genes (mRNA synthesis), it allows the synthesis some non-coding RNA (ncRNA) such as long non-coding RNAs (lncRNAs), small-nuclear RNAs (snRNAs) and small-nucleolar RNAs (snoRNA).

snoRNAs are a class of ncRNAs that can be found in any studied eukaryotic genome and can be distinguished in two main classes: box C/D and box H/ACA according to some conserved motifs they harbour. Their length spans between 60 and 300 nucleotides. snoRNAs are important mediators of the ribosomal RNA (rRNA) processing and nucleotide modification like 2'-O methylation and pseudo uridylation of the rRNA (Huang et al., 2022). snoRNA are mainly localised in the nucleus and the nucleoli where they can exert their functions in the ribosome biogenesis. To accomplish this defined localization, snoRNAs undergo specific pathways of maturation (Grzechnik et al., 2018).

snoRNAs form several secondary structures that are bound by specific RNA-binding proteins to generate the functional ribonucleoprotein (snoRNP) (Massenet et al., 2017).

At the 5' end, the pre-snoRNA are synthesised with a 150-200 nucleotides extension and, as any other Pol II transcript, they are modified with the addition of the m^7G cap but, contrarily to mRNAs, snoRNAs are either decapped (box C/D snoRNAs) by the RNase III (Rnt1)-mediated cleavage on the 5' extension or, especially for the box H/ACA snoRNAs, their m^7G cap trimethylated (TMG cap). The unprotected 5' end of the nascent snoRNAs are then trimmed by Rat1 (Xrn2 in mammals) to complete the 5' end maturation of the precursors (Grzechnik et al., 2018).

The transcription termination of the pre-snoRNAs is mediated by the Nrd1-Nab3-Sen1 (NNS) complex that in turn recruits the Trf4-Air2-Mtr4p Polyadenylation (TRAMP) complex responsible for the oligo-

adenylation of the pre-snoRNA 3' end. The oligo-adenylated pre-snoRNA can then undergo the 3'-5' trimming mediated by the nuclear exosome (M. Kim et al., 2006).

The Rnt1-mediated removal of the m^{7}G cap has been shown to be fundamental for the 3' end processing and localization of the box C/D snoRNAs. The persistence of the cap, indeed, was associated with the entry into the mRNA-like transcription termination pathway with consequent polyadenylation and cytoplasmic accumulation of the box C/D snoRNAs in *S. cerevisiae*, expanding the role of the m^{7}G cap recognising it as the marker for the identification of different Pol II-transcribed RNAs (Grzechnik et al., 2018).

To further investigate the role of the m^{7}G cap and the capping enzymes in the gene expression of the Pol II products, I employed the anchor away system for the depletion of the capping enzymes and assessed their role in the expression of the snoRNAs in yeast.

3.4.2 snoRNAs are not affected by the lack of the m^{7}G cap

The snoRNAs are a class of functional non-coding RNAs transcribed by the Pol II. As for any Pol II-transcribed RNA, snoRNAs precursors are modified by the synthesis of an m^{7}G cap at their 5'end. Unlike protein-coding genes though, snoRNAs undergo a different termination step and subsequent maturation to be functional (Kufel & Grzechnik, 2019). A crucial role for the general snoRNA fate, is the removal of the m^{7}G cap mediated by Rnt1 as affecting this process, the snoRNAs undergo a mRNA-like processing affecting their cellular levels and localization (Grzechnik et al., 2018)

snoRNAs represent therefore a class of Pol II transcript remaining stable in the cell even harbouring an unprotected 5' end. This stability is ensured by the formation of RNA secondary structures and the binding of RNA binding proteins that brings to the formation of the snoRNP.

To expand our view on the effects of the RNA capping on the expression of the Pol II products, I tested the stability of the snoRNA class in Ceg1 depleted cells.

I used the RNA-seq dataset to provide a systematic analysis of the RNA levels of the 77 snoRNAs fund in the genome of *S. cerevisiae*. The overall snoRNA levels resulted unchanged in Ceg1-depleted cells compared to the control (Fig. 3.40A) confirming the stability of such Pol II product when not capped. Only two snoRNAs resulted differentially expressed from the analysis; snoRNA *snR18*, intronic of the gene Elongation Factor Beta (EFB1), displayed log2 fold change of -2.1 (padj = 2×10^{-2}) possibly following the differential expression of the host gene (log2 fold change -2.7, padj = 2.16×10^{-12}) and the snoRNA *snR128* with log2 fold change of -1.2 (padj = 2×10^{-2}).

A

ceg1-AA (DMSO) vs ceg1-AA (Rap)

snoRNAs

snR4

snR128

snR50

snR18

-cap

WT

B

snR4

snR50

DMSO 603

Rap 603

DMSO 424

Rap 424

snR18   EFB1

snR128

DMSO 72

Rap 72

DMSO 117

Rap 117
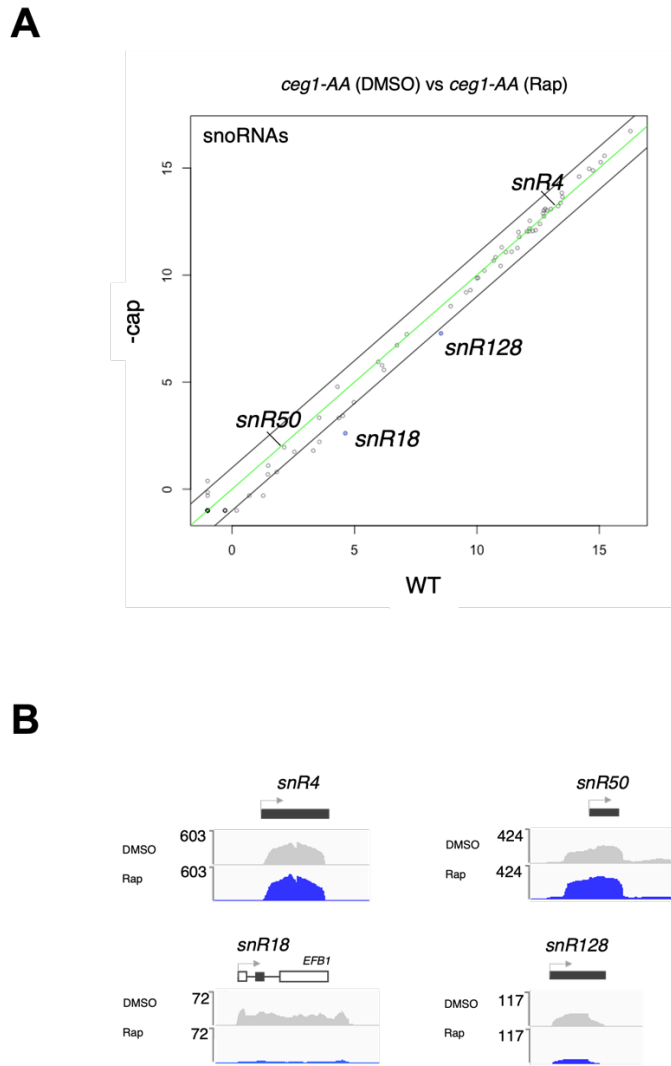
Figure 3.40 - The snoRNA expression is not affected upon Ceg1-depletion.

*A) The plot showing the differential expression of snoRNA genes in cells depleted of Ceg1 for 45 minutes (-cap) compared to the control (WT). The results are the average of two replicates. Each dot represents one gene. The gene whose abundance was decreased with log2 fold change < -1 and padj < 0.05 upon Ceg1 depletion are in blue. The genes whose abundance was increased with log2 fold change >1 and padj < 0.05 upon Ceg1 depletion are in red. Genes labelled in grey were considered not differentially expressed. The green line shows the zero-change line, the black lines delimit the ±1 log2 fold change. **B)** The genome viewer tracks showing RNA-seq reads in ceg1-AA (DMSO) and ceg1-AA (Rap) for the snoRNAs snR4 and snR50, not affected; snR18, intronic, decreased and snR128 decreased. The thick green boxes represent the snoRNA genes. The thin lines are introns, and the empty boxes show the exons of the host gene EFB1.*

166

For the so-called box C/D snoRNAs, the m^{7}G cap removal is an important step to ensure an efficient 3' end processing. snoRNAs are synthesised as extended precursors whose ends are trimmed by Rat1 (5'-3') and the nuclear exosome (3'-5') to generate the functional transcript (M. Kim et al., 2006; Kufel & Grzechnik, 2019). To test whether the m^{7}G cap synthesis or the capping enzymes are required to guarantee the Rnt1-mediated cleavage and the consequent correct 5' and 3' end processing, I inspected the coverage profile obtained from the alignment of the RNA-seq reads. The Ceg1 depletion, and therefore the lack of the m^{7}G cap did not affect neither the 5' nor the 3' end formation of the snoRNAs (Fig. 3.40B) as I could not observe 5' or 3' end extended species of the transcripts. This data suggests that, even if the cap removal is essential for the 5' trimming and the 3' end formation, as well as for the downstream maturation and processing of these RNA species, the complete lack of the capping enzyme and by consequence the lack of the m^{7}G cap, are not needed to ensure the correct trimming of the snoRNAs.

3.4.3 *RRP6* deletion does not affect the non-capped snoRNAs stability and 5' end processing

An important step in the fate of the snoRNAs is the trimming of the extended precursors mediated by the nuclear exosome and the associated 3'-5' exonuclease Rrp6. To further investigate the role of the m^{7}G cap in the expression of the Pol II product, I generated the mutant harbouring the deletion of the *RRP6* gene in the background of the *ceg1-AA* strain (*ceg1-AA/rrp6Δ*). The endogenous *RRP6* locus was disrupted by the insertion of the aminoglycoside phosphotransferase coding sequence (*kanMX*) used for the selection of the transformant cells due to the resistance to kanamycin and neomycin its products confer (Fig. 3.41A). I tested the correct insertion of the marker gene via PCR (Fig. 3.41B). Next, I performed growth spot assay to test the cell viability upon *RRP6* deletion and in parallel I tested the activation of the anchor away system for Ceg1 in the *ceg1-AA/rrp6Δ* strain, growing cells on media enriched with rapamycin at the final concentration of 1 μg/mL (Fig. 3.41C). The mutant *ceg1-AA/rrp6Δ* displayed a viable phenotype on media

containing DMSO, showing only mild growth defects compared to the parental strains HHY168 and *ceg1-AA*. In the presence of rapamycin, the strain exhibited a lethal phenotype as consequence of the Ceg1 depletion. I sequenced the *ceg1-AA/rrp6Δ* strain upon 45 minutes of rapamycin treatment and the relative control condition (DMSO).
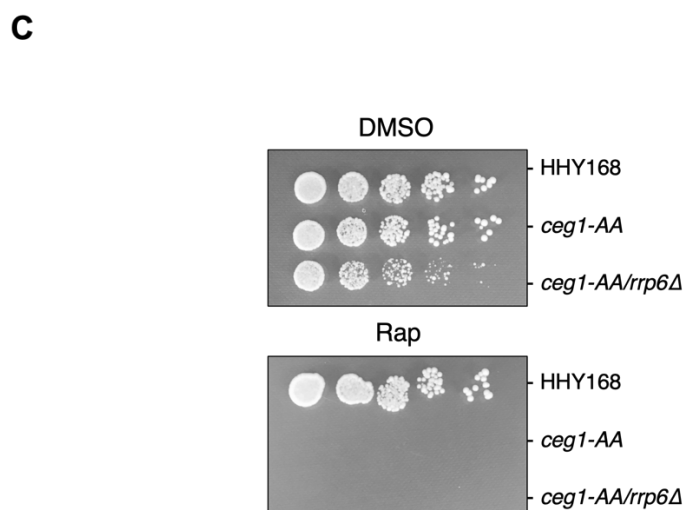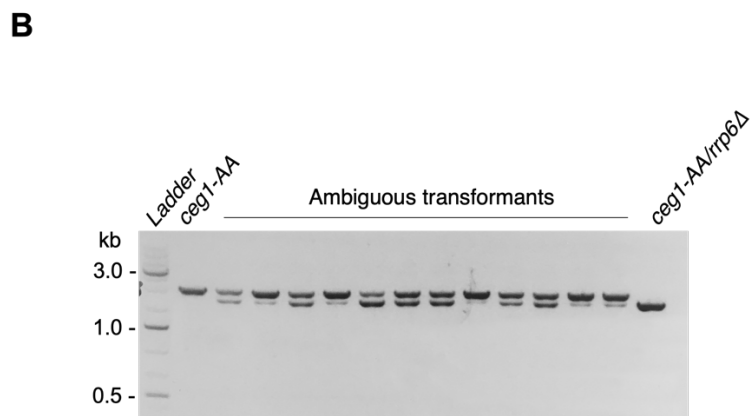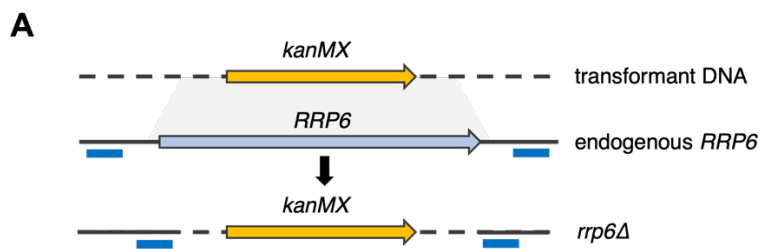
Figure 3.41 - The generation and characterisation of the ceg1-AA/rrp6Δ (rrp6Δ) mutant.

A) The scheme showing the gene targeting inducing the deletion of the RRP6 locus from the yeast genome and concomitant introduction of the kanMX selection marker gene. The blue horizontal bars under the

*genetic locus depicted, show the position of the PCR primers used to screen the correct introduction of the transgene. **B)** The PCR products run on 1% agarose gel showing the correct introduction of the transgene kanMX in the genomic locus targeted using the oligos located as in A. From the left to the right: molecular ladder, the parental strain ceg1-AA, ambiguously transformed clones and the mutant ceg1-AA/rrp6Δ. **C)** The spot growth assay of the strains HHY168, ceg1-AA and ceg1-AA/rrp6Δ. Serial dilutions of the liquid culture of each strain spotted on YPD media enriched with DMSO or rapamycin (Rap). The strain ceg1-AA is viable on DMSO but shows a lethal phenotype on Rap as consequence of the activation of Ceg1 nuclear depletion. The strain ceg1-AA/rrp6Δ shows a lethal phenotype on Rap as consequence of Ceg1 depletion.*

From the sequencing analysis, the knockout of the *RRP6* expression was further confirmed measuring the

normalised reads expressed as TPM counted in the *ceg1-AA* strain compared to the *ceg1-AA/rrp6Δ* mutant

DMSO treated (Fig. 3.42). As expected, the gene targeting completely abolished the expression of the

RRP6 locus bringing the TPM from an average of 195 in *ceg1-AA* to zero in *ceg1-AA/rrp6Δ*.
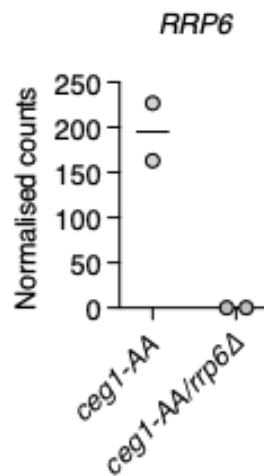
Figure 3.42 - The knockout of the RRP6 gene.

The plot showing the expression levels expressed as normalised counts (TPM) for RRP6 in ceg1-AA and
ceg1-AA/rrp6Δ in control conditions (DMSO). The depletion of the single copy of the endogenous RRP6
is sufficient to completely abrogate its expression generating 0 (zero) reads in the RNA-seq experiment.

I tested the cellular snoRNAs levels in the *ceg1-AA/rrp6Δ* (-Rrp6) compared to the *ceg1-AA* strain under

control conditions (WT) employing the RNA-seq dataset (Fig. 3.43A). As expected, the snoRNAs

abundance remained unchanged for most of the snoRNAs, except for seven out of 77 transcripts (Table

3.5 and Fig. 3.43A) suggesting that the stability of this subset of snoRNAs might be affected either

positively or negatively by the lack of the exonuclease Rrp6. I also investigated the coverage profile of the

reads aligned to the yeast genome during the phase of RNA-seq data analysis and observed the expected enrichment in the 3' extended snoRNA isoforms (Fig. 3.43B) caused by the defective processing of the snoRNAs in *RRP6* deleted cells (Fox et al., 2015; Kufel & Grzechnik, 2019).

Interestingly, the two intronic snoRNAs reported to be differentially expressed in the mutant *ceg1-AA/rrp6Δ*: *snR18* and *snR44*, displayed a log2 fold change with the opposite sign compared to their host genes, respectively *EFB1* and *RPS22B* (Fig. 3.43C), possibly a compensatory effect to the misregulation of the host gene expression. As described above, *snR18* was also differentially expressed in Ceg1-depleted cells with a functional Rrp6, where it was shown to be downregulated following the decrease of its host gene *EFB1*. In the Rrp6-depleted cells, instead, *snR18* was upregulated (Table 3.5) while *EFB1* was downregulated (log2 fold change -1.1) suggesting a more direct effect of the lack of Rrp6 on *snR18* compared to the effects observed in the Ceg1-depleted cells where it was possible that the differential expression of the snoRNA was related to the differential expression of *EFB1*.

Table 3.5 - The snoRNAs differentially expressed in ceg1-AA/rrp6Δ. *The table shows the name and the differential expression of the snoRNAs, the statistics about the differential expression are expressed as padj (adjusted p-value). The snoRNA type and their location are shown.*

Name	Log2 FC	padj	Type	Intronic
snR44	-2.2	2.9×10^{-19}	box H/ACA	Yes
snR49	-1.2	1×10^{-3}	box H/ACA	No
snR86	-1.1	3.3×10^{-5}	box H/ACA	No
snR13	1.4	2×10^{-8}	box C/D	No
snR18	1.4	1×10^{-2}	box C/D	Yes
snR77	2.1	5×10^{-3}	box C/D	No
snR73	2.8	5.3×10^{-5}	box C/D	No
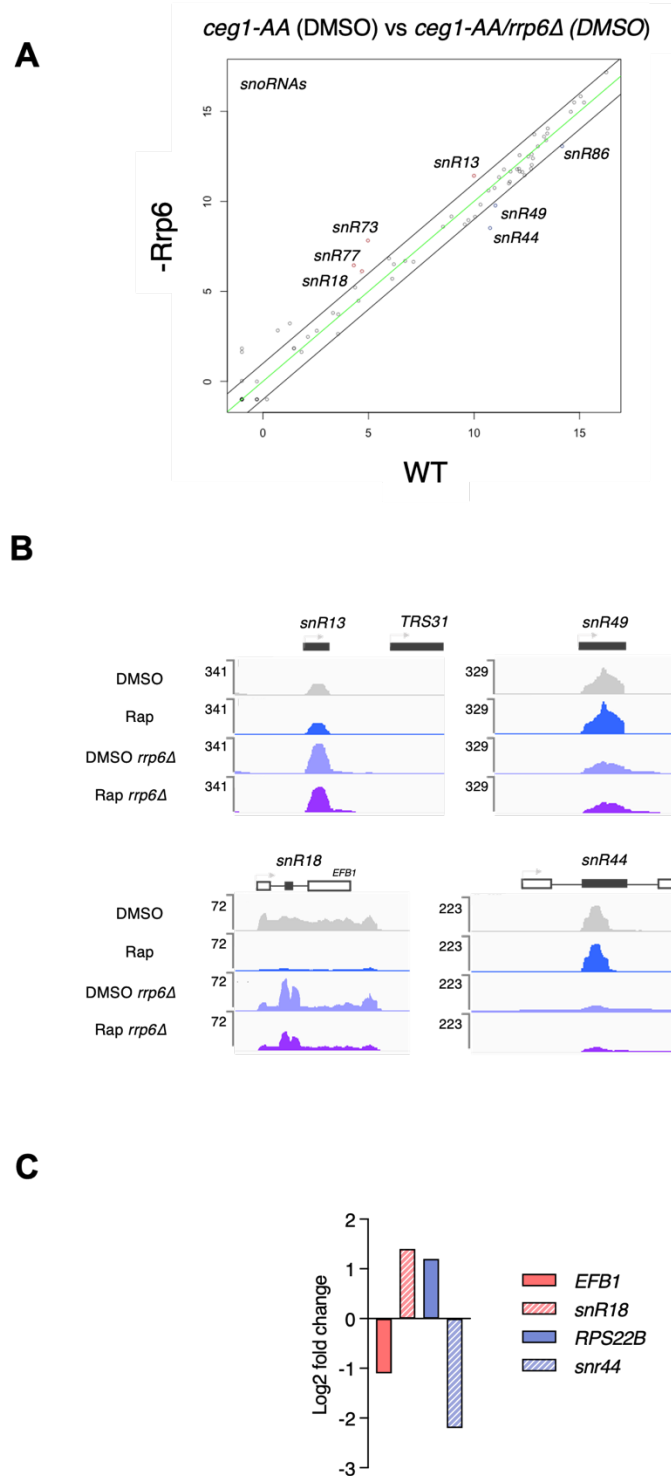
Figure 3.43 - The snoRNA expression is not affected in the RRP6 mutants.

A) The plot showing the differential expression of snoRNA genes in cells carrying the deletion of the RRP6 gene (-Rrp6) compared to the control (WT). The results are the average of two replicates. Each dot

represents one gene. The gene whose abundance was decreased with log2 fold change < -1 and padj < 0.05 are in blue. The genes whose abundance was increased with log2 fold change >1 and padj < 0.05 are in red. Genes labelled in grey were considered not differentially expressed. The green line shows the zero-change line, the black lines delimit the ±1 log2 fold change. **B)** *The genome viewer tracks showing RNA-seq reads in ceg1-AA (DMSO), ceg1-AA (Rap), ceg1-AA/rrp6Δ (DMSO) and ceg1-AA/rrp6Δ (Rap) for the snoRNAs snR13, 49, 18 and 44. The thick green boxes represent the snoRNA genes. The thin lines are introns, and the empty boxes show the exons of the host genes EFB1 and RPS22B.* **C)** *The differential expression of the host gene (solid fill) and the relative intronic snoRNA (stripes).*

Some 3' extended snoRNAs species have been shown to be stable and functional (Fox & Mosley, 2016). To assess the role of the cap and the capping enzymes in the early phases of snoRNA transcription before the decapping maturation step, I tested their stability in Ceg1-depleted cells.

I compared the global snoRNA expression levels in the *ceg1-AA/rrp6Δ* strain in the presence (DMSO) or depletion (Rap) of Ceg1 (Fig. 3.44). In this case none of the 77 snoRNA found in the genome of *S. cerevisiae* was differentially expressed upon 45 minutes of Ceg1 depletion compared to the control. This result indicated that the presence of the capping enzyme and the m^7G cap, even if it is intended to be co-transcriptionally removed in the subsequent stages of maturation, is not required in the early stages of transcription to guarantee the stability of the 3' end extended snoRNAs generated by the depletion of Rrp6.
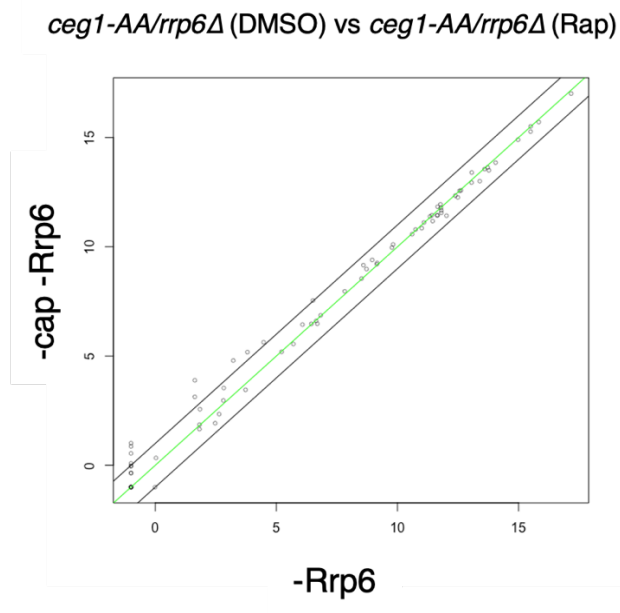
Figure 3.44 - The snoRNA expression is not affected in the RRP6 mutants depleted of Ceg1.

The plot showing the differential expression of snoRNA genes in Ceg1 depleted cells carrying the deletion of the RRP6 gene (-cap -Rrp6) compared to the control (-Rrp6). The results are the average of two replicates. Each dot represents one gene. The gene whose abundance was decreased upon Ceg1 depletion with log2 fold change < -1 and padj < 0.05 are in blue. The genes whose abundance was increased upon Ceg1 depletion with log2 fold change >1 and padj < 0.05 are in red. Genes labelled in grey were considered not differentially expressed. The green line shows the zero-change line, the black lines delimit the ±1 log2 fold change.

3.4.4 Discussion

In this chapter, the role of the m^{7}G cap and the capping enzymes in the expression of the snoRNAs was further investigated. The class of the snoRNAs represent an essential group of functional Pol II transcripts and therefore it is important to understand the mechanisms governing their stability and expression.

I employed the anchor away system to induce the fast depletion of the capping enzyme Ceg1 and, as shown in the previous chapter, efficiently obtain capping defects in *S. cerevisiae*.

The systematic analysis of their expression levels revealed that the lack of the capping enzymes and, by direct consequence, the lack of the m^{7}G cap in the early stage of transcription, does not affect the stability of the snoRNA (Fig. 3.40A). This result was expected at least for the box C/D class considering previous results obtained with the use of the temperature sensitive mutants for the capping enzymes (Grzechnik et al., 2018). The conformation of the snoRNAs can represent a crucial feature conferring stability to the unprotected ends of the transcript. Indeed, the constitution of secondary structures and the binding of the specific RNA binding proteins in the assembling of the snoRNP, can shield the uncapped snoRNA from the exonucleolytic degradation. In fact, box C/D snoRNAs have been shown to require the 5' decapping step of maturation to be functional (Grzechnik et al., 2018). Our results employing the Ceg1 nuclear depletion added an important piece of information to the description of these dynamics indeed, I posed the basis to hypothesise that the lack of the m^{7}G cap on the nascent snoRNA is sufficient to guarantee its correct 5' end processing (Fig. 3.40B) and it is not strictly required to stimulate the important Rnt1-mediated cleavage and subsequent Rat1 trimming as I did not recognise aberrant 5' extended species in Ceg1-depleted cells.

A strict connection between the 5' and 3' end snoRNA maturation has also been shown (Grzechnik et al., 2018; Kufel & Grzechnik, 2019). Our results confirm the marginal role exerted by the presence of either the cap or the capping enzymes in the early phases of the snoRNA transcription in this concern. The analysis of the gene coverage obtained by the RNA-sequencing did not bring to the light 3' extended

snoRNA in Ceg1-depleted cells (Fig. 3.40B). To further confirm this outcome, I investigated the snoRNA levels and processing in Ceg1-depleted cells where the function of the exonuclease Rrp6 was abrogated inducing 3' end processing defects in the snoRNAs (Fig. 3.43A and B). Again, the Ceg1 depletion had no visible effect on the stability and the processing of the 3' extended snoRNAs.

Overall, the results discussed here indicate that even if the decapping is an important step in the metabolism of the snoRNAs, its initial synthesis as well as the presence of the capping enzymes on the transcriptional machinery are not required to ensure the functional processing of this class of Pol II transcripts. Rather, the m^7G absence is per se sufficient to permit all the downstream events bringing to the correct maturation of the snoRNA. The observations reported here are certainly prelaminar and a deeper investigation would certainly shed light on more interesting details. Nevertheless, in this chapter, I described outcomes in line with the basis posed to clarify that the direct role of the m^7G cap cannot be restricted the passive RNA protection from the nucleolytic degradation (Grzechnik et al., 2018) but rather it is plausible to consider that such modification is applied to any Pol II transcript regardless its function and class and, subsequently, its presence or absence will direct the different transcripts to their relative pathways to globally optimise the cellular RNA sorting.

3.5 *CUT542 (aka NEL025c)*: an example of ncRNA affected in Ceg1-depleted cells

3.5.1 Introduction

The Pol II-dependent transcription has been shown to produce a number of ncRNAs whose function is yet to be clarified. Some of these transcripts rise from the unannotated regions and a subgroup of them can be detected by the current transcriptomic methods, only upon certain conditions involving the inactivation of the cellular RNA degradation systems. The class of ncRNAs investigated in this chapter is called cryptic unstable transcripts (CUTs). CUTs have been observed as short and transient RNA species upon the inactivation of the Rrp6 3'-5' exonuclease associated with the nuclear exosome (Xu et al., 2009). As any other Pol II product, CUTs are capped by the addition of an m^7G cap. The transcription termination of these ncRNAs is mediated by the Nrd1-Nab3-Sen1 (NNS) pathway which in turn, induces the CUTs 3'end oligo-adenylation TRAMP-mediated and consequent targeting and degradation accomplished by the nuclear exosome. Therefore, CUTs are maintained at low concentration and remain mainly nuclear transient species (Thiebaut et al., 2006; Wyers et al., 2005). The function of these ncRNAs is yet to be clarified but some of them are involved in the regulation of gene expression, the aim of their synthesis is still debated if being a consequence of the Pol II pervasive transcription of the genome or if they are synthesised from functional loci (Colin, Libri, & Porrua, 2011; Neil et al., 2009).

In this chapter I will describe the effects of the Ceg1 depletion on the well characterised *CUT542* (aka *NEL025c*) expression.

3.5.2 Ceg1 depletion stabilises NEL025c RNA inducing its accumulation over time

The model CUT *NEL025c* has been largely employed in the investigation of the CUTs RNA metabolism, for the control and the processing of this ncRNA class. *NEL025c* is transcribed by the Pol II and therefore

it's subjected to 5' capping (Thiebaut et al., 2006; Wyers et al., 2005). Therefore, I employed our system for the rapid depletion of the capping enzyme Ceg1 and investigated the role of the m^7G cap on the CUT's expression. I have previously shown that protein-coding genes expression is affected in Ceg1-depleted cells according to their basal expression levels. On the other hand, the lack of the capping enzymes in the nucleus and the consequent capping defects, have only a minimal or null impact on the abundance and the 5'/3' end processing of the snoRNAs. Investigating the effects of the capping defects on CUTs, can further expand the knowledge of the function and the pathways involving the m^7G cap on gene expression. I measured the RNA levels of *NEL025c* over time upon Ceg1 depletion via RT-qPCR (Fig. 3.45A). I observed that the lack of m^7G cap strongly stabilised *NEL025c* that showed an up to seven-fold increase in its RNA levels. The accumulation was not observed upon the depletion of the CBC-component Cbp20 in the *cbp20-AA* strain (Fig. 3.45A). This result confirmed that the accumulation was caused by the lack of the m^7G cap and not by the consequent CBC defects. To test whether the accumulation of the uncapped *NEL025c* was specific for the Ceg1-depleted cells or a common feature of the uncapped transcript, I assessed via RT-qPCR its levels in the temperature sensitive mutants for Ceg1 (ceg1-63 and ceg1-63/rat1-1) moved to the non-permissive temperature of 37 °C (Fig. 3.45B). Similarly to the *ceg1-AA* strain, the activation of the capping defects in the *ceg1-63* mutant induced the strong and progressive accumulation of *NEL025c* over time. A mild accumulation was observed in the WT strain after 45 minutes at 37 °C. In these conditions, the capped *NEL025c* levels reached the value of about 1.8 and remained stable up to 120 minutes.
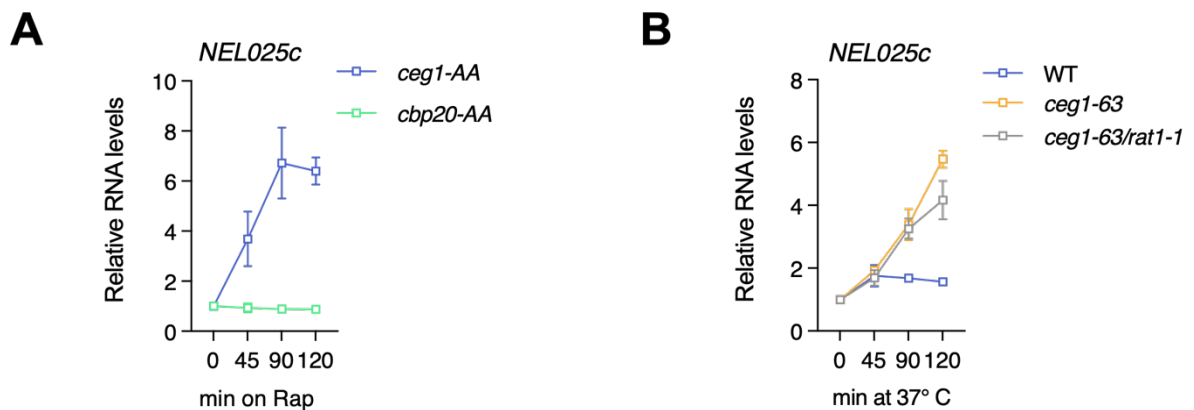
Figure 3.45 - The NEL025c RNA levels increase upon Ceg1 inactivation but not upon CBC depletion.

A) Plot showing the RNA levels measured with RT-qPCR for NEL025c compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment to deplete Ceg1 or Cbp20. The error bars show the standard deviation of three independent experiments. B) Plot showing the RNA levels measured with RT-qPCR for NEL025c compared to control condition (25 °C) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: 25 °C, then 45, 90 and 120 minutes at 37 °C. The error bars show the standard deviation of three independent experiments.

CUTs RNA levels are maintained low by their constant degradation mediated by the nuclear exosome. Therefore, to better understand the accumulation of *NEL025c* observed in the Ceg1 depleted cells and in the temperature sensitive mutants, I investigated the effects of capping defects in cells lacking the functional Rrp6, the 3'-5' exonuclease associated to the nuclear exosome whose defects were shown to induce an accumulation in the cellular CUTs levels as a consequence of their stabilization (Marquardt, Hazelbaker, & Buratowski, 2011; Xu et al., 2009). I expanded the analysis investigating the fate of the uncapped *NEL025c* upon the inactivation of the other either nuclear or cytoplasmic RNA degradation pathways. I employed the temperature sensitive mutant *ceg1-63/rat1-1* and the *ceg1-AA/xrn1Δ* strain both defective for the 5'-3' degradation systems.

I first confirmed the effects of the deletion of *RRP6* and *XRN1* genes (*rrp6Δ* and *xrn1Δ*) on the basal levels of capped *NEL025c*. I employed the strain *ceg1-AA/rrp6Δ* and *ceg1-AA/xrn1Δ* to measure via RT-qPCR the RNA levels of *NEL025c* compared to *ceg1*-AA (WT), all strains were grown in the presence of DMSO therefore without introducing capping defects (Fig. 3.46A). I observed that the lack of the functional Rrp6 caused a seven-fold accumulation of the capped *NEL025c*, this result was expected and in line with previous studies (Xu et al., 2009) and confirmed the inactivation of the nuclear exosome-mediated degradation system. The inactivation of Xrn1 induced the increase of the basal expression levels for *NEL025c* limited to an about two-fold change increase. Strikingly, the inactivation of Rrp6 and Xrn1 abolished the accumulation of the uncapped *NEL025c* upon Ceg1 depletion (Fig. 3.46B). The same effect was not recapitulated in the temperature sensitive mutant *ceg1-63/rat1-1* where the inactivation of the nuclear exonuclease Rat1 had a minimal effect on the accumulation of the uncapped *NEL025c* which ultimately accumulated over time at the non-permissive temperature (Fig. 3.45B). These results indicate a role of Rrp6 and Xrn1 in the regulation of the levels of the uncapped *NEL025c* and at the same time exclude the Rat1 nuclear exonuclease from the circuit.
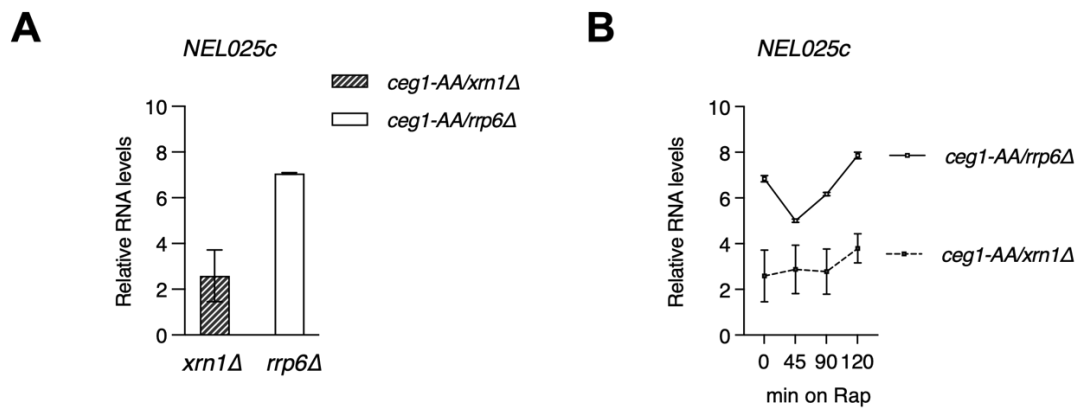
Figure 3.46 - The deletion of XRN1 and RRP6 induce the increase in the basal levels of NEL025c but abrogates its accumulation upon Ceg1 depletion.

A) Plot showing the RNA levels measured with RT-qPCR for NEL025c in the ceg1-AA/xrn1Δ and ceg1-AA/rrp6Δ normalised against the ribosomal RNA (rRNA 25S) and compared to WT (ceg1-AA on DMSO). The error bars show the standard deviation of three independent experiments. B) Plot showing the RNA levels measured with RT-qPCR for NEL025c upon Ceg1 depletion compared to control condition (DMSO) and normalised against uncapped ribosomal RNA (rRNA 25S). Timepoints represent 0: no rapamycin (Rap), then 45, 90 and 120 minutes of Rap treatment to deplete Ceg1 or Cbp20. The error bars show the standard deviation of three independent experiments.

As observed for the mRNAs, the differential expression in the Ceg1-depleted cells is a function of the basal expression levels. Therefore, I hypothesised that the low expression levels of *NEL025c* were the cause of its accumulation in our experiments upon Ceg1 depletion and asked whether reducing its synthesis further via 6-AU treatment could enhance the *NEL025c* accumulation. To answer this question, I treated *ceg1-AA* cells with 6-AU for 30 minutes to induce Pol II processivity and elongation rate defects (NH$_4$OH was used as negative control) and then subjected to rapamycin treatment to activate the Ceg1 anchor away. I measured the levels of *NEL025c* over time via RT-qPCR (Fig. 3.47). I observed that slower

RNA synthesis did not affect the overall *NEL025c* accumulation upon Ceg1 depletion as the accumulation

pattern of cell treated with 6-AU mirrored the control.
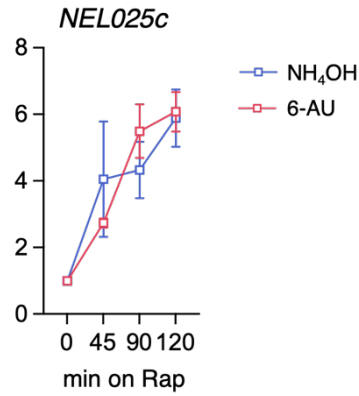
Figure 3.47 - 6-azauracil (6-AU) treatment does not affect the accumulation of the uncapped NEL025c in Ceg1-depleted cells.

RT-qPCR levels of the NEL025c mRNA in cells treated with 6-AU/ NH₄OH for 30 minutes followed by DMSO/Rap treatment. The error bars show the standard deviation of three independent experiments.

3.5.3 Discussion

The Pol II-mediated transcription generates a variety of RNA species, from the mRNAs coding for proteins to a plethora of ncRNAs either stable and functional, often essential (e. g. snRNAs and snoRNAs), or transient and poorly understood like CUTs (Carter & Drouin, 2009). The m^{7}G cap is added to all Pol II products and can be retained or removed to obtain the mature transcript according to the gene transcribed and its relative functions (Grzechnik et al., 2018; Topisirovic et al., 2011a). CUTs are a class of ncRNAs synthesised by Pol II, they are a transient group of RNAs generated from the pervasive genomic transcription started from the widespread nucleosome free regions (NFR) fund in the yeast genome. They share the promoter region, often with protein-coding genes or are generated from the anti-sense transcription of other annotated loci (Xu et al., 2009). The biological meaning of the existence of CUTs is yet to be clarified as only for a few of those a function has been described (Marquardt et al., 2011). CUTs are capped and therefore, to expand the knowledge on the m^{7}G cap function on gene expression, it is important to understand the role that this RNA modification has on this class of ncRNA. This, in turn, will help to clarify the possible functions CUTs might exert. In this chapter I have shown that the lack of the m^{7}G cap induce the accumulation of the model CUT *NEL025c* (Fig. 3.45A and B). The effect was observed employing the Ceg1 nuclear depletion, therefore in the absence of the capping enzyme in the nucleus, but also in the temperature-sensitive mutant *ceg1-63* where the capping enzyme is catalytically inactivated and degraded (Fresco & Buratowski, 1996) at the non-permissive temperature. The lack of the nuclear CBC did not exert the same effect on the levels of the capped *NEL025c* suggesting that the differential expression was directly related to the 5' cap and presumably mediated by the pathways involving its functions. CUTs cellular levels are controlled by their constant degradation operated by the nuclear exosome (Marquardt et al., 2011; Thiebaut et al., 2006). The data reported here showed that the lack of the m^{7}G cap is sufficient to permit *NEL025c* to evade this control and strongly accumulate in cap-defective cells. Intriguingly, the inactivation of the 3'-5' exonuclease associate with the nuclear exosome,

Rrp6 or the cytoplasmic 5'-3' exonuclease Xrn1 abolished the accumulation of the uncapped *NEL025c* together with the increase of its basal levels (Fig. 3.46A and B). This observation directly links the two exonucleases and/or the basal overexpression with the accumulation of the uncapped CUT *NEL025c*. At the same time, the experiments with the temperature-sensitive mutants (*ceg1-63/rat1-1*), excluded the nuclear exonuclease Rat1 from the control of the levels of *NEL025c* in cap-defective cells (Fig. 3.45B). The transcriptional termination and the nuclear degradation of the CUTs are linked by the NNS termination pathway. The failure of the NNS-mediated transcription termination induces Pol II readthrough with consequent generation of extended CUTs (eCUTs) species. eCUTs are polyadenylated and are committed into the mRNA-like metabolism and degraded by the Xrn1 cytoplasmic activity (Marquardt et al., 2011).

In light of the data reported here, along with the insight arose from the study on the uncapped mRNAs in Ceg1- depleted cells is possible to speculate as follow: i) the m^7G cap might be required for the CUT to be terminated via the NNS complex. The elements of the NNS complex indeed have been shown to interact with the CBC (Vasiljeva & Buratowski, 2006) therefore the lack of 5' cap might affect this interaction and eventually retard the transcription termination NNS-mediated. This would induce the synthesis of more copies of its 3' end extended species (eNEL025c) which could be transported to the cytoplasm following a mRNA-like fate. In the cytoplasm the uncapped *eNEL025c*, being a low expressed RNA specie, will be accumulated following the dependency between the basal expression and the abundance redistribution described for the uncapped mRNA in Ceg1-depleted cells. This hypothesis is still in line with our data where I did not appreciate the accumulation of *NEL025c* upon CBC depletion (*cbp20-AA*). In this strain indeed, the e*NEL025c* is capped and therefore its RNA levels can be buffered in the cytoplasm by the action of Xrn1 degradation. ii) The inactivation of the exonucleases Rrp6 and Xrn1 abolished the accumulation of the uncapped *NEL025c* in Ceg1-depleted cells inducing an overall increase of its basal expression levels. In *ceg1-AA/rrp6Δ* strain, uncapped *eNEL025c* might be exported in the

cytoplasm and not be subjected to the physiological levels buffering due to the lack of cap but rather fall in the redistribution related to the basal gene expression. In this case, the increased basal expression induced by the *rrp6Δ* mutation will bring to a balance between its synthesis and degradation which will not allow the detection of a further accumulation induced by the lack of cap. Similarly, in the condition missing the Xrn1 activity, the increased basal expression levels of the uncapped *eNEL025c* might lead to the same balance between synthesis and degradation masking its accumulation upon Ceg1 depletion. In this scenario it is possible to explain the outcomes of the expression of the uncapped *NEL025c* following the rules observed for the differential expression of the uncapped mRNA in Ceg1-depleted cells discussed before. A direct analysis of the isoforms and their cellular distribution will clarify the dynamics regulating the expression of the uncapped *NEL025c* to confirm or deny these hypotheses. The systematic application of the methods would eventually provide insights on possible common mechanisms regulating the expression of cellular CUTs and the role of the m^{7}G cap in this concern.

3.6 Uncapped mRNAs are not prematurely terminated in Ceg1-depleted cells

3.6.1 Introduction

Eukaryotic cells have developed their complexity with the specialization and the compartmentalisation of molecules and organelles. This complexity is also achieved via the elaborate maturation pathways involving RNAs. A direct consequence of such a finely orchestrated sequence of reactions is that defective RNAs are continuously produced by the cellular synthetic apparatus. The accumulation of aberrant RNA species has been shown to negatively affect the cell physiology and therefore control mechanisms have evolved in parallel (Bresson & Tollervey, 2018).

The addition of the m^{7}G cap at the 5' end of the Pol II transcripts is an RNA modification shown to br subjected to a quality control mechanism. Cap quality control, ensures the rapid clearance of uncapped or unmethylated capped RNAs (Jiao et al., 2010; Zhai & Xiang, 2014). The quick degradation of cap-defective transcripts is achieved by the Rai1-Rat1 complex. Rai1 exerts either decapping and pyrophospohydrolase activity therefore it can remove unmethylated caps or modify 5' triphosphate uncapped RNAs into 5' monophosphate RNAs. Both reactions generate a 5' monophosphate RNA product which is substrate for the main cellular 5' exonucleases: the nuclear Rat1 (XRN2 in mammals) and the mainly cytoplasmic Xrn1(Jiao et al., 2010; Xiang et al., 2009). In turn, the exonuclease Rat1 co-transcriptionally degrades the nascent uncapped RNA displacing the transcribing Pol II from the DNA template. The cap quality control has been shown to induce transcription termination upstream of the canonical transcription termination site (TES) leading to a process referred to as premature transcription termination (PTT). Therefore, capping defects have been shows to induce PTT upon certain conditions (Brannan et al., 2012; Jimeno-González et al., 2010). Nevertheless, from both: the same and independent studies, it was observed that the only exonucleolytic activity exerted by Rat1/Xrn2 is not sufficient to

induce the PTT that rather needs more synchronised events to occur (Brannan et al., 2012; Dengl & Cramer, 2009; Eaton et al., 2020; Jimeno-González et al., 2010).

Most of the studies aimed to elucidate and describe the cap quality control were performed employing temperature-sensitive mutants therefore, to further investigate the effects of capping defects on transcribing Pol II, in this chapter I will induce the nuclear depletion of the capping enzymes and assess the Pol II occupancy over protein-coding genes in different conditions.

3.6.2 Uncapped mRNAs are not prematurely terminated upon capping enzymes nuclear depletion

To investigate the effects of nuclear depletion of the capping enzymes on transcription, I started assessing the Pol II occupancy over protein-coding genes. I employed the well-established chromatin immunoprecipitation (ChIP) method (Collas, 2010). Due to the important role of the DNA fragmentation for the quality and the reliability of ChIPs, the sonicated chromatin was assessed for the size distribution of the resulting fragments on agarose gel (Fig. 3.48) aiming to a range between 100 and 1000 nt.
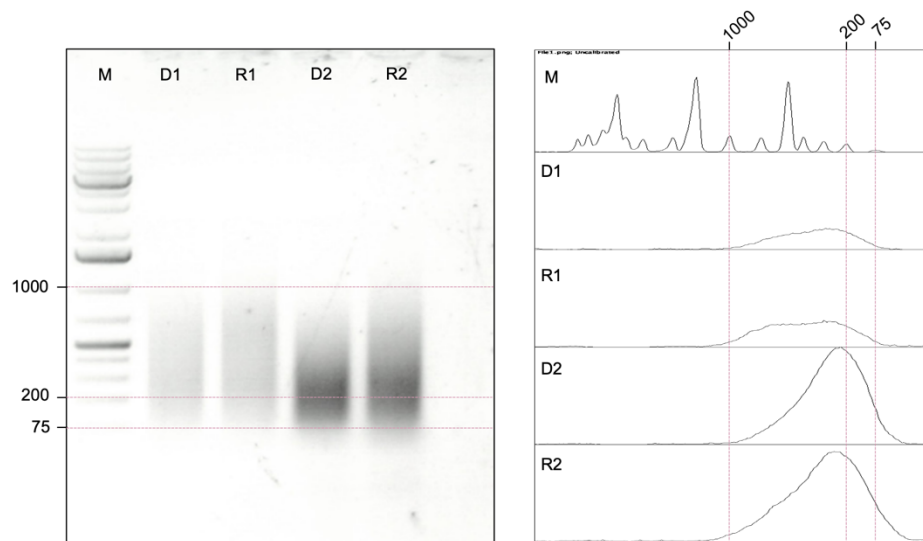
Figure 3.48 - The control of the chromatin fragment size distribution.

On the left. The fragmented chromatin run on 1% agarose gel. The samples were loaded from the left to the right as follow: Molecular ladder (M), DMSO treated replicate 1 (D1), rapamycin treated replicate 1 (R1), DMSO treated replicate 2 (D2), rapamycin treated replicate 2 (R2). The DNA size of 75, 200 and 100 nucleotides are labelled. On the right. The plot showing the band intensities from the gel on the left. DNA fragments peak between 100 and 1000 nucleotides. The samples and the DNA sizes are labelled as shown before, from top to bottom.

Next, employing ChIP followed by qPCR (ChIP-qPCR), I assessed the Pol II occupancy over the highly expressed gene *ADH1*, after 45 minutes of Ceg1 depletion compared to control (Fig. 3.49A). I started immunoprecipitating the fragmented chromatin with the 8WG16 antibody widely employed and known to bind preferentially the unphosphorylated isoform of the Pol II CTD. I observed a partial depletion of Pol II in Ceg1-depleted cells compared to the control which resulted statistically negligible; the effect was prevalent at the 5' end of the gene. I expanded the analysis investigating the serine 5 (Ser5) and serine 2 (Ser2) phosphorylated CTD isoforms as hallmarks of actively transcribing Pol II (Fig. 3.49A). In this case, the Pol II occupancy in Ceg1-depleted cells resulted minimally affected compared to the control.

To better understand the effects of the different capping enzymes on Pol II transcription, I performed the same ChIP analysis on Cet1 and Abd1-depleted cells (Fig. 3.49B and C). The effects of the capping enzymes depletion were even less pronounced compared to Ceg1 for both Cet1 and Abd1-depleted cells as the pattern of the Pol II coverage did not show any difference compared to the control. Ceg1 and Cet1 depletion showed the same effect on cell survival and mRNA levels of the gene tested (see previous results), nevertheless their depletion affected the Pol II distribution over *ADH1* differently suggesting an independency between cap deficiency and the Pol II distribution over the transcriptional unit. On the other hand, Abd1 depletion resulted viable and not sufficient to affect the RNA levels of the genes tested, therefore a different effect on transcription, compared to Ceg1, was expected.

Figure 3.49 - The Pol II occupancy over ADH1 is marginally affected by Ceg1, Cet1 or Abd1 depletion.

*A) The scheme showing the ADH1 gene. The red lines show the position of the oligos used in the ChIP-qPCR. The distribution of total/unphosphorylated Pol II (total/unph Pol II), serine 5-phosphorylated (Ser5-P Pol II) and serine 2-phosphorylated (Ser2-P Pol II) Pol II isoforms over ADH1 in control (DMSO) and after 45 min of rapamycin treatment (Rap) in ceg1-AA (**A**), cet1-AA (**B**) and abd1-AA (**C**) strains. The error bars show the standard deviation of three independent ChIP-qPCR experiments. "ns" indicates the statistical significance calculated via t-test (ns=p-value > 0.05; *=p-value ≤ 0.05; **=p-value ≤ 0.01; ***= p-value ≤ 0.001; ****= p-value ≤ 0.0001).*

I also tested the Pol II occupancy upon longer depletion of the capping enzyme Ceg1 (Fig. 3.50). I treated cells with rapamycin and sampled them at different timepoints (from 0 to 120 minutes). I performed ChIP-qPCR measuring the Pol II occupancy at the 3' end of three selected genes: *ADH1*, *PYK2* and *PMA1* which are in the high or medially expressed range of gene expression whose uncapped mRNA levels were shown to decrease upon Ceg1 depletion (see previous results). The outcome was comparable for all the genes assayed. Pol II occupancy showed 20% reduction in the first 45 minutes of Ceg1 depletion. Then the levels stabilised up to 120 minutes without any further fluctuations indicating that while the cap defects induce the progressive mRNA levels redistribution, the transcription is only marginally affected and does not contribute to the RNA levels drop described for these genes.
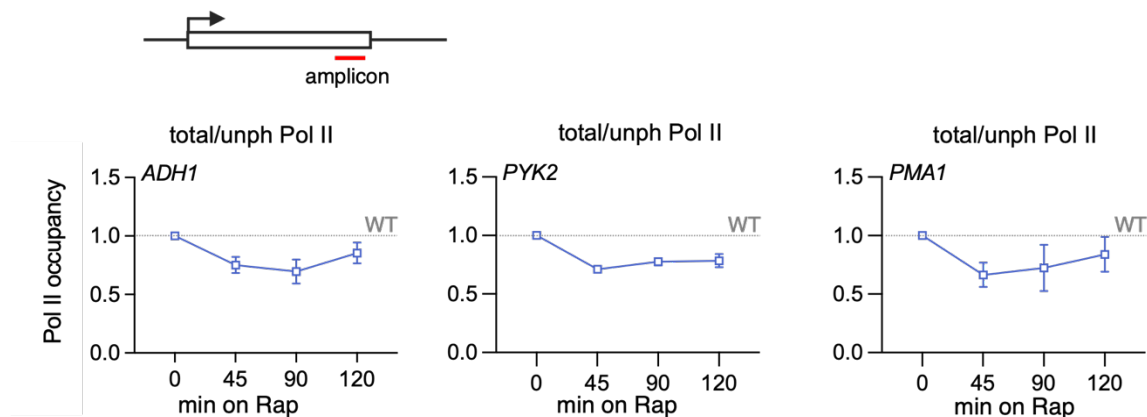
Figure 3.50 - The Pol II occupancy remains marginally affected upon Ceg1 depletion ate the 3' end of ADH1, PYK2 and PMA1.

The scheme showing the position of the oligos used in the ChIP-qPCR (red line). The occupancy of the total/unphosphorylated Pol II at the 3' end of ADH1, PYK2 and PMA1. Timepoints after 0, 45, 90 and 120 minutes on rapamycin compared to control (DMSO) which Pol II occupancy was arbitrarily set to 1. The error bars show the standard deviation of three independent ChIP-qPCR experiments.

ADH1 is a highly expressed gene, its RNA levels are strongly decreased upon Ceg1 depletion (see previous results). The results presented here show that the decreased RNA levels for the uncapped *ADH1* are mainly caused by the increased degradation rather than the reduction of its synthesis. To complete our observation of the effects of the Ceg1 depletion on transcription of protein-coding genes, I investigated the lowly expressed *PML39* whose uncapped mRNA levels increased compared to the control (see previous results). As for *ADH1*, ChIP-qPCR was performed for unphosphorylated, Ser2 and Ser5 phosphorylated isoform of the Pol II CTD over the lowly expressed gene *PML39* (Fig. 3.51). The results for *PLM39* recapitulated what seen for *ADH1*. Indeed, 45 minutes of Ceg1 depletion did not clearly affect the Pol II occupancy over the transcriptional unit. The output was in line among the three CTD phosphoisoforms tested. This result indicated that for lowly expressed genes, the uncapped mRNA levels increase upon Ceg1-depletion, is the result of the increased stability rather than the increased synthesis.

Importantly, ChIP analyses did not reveal any Pol II drop upstream of the TES indicating that the Pol II engaged in the transcription of uncapped mRNAs, is not prematurely terminated upon the nuclear depletion of the capping enzymes.
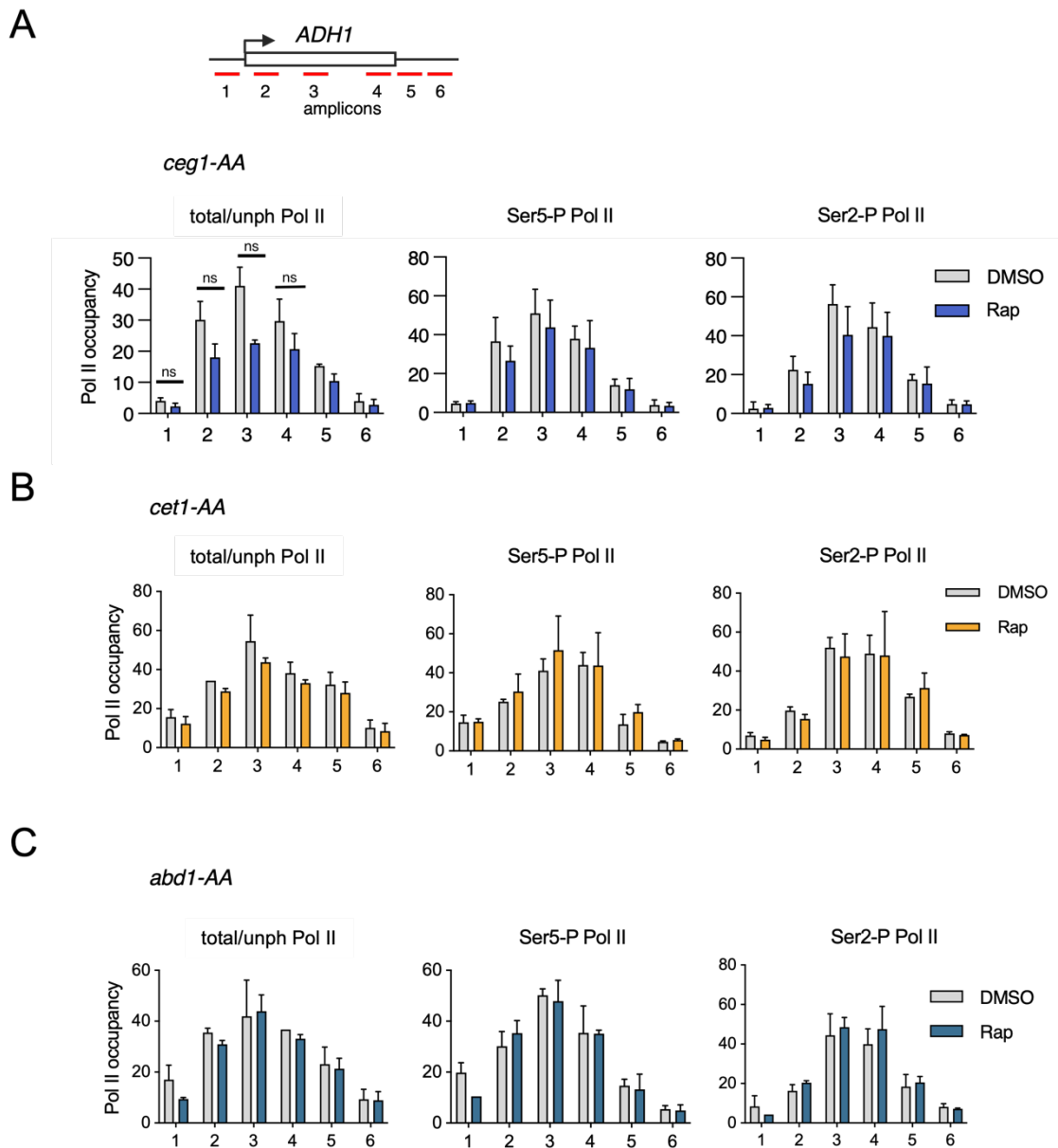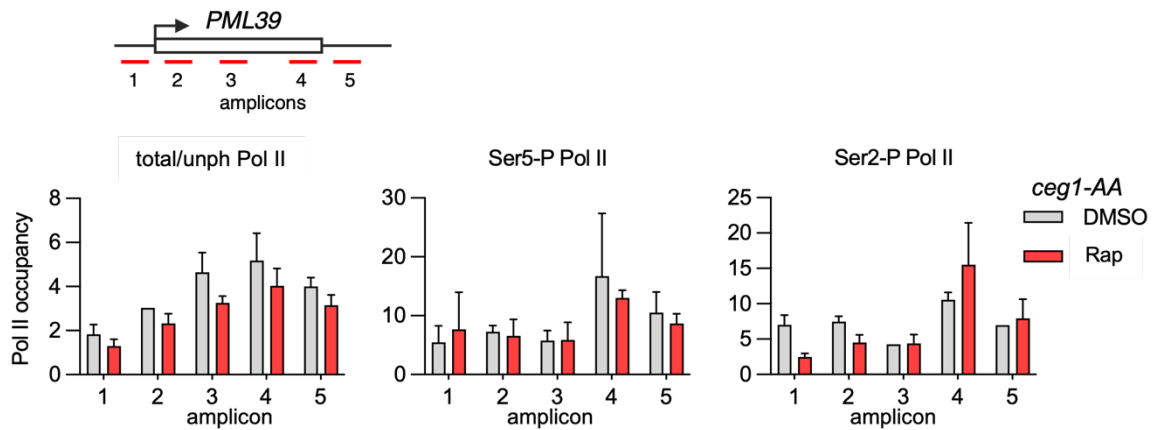
Figure 3.51 - The Pol II occupancy over PML39 is marginally affected by Ceg1 depletion.

The scheme showing the PML39 gene. The red lines show the position of the oligos used in the ChIP-qPCR. The distribution of total/unphosphorylated Pol II (total/unph Pol II), serine 5-phosphorylated (Ser5-P Pol II) and serine 2-phosphorylated (Ser2-P Pol II) Pol II isoforms over PML39 in control (DMSO) and after 45 min of rapamycin treatment (Rap) in ceg1-AA strain. The error bars show the standard deviation of three independent ChIP-qPCR experiments.

I also tested the Pol II occupancy over *FMP27* upon Ceg1 depletion. The PTT upon capping defect has been shown for *FMP27* which is a gene whose length (7.887 kb) is largely above the average gene length of 1.4 kb (Fig. 3.52A) measured in *S. cerevisiae* (Hurowitz and Brown, 2003). I performed ChIP-qPCR immunoprecipitating the chromatin from Ceg1-depleted cells for 45 minutes, using the 8WG16 antibody measuring the Pol II occupancy down up to 4 kb in the gene body (Fig. 3.52B). The capping defects induced by the nuclear depletion of the capping enzymes did not recapitulate the PTT shown in the literature employing temperature-sensitive mutants for Ceg1 (Jimeno-González et al., 2010) showing only a minimal redistribution of the Pol II pattern over the coding region of *FMP27*.
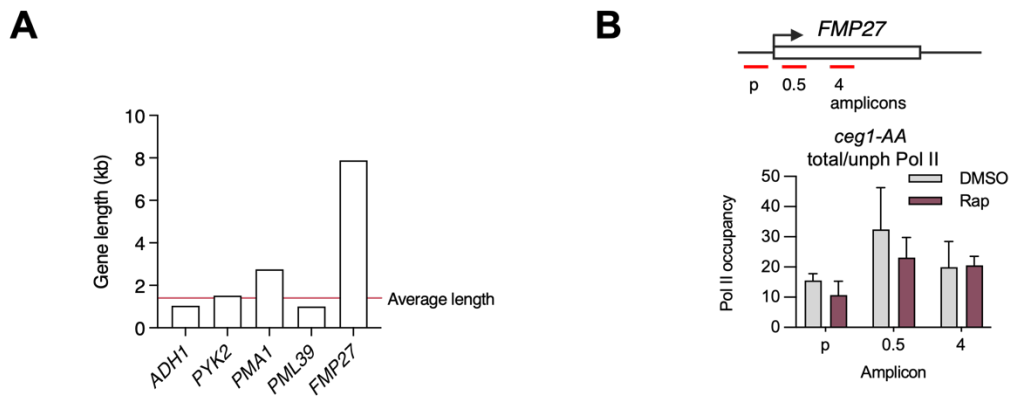
Figure 3.52 - FMP27 is an extra-long gene whose Pol II occupancy is not affected upon Ceg1 depletion.

A) Chart showing the length of the genes investigated. The red line represents the average gene length in S. cerevisiae (1.4 kb). B) The scheme showing the FMP27 gene. The red lines show the position of the oligos used in the ChIP-qPCR. The distribution of total/unphosphorylated Pol II (total/unph Pol II)§ over FMP27 in control (DMSO) and after 45 min of rapamycin treatment (Rap) in ceg1-AA strain. The error bars show the standard deviation of three independent ChIP-qPCR experiments.

To better understand the co-transcriptional events occurring upon Ceg1 depletion, I assessed the levels of newly synthesised uncapped mRNAs. To accomplish this task, I purified the 5 minutes, 4-thiouracil (4tU) labelled fraction from the total RNA of cells depleted of Ceg1 for 45 minutes. I first confirm the enrichment in newly synthesised RNA in the 4tU selected fraction. I employed RT-qPCR comparing the levels of the unspliced *BET1* in the newly synthesised RNA and the total RNA from *ceg1-AA* cells grown in DMSO and treated for 5 minutes with 4tU (Fig. 3.53A). The pull-down efficiency was taken in account for the downstream calculations normalising the RNA levels against the *MNH1* gene (MAGO) found in the spike-in genome from *Schizosaccharomyces pombe* but not present in *S. cerevisiae*.

To investigate the role of capping defects on newly synthesised RNA, cells were first treated with DMSO or rapamycin for 40 minutes and then 4tU was added to the media for 5 minutes (Fig. 3.53B). The 4tU

fraction was then purified and the newly synthesised RNA level assessed by RT-qPCR (Fig. 3.53C). For the two protein-coding genes assayed (*ADH1* and *ACT1*), the newly synthesised RNA levels were decreased in Ceg1-depleted cells, similarly to the total RNA fraction. This result was unexpected in the light of the previous results from the ChIP experiments and can reflect the fast degradation of the uncapped mRNAs tested. A statistically significant stronger reduction was detected in the fraction of the uncapped 4tU labelled RNAs compared to total RNA in Ceg1-depleted cells, possibly caused by the contamination, in the total RNA fraction, of pre-existing capped species.

On the other hand, for the ncRNA *NEL025c*, while the total RNA levels increased upon Ceg1 depletion, the newly synthesised RNA fraction did not show significant variation. This confirms that the accumulation of the uncapped *NEL025c* was mainly caused by the increased stability of the uncapped RNA rather than a boost in its synthesis.
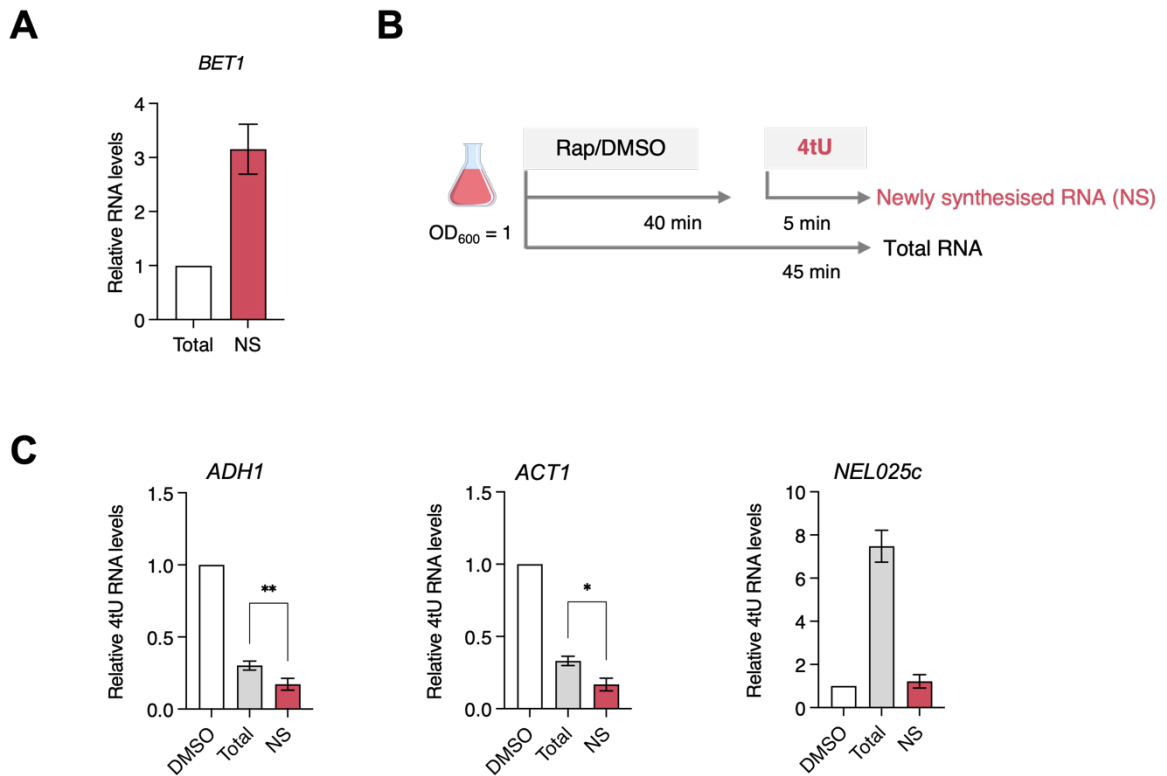
Figure 3.53 - The 4-thiouracil (4tU) treatment for the investigation of cap deficiency in the newly synthesised RNA.

A) The RT-qPCR obtained enrichment in unspliced BET1 in the 4tU purified fraction (NS) compared to total RNA (Total). The RNA levels where normalised against the MNH1 RNA levels from the spike-in S. pombe total RNA. The error bars show the standard deviation of three independent experiments. B) The schematic of the 4tU labelling experiments. Cells were treated with DMSO or rapamycin (Rap) for 40 minutes. Then, 4tU was added (NS) or not (Total) before sampling 45 minutes after the addition of DMSO or Rap. C) The Total or newly synthesised (NS) RNA levels obtained via RT-qPCR for ADH1, ACT1 and NEL025c upon 45 minutes of Ceg1 depletion. The RNA level was normalised against the MNH1 gene from spike-in total RNA and control condition (DMSO) arbitrarily set to 1. The error bars show the standard deviation of three independent experiments. The asterisks (*) indicate the statistical significance calculated via t-test (ns=p-value > 0.05; *=p-value ≤ 0.05; **=p-value ≤ 0.01; ***= p-value ≤ 0.001; ****= p-value ≤ 0.0001).

Overall, the data presented here indicate that the lack of the m^{7}G cap induced via the nuclear depletion of Ceg1 is not sufficient to induce PTT of the gene tested. Also, the redistribution in the abundance for uncapped Pol II products is not regulated by changes in their synthesis but it depends on the reassessed RNA stability. Nevertheless, the newly synthesised RNA fraction of highly expressed protein-coding genes, seems to be affected similarly to the mature, total RNA fraction. I cannot exclude that this effect was obtained from an artefact in the method.

3.6.3 Discussion

The results described in this chapter provide further insights on the transcriptional dynamics of the synthesis of uncapped Pol II products. In the previous chapters, I have shown a diverse redistribution of either protein-coding of non-coding RNA upon Ceg1 nuclear depletion and consequent capping defects. Such a redistribution observed at RNA levels (accumulation or decrease), could be caused by events affecting the RNA stability or RNA synthesis. Besides, previous data from the literature investigating capping defect and the nuclear cap quality control, showed that uncapped transcripts are subjected to co-transcriptional degradation with consequent PTT of the aberrant RNA. All the data published were generated *in vitro* or in the presence of catalytically inactive enzymes (Jiao et al., 2010; Jimeno-González et al., 2010; Zhai & Xiang, 2014). Here I sought to show the effects obtained by the nuclear depletion of the capping enzymes on the transcription of the resulting uncapped RNAs.

I observed that the Ceg1 depletion had a minimal impact on the transcription of the protein-coding genes whose RNA were instead strongly affected upon the same conditions. Specifically, I obtained a 20% reduction in the Pol II occupancy upon Ceg1 depletion. The reduction was obtained in the first 45 minutes of treatment aimed to induce the capping defects, to then stabilise up to 120 minutes of treatment. Methods aimed to investigate the Pol II distribution over all the transcriptional unit indicated that the main effect was exerted at the 5' end of the gene when the antibody 8WG16, specific for unphosphorylated Pol II, was used. This drop could be caused by a direct effect of the lack of Ceg1 on the PIC, affecting the release of the PIC itself into productive elongation. Another explanation to the effect relies on the specificity of 8WG16 in recognising its targets. Indeed, the signal obtained from this antibody, can represent unphosphorylated or hypophosphorylated Pol II that are usually prematurely terminated as unable to be engaged into productive elongation, the lack of Ceg1 could have promoted the turnover of these aberrant Pol II phosphoisoforms from the promoter region. The same effect was not replicated in the experiments employing antibodies specific for Ser5 and Ser2 phosphorylated CTD (Fig. 3.49-3.52).

Importantly, the ChIP experiments did not detect sudden drops in the Pol II distribution upstream of the canonical TES indicating that the transcription of uncapped RNAs is not prematurely terminated in Ceg1 depleted cells (the same results were obtained in cells depleted of Cet1).

To clarify the co-transcriptional events occurring in Ceg1-depleted cells, I investigated the effects on the newly synthesised-4tU labelled RNAs upon Ceg1 depletion (Fig. 3.53C). Surprisingly the newly synthesised RNA was affected by Ceg1 depletion recapitulating the effects observed in the total RNA fraction. These data could indicate a quick or co-transcriptional degradation of the uncapped transcripts in the absence of PTT but are in contrast with the observation of uncapped Pol II transcripts (either coding or noncoding) accumulating upon Ceg1 depletion (see previous results). The co-transcriptional RNA degradation hypothesis in Ceg1-depleted cells is also in contrast with our results obtained from the temperature-sensitive mutants and the *rai1Δ* mutants (see previous results) and with data from the literature showing the accumulation on uncapped mRNAs in the (cytoplasmic) *xrn1Δ* mutants (Schwer et al., 1998). At the light of this, it is possible that the effect observed on the 4tU RNA in Ceg1-depleted cells was generated by excessive labelling time (5 minutes). Due to the short average gene length of yeast genes (1.4 kb on average) and the high expression of the genes selected (Pelechano, Wei, & Steinmetz, 2013), it is possible that the 4tU labelled fraction was enriched in mature mRNAs and therefore the results were not able to properly represent the newly synthesised or nascent RNA fraction.

According to the results presented here, at least 80% of the Pol II holoenzymes engaged in the transcription of uncapped mRNA can reach the end of the transcriptional unit upon the depletion of the capping enzymes Ceg1 or Cet1. This indicates that the condition realised are not sufficient to induce the PTT of the transcription of cap defective mRNAs. Nevertheless, the existence of the cap quality control and the PPT Rat1-mediated upon certain experimental conditions have been demonstrated (Jiao et al., 2010; Jimeno-González et al., 2010; Zhai & Xiang, 2014).

I speculate that the cap quality control and the consequent PTT might require additional signals to occur. The presence of the capping enzymes on the PIC can be a condition required in this concern. Capping enzymes might directly or indirectly recruit the elements of the cap quality control (Rai1 and Rat1) at the promoter level or induce their activity. This can explain the distinct results obtained employing catalytically inactivated capping enzymes at sub-permissive temperatures (Jimeno-González et al., 2010) or their complete nuclear depletion (*ceg1-AA* or *cet1-AA*). While temperature-sensitive mutants can maintain their interactions and are still located in the nucleus, the nuclear depleted enzymes via anchor away system accumulate in the cytoplasm abrogating all their interactions (with the other proteins in the PIC in the case of the capping enzymes).

Therefore, according to the data presented, the uncapped mRNAs in Ceg1-depleted cells are able to evade the cap quality control, be completely transcribed and exported to the cytoplasm where they will follow their metabolic fate.

4. Conclusions

The study of the RNA m^{7}G cap has clarified its structure, the biochemical pathway aiding to its synthesis and has described its multiple interactions. Both, the cap and its interactors are involved in many biological processes and have been shown to mediate several steps related to the regulation of gene expression (Galloway & Cowling, 2019; Gonatopoulos-Pournatzis & Cowling, 2014; Lenasi et al., 2011; Ramanathan et al., 2016). Nevertheless, most of the research on capping has been conducted either *in vitro* or employing single gene methods. This pointed the focus on a restricted number of genes, often model genes with similar characteristics. In other cases, the outcomes were provided relying on synthetic constructs and transcription systems (Gonatopoulos-Pournatzis & Cowling, 2014; Ramanathan et al., 2016).

To date, a genome wide analysis of the effects of the RNA capping on gene expression was missing. The lack of an investigation on the role of the m^{7}G cap on the cellular transcriptome might have obscured further roles of this essential RNA modification, known to be conserved in the eukaryotic kingdom, exploited by other biological entities like viruses and also observed in prokaryotes (Cahová et al., 2015; Decroly et al., 2012; Ramanathan et al., 2016). Besides, the role of the cap in the global cellular processes might have been underestimated by single-gene methods.

The project, therefore, aimed to provide a global description of the effects of the lack of the m^{7}G cap in the cell and through this, describe the processes where the cap is involved but that have yet not been described.

4.1 S. cerevisiae **as a model organism for studying capping**

To conduct this investigation, I employed the model organism *Saccharomyces cerevisiae* and assayed the role of the cap by impairing its synthesis via the depletion of the elements of the capping machinery. The yeast *S. cerevisiae* was preferred as, unlike mammalian cells, the defects on cap binding complex CBC induced indirectly by the capping defects, are viable (Gonatopoulos-Pournatzis & Cowling, 2014; Izaurralde & Adam, 1998). The RNA capping is an essential process therefore, the introduction m^7G cap defects require an inducible system and cannot be sustained for a long time in a living biological system. In order to obtain a fast-induced phenotype, I applied the anchor away system (Haruki et al., 2008) for the nuclear depletion of the capping enzymes. The anchor away was fully characterised upon the experimental condition I applied and resulted ideal to obtain the fast nuclear depletion of the target protein with consequent generation of the capping-defective phenotype within less than hour. The application of the anchor away to the proteins of the CBC, confirmed the negligible overlap between the effects of the disruption of the m^7G cap or the CBC excluding the detection of secondary/indirect effects.

The use of the anchor away, therefore, allowed me to restrict my observations to the effects induced by capping defects only.

4.2 Genome wide analysis showed a complex cap-defective phenotype

The analysis of the gene expression of the entire yeast Pol II transcriptome was performed on cells depleted of the capping enzyme Ceg1 observed to induce a strong reduction of the capped RNA species in *S. cerevisiae*.

The outcome of the RNA sequencing was surprising in light of the literature available in the field. One of the roles assigned to the m^7G cap is the protection of the RNA from the exonucleolytic degradation (Ramanathan et al., 2016). The lack of cap has been associated to the increased degradation rate and

reduction of the half-life of the uncapped RNAs (Schwer et al., 1998). Therefore, a global decrease in the mRNA levels was expected as the result of a generalised degradation of the uncapped transcripts. Strikingly, the global transcriptomic analysis showed that the lack of cap induces the reduction in the levels of a subgroup of genes alongside the unexpected accumulation of other genes. The effect was of a comparable amplitude between the two groups and was related to the basal expression levels of the genes affected and their stability. Indeed, the highest expressed genes resulted downregulated while the lower expressed and more stable genes were overall upregulated upon globalcap deficiency.

A similar effect was observed in our mutant lacking the cytoplasmic 5'-3' exonuclease Xrn1 (*xrn1Δ*). Xrn1 is the one of the main components of the cytoplasmic RNA degradation system and is directly involved in the mRNA turnover and in the clearance of defective transcripts (Nagarajan. et al., 2013). Surprisingly, the *xrn1Δ* mutant showed an enrichment on lowly expressed genes associated with the decrease of the levels of highly expressed genes. The comparison of the differentially expressed genes in Ceg1-depleted cells and the *xrn1Δ* cells, indicated that the same group of genes was perturbed similarly between the two conditions.

Importantly, the introduction of capping defects via the activation of the anchor away in the *xrn1Δ* mutant did not result in a further re-arrangement of gene expression suggesting an overlap between the pathways affected by the lack of cap and Xrn1.

This also implies the fact that the uncapped RNAs follow the global degradation dynamics. Therefore, the highly expressed transcripts might be favoured in the degradation process when uncapped saturating the machinery responsible for RNA degradation (Bitton et al., 2011; Frenk, Oxley, & Houseley, 2014). This, in turn would favour the accumulation observed for the lowly expressed and uncapped transcripts.

The accumulation of uncapped RNA species in the context of a functional 5'-3' RNA degradation system, allowed to speculate that the m^7G cap is not essential for RNA stability in Ceg1-depleted cells. Importantly, the uncapped RNA generated harbour 5' triphosphate end that, together with aberrantly

capped RNAs, are not substrate of the cellular exonucleases and these results sufficient to inhibit the 5'-3' exonucleolytic degradation (Bresson & Tollervey, 2018). In support of this consideration, mechanisms for the conversion of uncapped or aberrantly capped RNAs into exonuclease substrates have been described. In the nucleus, the complex Rai1-Rat1 has been shown to co-transcriptionally process and degrade cap defective RNAs (Jiao et al., 2010; Xiang et al., 2009). The depletion of Rai1 and the temperature inactivation of Rat1, did not rescue the RNA levels of the uncapped RNAs suggesting the minimal function of these proteins in the degradation of uncapped RNAs in Ceg1-depleted cells. This can also indicate the existence of a role of the capping enzymes or the capping process in the recruitment or the activation of the Rai1-Rat1 complex in the cap quality control. This is also in line with the existence of stable uncapped RNA species (rRNAs and tRNAs) synthesised by the Pol I and III. These RNA polymerases (Pol I and Pol III) are known not to interact with capping enzymes and, by consequence, generate uncapped transcripts that are not subjected to cap quality control (Akiyama, Eiler, & Kieft, 2017; Ramanathan et al., 2016). However, the accumulation of non-functional mRNAs may be toxic for the cell. Therefore, the capping enzymes or the process of m^7G cap synthesis might be the condition required for the cap quality control pathway to be functional.

4.3 The role of the m^7G cap in the homeostasis of the cellular mRNA levels

Xrn1 has been shown to sense and regulate the balance between mRNA synthesis and degradation. Different effects have been observed upon the Xrn1 depletion, but all the results agreed on the crucial role exerted by the interaction between Xrn1 itself and the decapped mRNA being processed (Haimovich et al., 2013; Sun et al., 2013). Xrn1 interacts with decapping enzymes to regulate the degradation of the mRNA during the normal turnover (Schoenberg and Maquat, 2012). Therefore, the decapping as well as

a functional turnover degradation pathway are crucial to permit the Xrn1-mediated regulation of the cellular mRNA levels.

The common redistribution of the mRNA levels observed between Ceg1-depleted cells and *xrn1Δ* mutants, suggests a role for the m^{7}G cap in the buffering of the global mRNA levels. It is possible to speculate that non-capped mRNAs might not be identified as coding transcripts, and this can disturb the correct buffering of mRNA levels (Fig. 4.1).

The m^{7}G cap has already shown to mark the Pol II transcript for a specific pathway. The enzymatic cap removal defines the fate of yeast box C/D snoRNAs, a class of RNAs which do not possess the m^{7}G cap as mature species. The m^{7}G cap retention marks these snoRNAs for an mRNA processing pathway. By consequence, they are polyadenylated at the 3' end and exported to the cytoplasm (Grzechnik et al., 2018). Similarly, the cap structure may mark the mRNAs as coding transcripts facilitating the homeostatic adjustments in their levels by degradation and synthesis.
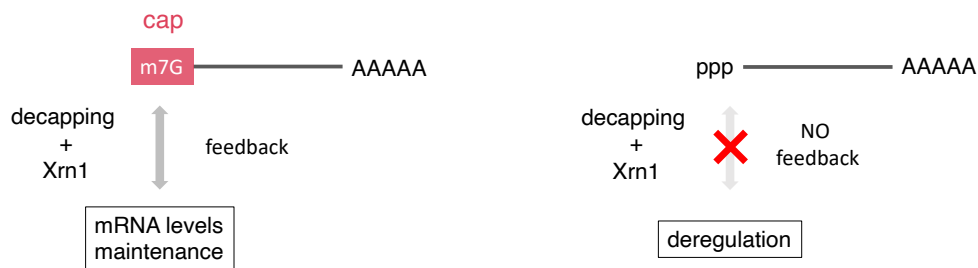

Figure 4.1 - A model of cap-dependent regulation of mRNA levels.

The interaction of Xrn1 with the processed mRNA is crucial for the buffering of the cellular mRNA levels Xrn1-mediated. The m^{7}G cap is therefore a modification which aids to mark mRNAs as coding genes to be sensed.

4.4 Uncapped mRNAs are not prematurely terminated

Previous results have suggested that the Pol II engaged in the transcription of cap-defective pre-mRNAs is subjected to premature transcription termination following the "torpedo model" described for the canonical termination of the protein-coding genes. According to the model, the 5'-3' exonuclease Rat1/Xrn2 catalyses the co-transcriptional degradation of the nascent RNA from the 5' unprotected end. The processive RNA degradation leads to the displacement of the transcribing polymerase from the DNA template (Eaton et al., 2020; Kim et al., 2004; West, Gromak, and Proudfoot, 2004).

ChIP analysis performed to evaluate the Pol II occupancy over some of the protein-coding genes differentially expressed upon Ceg1 depletion, showed that the depletion of the capping enzymes, and by direct consequence, the lack of m^7G cap on the nascent RNA was not sufficient to induce premature transcription termination.

Nevertheless, the premature transcription termination has been shown after decapping of the nascent RNA with consequent arrest of the transcribing Pol II and Rat1-mediated termination (Brannan et al., 2012; Contreras, Benkirane, & Kiernan, 2013; Eaton et al., 2020; Eaton & West, 2020).

Such condition is still in line with the hypothesis proposed here suggesting that the major role of the cap is marking the mRNA as a coding product of the Pol II needed to be exported and translated in the cytoplasm. The presence of the mRNA cap is also associated with the correct turnover of the coding RNAs and the regulation of the cellular RNA levels.

Future work on the effects of the RNA capping on gene expression will cast light on the aspects that remain unclear, and the questions opened after these results.

The global investigation of the transcriptional dynamics and Pol II occupancy might also confirm the marginal role of the m^7G cap on transcription, a data reported here only for a restricted group of genes.

An important question concerning the cap quality control and the role of the capping enzymes in this context remains open. Interaction analysis and studies on the transcriptional initiation might clarify the role of the capping process in modulating the synthesis of uncapped RNAs.

In summary, the outcomes obtained suggest a complex role of the m^{7}G RNA cap in the cellular environment for the control of gene expression. Its biological meaning is not limited to guarantee the interactions of the RNA with the CBC and the translatability of the mRNAs. The m^{7}G cap offers a further layer of complexity to distinguish the cellular RNA products and regulate the RNA processing and gene expression. It marks the mRNA for its cellular fate and cooperates in the sensing of the cellular mRNA levels. The protection from of exonucleolytic degradation could be an accessory function as shown by the presence and the stability of certain uncapped RNA species.

5. References

Aguilera, A. (2005). Cotranscriptional mRNP assembly: From the DNA to the nuclear pore. *Current Opinion in Cell Biology*, *17*(3), 242–250. https://doi.org/10.1016/j.ceb.2005.03.001

Akiyama, B. M., Eiler, D., & Kieft, J. S. (2017). *Structured RNAs that evade or confound exonucleases: function follows form*. 40–47. https://doi.org/10.1016/j.sbi.2015.12.006.Structured

Archer, N., Walsh, M. D., Shahrezaei, V., & Hebenstreit, D. (2016). Modeling Enzyme Processivity Reveals that RNA-Seq Libraries Are Biased in Characteristic and Correctable Ways. *Cell Systems*, *3*(5), 467-479.e12. https://doi.org/10.1016/j.cels.2016.10.012

Archer, S. K., Shirokikh, N. E., & Preiss, T. (2014). Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage. *BMC Genomics*, *15*(1), 1–9. https://doi.org/10.1186/1471-2164-15-401

Arimbasseri, A. G., Rijal, K., & Maraia, R. J. (2013). Transcription termination by the eukaryotic RNA polymerase III. *Biochimica et Biophysica Acta*, *1829*(3–4), 318–330. https://doi.org/10.1016/j.bbagrm.2012.10.006

Arndt, K. M., & Reines, D. (2015). Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annual Review of Biochemistry*, *84*, 381—404. https://doi.org/10.1146/annurev-biochem-060614-034457

Badran, B. M., Wolinsky, S. M., Burny, A., & Willard-Gallo, K. E. (2002). Identification of three NFAT binding motifs in the 5'-upstream region of the human CD3gamma gene that differentially bind NFATc1, NFATc2, and NF-kappa B p50. *The Journal of Biological Chemistry*, *277*(49), 47136–47148. https://doi.org/10.1074/jbc.M206330200

Baillat, D., Hakimi, M.-A., Näär, A. M., Shilatifard, A., Cooch, N., & Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of

RNA polymerase II. *Cell*, *123*(2), 265–276. https://doi.org/10.1016/j.cell.2005.08.019

Barba-Aliaga, M., Alepuz, P., & Pérez-Ortín, J. E. (2021). Eukaryotic RNA Polymerases: The Many Ways to Transcribe a Gene. *Frontiers in Molecular Biosciences*, *8*(April), 1–8. https://doi.org/10.3389/fmolb.2021.663209

Begley, V., Corzo, D., Jordán-Pla, A., Cuevas-Bermúdez, A., de Miguel-Jiménez, L., Pérez-Aguado, D., … Chávez, S. (2019). The mRNA degradation factor Xrn1 regulates transcription elongation in parallel to Ccr4. *Nucleic Acids Research*, *47*(18), 9524–9541. https://doi.org/10.1093/NAR/GKZ660

Bélanger, F., Stepinski, J., Darzynkiewicz, E., & Pelletier, J. (2010). Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. *The Journal of Biological Chemistry*, *285*(43), 33037–33044. https://doi.org/10.1074/jbc.M110.155283

Bharati, A. P., Singh, N., Kumar, V., Kashif, M., Singh, A. K., Singh, P., … Akhtar, M. S. (2016). The mRNA capping enzyme of Saccharomyces cerevisiae has dual specificity to interact with CTD of RNA Polymerase II. *Scientific Reports*. https://doi.org/10.1038/srep31294

Bird, G., Fong, N., Gatlin, J. C., Farabaugh, S., & Bentley, D. L. (2005). Ribozyme cleavage reveals connections between mRNA release from the site of transcription and pre-mRNA processing. *Molecular Cell*, *20*(5), 747–758. https://doi.org/10.1016/j.molcel.2005.11.009

Bird, J. G., Zhang, Y., Tian, Y., Panova, N., Barvík, I., Greene, L., … Nickels, B. E. (2016). The mechanism of RNA 5′ capping with NAD+, NADH and desphospho-CoA. *Nature*, *535*(7612), 444–447. https://doi.org/10.1038/nature18622

Birse, C. E., Minvielle-Sebastia, L., Lee, B. A., Keller, W., & Proudfoot, N. J. (1998). Coupling termination of transcription to messenger RNA maturation in yeast. *Science (New York, N.Y.)*, *280*(5361), 298–301. https://doi.org/10.1126/science.280.5361.298

Bitton, D. A., Grallert, A., Scutt, P. J., Yates, T., Li, Y., Bradford, J. R., … Miller, C. J. (2011).

Programmed fluctuations in sense/antisense transcript ratios drive sexual differentiation in S. pombe. *Molecular Systems Biology*, *7*, 559. https://doi.org/10.1038/msb.2011.90

Börner, T., Aleynikova, A. Y., Zubo, Y. O., & Kusnetsov, V. V. (2015). Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochimica et Biophysica Acta*, *1847*(9), 761–769. https://doi.org/10.1016/j.bbabio.2015.02.004

Bosch, P. S., Pepperl, J., & Basler, K. (2020). Anchor Away - A Fast, Reliable and Reversible Technique To Inhibit Proteins in Drosophila melanogaster. *G3 (Bethesda, Md.)*, *10*(5), 1745–1752. https://doi.org/10.1534/g3.120.401055

Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., … Bentley, D. L. (2012). mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. *Molecular Cell*, *46*(3), 311–324. https://doi.org/10.1016/j.molcel.2012.03.006

Braun, J. E., Truffault, V., Boland, A., Huntzinger, E., Chang, C. Te, Haas, G., … Izaurralde, E. (2012). A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5′ exonucleolytic degradation. *Nature Structural and Molecular Biology*, *19*(12), 1324–1331. https://doi.org/10.1038/nsmb.2413

Bresson, S., & Tollervey, D. (2018). Surveillance-ready transcription: Nuclear RNA decay as a default fate. *Open Biology*, *8*(3). https://doi.org/10.1098/rsob.170270

Brosius, J., & Raabe, C. A. (2016). What is an RNA? A top layer for RNA classification. *RNA Biology*, *13*(2), 140–144. https://doi.org/10.1080/15476286.2015.1128064

Brown, C. E., & Sachs, A. B. (1998). Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation. *Molecular and Cellular Biology*, *18*(11), 6548–6559. https://doi.org/10.1128/MCB.18.11.6548

Cahová, H., Winz, M. L., Höfer, K., Nübel, G., & Jäschke, A. (2015). NAD captureSeq indicates NAD

as a bacterial cap for a subset of regulatory RNAs. *Nature, 519*(7543), 374–377. https://doi.org/10.1038/nature14020

Carter, R., & Drouin, G. (2009). Structural differentiation of the three eukaryotic RNA polymerases. *Genomics*, *94*(6), 388–396. https://doi.org/10.1016/j.ygeno.2009.08.011

Casolari, J. M., Brown, C. R., Komili, S., West, J., Hieronymus, H., & Silver, P. A. (2004). Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell, 117*(4), 427–439. https://doi.org/10.1016/S0092-8674(04)00448-9

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., … Hentze, M. W. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell, 149*(6), 1393–1406. https://doi.org/10.1016/j.cell.2012.04.031

Cech, T. R., & Steitz, J. A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. *Cell, 157*(1), 77–94. https://doi.org/10.1016/j.cell.2014.03.008

Chang, J. H., Jiao, X., Chiba, K., Oh, C., Martin, C. E., Kiledjian, M., & Tong, L. (2012). Dxo1 is a new type of eukaryotic enzyme with both decapping and 5'-3' exoribonuclease activity. *Nature Structural & Molecular Biology, 19*(10), 1011–1017. https://doi.org/10.1038/nsmb.2381

Chen, J., Chiang, Y.-C., & Denis, C. L. (2002). CCR4, a 3'-5' poly(A) RNA and ssDNA exonuclease, is the catalytic component of the cytoplasmic deadenylase. *The EMBO Journal, 21*(6), 1414–1426. https://doi.org/10.1093/emboj/21.6.1414

Chen, Y. G., Kowtoniuk, W. E., Agarwal, I., Shen, Y., & Liu, D. R. (2009). LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nature Chemical Biology, 5*(12), 879–881. https://doi.org/10.1038/nchembio.235

Chernyakov, I., Whipple, J. M., Kotelawala, L., Grayhack, E. J., & Phizicky, E. M. (2008). Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. *Genes and Development, 22*(10), 1369–1380.

https://doi.org/10.1101/gad.1654308

Christofori, G., & Keller, W. (1989). Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA in vitro. *Molecular and Cellular Biology*, *9*(1), 193–203. https://doi.org/10.1128/mcb.9.1.193-203.1989

Cole, S. E., LaRiviere, F. J., Merrikh, C. N., & Moore, M. J. (2009). A convergence of rRNA and mRNA quality control pathways revealed by mechanistic analysis of nonfunctional rRNA decay. *Molecular Cell*, *34*(4), 440–450. https://doi.org/10.1016/j.molcel.2009.04.017

Colin, J., Libri, D., & Porrua, O. (2011). Cryptic Transcription and Early Termination in the Control of Gene Expression. *Genetics Research International*, *2011*, 1–10. https://doi.org/10.4061/2011/653494

Collas, P. (2010). The Current State of Chromatin Immunoprecipitation. *Molecular Biotechnology*, *45*(1), 87–100. https://doi.org/10.1007/s12033-009-9239-8

Contreras, X., Benkirane, M., & Kiernan, R. (2013). Premature termination of transcription by RNAP II: The beginning of the end. *Transcription*, *4*(2). https://doi.org/10.4161/trans.24148

Corden, J. L. (2016). Pol II CTD Code Light. *Molecular Cell*, *61*(2), 183–184. https://doi.org/10.1016/j.molcel.2016.01.005

Cortazar, M. A., Sheridan, R. M., Erickson, B., Fong, N., Glover-cutter, K., Brannan, K., … Glover-cutter, K. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a Article Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "' Sitting Duck Torpedo '" Mechanism. *Molecular Cell*, 1–13. https://doi.org/10.1016/j.molcel.2019.09.031

Cougot, N., Van Dijk, E., Babajko, S., & Séraphin, B. (2004). "Cap-tabolism." *Trends in Biochemical Sciences*, *29*(8), 436–444. https://doi.org/10.1016/j.tibs.2004.06.008

Cowling, V. H. (2010). Regulation of mRNA cap methylation. *Biochemical Journal*, *425*(2), 295–302.

https://doi.org/10.1042/BJ20091352

Dastidar, R. G., Hooda, J., Shah, A., Cao, T. M., Henke, R. M., & Zhang, L. (2012). The nuclear localization of SWI/SNF proteins is subjected to oxygen regulation. *Cell and Bioscience*, *2*(1), 1. https://doi.org/10.1186/2045-3701-2-30

Daugeron, M. C., Mauxion, F., & Séraphin, B. (2001). The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. *Nucleic Acids Research*, *29*(12), 2448–2455. https://doi.org/10.1093/nar/29.12.2448

Davidson, L., Muniz, L., & West, S. (2014). 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes and Development*, *28*(4), 342–356. https://doi.org/10.1101/gad.231274.113

de Vries, H., Rüegsegger, U., Hübner, W., Friedlein, A., Langen, H., & Keller, W. (2000). Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *The EMBO Journal*, *19*(21), 5895–5904. https://doi.org/10.1093/emboj/19.21.5895

Decker, C. J., & Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes & Development*, *7*(8), 1632–1643. https://doi.org/10.1101/gad.7.8.1632

Decroly, E., Ferron, F., Lescar, J., & Canard, B. (2012). Conventional and unconventional mechanisms for capping viral mRNA. *Nature Reviews Microbiology*, *10*(1), 51–65. https://doi.org/10.1038/nrmicro2675

Dengl, S., & Cramer, P. (2009). Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in Vitro. *Journal of Biological Chemistry*, *284*(32), 21270–21279. https://doi.org/10.1074/jbc.M109.013847

Denis, C. L., & Chen, J. (2003). The CCR4-NOT complex plays diverse roles in mRNA metabolism. *Progress in Nucleic Acid Research and Molecular Biology*, *73*, 221–250.

https://doi.org/10.1016/s0079-6603(03)01007-9

Ding, L., Laor, D., Weisman, R., & Forsburg, S. L. (2014). Rapid regulation of nuclear proteins by rapamycin- induced translocation in fission yeast. *Yeast (Chichester, England)*, (31), 253–264. https://doi.org/10.1002/yea

Dlakić, M. (2000). Functionally unrelated signalling proteins contain a fold similar to Mg2+-dependent endonucleases. *Trends in Biochemical Sciences*, *25*(6), 272–273. https://doi.org/10.1016/s0968-0004(00)01582-6

Doamekpor, S. K., Gozdek, A., Kwasnik, A., Kufel, J., & Tong, L. (2020). A novel 5′-hydroxyl dinucleotide hydrolase activity for the DXO/Rai1 family of enzymes. *Nucleic Acids Research*, *48*(1), 349–358. https://doi.org/10.1093/nar/gkz1107

Doma, M. K., & Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, *440*(7083), 561–564. https://doi.org/10.1038/nature04530

Douglas Maya, Maria José Quintero, María de la Cruz Muñoz-Centeno, S. n C. vez. (2008). *Systems for applied gene control in Saccharomyces cerevisiae*. 979–987. https://doi.org/10.1007/s10529-008-9647-z

Dunckley, T, & Parker, R. (1999). The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. *The EMBO Journal*, *18*(19), 5411–5422. https://doi.org/10.1093/emboj/18.19.5411

Dunckley, Travis, Tucker, M., & Parker, R. (2001). Two related proteins, Edc1p and Edc2p, stimulate mRNA decapping in Saccharomyces cerevisiae. *Genetics*, *157*(1), 27–37.

Eaton, J. D., Francis, L., Davidson, L., & West, S. (2020). *A unified allosteric / torpedo mechanism for transcriptional termination on human protein-coding genes*. 1–14. https://doi.org/10.1101/gad.332833.119.

Eaton, J. D., & West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! *Trends in*

Genetics, *36*(9), 664–675. https://doi.org/10.1016/j.tig.2020.05.008

Eddy, S. R. (2001). Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, *2*(12), 919–929. https://doi.org/10.1038/35103511

Egloff, S., & Murphy, S. (2008). *Cracking the RNA polymerase II CTD code*. (May). https://doi.org/10.1016/j.tig.2008.03.008

Eulalio, A., Behm-Ansmant, I., & Izaurralde, E. (2007). P bodies: at the crossroads of post-transcriptional pathways. *Nature Reviews. Molecular Cell Biology*, *8*(1), 9–22. https://doi.org/10.1038/nrm2080

Fianu, I., Chen, Y., Dienemann, C., Dybkov, O., Linden, A., Urlaub, H., & Cramer, P. (2021). Structural basis of Integrator-mediated transcription regulation. *Science (New York, N.Y.)*, *374*(6569), 883–887. https://doi.org/10.1126/science.abk0154

Fortes, P., Kufel, J., Fornerod, M., Polycarpou-Schwarz, M., Lafontaine, D., Tollervey, D., & Mattaj, I. W. (1999). Genetic and physical interactions involving the yeast nuclear cap-binding complex. *Molecular and Cellular Biology*, *19*(10), 6543–6553. https://doi.org/10.1128/MCB.19.10.6543

Fox, M. J., Gao, H., Smith-Kinnaman, W. R., Liu, Y., & Mosley, A. L. (2015). The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway. *PLoS Genetics*, *11*(2), e1004999. https://doi.org/10.1371/journal.pgen.1004999

Fox, M. J., & Mosley, A. L. (2016). Rrp6: Integrated roles in nuclear RNA metabolism and transcription termination. *Wiley Interdisciplinary Reviews. RNA*, *7*(1), 91–104. https://doi.org/10.1002/wrna.1317

Frenk, S., Oxley, D., & Houseley, J. (2014). The nuclear exosome is active and important during budding yeast meiosis. *PloS One*, *9*(9), e107648. https://doi.org/10.1371/journal.pone.0107648

Fresco, L. D., & Buratowski, S. (1996). Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing. *Rna*, Vol. 2, pp. 584–596.

Fu, X.-D. (2014). Non-coding RNA: a new frontier in regulatory biology. *National Science Review*, *1*(2), 190–204. https://doi.org/10.1093/nsr/nwu008

Fujiwara, R., Damodaren, N., Wilusz, J. E., & Murakami, K. (2019). The capping enzyme facilitates promoter escape and assembly of a follow-on preinitiation complex for reinitiation. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(45), 22573–22582. https://doi.org/10.1073/pnas.1905449116

Furuichi, Y. (2015). Discovery of m 7 G-cap in eukaryotic mRNAs. *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, *91*(8), 394–409.

Gagliardi, D., & Dziembowski, A. (2018). *Degradation : From Safeguards To Executioners*. *6*, 1–5.

Galloway, A., & Cowling, V. H. (2019). mRNA cap regulation in mammalian cell function and fate. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2018.09.011

Garí, E., Piedrafita, L., Aldea, M., & Herrero, E. (1997). A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. *Yeast (Chichester, England)*, *13*(9), 837–848. https://doi.org/10.1002/(SICI)1097-0061(199707)13:9<837::AID-YEA145>3.0.CO;2-T

Gebauer, F., & Hentze, M. W. (2004). Molecular mechanisms of translational control. *Nature Reviews Molecular Cell Biology*, *5*(10), 827–835. https://doi.org/10.1038/nrm1488

Geerlings, Vos, & Raué. (2000). The final step in the formation of 25S rRNA in Saccharomyces cerevisiae is performed by $5' \rightarrow 3'$ exonucleases. *RNA*, *6*(12), 1698–1703. https://doi.org/DOI: 10.1017/S1355838200001540

Geisberg, J. V., Moqtaderi, Z., Xiaochun, F., Ozsolak, F., & Struhl, K. (2014). Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. *Cell*, *156*(4), 812–824. https://doi.org/doi:10.1016/j.cell.2013.12.026

Gil, A., & Proudfoot, N. J. (1984). A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3′-end formation. *Nature*, *312*(5993), 473–474. https://doi.org/10.1038/312473a0

Gil, A., & Proudfoot, N. J. (1987). Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. *Cell*, *49*(3), 399–406. https://doi.org/10.1016/0092-8674(87)90292-3

Gilbert, W. V, Zhou, K., Butler, T. K., & Doudna, J. A. (2007). Cap-Independent Translation Is Required for Starvation-Induced Differentiation in Yeast. *Science*, (August), 1224–1227.

Gonatopoulos-Pournatzis, T., & Cowling, V. H. (2014). Cap-binding complex (CBC). *Biochemical Journal*, *457*(2), 231–242. https://doi.org/10.1042/BJ20131214

Goodfellow, I., Chaudhry, Y., Gioldasi, I., Gerondopoulos, A., Natoni, A., Labrie, L., … Roberts, L. (2005). Calicivirus translation initiation requires an interaction between VPg and eIF 4 E. *EMBO Reports*, *6*(10), 968–972. https://doi.org/10.1038/sj.embor.7400510

Grousl, T., Opekarová, M., Stradalova, V., & Hasek, J. (2015). *Evolutionarily Conserved 5 ' -3 ' Exoribonuclease Xrn1 Accumulates at Plasma Membrane-Associated Eisosomes in Post- Diauxic Yeast*. 1–19. https://doi.org/10.1371/journal.pone.0122770

Grzechnik, P., Szczepaniak, S. A., Dhir, S., Pastucha, A., Parslow, H., Matuszek, Z., … Proudfoot, N. J. (2018). Nuclear fate of yeast snoRNA is determined by co-transcriptional Rnt1 cleavage. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04094-y

Guiguen, A., Soutourina, J., Dewez, M., Tafforeau, L., Dieu, M., Raes, M., … Hermand, D. (2007). Recruitment of P-TEFb (Cdk9-Pch1) to chromatin by the cap-methyl transferase Pcm1 in fission yeast. *EMBO Journal*, *26*(6), 1552–1559. https://doi.org/10.1038/sj.emboj.7601627

Haimovich, G., Medina, D. A., Causse, S. Z., Garber, M., Millán-Zambrano, G., Barkai, O., … Choder, M. (2013). Gene expression is circular: Factors for mRNA degradation also foster mRNA synthesis. *Cell*, *153*(5), 1000–1011. https://doi.org/10.1016/j.cell.2013.05.012

Hammet, A., Pike, B. L., & Heierhorst, J. (2002). Posttranscriptional regulation of the RAD5 DNA repair gene by the Dun1 kinase and the Pan2-Pan3 poly(A)-nuclease complex contributes to survival of

replication blocks. *The Journal of Biological Chemistry*, *277*(25), 22469–22474. https://doi.org/10.1074/jbc.M202473200

Han, Z., Libri, D., & Porrua, O. (2017). Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Research*, *45*(3), 1355–1370. https://doi.org/10.1093/nar/gkw1230

Harlen, K. M., & Churchman, L. S. (2017). The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology*, *18*(4), 263–273. https://doi.org/10.1038/nrm.2017.10

Harlen, K. M., Trotta, K. L., Smith, E. E., Mosaheb, M. M., Stephen M. Fuchs, 2, & Churchman, L. S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Reports*, *15*(10), 2147–2158. https://doi.org/10.1016/j.celrep.2016.05.010

Hartwell, L. H. (1967). Macromolecule synthesis in temperature-sensitive mutants of yeast. *Journal of Bacteriology*, *93*(5), 1662–1670. https://doi.org/10.1128/jb.93.5.1662-1670.1967

Haruki, H., Nishikawa, J., & Laemmli, U. K. (2008). The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes. *Molecular Cell*, *31*(6), 925–932. https://doi.org/10.1016/j.molcel.2008.07.020

Hausch, F., Kozany, C., Theodoropoulou, M., & Fabian, A.-K. (2013). FKBPs and the Akt/mTOR pathway. *Cell Cycle (Georgetown, Tex.)*, *12*(15), 2366–2370. https://doi.org/10.4161/cc.25508

Hausmann, S., Pei, Y., & Shuman, S. (2003). Homodimeric quaternary structure is required for the in vivo function and thermal stability of Saccharomyces cerevisiae and Schizosaccharomyces pombe RNA triphosphatases. *Journal of Biological Chemistry*, *278*(33), 30487–30496. https://doi.org/10.1074/jbc.M303060200

Hellen, C. U. T., & Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules.

Genes and Development, *15*(13), 1593–1612. https://doi.org/10.1101/gad.891101

Houseley, J., LaCava, J., & Tollervey, D. (2006). RNA-quality control by the exosome. *Nature Reviews Molecular Cell Biology*, *7*(7), 529–539. https://doi.org/10.1038/nrm1964

Houseley, J., & Tollervey, D. (2009). The Many Pathways of RNA Degradation. *Cell*, *136*(4), 763–776. https://doi.org/10.1016/j.cell.2009.01.019

Hsu, C. L., & Stevens, A. (1993). Yeast cells lacking 5'-->3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Molecular and Cellular Biology*, *13*(8), 4826–4835. https://doi.org/10.1128/mcb.13.8.4826-4835.1993

Huang, Z., Du, Y., Wen, J., Lu, B., & Zhao, Y. (2022). snoRNAs: functions and mechanisms in biological processes, and roles in tumor pathophysiology. *Cell Death Discovery*, *8*(1), 259. https://doi.org/10.1038/s41420-022-01056-8

Hurowitz, E. H., & Brown, P. O. (2003). Genome-wide analysis of mRNA lengths in Saccharomyces cerevisiae. *Genome Biology*, *5*(1), R2. https://doi.org/10.1186/gb-2003-5-1-r2

Iizuka, N., Najita, L., Franzusoff, A., & Sarnow, P. (1994). Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *14*(11), 7322–7330. https://doi.org/10.1128/mcb.14.11.7322-7330.1994

Ikushima, S., Zhao, Y., & Boeke, J. D. (2015). Development of a Tightly Controlled Off Switch for Saccharomyces cerevisiae Regulated by Camphor, a Low-Cost Natural Product. *G3 (Bethesda, Md.)*, *5*(10), 1983–1990. https://doi.org/10.1534/g3.114.012765

Isken, O., & Maquat, L. E. (2008). The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nature Reviews Genetics*, *9*(9), 699–712.

Izaurralde, E., & Adam, S. (1998). Transport of macromolecules between the nucleus and the cytoplasm. *RNA (New York, N.Y.)*, *4*(4), 351–364.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., & Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell*, *78*(4), 657–668. https://doi.org/10.1016/0092-8674(94)90530-4

Jagodnik, J., & Gourse, R. L. (2020). Deciphering the RNA capping process in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(9), 4445–4446. https://doi.org/10.1073/pnas.2000341117

Jeronimo, C., Collin, P., & Robert, F. (2016). The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain. *Journal of Molecular Biology*, *428*(12), 2607–2622. https://doi.org/10.1016/j.jmb.2016.02.006

Jiao, X., Chang, J. H., Kilic, T., Tong, L., & Kiledjian, M. (2013). A Mammalian Pre-mRNA 5' End Capping Quality Control Mechanism and an Unexpected Link of Capping to Pre-mRNA Processing. *Molecular Cell*, *50*(1), 104–115. https://doi.org/10.1016/j.molcel.2013.02.017

Jiao, X., Doamekpor, S. K., Bird, J. G., Nickels, B. E., Tong, L., Hart, R. P., & Kiledjian, M. (2017). 5' End Nicotinamide Adenine Dinucleotide Cap in Human Cells Promotes RNA Decay through DXO-Mediated deNADding. *Cell*, *168*(6), 1015-1027.e10. https://doi.org/10.1016/j.cell.2017.02.019

Jiao, X., Xiang, S., Oh, C., Martin, C. E., Tong, L., & Kiledjian, M. (2010). Identification of a quality-control mechanism for mRNA 5'-end capping. *Nature*, *467*(7315), 608–611. https://doi.org/10.1038/nature09338

Jimeno-González, S., Haaning, L. L., Malagon, F., & Jensen, T. H. (2010). The Yeast 5'-3' Exonuclease Rat1p Functions during Transcription Elongation by RNA Polymerase II. *Molecular Cell*, *37*(4), 580–587. https://doi.org/10.1016/j.molcel.2010.01.019

Jinek, M., Scott M. Coyle, & Doudna, and J. A. (2014). Coupled 5' nucleotide recognition and processivity in Xrn1- mediated mRNA decay. *Bone*, *23*(1), 1–7.

https://doi.org/10.1016/j.molcel.2011.02.004.Coupled

Johnson, A. W., & Kolodner, R. D. (1991). Strand exchange protein 1 from Saccharomyces cerevisiae. A novel multifunctional protein that contains DNA strand exchange and exonuclease activities. *The Journal of Biological Chemistry*, *266*(21), 14046–14054.

Khatter, H., Vorländer, M. K., & Müller, C. W. (2017). RNA polymerase I and III: similar yet unique. *Current Opinion in Structural Biology*, *47*, 88–94. https://doi.org/https://doi.org/10.1016/j.sbi.2017.05.008

Kilchert, C., Wittmann, S., & Vasiljeva, L. (2016). The regulation and functions of the nuclear RNA exosome complex. *Nature Reviews Molecular Cell Biology*, *17*(4), 227–239. https://doi.org/10.1038/nrm.2015.15

Kiledjian, M. (2018). Eukaryotic RNA 5′-End NAD+ Capping and DeNADding. *Trends in Cell Biology*, *28*(6), 454–464. https://doi.org/10.1016/j.tcb.2018.02.005

Kim, H., Jeong, S., Heo, J., Kim, S., Youn, H., Han, J., … Jeong, S. (2004). *mRNA Capping Enzyme Activity Is Coupled to an Early Transcription Elongation mRNA Capping Enzyme Activity Is Coupled to an Early Transcription Elongation.* *24*(14), 6184–6193. https://doi.org/10.1128/MCB.24.14.6184

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., & Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, *432*(7016), 517–522. https://doi.org/10.1038/nature03041

Kim, M., Vasiljeva, L., Rando, O. J., Zhelkovsky, A., Moore, C., & Buratowski, S. (2006). Distinct pathways for snoRNA and mRNA termination. *Molecular Cell*, *24*(5), 723–734. https://doi.org/10.1016/j.molcel.2006.11.011

Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I., & Liu, D. R. (2009). A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proceedings of the National*

Academy of Sciences of the United States of America, *106*(19), 7768–7773. https://doi.org/10.1073/pnas.0900528106

Kuehner, J. N., Pearson, E. L., & Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews Molecular Cell Biology*, *12*(5), 283–294. https://doi.org/10.1038/nrm3098

Kufel, J., & Grzechnik, P. (2019). Small Nucleolar RNAs Tell a Different Tale. *Trends in Genetics*, *35*(2), 104–117. https://doi.org/10.1016/j.tig.2018.11.005

Kyburz, A., Sadowski, M., Dichtl, B., & Keller, W. (2003). The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. *Nucleic Acids Research*, *31*(14), 3936–3945. https://doi.org/10.1093/nar/gkg478

Lahudkar, S., Durairaj, G., Uprety, B., & Bhaumik, S. R. (2014). A novel role for Cet1p mRNA 5′-triphosphatase in promoter proximal accumulation of RNA polymerase II in Saccharomyces cerevisiase. *Genetics*, *196*(1), 161–176. https://doi.org/10.1534/genetics.113.158535

Lahudkar, S., Shukla, A., Bajwa, P., Durairaj, G., Stanojevic, N., & Bhaumik, S. R. (2011). The mRNA cap-binding complex stimulates the formation of pre-initiation complex at the promoter via its interaction with Mot1p in vivo. *Nucleic Acids Research*, *39*(6), 2188–2209. https://doi.org/10.1093/nar/gkq1029

Larimer, F. W., & Stevens, A. (1990). Disruption of the gene XRN1, coding for a 5′→3′ exoribonuclease, restricts yeast cell growth. *Gene*, *95*(1), 85–90. https://doi.org/10.1016/0378-1119(90)90417-P

Lenasi, T., Peterlin, B. M., & Barboric, M. (2011). Cap-binding protein complex links pre-mRNA capping to transcription elongation and alternative splicing through positive transcription elongation factor b (P-TEFb). *Journal of Biological Chemistry*, *286*(26), 22758–22768. https://doi.org/10.1074/jbc.M111.235077

Li, Y., Song, M., & Kiledjian, M. (2011). Differential utilization of decapping enzymes in mammalian

mRNA decay pathways. *RNA (New York, N.Y.)*, *17*(3), 419–428. https://doi.org/10.1261/rna.2439811

Lidschreiber, M., Leike, K., & Cramer, P. (2013). Cap Completion and C-Terminal Repeat Domain Kinase Recruitment Underlie the Initiation-Elongation Transition of RNA Polymerase II. *Molecular and Cellular Biology*, *33*(19), 3805–3816. https://doi.org/10.1128/MCB.00361-13

Liu, H., & Kiledjian, M. (2006). Decapping the message: a beginning or an end. *Biochemical Society Transactions*, *34*(Pt 1), 35–38. https://doi.org/10.1042/BST20060035

Luciano, D. J., & Belasco, J. G. (2015). NAD in RNA: unconventional headgear. *Trends in Biochemical Sciences*, *40*(5), 245–247. https://doi.org/10.1016/j.tibs.2015.03.004

Luciano, D. J., & Belasco, J. G. (2020). Np(4)A alarmones function in bacteria as precursors to RNA caps. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(7), 3560–3567. https://doi.org/10.1073/pnas.1914229117

Luciano, D. J., Levenson-Palmer, R., & Belasco, J. G. (2019). Stresses that Raise Np(4)A Levels Induce Protective Nucleoside Tetraphosphate Capping of Bacterial RNA. *Molecular Cell*, *75*(5), 957-966.e8. https://doi.org/10.1016/j.molcel.2019.05.031

Luo, J., & Hall, B. D. (2007). A Multistep Process Gave Rise to RNA Polymerase IV of Land Plants. *Journal of Molecular Evolution*, *64*(1), 101–112. https://doi.org/10.1007/s00239-006-0093-z

Luo, W., Johnson, A. W., & Bentley, D. L. (2006). The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: Implications for a unified allosteric-torpedo model. *Genes and Development*, *20*(8), 954–965. https://doi.org/10.1101/gad.1409106

Lykke-Andersen, J. (2002). Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. *Molecular and Cellular Biology*, *22*(23), 8114–8121. https://doi.org/10.1128/MCB.22.23.8114-8121.2002

Machineries, J., Takagi, T., Cho, E., Janoo, R. T. K., Polodny, V., Takase, Y., … Buratowski, S. (2002).

Divergent Subunit Interactions among Fungal mRNA. *1*(3), 448–457. https://doi.org/10.1128/EC.1.3.448

Mandal, S. S., Chu, C., Wada, T., Handa, H., Shatkin, A. J., & Reinberg, D. (2004). Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(20), 7572–7577. https://doi.org/10.1073/pnas.0401493101

Mao, X., Schwer, B., & Shuman, S. (1995). Yeast mRNA cap methyltransferase is a 50-kilodalton protein encoded by an essential gene. *Molecular and Cellular Biology*, *15*(8), 4167–4174. https://doi.org/10.1128/mcb.15.8.4167

Marquardt, S., Hazelbaker, D. Z., & Buratowski, S. (2011). Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. *Transcription*, *2*(3), 145–154. https://doi.org/10.4161/trns.2.3.16298

Martinez-Rucobo, F. W., Kohler, R., van de Waterbeemd, M., Heck, A. J. R., Hemann, M., Herzog, F., … Cramer, P. (2015). Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Molecular Cell*, *58*(6), 1079–1089. https://doi.org/10.1016/j.molcel.2015.04.004

Mason, P. B., & Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Molecular Cell*, *17*(6), 831–840. https://doi.org/10.1016/j.molcel.2005.02.017

Massenet, S., Bertrand, E., & Verheggen, C. (2017). Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biology*, *14*(6), 680–692. https://doi.org/10.1080/15476286.2016.1243646

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V, Patil, D. P., … Jaffrey, S. R. (2017). Reversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nature*, *541*(7637), 371–375. https://doi.org/10.1038/nature21022

Mazza, C., Ohno, M., Segref, A., Mattaj, I. W., & Cusack, S. (2001). Crystal structure of the human

nuclear cap binding complex. *Molecular Cell*, *8*(2), 383–396. https://doi.org/10.1016/s1097-2765(01)00299-4

McClure, W. R. (1985). Mechanism and control of transcription initiation in prokaryotes. *Annual Review of Biochemistry*, *54*, 171–204. https://doi.org/10.1146/annurev.bi.54.070185.001131

Medina, D. A., Jordán-Pla, A., Millán-Zambrano, G., Chávez, S., Choder, M., & Pérez-Ortín, J. E. (2014). Cytoplasmic 5'-3' exonuclease Xrn1p is also a genome-wide transcription factor in yeast. *Frontiers in Genetics*, *5*(FEB), 1–10. https://doi.org/10.3389/fgene.2014.00001

Mitchell, P., & Tollervey, D. (2003). An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. *Molecular Cell*, *11*(5), 1405–1413. https://doi.org/10.1016/s1097-2765(03)00190-4

Moore, M. J. (2002). Nuclear RNA turnover. *Cell*, *108*(4), 431–434. https://doi.org/10.1016/S0092-8674(02)00645-1

Moreira, A., Wollerton, M., Monks, J., & Proudfoot, N. J. (1995). Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. *The EMBO Journal*, *14*(15), 3809–3819. https://doi.org/10.1002/j.1460-2075.1995.tb00050.x

Moser, M. J., Holley, W. R., Chatterjee, A., & Mian, I. S. (1997). The proofreading domain of Escherichia coli DNA polymerase I and other DNA and/or RNA exonuclease domains. *Nucleic Acids Research*, *25*(24), 5110–5118. https://doi.org/10.1093/nar/25.24.5110

Moss, T., & Stefanovsky, V. Y. (2002). At the Center of Eukaryotic Life. *Cell*, *109*(5), 545–548. https://doi.org/https://doi.org/10.1016/S0092-8674(02)00761-4

Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J., & Bordonné, R. (2002). Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Molecular Cell*, *9*(4), 891–901. https://doi.org/10.1016/s1097-2765(02)00484-7

Mugridge, J. S., Coller, J., & Gross, J. D. (2018). Structural and molecular mechanisms for the control of eukaryotic 5′–3′ mRNA decay. *Nature Structural and Molecular Biology*, *25*(12), 1077–1085. https://doi.org/10.1038/s41594-018-0164-z

Munoz-Tello, P., Rajappa, L., Coquille, S., & Thore, S. (2015). Polyuridylation in Eukaryotes: A 3′-End Modification Regulating RNA Life. *BioMed Research International*, *2015*, 968127. https://doi.org/10.1155/2015/968127

Nagarajan, V. K., Jones, C. I., Newbury, S. F., & Green, P. J. (2013). XRN 5′→3′ exoribonucleases: Structure, mechanisms and functions. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1829*(6), 590–603. https://doi.org/https://doi.org/10.1016/j.bbagrm.2013.03.005

Natsume, T., Kiyomitsu, T., Saga, Y., & Kanemaki, M. T. (2016). Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Reports*, *15*(1), 210–218. https://doi.org/10.1016/j.celrep.2016.03.001

Neil, H., Malabat, C., D'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L. M., & Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, *457*(7232), 1038–1042. https://doi.org/10.1038/nature07747

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., & Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods*, *6*(12), 917–922. https://doi.org/10.1038/nmeth.1401

Noe Gonzalez, M., Blears, D., & Svejstrup, J. Q. (2021). Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nature Reviews Molecular Cell Biology*, *22*(1), 3–21. https://doi.org/10.1038/s41580-020-00308-8

Noe Gonzalez, M., Sato, S., Tomomori-Sato, C., Conaway, J. W., & Conaway, R. C. (2018). CTD-dependent and -independent mechanisms govern co-transcriptional capping of Pol II transcripts. *Nature Communications*, *9*(1), 3392. https://doi.org/10.1038/s41467-018-05923-w

Nojima, T., Rebelo, K., Gomes, T., Grosso, A. R., Proudfoot, N. J., & Carmo-Fonseca, M. (2018). RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Molecular Cell*, *72*(2), 369-379.e4. https://doi.org/10.1016/j.molcel.2018.09.004

Pabis, M., Neufeld, N., Shav-Tal, Y., & Neugebauer, K. M. (2010). Binding properties and dynamic localization of an alternative isoform of the cap-binding complex subunit CBP20. *Nucleus (Austin, Tex.)*, *1*(5), 412–421. https://doi.org/10.4161/nucl.1.5.12839

Parker, R. (2012). RNA degradation in Saccharomyces cerevisae. *Genetics*, *191*(3), 671–702. https://doi.org/10.1534/genetics.111.137265

Parker, R., & Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. *Nature Structural and Molecular Biology*, *11*(2), 121–127. https://doi.org/10.1038/nsmb724

Peguero-Sanchez, E., Pardo-Lopez, L., & Merino, E. (2015). IRES-dependent translated genes in fungi: Computational prediction, phylogenetic conservation and functional association. *BMC Genomics*, *16*(1), 1–15. https://doi.org/10.1186/s12864-015-2266-x

Pelechano, V., Chavez, S., & Perez-Ortın, J. E. (2010). A Complete Set of Nascent Transcription Rates for Yeast Genes. *PloS One*, *5*(11), 1–10.

Pelechano, V., Wei, W., & Steinmetz, L. M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*. https://doi.org/10.1038/nature12121

Pellegrini, O., Mathy, N., Condon, C., & Bénard, L. (2008). In vitro assays of 5' to 3'-exoribonuclease activity. *Methods in Enzymology*, *448*, 167–183. https://doi.org/10.1016/S0076-6879(08)02609-8

Perrotta, A. T., Shih, I., & Been, M. D. (1999). Imidazole Rescue of a Cytosine Mutation in a Self-Cleaving Ribozyme. *Science*, *286*(5437), 123–126. https://doi.org/10.1126/science.286.5437.123

Postmus, J., Canelas, A. B., Bouwman, J., Bakker, B. M., van Gulik, W., de Mattos, M. J. T., … Smits, G. J. (2008). Quantitative analysis of the high temperature-induced glycolytic flux increase in

Saccharomyces cerevisiae reveals dominant metabolic regulation. *The Journal of Biological Chemistry*, *283*(35), 23524–23532. https://doi.org/10.1074/jbc.M802908200

Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell*, *108*(4), 501–512. https://doi.org/10.1016/S0092-8674(02)00617-7

Qu, L. H., Henras, A., Lu, Y. J., Zhou, H., Zhou, W. X., Zhu, Y. Q., … Bachellerie, J. P. (1999). Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Molecular and Cellular Biology*, *19*(2), 1144–1158. https://doi.org/10.1128/MCB.19.2.1144

Ramachandran, A., Nandakumar, D., Deshpande, A. P., Lucas, T. P., R-Bhojappa, R., Tang, G.-Q., … Patel, S. S. (2016). The Yeast Mitochondrial RNA Polymerase and Transcription Factor Complex Catalyzes Efficient Priming of DNA Synthesis on Single-stranded DNA. *The Journal of Biological Chemistry*, *291*(32), 16828–16839. https://doi.org/10.1074/jbc.M116.740282

Ramanathan, A., Robb, G. B., & Chan, S. H. (2016). mRNA capping: Biological functions and applications. *Nucleic Acids Research*, *44*(16), 7511–7526. https://doi.org/10.1093/nar/gkw551

Rasmussen, E. B., & Lis, J. T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(17), 7923–7927. https://doi.org/10.1073/pnas.90.17.7923

Reeder, R. H. (1998). *Regulation of RNA Polymerase I Transcription in Yeast and Vertebrates* (K. B. T.-P. in N. A. R. and M. B. Moldave, Ed.). https://doi.org/https://doi.org/10.1016/S0079-6603(08)60511-5

Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crépin, T., … Cusack, S. (2014). Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature*, *516*(7531), 361–366. https://doi.org/10.1038/nature14009

Reinders, J., Zahedi RP FAU - Pfanner, N., Pfanner N FAU - Meisinger, C., Meisinger C FAU -

Sickmann, A., & Sickmann, A. (2006). *Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics. PG - 1543-54*. Protein Mass Spectrometry and Functional Proteomics Group, Rudolf-Virchow-Center for Experimental Biomedicine, Julius-Maximilians-Universität Würzburg, 97078 Würzburg, Germany. FAU - Zahedi, René P.

Rodríguez-Molina, J. B., West, S., & Passmore, L. A. (2023). Knowing when to stop: Transcription termination on protein-coding genes by eukaryotic RNAPII. *Molecular Cell*, 404–415. https://doi.org/10.1016/j.molcel.2022.12.021

Ryner, L. C., Takagaki, Y., & Manley, J. L. (1989). Multiple forms of poly(A) polymerases purified from HeLa cells function in specific mRNA 3'-end formation. *Molecular and Cellular Biology*, *9*(10), 4229–4238. https://doi.org/10.1128/mcb.9.10.4229-4238.1989

Sainsbury, S., Bernecky, C., & Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm3952

Schmid, M., & Jensen, T. H. (2018). Controlling nuclear RNA levels. *Nature Reviews Genetics*, *19*(8), 518–529. https://doi.org/10.1038/s41576-018-0013-2

Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, *3*(6), 1101–1108. https://doi.org/10.1038/nprot.2008.73

Schneider, S., Pei, Y., Shuman, S., & Schwer, B. (2010). Separable Functions of the Fission Yeast Spt5 Carboxyl-Terminal Domain (CTD) in Capping Enzyme Binding and Transcription Elongation Overlap with Those of the RNA Polymerase II CTD. *Molecular and Cellular Biology*, *30*(10), 2353–2364. https://doi.org/10.1128/mcb.00116-10

Schoenberg, D. R., & Maquat, L. E. (2012). Regulation of cytoplasmic mRNA decay. *Nature Reviews Genetics*, *13*(4), 246–259. https://doi.org/10.1038/nrg3160

Schroeder, S. C., Schwer, B., Shuman, S., & Bentley, D. (2000). Dynamic association of capping enzymes

with transcribing RNA polymerase II. *Genes and Development*, *14*(19), 2435–2440. https://doi.org/10.1101/gad.836300

Schroeder, S. C., Zorio, D. A. R., Schwer, B., Shuman, S., & Bentley, D. (2004). A Function of Yeast mRNA Cap Methyltransferase, Abd1, in Transcription by RNA Polymerase II. *Molecular Cell*, *13*(3), 377–387. https://doi.org/10.1016/S1097-2765(04)00007-3

Schwer, B., Mao, X., & Shuman, S. (1998). Accelerated mRNA decay in conditional mutants of yeast mRNA capping enzyme. *Nucleic Acids Research*, *26*(9), 2050–2057. https://doi.org/10.1093/nar/26.9.2050

Schwer, B., & Shuman, S. (1994). Mutational analysis of yeast mRNA capping enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(10), 4328–4332. https://doi.org/10.1073/pnas.91.10.4328

Sen, R., Kaja, A., Ferdoush, J., Lahudkar, S., Barman, P., & Bhaumik, S. R. (2017). An mRNA Capping Enzyme Targets FACT to the Active Gene To Enhance the Engagement of RNA Polymerase II into Transcriptional Elongation. *Molecular and Cellular Biology*, *37*(13), 1–19. https://doi.org/10.1128/mcb.00029-17

Shah, N., Maqbool, M. A., Yahia, Y., El Aabidine, A. Z., Esnault, C., Forné, I., … Andrau, J. C. (2018). Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.12.009

Shatkin, A J. (1976). Capping of eucaryotic mRNAs. *Cell*, *9*(4, Part 2), 645–653. https://doi.org/https://doi.org/10.1016/0092-8674(76)90128-8

Shatkin, Aaron J., & Manley, J. L. (2000). The ends of the affair: Capping and polyadenylation. *Nature Structural Biology*, *7*(10), 838–842. https://doi.org/10.1038/79583

She, M., Decker, C. J., Sundramurthy, K., Liu, Y., Chen, N., Parker, R., & Song, H. (2004). Crystal structure of Dcp1p and its functional implications in mRNA decapping. *Nature Structural &*

Molecular Biology, *11*(3), 249–256. https://doi.org/10.1038/nsmb730

Shibagaki, Y., Itoh, N., Yamada, H., Nagata, S., & Mizumoto, K. (1992). mRNA capping enzyme. Isolation and characterization of the gene encoding mRNA guanylyltransferase subunit from Saccharomyces cerevisiae. *Journal of Biological Chemistry*, *267*(14), 9521–9528.

Shimizu-Yoshida, Y., Sasamoto, M., Yoshida, A., Yoshioka, T., Matsumoto, A., & Sakai, A. (1999). Mouse CAF1, a mouse homologue of the yeast POP2 gene, complements the yeast pop2 null mutation. *Yeast (Chichester, England)*, *15*(13), 1357–1364. https://doi.org/10.1002/(SICI)1097-0061(19990930)15:13<1357::AID-YEA465>3.0.CO;2-Y

Sims, R. J., Belotserkovskaya, R., & Reinberg, D. (2004). Elongation by RNA polymerase II: The short and long of it. *Genes and Development*, *18*(20), 2437–2468. https://doi.org/10.1101/gad.1235904

Song, M.-G., Li, Y., & Kiledjian, M. (2010). Multiple mRNA decapping enzymes in mammalian cells. *Molecular Cell*, *40*(3), 423–432. https://doi.org/10.1016/j.molcel.2010.10.010

Sonntag, K.-C., & Darai, G. (1995). Evolution of viral DNA-dependent RNA polymerases. *Virus Genes*, *11*(2), 271–284. https://doi.org/10.1007/BF01728665

Staals, R. H. J., Bronkhorst, A. W., Schilders, G., Slomovic, S., Schuster, G., Heck, A. J. R., … Pruijn, G. J. M. (2010). *Dis3-like 1: a novel exoribonuclease associated with the human exosome PT - Journal Article*.

Stevens, A. (2001). 5'-exoribonuclease 1: Xrn1. *Methods in Enzymology*, *342*, 251–259. https://doi.org/10.1016/s0076-6879(01)42549-3

Suh, H., Ficarro, S. B., Kang, U. B., Chun, Y., Marto, J. A., & Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2015.12.021

Sun, M., Schwalb, B., Pirkl, N., Maier, K. C., Schenk, A., Failmezger, H., … Cramer, P. (2013). Global analysis of Eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels.

Molecular Cell, *52*(1), 52–62. https://doi.org/10.1016/j.molcel.2013.09.010

Takagaki, Y., & Manley, J. L. (1997). RNA recognition by the human polyadenylation factor CstF. *Molecular and Cellular Biology*, *17*(7), 3907–3914. https://doi.org/10.1128/MCB.17.7.3907

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., & Libri, D. (2006). Transcription Termination and Nuclear Degradation of Cryptic Unstable Transcripts: A Role for the Nrd1-Nab3 Pathway in Genome Surveillance. *Molecular Cell*, *23*(6), 853–864. https://doi.org/10.1016/j.molcel.2006.07.029

THOMPSON, M. K., KIOURLAPPOU, M., & DAVIS, I. (2020). Ribo-Pop: Simple, cost-effective, and widely applicable ribosomal RNA depletion. *Rna*, *26*(11), 1731–1742. https://doi.org/10.1261/rna.076562.120

Thore, S., Mauxion, F., Séraphin, B., & Suck, D. (2003). X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. *EMBO Reports*, *4*(12), 1150–1155. https://doi.org/10.1038/sj.embor.7400020

Topisirovic, I., Svitkin, Y. V., Sonenberg, N., & Shatkin, A. J. (2011a). Cap and cap-binding proteins in the control of gene expression. *Wiley Interdisciplinary Reviews: RNA*, *2*(2), 277–298. https://doi.org/10.1002/wrna.52

Topisirovic, I., Svitkin, Y. V., Sonenberg, N., & Shatkin, A. J. (2011b). Cap and cap-binding proteins in the control of gene expression. *Wiley Interdisciplinary Reviews: RNA*, *2*(2), 277–298. https://doi.org/10.1002/wrna.52

Tucker, M., Valencia-Sanchez, M. A., Staples, R. R., Chen, J., Denis, C. L., & Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. *Cell*, *104*(3), 377–386. https://doi.org/10.1016/s0092-8674(01)00225-2

Urrutia, A. O., & Hurst, L. D. (2003). The signature of selection mediated by expression on human genes.

Genome Research, *13*(10), 2260–2264. https://doi.org/10.1101/gr.641103

Van Dijk, E. L., Chen, C. L., Daubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., … Morillon, A. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature*, *475*(7354), 114–119. https://doi.org/10.1038/nature10118

van Hoof, A., Frischmeyer, P. A., Dietz, H. C., & Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science (New York, N.Y.)*, *295*(5563), 2262–2264. https://doi.org/10.1126/science.1067272

Vasiljeva, L., & Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3′ processing of RNA polymerase II transcripts. *Molecular Cell*, *21*(2), 239–248. https://doi.org/10.1016/j.molcel.2005.11.028

Wadkins, T. S., Perrotta, A. T., Ferré-D'Amaré, A. R., Doudna, J. A., & Been, M. D. (1999). A nested double pseudoknot is required for self-cleavage activity of both the genomic and antigenomic hepatitis delta virus ribozymes. *RNA (New York, N.Y.)*, *5*(6), 720–727. https://doi.org/10.1017/s1355838299990209

Walters, R. W., Matheny, T., Mizoue, L. S., Rao, B. S., Muhlrad, D., & Parker, R. (2017). Identification of NAD+ capped mRNAs in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(3), 480–485. https://doi.org/10.1073/pnas.1619369114

Wang, Z., Jiao, X., Carr-Schmid, A., & Kiledjian, M. (2002). The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 12663–12668. https://doi.org/10.1073/pnas.192445599

Warminski, M., Sikorski, P. J., Kowalska, J., & Jemielity, J. (2017). Applications of Phosphate Modification and Labeling to Study (m)RNA Caps. *Topics in Current Chemistry*, *375*(1), 16. https://doi.org/10.1007/s41061-017-0106-y

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences*, *24*(11), 437–440. https://doi.org/https://doi.org/10.1016/S0968-0004(99)01460-7

Werner, M., Purta, E., Kaminska, K. H., Cymerman, I. A., Campbell, D. A., Mittra, B., … Bujnicki, J. M. (2011). 2'-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. *Nucleic Acids Research*, *39*(11), 4756–4768. https://doi.org/10.1093/nar/gkr038

West, S., Gromak, N., & Proudfoot, N. J. (2004). Human 5′ → 3′ exonuclease Xm2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, *432*(7016), 522–525. https://doi.org/10.1038/nature03035

Wierzbicki, A. T., Haag, J. R., & Pikaard, C. S. (2008). Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell*, Vol. 135, pp. 635–648. https://doi.org/10.1016/j.cell.2008.09.035

Wong, C.-M., Qiu, H., Hu, C., Dong, J., & Hinnebusch, A. G. (2007). Yeast Cap Binding Complex Impedes Recruitment of Cleavage Factor IA to Weak Termination Sites. *Molecular and Cellular Biology*, *27*(18), 6520–6531. https://doi.org/10.1128/mcb.00733-07

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J. C., Dufour, M. E., Boulay, J., … Jacquier, A. (2005). Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*, *121*(5), 725–737. https://doi.org/10.1016/j.cell.2005.04.030

Wysoker, A., Fennell, T., Marth, G., Abecasis, G., Ruan, J., Li, H., … Handsaker, B. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J. L., & Tong, L. (2009). Structure and function of the 5′3′ exoribonuclease Rat1 and its activating partner Rai1. *Nature*, *458*(7239), 784–788. https://doi.org/10.1038/nature07731

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., … Steinmetz, L. M. (2009).

Bidirectional promoters generate pervasive transcription in yeast. *Nature*. https://doi.org/10.1038/nature07728

Xue, Y., Bai, X., Lee, I., Kallstrom, G., Ho, J., Brown, J., … Johnson, A. W. (2000). Saccharomyces cerevisiae RAI1 (YGL246c) Is Homologous to Human DOM3Z and Encodes a Protein That Binds the Nuclear Exoribonuclease Rat1p . *Molecular and Cellular Biology*, *20*(11), 4006–4015. https://doi.org/10.1128/mcb.20.11.4006-4015.2000

Yamashita, A., Chang, T.-C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.-Y. A., & Shyu, A.-B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nature Structural & Molecular Biology*, *12*(12), 1054–1063. https://doi.org/10.1038/nsmb1016

Yu, H.-B., Yurieva, M., Balachander, A., Foo, I., Leong, X., Zelante, T., … Ricciardi-Castagnoli, P. (2015). NFATc2 mediates epigenetic modification of dendritic cell cytokine and chemokine responses to dectin-1 stimulation. *Nucleic Acids Research*, *43*(2), 836–847. https://doi.org/10.1093/nar/gku1369

Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D., & Shatkin, A. J. (1997). Mammalian capping enzyme complements mutant Saccharomyces cerevisiae lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(24), 12898–12903. https://doi.org/10.1073/pnas.94.24.12898

Zhai, L. T., & Xiang, S. (2014). mRNA quality control at the 5′ end. *Journal of Zhejiang University: Science B*, *15*(5), 438–443. https://doi.org/10.1631/jzus.B1400070

Zhao, J., Hyman, L., & Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews : MMBR*, *63*(2), 405–445. https://doi.org/10.1128/MMBR.63.2.405-445.1999

Zhou, M., & Law, J. A. (2015). RNA Pol IV and V in gene silencing: Rebel polymerases evolving away

from Pol II's rules. *Current Opinion in Plant Biology*, *27*, 154–164. https://doi.org/10.1016/j.pbi.2015.07.005

Zou, Z., Tao, T., Li, H., & Zhu, X. (2020). mTOR signaling pathway and mTOR inhibitors in cancer: progress and challenges. *Cell & Bioscience*, *10*(1), 31. https://doi.org/10.1186/s13578-020-00396-1