A comparative analysis of nuclear cap binding complex proteins Cbc1, Cbc2 and putative cap binding protein Cbc3 in fission yeast



Doğukan Hazar Ülgen

Green Templeton College

University of Oxford

This dissertation is submitted for the degree of

Doctor of Philosophy

Trinity Term 2019

To my loving and supportive family

Abstract

Cap-binding complex (CBC) proteins Cbc1 and Cbc2 bind to the 7-methylguanosine (7mG) cap modification present at the 5' end of RNA polymerase II transcripts and have been linked to a variety of functions in RNA regulation. Despite the finding that Cbc1 and Cbc2 bind the 7mG cap synergistically, there are also reports suggesting that CBC proteins can also act independently. Furthermore, recently a new cap binder, named Cbc3, has been discovered, with its function in the context of CBC yet to be discovered. In this thesis, *S. pombe* Cbc1, Cbc2 and Cbc3 are compared with respect to their co-transcriptional recruitment, interacting proteins, and the subsequent effect of their depletion on the transcriptome. Cbc1 and Cbc2 depletions resulted in diverging differential expression profiles and changes in RNA processing. Differences could not be attributed to transcriptional changes. A genetic screen showed that these genes also have diverging interactions with RNA surveillance factors. Interestingly, Cbc1 depletion, and not Cbc2, was found to rescue readthrough transcription associated with a Dhp1 temperature-sensitive mutation. This thesis thus provides data supportive of Cbc1 having a Cbc2-independent, anti-cleavage role. On the other hand, *S. pombe* Cbc3 was found to facilitate the recruitment of elongation factor Paf1 to protein-coding (pc) genes that are longer than ~500 bases. Furthermore, Cbc3 deletion resulted in a mild hindrance of RNA export. Taken together, my findings point to inherent CBC-independent functions for Cbc1 and Cbc2. While Cbc3 is important for transcription elongation and nuclear export, it is unlikely that it forms the basis of an alternative CBC.

Contents

ABSTRACT ..... **II**

CONTENTS ..... **III**

LIST OF FIGURES ..... **VI**

LIST OF TABLES ..... **VIII**

ABBREVIATIONS ..... **IX**

ACKNOWLEDGMENTS ..... **XI**

DECLARATION ..... **XII**

CHAPTER 1 INTRODUCTION **1**

1.1 EMERGENCE AND LIFE OF AN RNA POLYMERASE II TRANSCRIPT 1

1.1.1 Initiation *2*

1.1.2 Capping *2*

1.1.3 Elongation *4*

1.1.4 Termination *5*

1.1.5 Nuclear export of mRNAs *6*

1.2 NUCLEAR CAP-BINDING COMPLEX (CBC) 9

1.2.1 Structure of CBC *10*

1.2.2 Roles of CBC in transcription *11*

1.2.3 CBC and RNA processing *12*

1.2.4 CBC and nuclear export *13*

1.3 RNA SURVEILLANCE 14

1.3.1 The nuclear exosome and its associated cofactors *14*

1.3.2 5'-3' degradation *17*

1.3.3 The interwoven RNA surveillance: Where does the CBC come in? *17*

1.4 CBC3: A NEW CAP BINDER? 19

1.5 OBJECTIVES OF THIS WORK 20

CHAPTER 2 INTERACTING PARTNERS AND GENOME-WIDE DISTRIBUTION PROFILES OF CBC1, CBC2 AND CBC3 ..... **23**

2.1 CBC1, CBC2 AND CBC3 ARE RECRUITED CO-TRANSCRIPTIONALLY, BUT SHOW DIFFERENT DISTRIBUTION PROFILES 23

2.2 PURIFICATION OF CBC COMPONENTS AND CBC3 27
2.3 DISCUSSION 31
2.3.1 Co-transcriptional recruitment of CBC components and Cbc3 to pc genes 31
2.3.2 CBC interacts with P-TEFb and splicing factors 32
2.3.3 Cbc3 interacts with export factors and a number of RNA helicases 34

CHAPTER 3 THE POSSIBILITY OF CBC1 AND CBC2 HAVING SEPARATE ROLES OUTSIDE THE CBC 37

3.1 DEPLETION OF CBC1 OR CBC2 RESULTS IN GROWTH DEFECTS, WITHOUT INFLUENCING THE STABILITY OF THE OTHER PROTEIN 38
3.2 DEPLETION STRAINS HAVE DIVERSE CHANGES IN TRANSCRIPTOME 40
3.3 CROSSTALK OF CBC WITH DIFFERENT MEMBERS OF RNA SURVEILLANCE PATHWAYS 45
3.3.1 Readthrough in dhp1-154 is suppressed upon Cbc1 depletion 46
3.3.2 How does the depletion of Cbc1 rescue the dhp1-154 readthrough? 48
3.4 DISCUSSION 52

CHAPTER 4 LOSS OF CBC3 AFFECTS RNA BIOGENESIS 59

4.1 PC GENES ARE GENERALLY DOWNREGULATED UPON CBC3 DELETION 59
4.2 CBC3 PROMOTES TRANSCRIPTION ELONGATION IN LONGER GENES BY FACILITATING PAF1 RECRUITMENT 62
4.3 CBC3 FACILITATES NUCLEAR EXPORT OF POLY(A)+ RNA LOCALIZATION 68
4.4 DISCUSSION 70

CHAPTER 5 GENERAL DISCUSSION AND CONCLUSIONS 77

5.1 CBC: IS THERE ROOM FOR ANOTHER? 78
5.2 FUNCTION OF CBC3 IN *S. POMBE* 79
5.3 CBC1/2 HAVING CBC-INDEPENDENT ROLES 81
5.4 CBC1 AND READTHROUGH PHENOTYPE RESCUE: ANTI-CLEAVAGE OR ANTI-DEGRADATION? 84

CHAPTER 6 MATERIALS AND METHODS 87

6.1 LISTS OF OLIGONUCLEOTIDES, VECTORS, *S. POMBE* STRAINS AND ANTIBODIES USED IN THIS STUDY 87
6.1.1 Oligonucleotides 87
6.1.2 Vectors 89
6.1.3 S. pombe strains 90
6.1.4 Antibodies 91
6.2 *S. POMBE* GROWTH CONDITIONS AND GENETIC MANIPULATIONS 92
6.2.1 Media and growth conditions 92

6.2.2 Strain construction and transformations *93*
6.2.3 Crosses, tetrad dissections and random spore analyses *93*
6.2.4 Spot tests and growth curves *94*
6.3 BIOCHEMISTRY TECHNIQUES 94
6.3.1 RNA extraction *94*
6.3.2 Protein extraction (TCA method) *95*
6.3.3 SDS polyacrylamide gel *96*
6.3.4 Coomassie and silver staining *96*
6.3.5 Western blotting *96*
6.3.6 Northern blotting *97*
6.3.7 Protein purification from yeast and mass spectrometry *98*
6.4 MOLECULAR BIOLOGY TECHNIQUES 99
6.4.1 Polymerase Chain Reaction (PCR) *99*
6.4.2 Quantitative Real-time PCR *100*
6.4.3 Chromatin Immunoprecipitation (ChIP) *100*
6.4.4 polyA· RNA Fluorescence in situ Hybridization (FISH) *101*
6.4.5 Library preparation *102*
6.5 BIOINFORMATICS ANALYSIS 103
6.5.1 Sequence alignment *103*
6.5.2 ChIP-seq analysis *104*
6.5.3 Differential expression analysis of RNA-seq alignments *104*
6.5.4 RNA polymerase II processivity measurement *104*

CHAPTER 7 REFERENCES 109

Appendix A Genetic screening of CBC depletion strains and RNA surveillance factors……...129

List of Figures

Figure 1.1: The chemical structure of the 7mG cap. 3

Figure 1.2: The summary of events described in section 1.1.. 9

Figure 1.3: The crystal structure of the cap-bound human CBC (PDB accession ID: 1H2T) by (Mazza *et al.*, 2001). 11

Figure 2.1: ChIP-seq for Cbc1-2-3 points to a co-transcriptional recruitment, but with different patterns.. 24

Figure 2.2: RNA requirement for the recruitment of Cbc1 and Cbc3.. 25

Figure 2.3: Metagene profile for the recruitment of Cbc1-2-3 in nc genes. 26

Figure 2.4: Native purifications for Cbc1, Cbc2 and Cbc3. 28

Figure 3.1: CBC Depletion, experimental setup and time courses. 38

Figure 3.2: Spot test for CBC depletion strains. 39

Figure 3.3: Depletion of one CBC protein does not affect the stability of the other in fission yeast. 40

Figure 3.4: Depletion of Cbc1 or Cbc2 results in distinct changes on the transcriptome. 41

Figure 3.5: (A) Percent overlap of significantly changing transcripts in Cbc1 depletion transcripts with Cbc2 depletion for pc, nc genes, CUTs and antisense transcripts. Upregulated genes are coloured in cyan; downregulated in light pink. (B) Significantly upregulated or downregulated pc genes found in only one of the depletion strains were determined, and percentage of intron-containing genes among them was calculated. 42

Figure 3.6: Representative northern blots for *phb2* or *pre8* with the depletion strains or splicing mutants *prp2-1* and *prp10-1*.....43

Figure 3.7: Metagene profile for Pol II occupancy across pc genes in depletion strains.44

Figure 3.8: Representative genome browser snapshots of RNA-seq reads from wt, *dhp1-154*, Cbc1 depletion, Cbc2 depletion, Cbc1 double mutant and Cbc2 double mutant at *tom22* and *wis2* loci. 47

Figure 3.9: Northern blots for *tom22* and *wis2* transcripts to confirm RNA-seq results. 48

Figure 3.10: Pol II metagene profiles across pc genes in wt, *dhp1-154*, Cbc1 depletion or double mutant.....49

Figure 3.11: poly(A) FISH with double mutants. 51

Figure 4.1: Scatterplots showing changes in pc or nc gene transcripts, or CUTs after Cbc3 deletion with respect to wt. 60

Figure 4.2: Shorter genes are less affected by Cbc3 loss. 61

Figure 4.3: Pol II occupancy metagene profiles in wt or *cbc3*Δ for all pc genes, genes downregulated, or not significantly changing genes upon Cbc3 deletion. 63

Figure 4.4: Ser5- and Ser2-phosphorylation levels of Pol II CTD in wt vs *cbc3Δ*. 64

Figure 4.5: Paf1 recruitment to pc genes are affected in Cbc3 deletion. 65

Figure 4.6: Paf1 recruitment to longer genes were more affected upon Cbc3 deletion. .67

Figure 4.7: Spot test and growth curves with wt, *cbc3Δ*, *tex1Δ* and *cbc3Δ, tex1Δ* strains. 68

Figure 4.8: poly(A) FISH with wt, *cbc3Δ*, *tex1Δ* and *cbc3Δ, tex1Δ*. Alexa488-conjugated oligo-d(T)$_{20}$ probe and DAPI staining was used to detect the relative localization of polyadenylated RNA in *S. pombe*.....70

Figure 5.1: Working model for the role of Cbc3 in *S. pombe*. 80

Figure A.1: Screening of genetic interaction between *rrp6Δ* and the depletion strains, grown at 25°C, 30°C or 37°C. 129

Figure A.2: Screening of genetic interaction between *dis3-54* and the depletion strains, grown at 25°C, 30°C or 37°C. Restrictive temperature for *dis3-54* is 25°C. 130

Figure A.3: Screening of genetic interaction between *mtl1-1* and the depletion strains, grown at 25°C, 30°C or 37°C. Restrictive temperature for *mtl1-1* is 25°C.....130

Figure A.4: Screening of genetic interaction between *red1Δ* and the depletion strains, grown at 25°C, 30°C or 37°C. 131

Figure A.5: Screening of genetic interaction between *pab2Δ* and the depletion strains, grown at 25°C, 30°C or 37°C. 131

Figure A.6: Screening of genetic interaction between *hrb1Δ* and the depletion strains, grown at 25°C, 30°C or 37°C. 132

Figure A.7: Screening of genetic interaction between *dhp1-154* and the depletion strains, grown at 25°C, 30°C or 37°C. Restrictive temperature for *dhp1-154* is 25°C.....132

List of Tables

Table 2.1: List of proteins that came up in the Cbc1 pulldown, with descending intensity scores. Proteins shown to be interacting with both Cbc1 and Cbc2 are highlighted in cyan. Intensity values are obtained from MaxQuant analysis. The results are an average of three independent replicates. 29

Table 2.2: List of proteins that came up in the Cbc2 pulldown, with descending intensity scores. Proteins shown to be interacting with both Cbc1 and Cbc2 are highlighted in cyan. Intensity values are obtained from MaxQuant analysis. The results are an average of three independent replicates. 30

Table 2.3: List of proteins that came up in the Cbc3 pulldown, with descending intensity scores. Intensity values are obtained from MaxQuant analysis. The results are an average of two independent replicates. 30

Table 3.1: Screening of genetic interactions between RNA surveillance factors and depletion strains. Interactions were tested in three temperatures (25°C, 30°C, 37°C), on EMMG (non-depletion) or YES (depletion) plates. A negative genetic interaction is indicated with "-", a positive with "+" and no apparent interaction with "=". 45

Table 6.1: List of oligonucleotides used in this study. Note that the L4 and L5 primers contain a 5' homology arm to the vector to be used. These homology arms are indicated at the end of the table. 87

Table 6.2: List of vectors used in this study. 89

Table 6.3: List of *S. pombe* strains used in this study. 90

Table 6.4: List of antibodies used in this study. 91

Abbreviations

7mG	7-methylguanosine
aa	amino acid
bp	base pair
CBC	Nuclear cap binding complex
cDNA	complementary DNA
ChIP	Chromatin immunoprecipitation
CTD	C-terminal domain of RNA Polymerase II
CRAC	Crosslinking and analysis of cDNA
CUT	Cryptic unstable transcript
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
EMMG	Edinburgh minimal medium + glutamate
FISH	Fluorescence *in situ* hybridization
h	hour(s)
IgG	immunoglobulin G
iCLIP	individual-nucleotide-resolution crosslinking and immunoprecipitation
mRNA	messenger RNA
MTREC	Mtl1-Red1 core
NET-seq	Native elongating transcript sequencing
NEXT	nuclear exosome targeting (complex)
nc	non-coding
S. pombe	Schizosaccharomyces pombe (fission yeast)
S. cerevisiae	Saccharomyces cerevisiae (budding yeast)
pA	polyadenylation

PAXT	poly(A) RNA exosome targeting (connection)
pc	protein-coding
PCR	polymerase chain reaction
Pol II	RNA polymerase II
P-TEFb	positive elongation factor b
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonucleoprotein
RRM	RNA recognition motif
SDS	sodium dodecyl sulfate
ssDNA	single-stranded salmon sperm DNA
TCA	trichloroacetic acid
TEV	tobacco edge virus
TRAMP	Trf4/5-Air1/2-Mtr4 polyadenylation (complex)
Tris	Tris(hydroxymethyl)aminomethane
ts	temperature-sensitive
TSS	Transcription start site
wt	wild-type
YES	Yeast extract broth + supplements

Acknowledgments

First and foremost, I would like to thank my supervisor, Lidia, for giving me the opportunity to work on this project. With her guidance, I expanded my scientific knowledge and gained valuable experience, which I believe will help me immensely further down the road, throughout my career.

I would also like to thank my family, for their ever-continuing support on my endeavours. No matter the distance, they were always there when I needed them the most, and I could always rely on them when I felt down. I feel lucky to having drawn them in the blind lottery of life and will always be indebted to these lovely people.

Being in a different country to undertake studies could be overwhelming from time to time, but I was fortunate to find amazing friends throughout my DPhil. First off, I would like to thank my "holy triad": Sina, Kris and Miles, to whom I owe what little sanity I had remaining in these years. Without you, I would not have the strength to soldier on and had probably given up. Additionally, I enjoyed the company of all my lab mates: Tea, Dong-hyuk and Adrien; not only were they great friends, but they also taught me tips and tricks of many techniques. Alongside, I am grateful to everyone in the third floor of Biochemistry for the interesting conversations, help and support.

This DPhil was made possible by TUBITAK's Graduate Scholarship (2213), an honour I was humbled to achieve. Thus, I would like to thank TUBITAK for giving me the opportunity to go overseas and improve myself.

A special thanks from the bottom of my heart to Melis who, despite being thousands of kilometres away, has never stopped supporting me in every way imaginable. Being away from you was the hardest part of all. Hopefully, it will not happen ever again. Not anymore.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except work indicated below:

- First replicate of the Cbc1-2 purification (section 2.2) was done under the valuable supervision of Dr. Tea Kecman, Vasilieva Lab.
- Metagene plots included in this thesis are drawn with the help of an R script coded by Dr. Sina Wittmann, Hyman Lab, Max Planck Institute in Dresden. Differential expression analyses for RNA-seq was also carried out with a modified version of Dr. Sina Wittmann's original R script.
- Tryptic digestion and tandem mass spectrometry for identification of interacting partners from purifications (section 2.2) are done by Ms. Svenja Hester from Proteomics Facility, Department of Biochemistry, University of Oxford.
- Northern blot for *tom22* readthrough transcript (section 3.3.1) was started and finished by me; longer film exposures were developed by Dr. Dong-hyuk Heo, Vasilieva Lab.
- Pol II ChIP-seq in depletion strains and *dhp1-154* double mutants (section 3.3.2) was carried out by me, and the DNA libraries I prepared were sequenced by Dr. Dong-hyuk Heo, Vasilieva Lab.
- Structural and biochemical characterization of Cbc3, as discussed in Chapter 5 General Discussion (section 5.1), was carried out entirely by Dr. Krzysztof Kus and Ms. Anna Wiecek, Vasilieva Lab.

Chapter 1 Introduction

Across all kingdoms, sustainability of cellular function necessitates regulated and timely expression of genes stored in the genetic material. The first step to achieve this is "writing" the genetic information on a more mobile, versatile material; namely RNA, in a process called transcription. Enzymes which use DNA as a template to "transcribe" RNA are collectively known as RNA polymerases. While prokaryotes have only one polymerase responsible for transcribing all genes, higher steps in evolution brought out three different polymerases which take on different responsibilities. Most of the non-coding (nc) genes including ribosomal (r)RNAs and U6 small nuclear (sn)RNA are transcribed by either RNA pol I or III. On the other hand, RNA pol II is responsible for synthesis of messenger (m)RNAs and a variety of ncRNAs including U1, U2, U4 and U5 snRNAs as well as small nucleolar (sno)RNAs.

This study revolves around two proteins that not only protect RNA pol II transcripts, but also regulate the downstream processing and mobilization of these RNAs. To underline the importance of these proteins in RNA biogenesis and regulation, I am going to first introduce the emergence of an RNA pol II transcript, followed by the "milestones" of its "life". I will then introduce the nuclear cap binding complex and describe its known roles in these events.

1.1 Emergence and life of an RNA polymerase II transcript

Every RNA polymerase II (Pol II) transcript begins its life through the three stages of transcription: initiation, elongation and termination.

The largest subunit of Pol II, Rpb1, has a protracted C-terminal domain (CTD) consisting of a number of repeats of the same heptad sequence: Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7.

The number of repeats vary between species, i.e. 52 in humans and 29 in *S. pombe* (Chapman *et al.*, 2008). All residues in the CTD are subject to post-translational modifications; Tyr1, Ser2, Thr4, Ser5 and Ser7 can be phosphorylated while the prolines can be isomerised (reviewed in Harlen & Churchman 2017). The CTD does not affect the catalytic function of Pol II, but is integral for the timely regulation of the transcription cycle, as well as facilitating the co-transcriptional recruitment of other important factors in RNA processing (Hsin and Manley, 2012). While this study does not focus on the CTD, some of its modifications and their occurrences during the transcription cycle need to be introduced.

1.1.1 Initiation

Pol II is recruited to a stretch of DNA, called the promoter, upstream of the gene. Here, Pol II binds other transcription factors TFIIA/B/D/E/F/H forming the pre-initiation complex (PIC). The DNA is unwound by the DNA helicase TFIIH, allowing the Pol II to sit on the template strand, ready to start transcription from the transcription start site (TSS) (reviewed in Sikorski & Buratowski 2009).

Pol II in the PIC is unphosphorylated, which has a high affinity for the Mediator complex (Myers *et al.*, 1998; Lu *et al.*, 2006). The initiation of transcription is coupled to the promoter escape of Pol II, and this is partly mediated by the Cdk7-(Mcs6 in *S. pombe*) dependent phosphorylation of Ser5 and Ser7, which disrupts the hydrogen bonding between Rpb1 and the Mediator (Søgaard and Svejstrup, 2007; Jeronimo and Robert, 2014). Thus, Pol II in the initiation stage is usually defined by CTD phospho-marks Ser5 and Ser7. Ser5 also provides the signal for the first RNA processing event: 5’ capping.

1.1.2 Capping

The emerging 5’ end of the nascent RNA is prone to degradation by 5’-3’ exonucleases (Meyer, Temme and Wahle, 2004); and hence, needs to be protected as soon as the transcription starts. Ser5-phosphorylated Pol II recruits proteins responsible for the

addition of a guanosine monophosphate that is methylated in its 7th position (7mG, Figure 1.1) in a process referred to as 5' capping (Ho and Shuman, 1999).

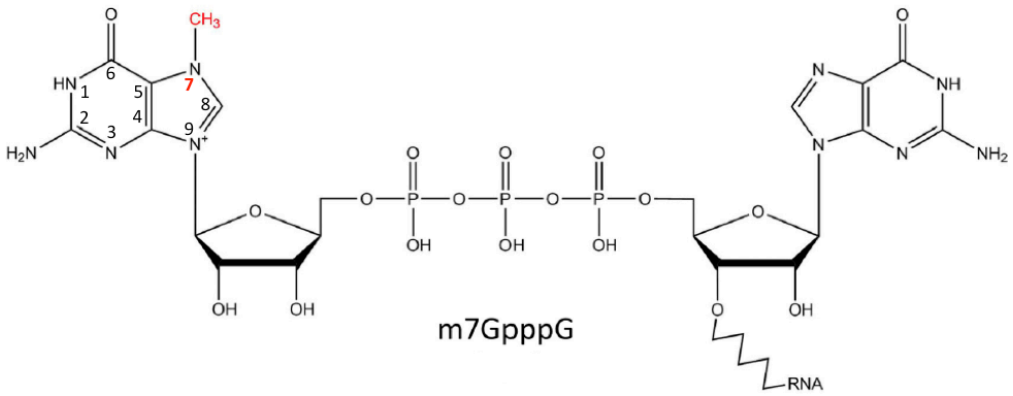


Figure 1.1: The chemical structure of the 7mG cap, here shown to be linked to a guanine. Adapted from (Ramanathan, Robb and Chan, 2016).

Capping is a three-stage process. First, 5′ -γ -phosphate of the first nucleotide is removed. Then, a guanine monophosphate is linked to the newly-formed 5' diphosphate tail. Finally, a methyl group is transferred to the N-7 position of the guanine to yield the 7mG cap (reviewed in Gonatopoulos-Pournatzis & Cowling 2014). While these activities are distributed to three distinct enzymes in *S. pombe* as Pct1, Ceg1 and Pcm1 in the aforementioned order; all functions are embodied in one complex in higher eukaryotes (Saha, Schwer and Shuman, 1999). Since the capping enzymes are recruited to the 5' of the emerging transcript through the interactions with the Ser5-phosphorylated CTD, the addition of the 7mG is co-transcriptional. Interestingly, chromatin immunoprecipitation (ChIP) analyses in budding yeast revealed that the triphosphatase and the guanylyltransferase are only associated with the initiation-stage Pol II while the methyltransferase crosslinks to further downstream of genes (Komarnitsky, Cho and Buratowski, 2000). It is yet unclear whether this mirroring in the order of enzymatic stages

of capping and the movement of Pol II over the gene is merely a representation of the transcription stages or has further biological implications.

1.1.3 Elongation

In higher eukaryotes, Pol II is paused upon transcribing the initial 20-60 nt (Rougvie and Lis, 1988). This pause is induced upon binding of two factors: DRB-sensitivity-inducing factor (DSIF; composed of Spt4/5) and negative elongation factor (NELF) (Yamaguchi *et al.*, 1999). The Pol II-DSIF-NELF ternary complex has a tilted DNA-RNA hybrid in the active site that impairs binding of the nucleoside triphosphate, providing a structural inference for the pausing phenomenon (Vos *et al.*, 2018). It is the general view that the promoter-proximal pausing helps recruit the RNA processing factors and elongation factors to the site of transcription. This pause also provides a window of opportunity for the capping enzyme to be recruited and the 5' cap to be established (section 1.1.2). The establishment of 5' cap is integral for the downstream RNA processing events such as splicing and 3' end cleavage and polyadenylation. This importance is perhaps best illustrated by the finding that a Ser5-to-Ala substitution on CTD is lethal but can be rescued by tethering the capping complex to Rpb1 (Schwer and Shuman, 2011). Therefore, the crucial function of Ser5 phosphorylation in budding yeast is recruiting the capping enzyme to the emerging transcript.

The inhibitory clamp of DSIF and NELF is broken by the positive transcription elongation factor (P-TEFb), which, upon phosphorylating NELF, causes them to dissociate from Pol II (Peterlin and Price, 2006). P-TEFb also phosphorylates Ser2 in Pol II CTD, thereby inducing a change in the phosphorylation pattern of CTD, which in turn results in the substitution of factors responsible for the initiation of transcription with the ones for the elongation and RNA processing (reviewed in Brés et al. 2008).

Promoter-proximal pausing was long thought to be a phenomenon exclusive to higher eukaryotes, in part due to the absence of NELF, but recent precision run-on sequencing (PRO-seq) experiments in *S. pombe* challenged this view (Booth *et al.*, 2016). The latter study points that the pausing behaviour required the conserved elongation factor Spt4 in

fission yeast. Furthermore, it is reported that inhibition of Cdk9, the kinase component of P-TEFb, shifts the Pol II profile toward the TSS, and the Pol II that manages to escape this pause has a severely reduced elongation rate, pointing to a similar initiation-to-elongation checkpoint in fission yeast (Booth *et al.*, 2018).

As the Pol II traverses along the gene body, a number of important factors are recruited to the elongating complex. Upon reaching an exon-intron boundary, spliceosome complex is recruited and introns are spliced co-transcriptionally (Will and Lührmann, 2011; Saldi *et al.*, 2016). A number of factors including but not limited to Paf1 and Spt4/5 are integral to proper elongation; loss of them drastically reduces the Pol II travel rate *in vivo* and elongation rate *in vitro* (Chen *et al.*, 2009; Kim, Guermah and Roeder, 2010; Hartzog and Fu, 2013). Once the Pol II reaches the polyadenylation (pA) site, usually a 6-nt signal of AAUAAA, mRNA is cleaved and polyadenylated (reviewed in Millevoi & Vagner 2009). All of these steps influence the elongation rate of Pol II (Alexander *et al.*, 2010; Carrillo Oesterreich, Preibisch and Neugebauer, 2010; Nojima *et al.*, 2015; Hujer *et al.*, 2016). Moreover, checkpoints have been described for both splicing and 3' end processing (Chathoth *et al.*, 2014; Laitem *et al.*, 2015), illustrating the tight regulation of RNA processing to ensure proper production of pre-mRNAs.

1.1.4 Termination

After the pre-mRNA is cleaved off, the Pol II that is still transcribing the coding strand must be terminated and recycled. Still under debate, there are two models for Pol II dislocation from the DNA.

The allosteric model argues that the dissociation of elongation factors and recruitment of termination factors cause a conformational change in the elongating Pol II, decreasing its processivity and eventually falling off the template (Logan *et al.*, 2006). In *E. coli*, termination factor Rho associates with bacterial RNA polymerase, wrapping the nascent RNA around its body, essentially "unwinding" the Pol II until it disengages from the DNA (Epshtein *et al.*, 2010). In higher eukaryotes, Pcf11 was shown to be capable of dismantling the elongating complex (EC) *in vitro*, in a CTD-dependent manner (Zhang,

Fu and Gilmour, 2005). Furthermore, in HeLa cells, depletion of Pcf11 resulted in reduced termination efficiency (West and Proudfoot, 2008).

The second model, the so-called "torpedo model", postulates that cleavage of the nascent RNA creates an uncapped 5' end, which can then be degraded by the 5'-3' exonuclease Rat1 (Dhp1 in *S. pombe*). Dhp1 degrades RNA faster than Pol II transcribes it, eventually reaching the enzyme and "torpedoing" it off the DNA template (Connelly and Manley, 1988). Indeed, the kinetics of the model was validated when slow Pol II mutants were found to terminate closer to the pA site (Fong *et al.*, 2015). Furthermore, mutants of Rat1 show readthrough transcription *in vivo* (Kim, Krogan, *et al.*, 2004). However, there is some evidence arguing against the model. Solely, the degradation of nascent RNA was not sufficient to destabilize the EC (Gu, Wind and Reines, 2002). Another study found that cleavage after the pA site was not required for termination (Osheim, Proudfoot and Beyer, 1999). Finally, an *in vitro* model of termination showed that Rat1 and its associated factor Rai1 were insufficient to dislodge the EC (Dengl and Cramer, 2009), justifying the doubt that an exonuclease would be able to exert enough force to disrupt the highly stable EC.

Therefore, a combined allosteric-torpedo model was proposed (Luo, Johnson and Bentley, 2006). Luo et al. bring forward the idea that the conformational change resulted from substitution of elongation factors with termination factors, as well as Dhp1 is required for timely and efficient termination. Addition of α-amanitin, a selective inhibitor of Pol II that drastically reduces its processivity rate, blocks termination completely, implying that the conformational change is also necessary for Dhp1-mediated termination (Zhang, Rigo and Martinson, 2015).

1.1.5 Nuclear export of mRNAs

Transcribed and processed mRNAs need to translocate to cytoplasm where they can be translated into proteins by ribosomes. Nuclear export is yet another layer of regulation that ensures that the mRNA is properly transcribed, capped, its introns removed, polyadenylated, and is ready to be translated.

Though the nuclear envelope is perforated by a large complex of proteins that are collectively known as the nuclear pore complex (NPC), the inside of these pores is rich in Phe and Gly residues, making them hydrophobic and thus non-permissive to RNA diffusion (Strambio-De-Castillia, Niepel and Rout, 2010). Therefore, RNA needs to be decorated by a set of proteins to permeate this barrier. Perhaps the most important of these proteins are Mex67-Mtr2 (Nxf1-Nxt1 in metazoans). In *S. cerevisiae*, a thermo-sensitive Mex67 mutant shows accumulated poly(A)$^+$ RNA in the nucleus at the non-permissive temperature (Segref *et al.*, 1997). Both being RNA-binding proteins, they non-selectively bind RNA *in vitro* (Segref *et al.*, 1997; Santos-Rosa *et al.*, 2015), and require other factors *in vivo* to pick the correct RNAs to bind and export. This purpose is realized by the evolutionarily conserved transcription-export (TREX) complex, composed of the THO subcomplex of Hpr1, Tho2, Mft1, Thp2 and Tex1 (Hpr1, Thoc1, Thoc7, Thoc5, Thoc6 and Thoc3 in higher eukaryotes), Uap56 and Yra1 (ALY/REF in mammals) (Sträßer *et al.*, 2002). Uap56 and Yra1 act as adaptor proteins bridging TREX to Mex67-Mtr2 (Sträßer and Hurt, 2001; Sträßer, 2002). In mammals, hnRNP C tetramer was reported to measure the RNA to distinguish between short (snRNAs) and longer (usually mRNAs), which will then take up different methods of nuclear export (McCloskey *et al.*, 2012). It is not known whether the yeast orthologue, Nab3, possesses a similar function.

TREX components have a co-transcriptional recruitment whereby the signal increases over the gene body and drops close to the pA site. Meinel et al. reported that this is via direct interaction of THO subcomplex with Ser5P-Ser2P diphosphorylated CTD, explaining the recruitment profiles (Meinel *et al.*, 2013), and once again attesting to the importance of CTD in the orchestration of RNA biogenesis.

In both yeast and higher eukaryotes, some mechanisms, interconnected with nuclear export, are present to verify splicing status of introns, so that unspliced transcripts do not end up in the cytoplasm. In yeast, this is done by two serine-arginine rich (SR)-proteins, Hrb1 and Gbp2, which bind to pre-mRNAs and the spliceosome during splicing, and assess the splicing state. If splicing is completed successfully, they recruit Mex67 and allow for nuclear export. If, however, the pre-mRNA is unspliced, they instead activate RNA surveillance in the nucleus (introduced in section 1.3), degrading the faulty transcript

(Krebber *et al.*, 2014). In higher eukaryotes, this is done by the exon-junction complex (EJC), which decorates the exon-exon boundaries and form higher-order multimers with their associated SR-proteins (Singh *et al.*, 2012).

After the recognition of the correct composition of RNA and associated factors, collectively known as messenger ribonucleoprotein (mRNP), Yra1 and Uap56 bind the mRNA to be transported, subsequently recruiting Mex67-Mtr2. The RNA is unwound by the activity of the DExD/H-box-type RNA helicase Uap56 (Luo *et al.*, 2001; Sträßer and Hurt, 2001). The ATP-bound Uap56 interacts with RNA and upon binding of Yra1, RNA is unwound upon ATP hydrolysis, followed by Uap56 dissociation from the RNA, ensuring the retention of the protein in the nucleus, ready for another round of RNA export (Taniguchi and Ohno, 2007). Another RNA helicase situated in the nuclear periphery and cytoplasm, Dbp5 (DDX19 in mammals), then dislocates the remaining mRNP-bound export factors, allowing them to move back into nucleus (reviewed in Tieg & Krebber 2013).

Unlike mRNAs, snRNAs requiring nuclear export do not utilize Yra1 as their adapter, instead opting for phosphorylated adapter of RNA export (PHAX). PHAX binding can be competitively inhibited by hnRNP C tetramer (McCloskey *et al.*, 2012), thereby modulating which export adapter is used by mRNAs vs. snRNAs (explained in more detail in section 1.2.4).

Figure 1.2 summarizes the phenomena described in section 1.1, from transcription stages to different RNA processing events that take place, followed by the mRNP recognition by Mex67-Mtr2 and export. It underlines that the mRNPs formation starts as soon as the transcription by Pol II starts, and the composition of the mRNP changes drastically at each stage, eventually converging at the content which is recognized as a “ready-to-go” mRNP by the export factors.

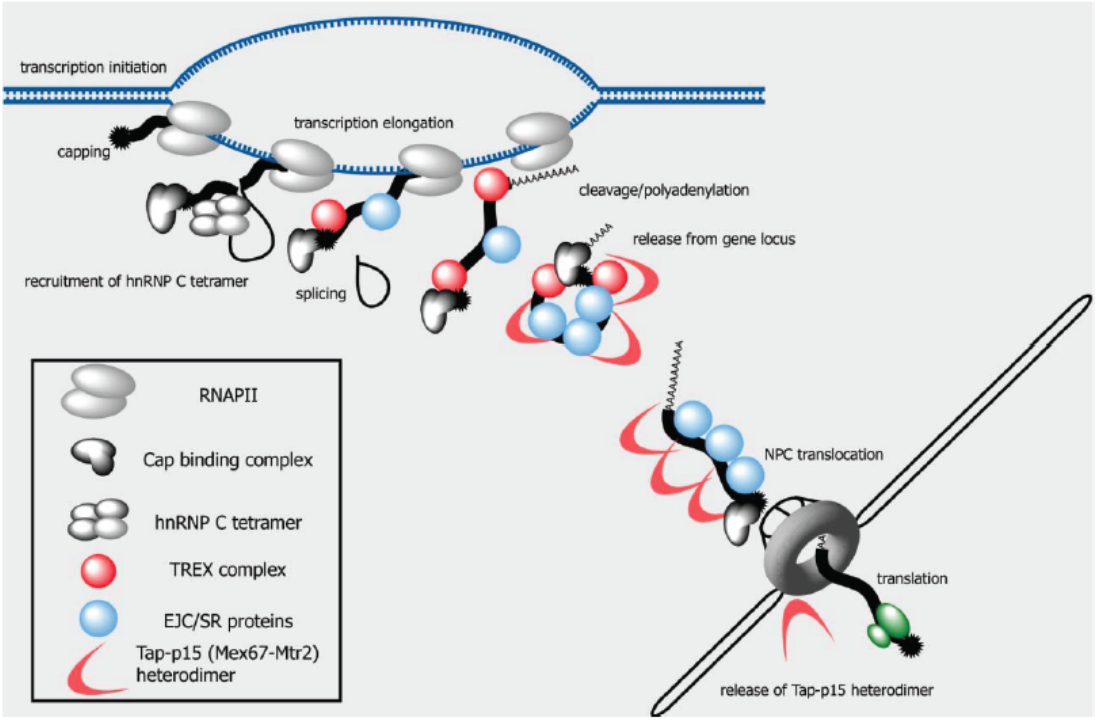


Figure 1.2: The summary of events described in section 1.1. Initiation of transcription by Pol II (RNAPII) is shortly followed by the establishment of the 5' cap. Upon escape from promoter-proximal pausing, Pol II moves along the gene body, all the while RNA processing such as splicing continues on the emerging transcript. The EJC/SR proteins, as well as the TREX components, are recruited co-transcriptionally. After the pA signal was transcribed, the nascent RNA is cleaved at this position and loaded into Mex67-Mtr2, which then facilitates its export via the NPC. Adapted from (Katahira, 2015).

1.2 Nuclear cap-binding complex (CBC)

In 1994, two predominantly nuclear proteins with high affinity to 7mG were identified in HeLa cells, and named Cbp80 and Cbp20 (Cbc1 and Cbc2 in yeast, respectively) (Izaurralde *et al.*, 1994). From here on, I will refer to them with their yeast names.

1.2.1 Structure of CBC

S. pombe CBC is a heterodimer of the 90-kDa Cbc1 and the 21-kDa Cbc2. Cbc2 possesses an RNA recognition motif (RRM) domain spanning residues 34-102 (Mazza *et al.*, 2001), with a relatively disordered C-terminal domain that is important for Cbc1 interaction. Cbc1, on the other hand, does not contain an RRM, and instead accommodates three MIF4G-like domains. These are Armadillo repeat-type folds that act as adapter sequences that encourage interactions between akin domains (Tewari *et al.*, 2010).

Two Tyr residues in *S. pombe* Cbc2, in positions 30 and 41, facilitates the 7mG binding. The aromatic rings of the tyrosines envelop the purine ring of 7mG, "sandwiching" the nucleotide between them, in an interaction referred to as "π -stacking" (Stolarski *et al.*, 1996; Mazza *et al.*, 2002). Alanine substitutions for the human counterparts of these residues, Tyr20 and 43, completely abolish binding to 7mG (Mazza *et al.*, 2001). These residues, as well as the acidic amino acids interacting with 7mG are highly conserved among all eukaryotes. In contrast, Cbc1 shows less conservation; and is considered to have evolved with its other interactors that contain MIF4G-like motifs (Mazza *et al.*, 2001; Worch *et al.*, 2005). As such, the C-terminal domain of Cbc2 is also less conserved, possibly to accommodate for changes in Cbc1.

The 2.1Å-resolution crystal structure of human CBC bound to the cap analog m7GpppG by Mazza et al. can be seen in Figure 1.3. The structure comprises only the middle section of Cbc2, containing the 7mG binding pocket and the RRM domain, while the more disordered C-terminal region was mostly excluded. Nevertheless, the Cbc1-interacting surface of the protein can be observed. The outward-facing surface of Cbc1 acts as a landing platform for its numerous interactors.

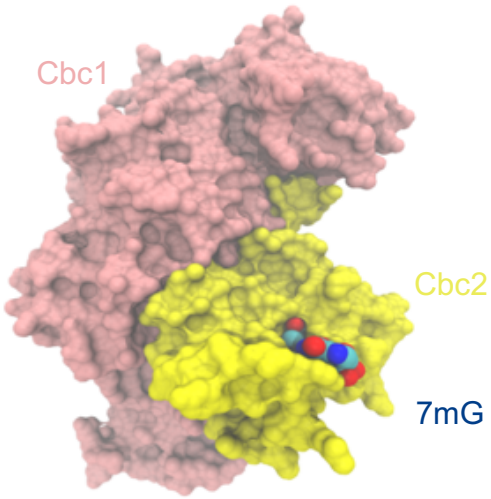


Figure 1.3: The crystal structure of the cap-bound human CBC (PDB accession ID: 1H2T) by (Mazza *et al.*, 2001). The 7mG analog modelled in ball-and-sticks can be seen bound to Cbc2, with Cbc1 interacting with the C-terminal region of Cbc2.

CBC binds the 7mG cap with an equilibrium association constant (K_{as}) as high as 10.8 x 10^7 M^{-1} (Worch *et al.*, 2005). Both proteins were reported to be required for 7mG binding, as Cbc2 alone binds with has about 100-fold less affinity to cap structures (Mazza *et al.*, 2002). The stability of Cbc2-7mG cap complex is strongly dependent on an interaction with the N-terminal tail of Cbc1 (Mazza *et al.*, 2001). Furthermore, Cbc1 lacks a nuclear localization signal (NLS), and thus, in *Arabidopsis thaliana*, Cbc2 directs Cbc1 to the nucleus via one of its NLSs, and is in turn stabilized by Cbc1 binding (Kierzkowski *et al.*, 2009). These data collectively suggest that Cbc1 and Cbc2 should be in a complex for proper binding to 7mG cap and its subsequent roles in RNA biogenesis.

1.2.2 Roles of CBC in transcription

ChIP experiments showed that CBC is recruited to chromatin co-transcriptionally, likely as soon as the 5' cap is established (Castello *et al.*, 2012; Lidschreiber, Leike and Cramer,

2013). In *S. cerevisiae*, CBC interacts with Mot1p, a transcription factor that selectively activates or inhibits transcription on a number of genes (Lahudkar *et al.*, 2011).

Perhaps the most prominent link between CBC and transcription is via P-TEFb. In mammals, CBC was reported to recruit P-TEFb to Pol II, thereby facilitating Ser2 phosphorylation and allowing promoter-proximal pause release (Lenasi, Peterlin and Barboric, 2011). siRNA-mediated depletion of Cbc2 reduces Cdk9 recruitment and led to a global reduction of Ser2P levels (Lenasi, Peterlin and Barboric, 2011). These findings point to a regulatory role for CBC, whereby the recruitment of CBC denotes completion of capping and facilitates the release of paused pol II into the gene body.

1.2.3 CBC and RNA processing

CBC also plays an integral role in subsequent RNA processing events such as splicing and 3' end formation. The finding that CBC was involved in splicing stemmed from the initial observation that cap analogs inhibited splicing *in vitro* (Izaurralde *et al.*, 1994). The complex plays a central role in U1 snRNP-mediated recognition of the 5' splice site (Lewis et al. 1996). When both Cbc1 and Cbc2 are depleted from HeLa cells, co-transcriptional accumulation of U4/U6-U5 tri-snRNP is decreased as well (Pabis *et al.*, 2013). Splicing extracts from *S. cerevisiae* where Cbc1 is immunodepleted show reduced splicing efficiency *in vitro* (Lewis et al. 1996). CBC promotes both the splicing commitment complex formation and removal of U1 snRNP upon association of the U4/U6-U5 tri-snRNP, and is necessary for the stepwise spliceosome assembly (Görnemann *et al.*, 2005). Initially, CBC was implicated in the splicing of only the first intron of a transcript, but more recently, another study showed that incompletely capped pre-mRNAs show inefficient splicing of all their introns (Jiao *et al.*, 2013a).

CBC participates in 3' end formation events, as well. Immunodepletion of CBC from HeLa nuclear extracts results in drastic reduction of endonucleolytic cleavage, which can be rescued by the addition of recombinant CBC (Flaherty *et al.*, 1997). Another strong interactor of CBC, Ars2 (Pir2 in yeast), pulls down with many components of cleavage and polyadenylation specificity and stimulatory factors (CPSF and CstF) (Egan *et al.*,

2014). In mammals, Ars2 stimulates cleavage through its interaction with Clp1, which in turn binds Pol II via Pcf11. The 3' cleavage-stimulatory effect of CBC-Ars2 was the strongest for cap-proximal polyadenylation signals. However, the interaction between Clp1 and CBC-Ars2 was found to be sensitive to RNase A treatment, and may therefore not be direct (Hallais *et al.*, 2013).

1.2.4 CBC and nuclear export

In mammals, there are two distinct export pathways described for snRNAs and mRNAs. CBC has been linked with both.

snRNAs that make up the spliceosome have to be processed and receive their protein components in the cytoplasm before translocating back into the nucleus, where they function (reviewed in Matera & Wang 2014). Initial evidence for CBC involvement in nuclear export of snRNAs came from *X. laevis* oocytes, where 7mG was required for their efficient export (Hamm and Mattaj, 1990; Izaurralde *et al.*, 1995). In higher eukaryotes, CBC interacts with PHAX, facilitating snRNA export (Ohno *et al.*, 2000; Segref, Mattaj and Ohno, 2001).

mRNA export is also promoted by CBC, but via a different mechanism. Here, both CBC proteins interact with the TREX member Yra1 (ALY/REF) (Cheng *et al.*, 2006; Nojima *et al.*, 2007). Yra1 is recruited to the 5' end of pre-mRNA in a CBC-dependent manner and thus the export of mRNAs is carried out in a 5'-3' direction (Cheng *et al.*, 2006), which is in line with electron micrography observations of transporting Balbiani ring mRNPs (Daneholt, 1997).

What distinguishes between spliceosomal snRNAs and mRNAs, both of which are transcribed by Pol II and undergo similar modifications such as 5' capping and polyadenylation? One differentiating feature is their size; with snRNAs averaging 150 nt as opposed to the much longer mRNAs. As described in section 1.1.5, hnRNP C can function as a "molecular ruler", selectively binding to RNA longer than 200-300 nt *in vitro*. hnRNP C tetramers interact with Cbc1, competitively inhibiting PHAX, but not Yra1, binding to CBC. Therefore, longer RNAs are directed to the TREX-mediated export

pathway (McCloskey *et al.*, 2012). On the other hand, iCLIP experiments with PHAX showed only a moderate specificity to pre-snRNAs and also displayed binding to pre-mRNAs *in vivo* (Giacometti *et al.*, 2017). Hence, further studies should be conducted to determine if, and how, RNA identity is assigned in the nucleus, which in turn actuate a certain export pathway.

1.3 RNA surveillance

Faulty RNA, such as not correctly transcribed or inefficiently spliced, can be devastating to the cellular processes if left unattended. Therefore, a plethora of mechanisms are in place to ensure these RNAs do not end up in cytoplasm or are not translated into proteins. Though there are also RNA surveillance mechanisms that take place in the cytoplasm, for the sake of this study, I will only focus on nuclear RNA surveillance.

Eukaryotic organisms have a prevalence of 5'-3' or 3'-5' exonucleases, rather than endonucleases; and thus, the protective modifications on RNA such as the 5' cap or the polyA tail provide a safeguard against degradation from their respective ends.

1.3.1 The nuclear exosome and its associated cofactors

The major RNA surveillance activity in the nucleus is presented by the nuclear exosome. Originally defined for its role in processing precursor rRNAs, sn-/ and snoRNAs (Allmang *et al.*, 1999), we now know that it can target a wide range of transcripts, from aberrant mRNAs and cryptic unstable transcripts (CUTs) to pervasive transcription products (Wyers *et al.*, 2005; Davis and Ares, 2006).

The exosome comprises a catalytically inert central hexameric ring with three additional subunits on the top, called the "barrel". This 9-subunit complex is collectively known as the exosome core, and creates a tunnel into which RNA can be fed. At the bottom, the core associates with a processive endo- and 3'-5' exonuclease, Dis3; and at the top, with another 3'-5' exonuclease called Rrp6 (reviewed in Chlebowski et al. 2013). RNA is fed through a channel to the barrel in 3'-5' orientation and reaches Dis3, leading to processive

degradation (Makino, Baumgärtner and Conti, 2013). A recent crystal structure revealed that Rrp6 also uses the same channel, raising the question how RNA chooses which direction (top/Rrp6 or bottom/Dis3) to go once it enters the barrel (Wasmuth, Januszyk and Lima, 2014). While the path selection is stochastic *in vitro*, it is possible that certain combination of cofactors *in vivo* might skew the path selection in favour of one exonuclease *in vivo*. Though there is a high overlap between Dis3 and Rrp6 targets, CRAC studies also revealed that Rrp6 was preferentially associated with more structured RNAs, raising the possibility that some RNAs are not unwound to enter the barrel and are instead degraded outright by Rrp6 (Schneider *et al.*, 2012).

Catalytic activity of the isolated exosome complex is much weaker as compared to what is observed *in vivo* (Liu, Greimann and Lima, 2006). Furthermore, Dis3 alone has a more rapid substrate turnover than Dis3 in complex with the exosome core (Wasmuth and Lima, 2012), suggesting that the core restricts access to Dis3. Entry into barrel is in turn expectedly regulated by RNA helicases, acting as regulators of the exosome.

1.3.1.1 TRAMP complex

One complex that can modify and feed the RNA to be degraded into the nuclear exosome is the Trf4/5-Air1/2-Mtr4 polyadenylation, or TRAMP, complex. Trf4/5 are poly(A) polymerases, Air1/2 are RNA-binding proteins, while Mtr4 is a DExH-box RNA helicase (LaCava *et al.*, 2005; Vanácová *et al.*, 2005; Hardwick and Luisi, 2013). Upon RNA binding, Air1/2 recruits Trf4/5, which adds a short oligo(A) tail to the 3' end, to which Mtr4 has high affinity. Mtr4 then unwinds the polyadenylated RNA and targets it to degradation via the exosome (LaCava *et al.*, 2005). This oligo(A) stretch is shorter than the poly(A) tail that is added to Pol II transcripts as part of the 3' end formation, the length of which is controlled by Mtr4 (Jia *et al.*, 2011).

TRAMP complex is conserved in *S. pombe* and higher eukaryotes; but its subcellular localization and reported function are somewhat different. Whereas Mtr4 is localized to both nucleoplasm and the nucleolus in *S. cerevisiae*, it is confined to only the nucleolus in *S. pombe* and higher eukaryotes (Wyers *et al.*, 2005; Lubas *et al.*, 2011), where it functions

in degrading aberrant rRNAs and tRNAs. In yeast, TRAMP complex also undertakes another role in the turnover of heterochromatic RNA, providing exosome-mediated heterochromatin silencing (Verdel *et al.*, 2004; Vasiljeva *et al.*, 2008). On the other hand, in *S. pombe* and higher eukaryotes, the RNA surveillance role of TRAMP complex is mediated by different complexes.

1.3.1.2 MTREC complex in *S. pombe*

In *S. pombe*, the Mtl1-Red1 core (MTREC) complex takes on the RNA surveillance role from the TRAMP complex. Here, the RNA helicase Mtl1 unwinds unspliced mRNAs, CUTs, and some meiotic mRNAs and marks them for degradation by the exosome (Sugiyama *et al.*, 2015). During vegetative growth of fission yeast, meiotic mRNAs are degraded by the MTREC complex with the help of Mmi1, a specificity factor that recognises sequence motifs called determinants of selective removal (DSR) (Harigaya *et al.*, 2006; Yamanaka *et al.*, 2010). These decay-promoting sequences can be present in introns, thereby targeting transcripts where splicing is not completed rapidly, for degradation (Kilchert *et al.*, 2015).

1.3.1.3 NEXT and PAXT complexes in *H. sapiens*

In humans, Mtr4 can associate with two RNA-binding proteins, RBM7 and ZCCHC8, to form another RNA surveillance complex, named the nuclear exosome targeting (NEXT) complex. RBM7 binds many Pol II transcripts, such as promoter-proximal transcripts (PROMPTs), enhancer RNAs, and pre-mRNAs (Lubas *et al.*, 2015). The RRM domain of RBM7 has a preference for U-rich pyrimidine sequences, and associates with 3’ ends of introns, and snoRNAs that derive from these introns (Hrossova *et al.*, 2015).

Interestingly, NEXT diverged from its yeast counterparts in that it does not require polyadenylation of the target RNA. Instead, Mtr4 can associate with yet another set of proteins, ZFC3H1 and a poly(A)-binding protein, PABPN1 (Pab2 in *S. pombe*) to form poly(A) RNA exosome targeting (PAXT) connection (Meola *et al.*, 2016). Similarly, two nuclear PABPs in *S. pombe*, Pab2 and Nab2, can target RNA for degradation or nuclear

export, respectively (Grenier St-Sauveur *et al.*, 2013). These discoveries highlight the function of the poly(A)-binding proteins as a determinant of RNA fate.

1.3.2 5'-3' degradation

Unlike the prevalence of 3' exonucleases, there are only two major 5' exonucleases in eukaryotes: The cytoplasmic Xrn1 (Exo2 in *S. pombe*) and the predominantly nuclear Rat1 (Dhp1). Presence of a 5' triphosphate inhibits most of their catalytic activity (Stevens and Poole, 1995). Furthermore, the addition of the 7mG cap, and subsequent binding by CBC block off degradation in this direction by restricting decapping enzyme access (Jiao *et al.*, 2013b), and thus, a targetable 5' monophosphate cannot be formed by the Rat1-accompanying enzyme Rai1 (Xiang *et al.*, 2009). Conversely, uncapped or defectively capped (i.e. 7^{th} residue is not methylated) mRNAs are rapidly decapped and phosphorylated by Rai1, recruiting Rat1 and subsequent 5'-3' degradation of the faulty RNA (Brannan *et al.*, 2012).

Another case where a 5' monophosphate is generated is during 3' end formation. When the nascent RNA is cleaved by the cleavage and polyadenylation machinery, it creates an entry point for Rat1. This forms the basis of the torpedo model of transcription termination (introduced in section 1.1.4).

1.3.3 The interwoven RNA surveillance: Where does the CBC come in?

RNA surveillance is intertwined with RNA biogenesis and processing. This ensures that, at any given time, if something goes wrong with the transcription or the concomitant processing stages, the RNA is rigorously degraded, so as not to give rise to further problems at a later time. What follows is a brief review of only a few points of entry into transcription, processing and export for RNA surveillance mechanisms.

As described in section 1.1.2, the 7mG cap is established very early during transcription. In the event that the addition of the 5' cap fails, RNA degradation follows by the 5' exonuclease Rat1 (Brannan *et al.*, 2012). The recruitment of CBC imparts further stability to the 7mG cap, as the complex inhibits binding of decapping enzymes (Jiao *et al.*, 2013b).

During elongation, both ends of the RNA are protected; 5' by the 7mG cap and 3' by the Pol II. However, Pol II can backtrack along the DNA if elongation is impeded, which can expose the 3' end (reviewed in Nudler 2012). If the backtracked Pol II is not rescued, the nuclear exosome can help degrade the aberrant RNA and promote transcription turnover (Lemay *et al.*, 2014).

As mentioned in section 1.1.3, after cleavage, polyadenylation of the 3' end occurs rapidly, and this is crucial for the protection of the transcript from degradation in 3'-5' direction. Specifically, the protein that decorates the poly(A) tail is important in RNA fate determination. In *S. pombe*, if Nab2 binds the tail, it facilitates nuclear export; conversely, Pab2 binding results in the recruitment of the nuclear exosome to the transcript (Lemay *et al.*, 2010; Grenier St-Sauveur *et al.*, 2013). The human homolog of Pab2, PABPN1, along with CBC, forms part of the PAXT connection, which can target RNAs for degradation (Meola *et al.*, 2016).

Unspliced transcripts are rapidly degraded via the exosome, so quickly that splicing mutants are usually combined with an exosome mutant to observe the intron-retained transcripts. In fission yeast, introns may contain decay-promoting sequences which mediate Mmi1-driven recruitment of exosome and subsequent degradation (Kilchert *et al.*, 2015). It is currently unknown how introns without these motifs could promote RNA surveillance, but one hypothesis states that the machinery responsible for the degradation of the excised intron could be the culprit, as a delayed splicing would expose the 3' end of the transcript, targetable by the exosome (Bresson and Tollervey, 2018). CBC indirectly influences this surveillance mechanism by facilitating splicing (Izaurralde *et al.*, 1994; Jiao *et al.*, 2013b).

The efficacy of the nuclear RNA surveillance, combined with the low number of specificity factors (i.e. proteins or sequences that promote degradation) discovered, has led to the idea of a "destruction timer" for transcripts. This hypothesis postulates that if a transcript spends too much time in the nucleus, it will eventually get targeted by the nuclear RNA surveillance pathways (Gudipati *et al.*, 2012). Therefore, the most efficient way for an RNA to escape decay is to be swiftly exported. CBC and Ars2 are known to promote

export via interacting with Yra1, competing with Mtr4, which would otherwise promote degradation (Zhou *et al.*, 2017). For RNAs that have nuclear roles, stabilizing proteins, including CBC and Nab2, delay the "destruction timer" by competing with the degradation machinery for binding the 5' and 3' ends, respectively.

1.4 Cbc3: A new cap binder?

A recent study proposed that an alternative CBC might form between Cbc1 and another protein containing a 7mG binding pocket, which was named NCBP3 in human nomenclature (Gebhardt *et al.*, 2015). This gene has an orthologue in *S. pombe*, an unannotated gene with the ID *SPCC16C4.16c*, while any apparent ortholog is missing in *S. cerevisiae*. For the sake of convenience, the product of this gene will be referred to as Cbc3.

In HeLa cells, Cbc3 was shown to selectively bind 7mG *in vitro* and to interact with Cbc1 *in vivo*. Human Cbc3 has a canonical RRM motif sharing a sequence identity of 16.4% with that of Cbc2; however, its mode of 7mG binding appears to be more similar to poly(A)-associated ribonuclease (PARN), where 7mG is trapped between an aromatic residue (W68) and the body of the protein. Knockdown of Cbc2 was not sufficient to abolish binding of Cbc1 to mRNAs; however, a double knockdown of Cbc2 and Cbc3 showed a drastic decrease in mRNA-bound Cbc1 levels. They proposed a division of labour between Cbc2 and Cbc3 in binding snRNAs, long interspersed ncRNAs and antisense RNAs; or mRNAs, respectively. They also demonstrated that Cbc3 interacts with TREX components Tho1-2, Tex1 and Uap56. Upon depletion of both Cbc2 and Cbc3, nuclear retention of poly(A)$^+$ RNA was observed. They concluded by showing that although *cbc3Δ* showed no apparent growth defect, these cells were more prone to viral infections (Gebhardt *et al.*, 2015).

This view of alternative vs. canonical CBC was challenged by a recent structural study showing that CBC-Ars2 can accommodate Cbc3 *in vitro* (Schulze *et al.*, 2018). Schulze et al. were also able to form a complex with CBC-Ars2-Cbc3 bound to the cap analog, m7GTP, indicating that cap binding of Cbc3 might not be the only means of interaction.

Another interesting observation from the latter study was that the CBC-Ars2-Cbc3 ternary complex was mutually exclusive from the CBC-Ars2-PHAX complex. This may be the means of recognition for mRNAs vs. snRNAs, exported via either the PHAX- or TREX-mediated mechanisms (described in section 1.2.4).

Unlike Cbc1 and Cbc2, the Cbc3 orthologue in *S. pombe* is a non-essential gene. Considering that the human counterpart becomes relevant under viral stress, it is possible that *S. pombe* Cbc3 might also take on a similar, stress-relieving role.

1.5 Objectives of this work

The importance of the CBC has long been established in *S. cerevisiae* and higher eukaryotes. However, the *S. cerevisiae* orthologues of Cbc1 and Cbc2 are non-essential. Furthermore, CBC, as well as the 5' cap, are dispensable for most mRNA export in this organism (Fresco and Buratowski, 1996). Combined with the fact that *S. pombe* carries many more intron-containing genes as compared to the budding yeast, this puts the *S. pombe* CBC in an interesting position, situated between the higher and basic eukaryotes. Thus, exploring the *S. pombe* CBC might shed light on the evolution of this important complex and its roles. Moreover, the emergence of a new potential cap binder, Cbc3, encourages the investigation of the function of this protein, as well as its relation to the CBC proper, in fission yeast.

In Chapter 2, I conduct studies to understand the role of CBC components in RNA regulation and identify their interacting partners. Comparing these profiles among the three proteins reveal some differences between Cbc1 and Cbc2, and provides a heavy contrast between Cbc3 and the CBC.

In Chapter 3, I set out to investigate the differences between Cbc1 and Cbc2. Previously known to be exclusively in the same complex and deemed to require each other to work, I investigate the transcriptome changes upon depletion of the two proteins, revealing some inherent differences between the two depletion strains. To explain these variations, I check the genetic interactions between CBC and RNA regulatory pathways. Curiously, Cbc1

depletion has an ameliorating effect on a Dhp1 mutant. Readthrough transcription is suppressed upon Cbc1, but not Cbc2, depletion. Therefore, this thesis provides the first clear evidence, to my knowledge, of Cbc1 having a role separate from Cbc2.

Finally, in Chapter 4, I undertake functional analysis of *S. pombe* Cbc3. Deletion of the gene results in a general downregulation of many protein-coding, but not non-coding, transcripts. Interestingly, longer genes appear to be more affected by the Cbc3 loss. The mammalian findings on Cbc3 having a role in mRNA export are corroborated, and an additional function for Cbc3 in Pol II transcription is described. Cbc3 deletion yields a slower elongating Pol II, and this might be due to reduced recruitment of elongation factor Paf1 to longer genes.

Chapter 2 Interacting partners and genome-wide distribution profiles of Cbc1, Cbc2 and Cbc3

As explained above, CBC is extensively studied among many different organisms. However, fission yeast has been excluded from these studies. Keeping in mind the evolutionary state of *S. pombe*; being more similar to higher eukaryotes in terms of transcription regulation (Hoffman, Wood and Fantes, 2015), investigating Cbc1 and Cbc2 in this organism, and comparing them to budding yeast vs. higher eukaryotes could provide valuable insights. To this end, it is relevant to check the interacting partners of both proteins, as well as their recruitment profiles. Furthermore, a comparison of interactors for the two CBC proteins should be informative as to whether they have somewhat diverging roles, as well.

Cbc3 is a brand new cap binder found in HeLa cells (Gebhardt *et al.*, 2015), and has an orthologue in *S. pombe*: *SPCC16C4.16c*. Once again, checking the interacting partners would open avenues to explore. One question unanswered in Gebhardt et al.'s study was whether Cbc3 has co-transcriptional recruitment, similar to canonical CBC components. It makes sense for the recruitment of a cap binder to be coupled to transcription, as this would ensure protection from nuclear decapping enzymes Dcp1 and Dcp2 (Brannan *et al.*, 2012). If this is indeed the case, another question that arises is if there is a subset of genes that recruit Cbc3 versus Cbc2, and why.

2.1 Cbc1, Cbc2 and Cbc3 are recruited co-transcriptionally, but show different distribution profiles

To understand if the recruitment of Cbc1, Cbc2 and Cbc3 is coupled to transcription in *S. pombe*, the proteins were tagged C-terminally with a FLAG-TEV-Protein A (FTP) epitope.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) was carried out to assess co-transcriptional recruitment of the proteins. Briefly, chromatin-bound proteins were pulled down using rabbit IgG beads, which have affinity toward the Protein A portion of the tag. The DNA fragments obtained were then subjected to high-throughput sequencing on the Illumina NextSeq platform.

Heatmaps and metagenes for the three proteins (Figure 2.1) suggested that all were recruited co-transcriptionally to protein-coding (pc) genes, as compared to their respective input samples (data not shown). However, while Cbc1 and Cbc2 had similar recruitment profiles, peaking close to the TSS, Cbc3 signal increased over the gene body, peaking around the pA site.

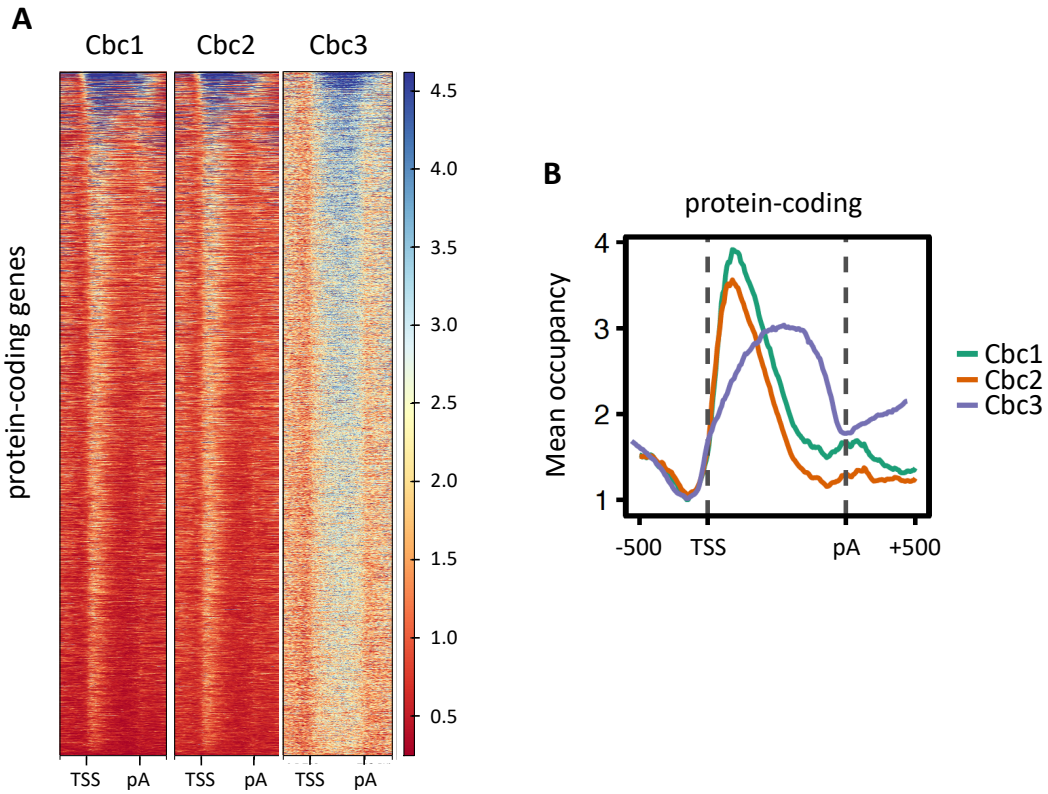


Figure 2.1: ChIP-seq for Cbc1-2-3 points to a co-transcriptional recruitment, but with different patterns. (A) Heatmaps for the three proteins in pc genes, with signals colour-coded from red

(lowest) to blue (highest). ChIP-seq alignments were binned to 0.5 kb before TSS, 1 kb for the gene body, and 0.5 kb after pA site. (B) Combined metagene profile for the three proteins, in pc genes. Cbc1, Cbc2 and Cbc3 signals are represented in green, red, and purple, respectively. Data are binned as in panel A. Signal is represented as mean from three independent replicates.

In a ChIP-seq experiment, there are two possibilities to explain the signal: 1. the protein is bound to the nascent RNA; or 2. it is recruited to the chromatin, or a chromatin-bound protein. Only the former case could be deemed "co-transcriptional recruitment", as the latter might be a protein that can simply bind DNA. To distinguish whether the proteins are directly recruited via RNA or direct binding to the transcription machinery, ChIP-seq for Cbc1 and Cbc3 were repeated, either with or without RNase A treatment. It can be seen in Figure 2.2 that the Cbc1 signal next to the TSS was lost completely upon RNase A treatment before the IP, attesting that all ChIP-seq signal was by RNA-bound Cbc1. On the other hand, signal for Cbc3 was only halved in the RNase-treated pulldown, indicating that Cbc3 recruitment is only partially dependent on RNA presence.

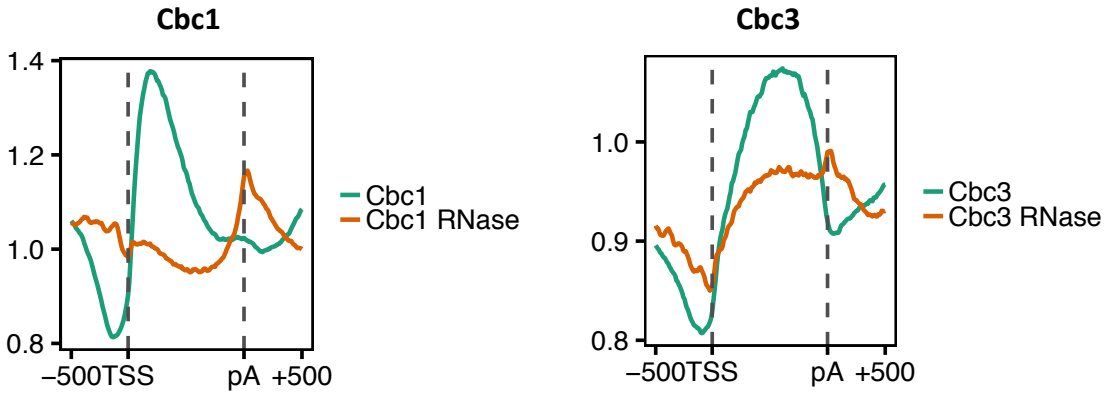


Figure 2.2: RNA requirement for the recruitment of Cbc1 and Cbc3. ChIP-seq for Cbc1 (left) or Cbc3 (right) was performed with or without RNase A treatment before the pulldown, and metagene profiles for pc genes were plotted as in Figure 2.1B. RNase A-treated sample (red) show a

reduction of signal compared to the untreated (green); while Cbc1 recruitment is entirely RNA-dependent, Cbc3 recruitment still occurs to a lower extent.

Since Gebhardt et al. put forward Cbc3 as an alternative cap binder to Cbc2, it would make sense for these proteins to be recruited to different subsets of genes. This, however, did not appear to be the case for pc genes. Only about a quarter of the pc genes assessed (n = 1078 out of 4104) had significant Cbc3 signal across the replicates, and all of them showed recruitment of CBC, as well. On the other hand, non-coding (nc) genes had a lack of Cbc3 signal, while they contained CBC across the gene body (Figure 2.3). The odd distribution of the canonical CBC proteins may be attributed to the fact that nc genes are quite short in fission yeast, and the low resolution of ChIP experiment may not be sufficient to show an accurate distribution profile.

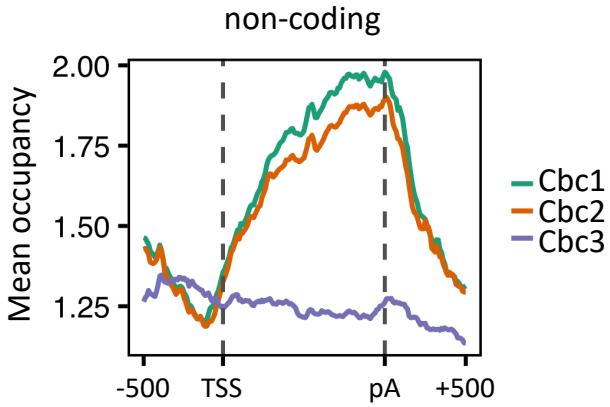


Figure 2.3: Metagene profile for the recruitment of Cbc1-2-3 in nc genes. Data were binned as Figure 2.1B, and Cbc1 (green), Cbc2 (red) and Cbc3 (purple) signals were represented as mean occupancy versus no-tag control. Signal is represented as mean from three independent replicates.

2.2 Purification of CBC components and Cbc3

Using a tandem affinity approach, Cbc1-2 were purified under physiological salt conditions (150 mM NaCl). These IPs were initiated with an IgG pulldown, followed by TEV cleavage and subsequent FLAG purification; eluted by competition with 3x FLAG peptide. In contrast, Cbc3 was purified with a 1-step IgG purification under stringent conditions (500 mM NaCl, 1 M urea).

Silver staining gels demonstrate that one CBC component almost exclusively eluted with the other, with not many other prominent bands (Figure 2.4A). On the other hand, Cbc3 elution showed several bands between 50-75 kDa, and one faint band above 37 kDa (Figure 2.4B). No band corresponding to Cbc3 was visible in Cbc1 or Cbc2 pulldowns. Western blots confirmed efficient pulldown of all three proteins, with very little remaining in the flowthrough (Figure 2.4C). Note that for Cbc1 and Cbc2 purifications, eluted band is smaller in size due to TEV cleavage. For Cbc3, TEV cleavage is not utilized as part of the purification, and thus, the eluate band is the same size.

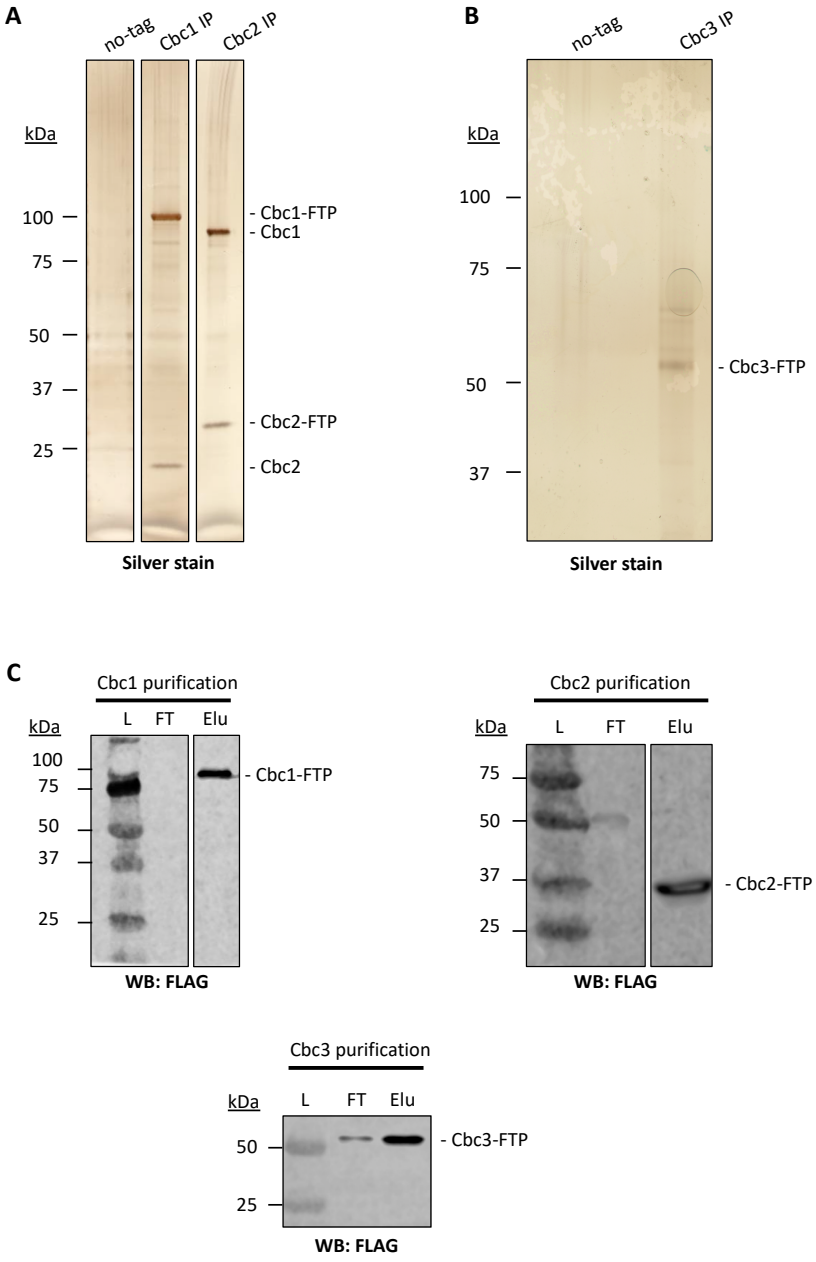


Figure 2.4: Native purifications for Cbc1, Cbc2 and Cbc3. Representative silver staining gels from Cbc1-2 (A) and Cbc3 (B) purifications. 4% of the elution fraction was loaded into each well. (C) Western blot images for the respective purifications. Images are representative of three purifications for Cbc1-2, and two purifications for Cbc3.

Pulldowns were then subjected to tandem liquid chromatography mass spectrometry (LC-MS/MS) with a 1 h-gradient to identify the interacting proteins present in the Cbc1-2-3 elutions. Three replicates of Cbc1 and Cbc2; and two replicates of Cbc3 purifications were analysed by MS and compared to the no-tag control IPs.

Table 2.1 provides a list of proteins that were enriched in the Cbc1 pulldown. Expectedly, the highest interactor was the other CBC member, Cbc2. Not surprisingly, the list included the P-TEFb components Pch1 and Cdk9, as well as RNA-binding proteins involved in splicing (i.e. Prp1/19/39, Srp2). Furthermore, some 3' end formation, cleavage and polyadenylation factors such as Cft1, Cft2, Pta1 and Rna14 were also present in two replicates out of three. The TRAMP complex RNA helicase Mtr4 could also be observed, albeit with lower intensity.

Table 2.1: List of proteins that came up in the Cbc1 pulldown, with descending intensity scores. Proteins shown to be interacting with both Cbc1 and Cbc2 are highlighted in cyan. Intensity values are obtained from MaxQuant analysis. The results are an average of three independent replicates.

Protein name	**Average Intensity**
Cbc1	1,38E+12
Cbc2	2,82E+10
Cdk9	2,39E+08
Prp1	1,47E+08
Pch1	1,18E+08
Pabp	1,15E+08
Prp19	3,99E+07
Prp39	3,16E+07
Srp2	2,10E+07
Mtr4	8,09E+06

Table 2.2 summarizes the interacting proteins eluted alongside Cbc2. As expected, Cbc1 was precipitated as the highest interactor, followed by RNA helicase Dbp2 and the P-TEFb components. All proteins pulled down with Cbc2 were also interactors of Cbc1.

Table 2.2: List of proteins that came up in the Cbc2 pulldown, with descending intensity scores. Proteins shown to be interacting with both Cbc1 and Cbc2 are highlighted in cyan. Intensity values are obtained from MaxQuant analysis. The results are an average of three independent replicates.

Protein name	Average intensity
Cbc1	5,22E+09
Cbc2	2,02E+09
Dbp2	1,01E+09
Cdk9	3,14E+07
Pch1	1,42E+07

Finally, Table 2.3 lists the accompanying proteins in Cbc3 pulldown. RNA-binding proteins that have nuclear export roles appear as prominent interactors. Examples include ATP-dependent RNA helicase Dbp5, nucleoporin component Rae1, as well as TREX member Uap56. The only shared protein in this pulldown, which also showed up in Cbc1 IP, was Pabp.

Table 2.3: List of proteins that came up in the Cbc3 pulldown, with descending intensity scores. Intensity values are obtained from MaxQuant analysis. The results are an average of two independent replicates.

Protein name	Average intensity
Cbc3	3,75E+09
Dbp5	6,25E+08
Dbp3	6,09E+08
Pabp	4,18E+08
Uap56	4,88E+07
Rae1	1,72E+07

2.3 Discussion

2.3.1 Co-transcriptional recruitment of CBC components and Cbc3 to pc genes

To start understanding the CBC in fission yeast, first the recruitment of CBC proteins was assessed. Not surprisingly, co-transcriptional recruitment close to the TSS was observed for both Cbc1 and Cbc2. This corroborates the data on budding yeast and human cells (Tuck and Tollervey, 2013; Giacometti *et al.*, 2017). No genes could be detected where Cbc1 or Cbc2 was recruited without the other, implying that, at least for co-transcriptional recruitment, both of the proteins were required.

In contrast, Cbc3 recruitment was observed over the gene body, peaking close to the pA site. This pattern resembles the recruitment of elongation factors, as well as THO components Tho2 and Gbp2 (Tuck and Tollervey, 2013; Fischl *et al.*, 2017). Two possibilities for the Cbc3 binding are as follows: 1. Cbc3 is recruited to the 7mG, but later in transcription, replacing Cbc2 and possibly the CBC proper; or, 2. Cbc3 is recruited on the RNA independently of the cap. The second instance makes it possible for Cbc3 and CBC to co-occupy an RNA. Nascent RNA-based techniques could be utilized to reveal where exactly Cbc3 binds on RNA.

Another intriguing finding was that Cbc3 was recruited to a subset of pc genes, whereas CBC was present near the TSS of most Pol II-transcribed genes. Cbc3-recruiting genes had no significant overlap with intron-containing genes, nor did they have any visible bias for gene length. A GO-term analysis shows that genes involved in metabolic processes were enriched in Cbc3-recruiting genes. This might be due to the fact that these genes are more highly transcribed and might be easier to pull down in a ChIP experiment. Combined with the low abundance of the protein (Carpy *et al.*, 2014), it is possible that I could not detect some of the genes that could bind Cbc3 due to technical limitations.

To understand the RNA dependency of the recruitment, RNase treatment was utilized before the IP. In line with previous reports, Cbc1 recruitment was RNA-dependent (Izaurralde *et al.*, 1994). In contrast, Cbc3 recruitment still occurred after RNase treatment. This implies that Cbc3 either interacts with the transcription machinery to stay close to the chromatin, or that it can directly interact with the chromatin. The loss of signal could not be attributed to a library size or IP efficiency difference due to the inclusion of *S. cerevisiae* spike-in; therefore, it can be inferred that co-transcriptional recruitment of Cbc3 is not entirely RNA-dependent. Finally, the peak close to the pA site observed in all RNase-treated IPs appears to be an artefact, as it can also be seen in RNase-treated input samples.

2.3.2 CBC interacts with P-TEFb and splicing factors

Our purification has identified several interacting partners of Cbc1 and Cbc2. We neglected hits coming from the non-tagged strain as background. It is worth noting that intensity values for our hits were the highest for Cbc1 and Cbc2, and all the remaining hits had relatively low intensities. This might be due to the fact that the interaction between Cbc1 and Cbc2 was the strongest among all the interactors, and everything else were sub-stoichiometric, likely reflecting the transient nature of these interactions. This would be hardly surprising as the known interactors for the CBC include pre-mRNA splicing factors, 3'end processing factors and RNA export proteins; all being very rapid processes. Moreover, considering how, at one point, a given capped transcript might be undergoing splicing, 3'end processing or mRNA export, it could be expected that a pool of pulled down CBC should include sub-stoichiometric levels of all such interactors, with the highest and the most confident interaction being the one between Cbc1 and Cbc2; as this is present on the 7mG cap in all the aforementioned cases.

Numerous splicing factors showed up in our MS studies, such as Prp1, Prp19, Prp39 and Srp2. These hits also possessed high intensity values, indicating that at any given time, it is more likely to find these proteins in complex with Cbc1 or Cbc2 than the other hits. Such interactions were already known for CBC for a long time (Edery and Sonenberg,

1985; Izaurralde *et al.*, 1994; Lewis *et al.*, 1996; Mayes *et al.*, 1999; Pabis *et al.*, 2013), indicating that our pulldowns were successful.

Another interesting group of hits were Pch1 and Cdk9. Subunits of P-TEFb, these proteins are integral to the transition from transcription initiation to elongation (Peterlin and Price, 2006). P-TEFb phosphorylates the Ser2 residues at the C-terminal domain (CTD) of RNA pol II, giving the signal for productive elongation (Ni *et al.*, 2008). A publication showed that, in HeLa cells, Cdk9 and Pch1 co-immunoprecipitated with Cbc1 and Cbc2; and that siRNA-mediated knockdown of Cbc2 resulted in reduced Cdk9 occupancy in promoter regions, decreased Ser2-P levels, and prolonged promoter-proximal pausing (Lenasi, Peterlin and Barboric, 2011). Both P-TEFb proteins were pulled down with Cbc1 and Cbc2, underlining the intensity and importance of the interaction.

RNA helicase Dbp2 was yet another shared interactor between Cbc1 and Cbc2, though it was more abundantly present in the latter pulldown. Dbp2 has a role in poly(A)+ mRNP assembly, as well as orchestrating transcriptional fidelity. Its loss is associated with inefficient 3' end formation and emergence of aberrant transcriptional initiation products (Cloutier *et al.*, 2012; Ma, Cloutier and Tran, 2013). It can thus be speculated that achieving this fidelity might require interaction with the CBC.

Interestingly, a few of the interacting proteins such as Prp1-19-39, Srp2 and Pabp were specific to only Cbc1, and not Cbc2. This raises the possibility that Cbc1 might also have separate roles where it might not require Cbc2, or cap binding. Indeed, it was previously reported that Cbc1, but not Cbc2, could be co-purified with eIF4E-bound transcripts (Rufener and Mühlemann, 2013). Another publication showed that Cbc1 can interact with poly(A)-specific ribonuclease (PARN), preventing its 3'-5' exonuclease activity, while Cbc2 did not possess such an effect (Balatsos *et al.*, 2006). Conversely, alongside its 7mG binding pocket, Cbc2 has an RRM domain that can be used to bind RNA regardless of the 7mG cap (Baejen et al. 2014). Furthermore, it was shown in mammalian cells that a splice variant of Cbc2 that lacks the 7mG binding domain as well as part of the RRM domain was still able to bind mRNA and localize to active transcription sites (Pabis *et al.*, 2010).

Thus, it is possible that the RRM domain of Cbc2 can be another means by which it can bind mRNA; a binding where it might not necessarily require 7mG, or the stabilization of the binding exerted by Cbc1. However, Cbc2 is a less abundant protein as compared to Cbc1, and as such, random sampling in MS might skew against low-abundance interactors, and instead may favour Cbc1 interactors. Another possibility is that, since Cbc2 is "sandwiched" between the nascent RNA and Cbc1 (Figure 1.3), it might be harder to pull down Cbc2 individually; and therefore, all pulled-down Cbc2 might be in contact with Cbc1, and consequently, with Cbc1's interacting partners.

2.3.3 Cbc3 interacts with export factors and a number of RNA helicases

Cbc3 pulldown in this study was done under stringent conditions, due to the high level of background obtained under physiological conditions. Another reason why we required stringent conditions was that the FLAG-tagged Cbc3 had a tendency to stick to M2 resin; and thus, the purification employed only the IgG pulldown. Therefore, I should emphasize that the list of interactors presented here is certainly not exhaustive. Nevertheless, comparing these results to the mammalian Cbc3 pulldowns would attest that the fission yeast Cbc3 carries some similarities.

S. pombe Cbc3 interacted with nucleoporin components Rae1, Nup61 and Nup82 (with lower intensity values), as well as the TREX adaptor Uap56. However, unlike Gebhardt et al.'s report, I failed to detect any THO complex members. This might be either due to the stringent purification conditions utilized or because this interaction was more transient. Moreover, a number of RNA helicases could be detected, which were missing from the no-tag control eluate. Dbp5 appeared to be the most promising among them, as it is known to play a role in nuclear export (introduced in section 1.1.5).

It is also important to underline that neither Cbc1 nor Cbc2 purifications yielded any peptides from Cbc3. Human Cbc1 was implicated to interact with Cbc3 to form an alternative CBC *in vivo* (Gebhardt *et al.*, 2015) and *in vitro* (Schulze *et al.*, 2018), but this does not appear to be the case in *S. pombe*. Conversely, for Cbc3 purification, the sole

prominent band in the silver staining was that of Cbc3 and MS analysis did not show any CBC components. However, it is still possible that such Cbc1-3 interaction is highly transient, and as such, could not be detected alongside the much more avid interaction between canonical CBC members.

Consequently, two conclusions can be made at the end of this chapter: Firstly, despite almost identical recruitment patterns, Cbc1 and Cbc2 appeared to have some distinct interacting proteins. Since previously the two CBC proteins were always thought to act together, I continued to further investigate these potential differences, to be reported in Chapter 3. Secondly, Cbc3 did not resemble any of the CBC components in terms of its recruitment pattern, RNA dependence or its interacting partners. I endeavoured to understand better the role of this protein in fission yeast, to be reported in Chapter 4.

Chapter 3 The possibility of Cbc1 and Cbc2 having separate roles outside the CBC

One overarching question in CBC field is whether Cbc1 and/or Cbc2 have separate roles from each other, outside the CBC. As mentioned in section 2.3.2, some earlier studies found Cbc1, and not Cbc2, to be interacting with eIF4E-bound transcripts or PARN (Balatsos *et al.*, 2006; Rufener and Mühlemann, 2013). The MIF4G-like domains of Cbc1 allow it to interact with other proteins containing the same motif; an interaction that might not necessarily require Cbc2. Conversely, Cbc2 also possesses an RRM domain, which can be used to bind RNA, and an isoform lacking the 7mG binding pocket was shown to localize to active transcription sites (Pabis *et al.*, 2010). In my pulldowns, Cbc1 appeared to have many more interacting partners, some of which were not shared with Cbc2. For instance, splicing factors were missing from the Cbc2 pulldown. Does this mean that only Cbc1 is important for CBC's role in splicing, or is this merely an artefact of the MS?

To answer this question in the context of RNA biogenesis and regulation, I depleted Cbc1 or Cbc2 from cells, observing changes in the transcriptome level. If both proteins are entirely dependent on each other, the expected outcome of such an experiment would be a high overlap of trends and changing genes, which did not turn out to be true. I then proceeded to dissect whether these changes were transcriptional or post-transcriptional, and which RNA surveillance pathways might be responsible for them. Interestingly, Cbc1, but not Cbc2, depletion was found to rescue the readthrough phenotype arising from *dhp1-154* temperature sensitive (ts) mutant. I hypothesize that this is due to more efficient cleavage in Cbc1 depletion, allowing for faster 5'-3' degradation; but due to RNA processing defects associated with Cbc1 depletion, the resulting RNA is retained in the nucleus. Cbc2 depletion, on the other hand, has a drastic effect on the stability of capped RNAs; a prominent phenotype that could not be compensated, ultimately resulting in cell death.

3.1 Depletion of Cbc1 or Cbc2 results in growth defects, without influencing the stability of the other protein

As both CBC genes are reported to be essential in fission yeast, we made use of a repressible promoter (*no message in thiamine 1*: NMT). When grown in thiamine-containing media, the genes are shut down, eventually resulting in protein depletion (Figure 3.1A). Endogenous Cbc1 or Cbc2 promoters were replaced with medium-strength NMT promoter. Expression of the proteins under NMT promoter was ~4-fold stronger than endogenous Cbc1, and ~2-fold than Cbc2 (Figure 3.3, lanes 1 and 3). A time course experiment revealed that lower than 20% of the initial protein amount was achieved after 16 h of depletion (Figure 3.1B); and hence, this time point was chosen for the later experiments.

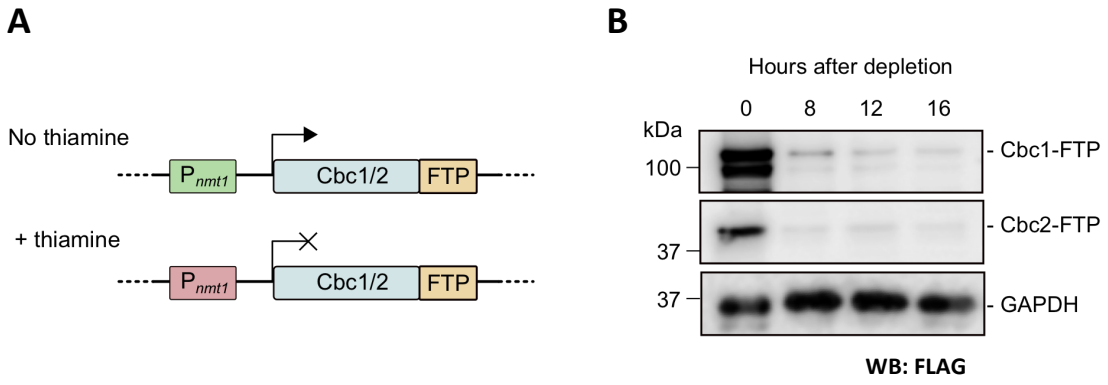


Figure 3.1: (A) Schematic for the depletion experiments. *nmt1* promoter shuts down in the presence of thiamine. (B) Representative Western blot for time course with depletion strains. Cells were grown in YES (+thiamine) for the indicated hours; proteins harvested using TCA precipitation; and detected with α-FLAG antibody. GAPDH was used as a loading control.

Next, the effect of Cbc1 or Cbc2 depletion on growth was tested. Both depletion strains had slower growth on YES media containing thiamine, with Cbc2 depletion having a more severe impairment (Figure 3.2).

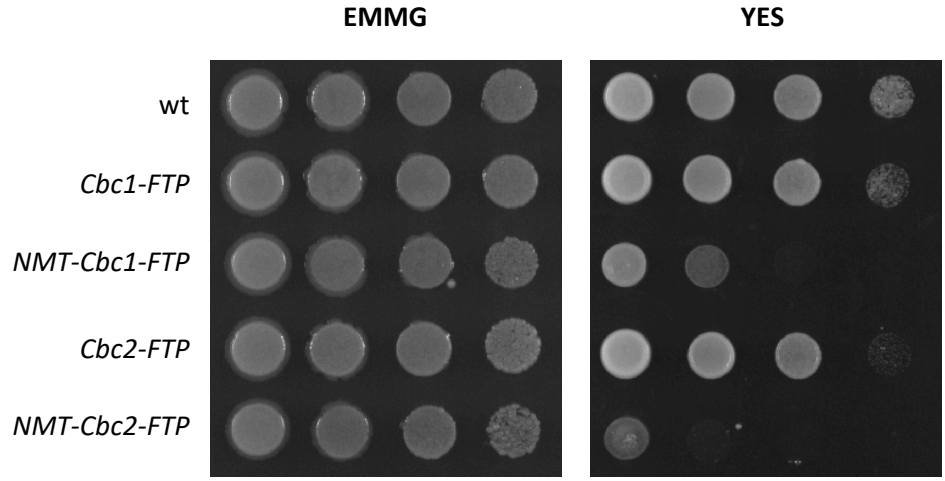


Figure 3.2: Spot test for CBC depletion strains. Indicated strains were spotted in a 10-fold dilution gradient, on either EMMG (no thiamine) or YES (+thiamine) plates and let grown for three days. Images are representative of three independent replicates.

Since it was previously reported that Cbc2 stability is affected by Cbc1 depletion (Fortes *et al.*, 1999), another set of strains where either Cbc1/2 was depleted while the other protein was tagged was generated. Surprisingly, after 16 h of depletion of one protein, levels of the other protein were unaffected (Figure 3.3). This finding suggests that *S. pombe* is a good model to study the effect of individual depletion of CBC proteins.

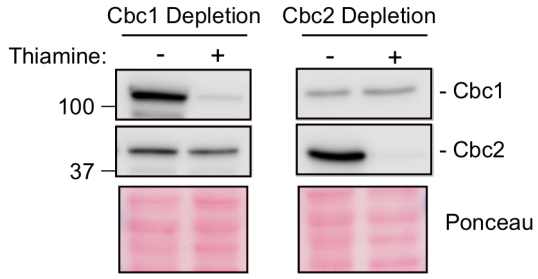


Figure 3.3: *Cbc1-HA-FLAG, NMT-Cbc2-FTP* and *Cbc2-HA-FLAG, NMT-Cbc1-FTP* strains were grown in either EMMG or YES for 16 h; proteins were precipitated using the TCA technique, and equal volumes were loaded per well. HA-FLAG- or FTP-tagged proteins were detected using α-HA and α-PAP antibody, respectively. Ponceau staining was used as a loading control. Images are representative of two independent replicates.

3.2 Depletion strains have diverse changes in transcriptome

To analyse the impact of Cbc1/2 depletion on fission yeast transcription, total RNA was obtained from the depletion strains after 16 h in YES, spiked in with *S. cerevisiae*, ribodepleted and made into cDNA libraries. Illumina next-generation sequencing was performed in duplicates, with wt RNA (tag-only strains grown in YES) as control. Differential expression analysis was carried out using the DESeq2 R package using the spike-in normalization.

Looking at the changes in pc genes led to the surprising observation that while Cbc2 depletion resulted in significant downregulation of most of the genes (n = 3519/4104), Cbc1 depletion showed upregulation of a number of genes (n = 977) alongside a subset of downregulated genes (n = 1197) (Figure 3.4A). Interestingly, the downregulated genes in both cases had high overlap (96%) with each other, whereas the genes upregulated in Cbc1 depletion had a meagre 14% overlap with the other depletion strain (Figure 3.4B). In fact, genes upregulated in Cbc1 depletion had more overlap with the genes downregulated in

Cbc2 depletion, indicating that the RNA regulation was affected differently in the two depletion strains.

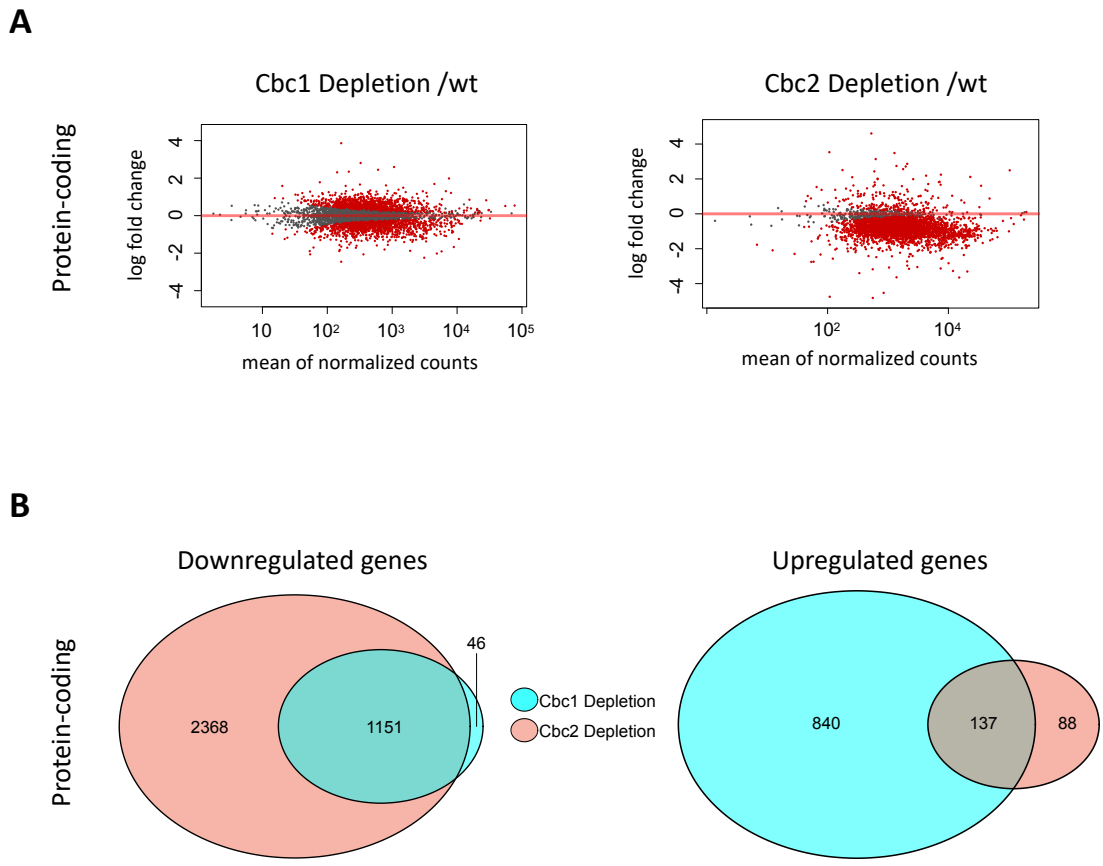


Figure 3.4: Depletion of Cbc1 or Cbc2 results in distinct changes on the transcriptome. (A) Scatterplots showing changes in pc genes transcripts in Cbc1 or Cbc2 depletion with respect to wt (their respective tag-only strains). Red dots indicate significantly changing genes (adjusted p-value < 0.005). (B) Venn diagrams illustrating the overlaps between significantly downregulated or upregulated pc genes in the two depletion strains. Cbc1 depletion is shown in cyan; Cbc2 depletion in light pink.

I then explored whether this trend holds true for other gene groups, such as nc genes, CUTs, or antisense transcripts. As seen in Figure 3.5A, downregulated genes tended to overlap much more as opposed to upregulated genes for all gene groups except CUTs,

where overlap for both is around 50%. Combined with the finding that most transcripts assessed were downregulated in Cbc2 depletion, this suggests that the main function of Cbc2 is to confer stability to the capped transcripts, while Cbc1 has a more intricate, subset-specific relationship with these transcripts.

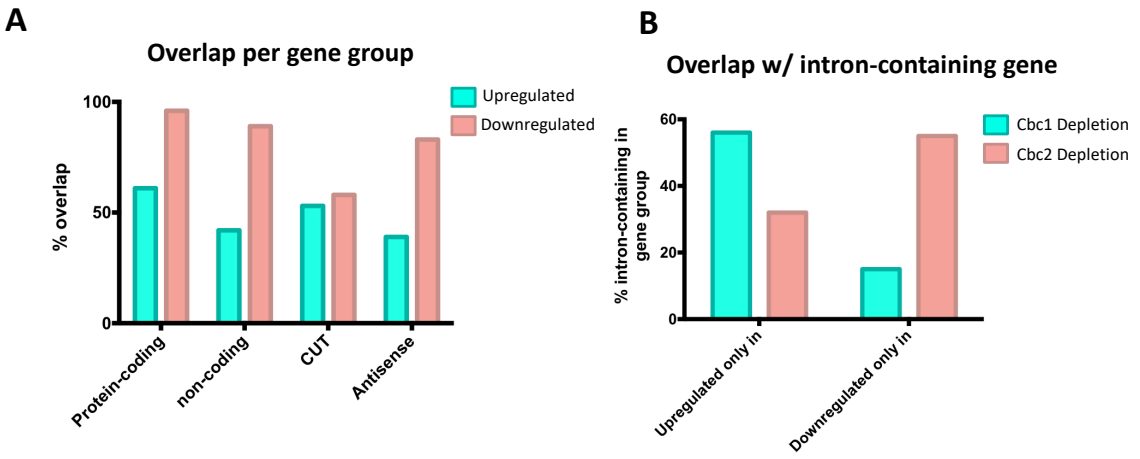


Figure 3.5: (A) Percent overlap of significantly changing transcripts (adjusted p-value < 0.005) in Cbc1 depletion transcripts with Cbc2 depletion for pc, nc genes, CUTs and antisense transcripts. Upregulated genes are coloured in cyan; downregulated in light pink. (B) Significantly upregulated or downregulated pc genes found in only one of the depletion strains were determined, and percentage of intron-containing genes among them was calculated. Cbc1 depletion was indicated in cyan; and Cbc2 depletion in light pink.

What could be a common property for genes upregulated upon Cbc1 depletion? To this end, I determined the subsets of pc genes that were up- or downregulated only in one depletion strain and asked how much of these gene groups were intron-containing genes; as opposed to intronless. Interestingly, intron-containing genes formed most of the upregulated genes in Cbc1 depletion and downregulated genes in Cbc2 depletion (Figure 3.5B). Furthermore, genes downregulated only in Cbc1 depletion were mostly intronless genes. However, it needs to be borne in mind that this group is composed of only 46 genes, and thus is more prone to selection bias. Nonetheless, that depletion strains have different

effects on intron-containing genes cements the possibility of Cbc1 and Cbc2 having separate roles affecting RNA biogenesis or regulation.

Is splicing affected only in Cbc1 depletion? Is the upregulation of the transcript due to more reads coming from the intron-retained transcripts? To answer these, I picked two representative genes; *phb2* and *pre8*, where Cbc1 depletion, but not Cbc2 depletion, results in a significant upregulation. Total RNA was obtained from wt or depletion strains, as well as splicing mutants *prp2-1* and *prp10-1* to detect the unspliced bands. For *phb2*, both unspliced and spliced products were increased upon Cbc1 depletion (Figure 3.6A, lane 2). Cbc2 depletion (lane 3) also showed a mild upregulation, with more unspliced transcript. In both depletion strains, the unspliced product contained not only the longer first intron, but all of them, corroborating the earlier finding that CBC is integral in splicing of all introns (Jiao *et al.*, 2013b).

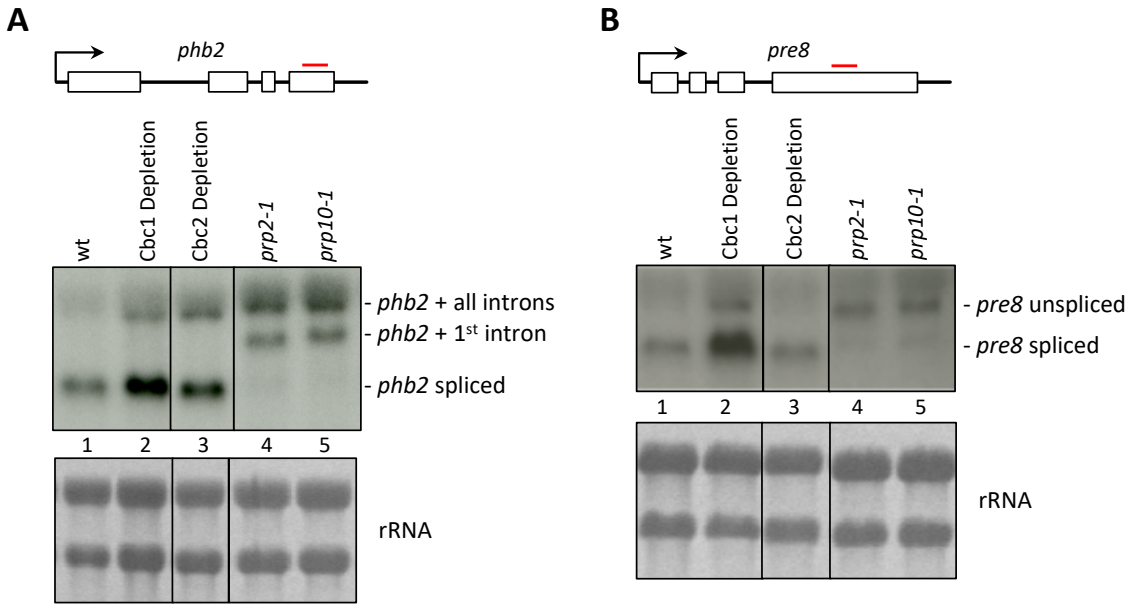


Figure 3.6: Representative northern blots for *phb2* (A) or *pre8* (B) with the depletion strains or splicing mutants *prp2-1* and *prp10-1*. Gene schematics above indicate the exons (rectangles) and

where the Northern probe (red line) is complementary to. Methylene blue staining was utilized to obtain rRNA bands as control. Images were cropped for viewing convenience.

On the other hand, only Cbc1 depletion resulted in intron retention in the *pre8* transcript (Figure 3.6B, lane 2). Once again, the spliced band was increased alongside the unspliced, implying that transcription might be induced, while RNA processing is affected. In contrast to Cbc1 depletion, *pre8* transcript levels remained unchanged in Cbc2 depletion, and no unspliced band was visible. Together, these results indicate that Cbc1 and Cbc2 have different effects on a number of genes and RNA processing.

Are the changes observed in genes upon depletion of Cbc1 or Cbc2 transcriptional or post-transcriptional? To answer this, ChIP of total Pol II was carried out using the monoclonal MABI0601 antibody, in either the tag-only strains (Cbc1/2-FTP) serving as wt controls, or depletion strains. Metagene for pc genes revealed that the Pol II profile exhibited hardly any change upon depletion of Cbc1 or Cbc2, except for a more pronounced bump at the TSS (Figure 3.7). It then follows suit that the transcriptome changes are likely to be at post-transcriptional level.

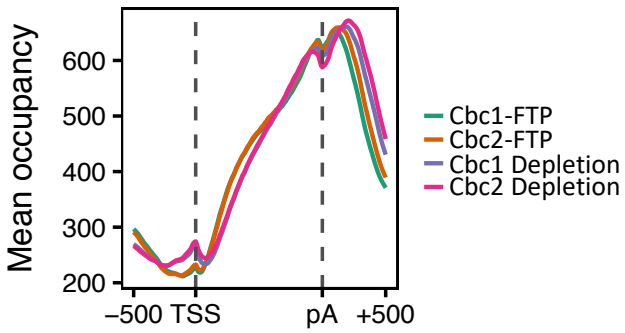


Figure 3.7: Metagene profile for Pol II occupancy across pc genes. Pol II-bound chromatin was immunoprecipitated from Cbc1-FTP (green), Cbc2-FTP (red), Cbc1 depletion (purple) and Cbc2 depletion (pink) strains and subjected to Illumina sequencing. ChIP-seq alignments were binned

to 0.5 kb before TSS, 1 kb for the gene body, and 0.5 kb after pA site. Data are represented as the average of two replicates.

3.3 Crosstalk of CBC with different members of RNA surveillance pathways

To investigate the possibility that Cbc1 and Cbc2 might be interacting with different RNA surveillance pathways, I then proceeded to make double mutants of a number of genes known to have a role in surveillance with the depletion strains. The aim here was to identify protein(s) that interacted with the two depletion strains differently, which might imply that the said protein might be involved in the differential changes observed in Cbc1/2 depletion.

Table 3.1 summarizes the genetic interactions observed with a number of RNA surveillance factors. The images for the individual spot tests can be found in Appendix A.

Table 3.1: Screening of genetic interactions between RNA surveillance factors and depletion strains. Interactions were tested in three temperatures (25°C, 30°C, 37°C), on EMMG (non-depletion) or YES (depletion) plates. A negative genetic interaction is indicated with "-", a positive with "+" and no apparent interaction with "=".

Factor	Cbc1 depletion			Cbc2 depletion		
	25°C	30°C	37°C	25°C	30°C	37°C
rrp6	-	-	-	-	=	+
dis3	-	-	-	-	-	--
mtl1	-	-	--	-	-	-
red1	-	--	--	-	-	+
pab2	-	--	-	+	+	++
hrb1	-	-	+	+	+	+
dhp1	=	+	+	=	-	-

Mutations in the 3' exonuclease at the bottom of the exosome, *dis3*, as well as some of its adapters *mtl1* and *red1*, showed deterioration of the growth defect when combined with either depletion strain. Curiously, *rrp6Δ, NMT-Cbc2-FTP* grew better as the temperature rises, as compared to the single mutant. On the other hand, a distinct difference between the two depletion strains could be observed when combined with *pab2Δ* and *hrb1Δ*. Cbc1 double mutants showed synthetic genetic interaction, while Cbc2 mutants showed suppression of the growth defect. In contrast, the ts mutant of the 5' exonuclease Dhp1, *dhp1-154*, showed faster growth when combined with Cbc1; and slower growth combined with Cbc2 depletion, at the restrictive temperature of 37°C. This screening alluded to the possibility that Cbc1 and Cbc2 crosstalk with different RNA surveillance pathways. Therefore, the effect of losing one member of CBC on the RNA levels might not be as straightforward to assess as previously thought.

I then focused on dissecting the interaction between *dhp1-154* and Cbc1 due to several reasons. Firstly, Dhp1 phenotype is easy to assess, as the loss of it results in global readthrough from Pol II genes (Kim, Krogan, *et al.*, 2004). This would allow me to check if the suppression of the growth defect in the double mutant is directly due to a rescue of the Dhp1 phenotype. Secondly, a role for CBC in termination was reported for cryptic pA sites that usually reside within the first intron (Hallais *et al.*, 2013; Kainov and Makeyev, 2020), but not at the 3' end, and thus was intriguing to look into. Finally, such a role for Cbc1, and not Cbc2, would explain why there was a group of upregulated genes upon depletion of the former.

3.3.1 Readthrough in *dhp1-154* is suppressed upon Cbc1 depletion

As explained above, the *dhp1-154* phenotype is defined by readthrough past the pA site in Pol II genes. To understand if the readthrough is rescued in a Cbc1 double mutant, RNA-seq was performed on single and double mutants. As expected, strong readthrough was detected for pc genes in the *dhp1-154* mutant. Strikingly, this phenotype appeared to have been reversed in the Cbc1 double mutant for many genes. In contrast, Cbc2 double mutant

still exhibited readthrough, indicating that this effect is Cbc1 depletion-specific. Two genes exemplifying the phenotype were chosen: *tom22* and *wis2* (Figure 3.8). *tom22* is an intron-containing gene showing readthrough, while *wis2* is chosen as an intronless gene with drastic readthrough spilling over to the neighbouring gene.

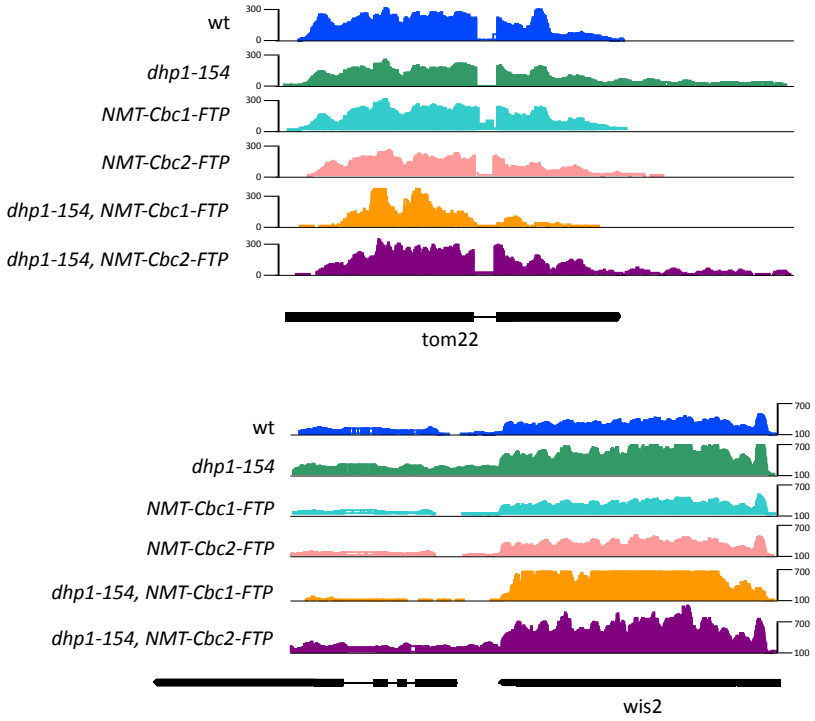


Figure 3.8: Representative genome browser snapshots of RNA-seq reads from wt (dark blue), *dhp1-154* (green), Cbc1 depletion (light blue), Cbc2 depletion (pink), Cbc1 double mutant (orange) and Cbc2 double mutant (purple) at *tom22* (upper) and *wis2* (lower) loci. Data were normalized with respect to *S. cerevisiae* spike-in. Readthrough can be observed in *dhp1-154* and Cbc2 double mutant and is rescued in the Cbc1 double mutant.

Next, these two genes were tested individually by Northern blots. The longer transcript including readthrough could be seen in *dhp1-154* mutant (Figure 3.9, lanes 2 and 7); and disappeared in the double mutant (lane 4). The same double mutant showed readthrough

when grown in non-depletion conditions (EMMG, lane 9), attesting that the rescue is linked to depletion of Cbc1.

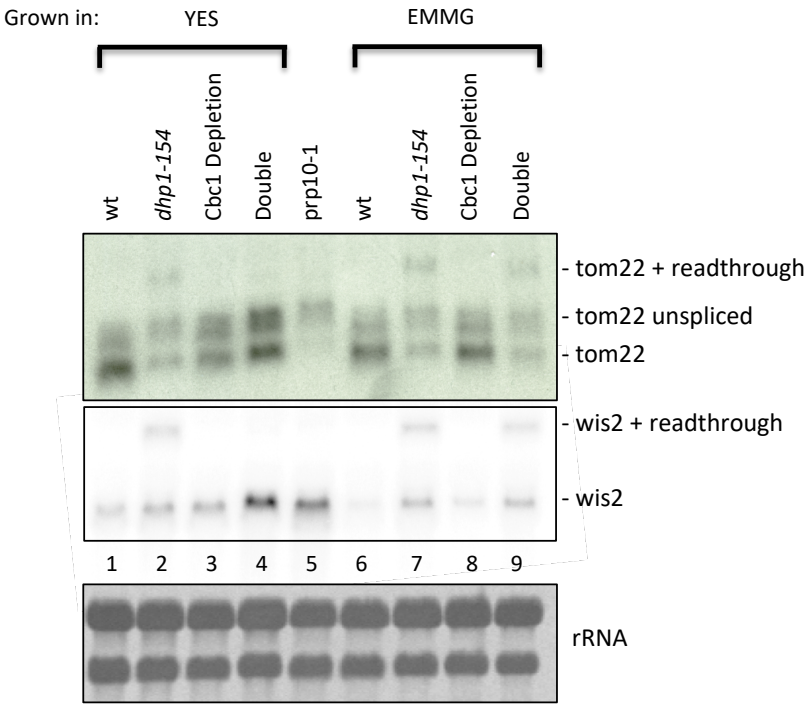


Figure 3.9: Northern blots for *tom22* and *wis2* transcripts to confirm RNA-seq results. Strains used in RNA-seq were grown either in YES (depletion) or EMMG (no depletion) and harvested. 15 µg of total RNA was loaded per well and probes complementary to the gene body of either *tom22* or *wis2* were utilized to detect different transcripts. Methylene blue staining was performed to detect rRNA as loading control.

3.3.2 How does the depletion of Cbc1 rescue the *dhp1-154* readthrough?

Rescue of the readthrough upon Cbc1 depletion could be explained in two ways: 1. Cbc1 might have an anti-cleavage role, and depletion causes more efficient cleavage and termination; or 2. Cbc1 stabilizes the 3' extended transcripts in the *dhp1-154* mutant. If the former hypothesis holds, Pol II profiles should show a reduction in Pol II signal around

the pA site. Therefore, Pol II occupancy was assessed in the single and double mutants by ChIP-seq. Indeed, Pol II signal in the double mutant had a more pronounced dip close to the pA site, as compared to *dhp1-154* (Figure 3.10, two-sided Wilcoxon rank sum test p-value < 0.001). This effect could also be seen in Cbc1 depletion single mutant as compared to wt, implying that depletion alone causes this. On the other hand, the Pol II signal beyond pA site persisted in the double mutant. Therefore, it is possible that Pol II termination is still defective due to loss of Dhp1 activity, but since cleavage is more effective, 5'-3' degradation of the readthrough is also encouraged in the double mutant.

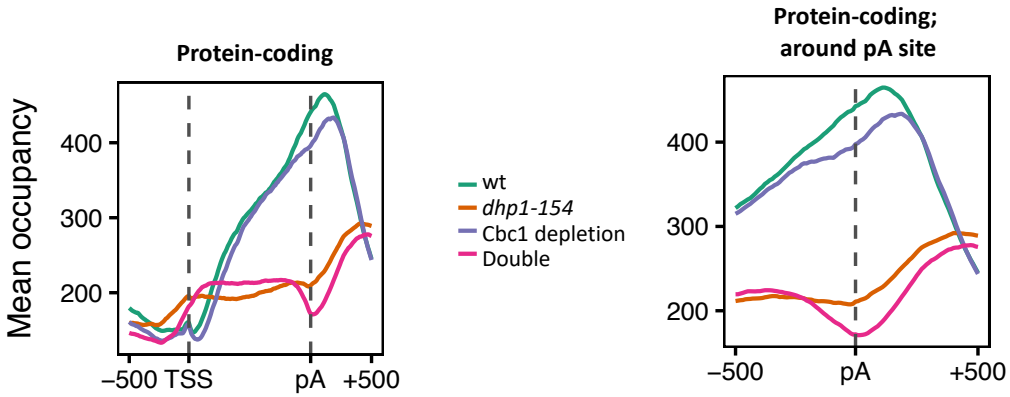


Figure 3.10: Pol II metagene profiles across pc genes in wt (green), *dhp1-154* (orange), Cbc1 depletion (purple) or double mutant (pink). Chromatin-bound Pol II was pulled down using MABI0601 antibody and DNA fragments were subjected to Illumina sequencing. On the left, alignments were binned to 0.5 kb before TSS, 1 kb for the gene body, and 0.5 kb after pA site. On the right, Pol II signal -0.5 kb to 0.5 kb around the pA site of pc genes were plotted. Data are represented as the average of two replicates.

What happens to the cleaved transcript in the double mutant? One possibility is that the pre-mRNA is not exported, staying in the nucleus. If Cbc1 is also responsible for the recruitment of RNA surveillance machinery to the nuclear-retained RNA, the double mutant should have an accumulation of RNA in the nucleus. To assess this, fluorescence *in situ* hybridization (FISH) was performed using an oligo-d$(T)_{20}$ probe conjugated to

Alexa488 fluorophore, which would detect polyadenylated transcripts. DAPI staining was used to visualize nuclei and the relative localization of poly(A)$^+$ transcripts was determined.

The double mutant of *dhp1-154* and Cbc1 depletion showed a mild nuclear retention phenotype as compared to the single mutants (Figure 3.11). Retention was not observed when the cells were grown in EMMG (data not shown). In contrast, Cbc2 double mutant resulted in a grossly deformed cell shape, making RNA localization hard to assess. Another point to emphasize was the way cell morphology changed in the two double mutants. Cells of Cbc1 double mutant resemble the Cbc1 depletion cells more than the smaller, shrunk cells of *dhp1-154*. As expected, when grown in EMMG, Cbc1 double mutant cells closely resembled that of *dhp1-154* single mutant. On the other hand, Cbc2 double mutant was even smaller than *dhp1-154*. Most likely, these cells are on the brink of cell death, and therefore, RNA localization holds less of an importance. All in all, depletion of Cbc1 rescued the *dhp1-154* cell morphology but resulted in a mild RNA retention in the nucleus.

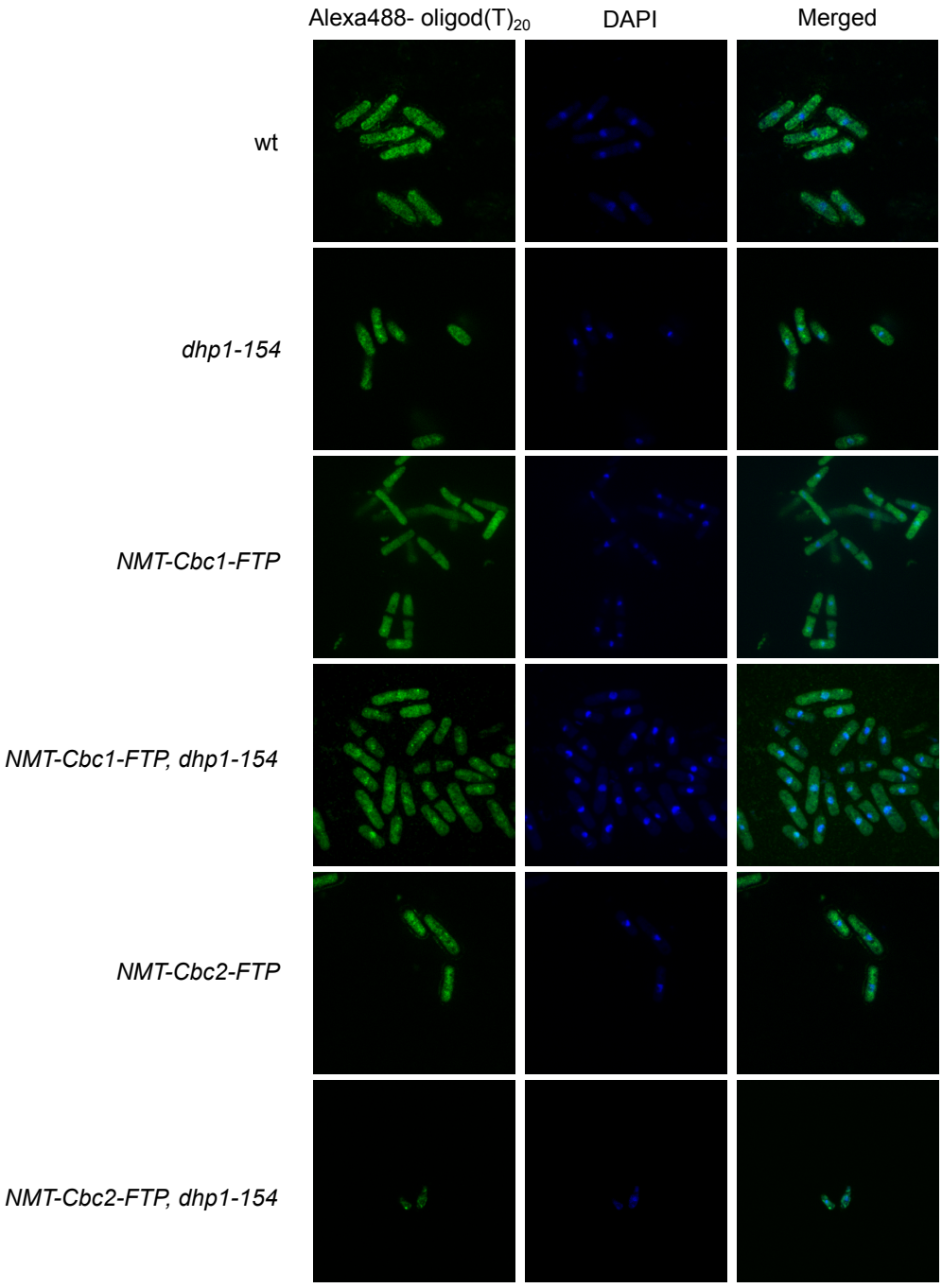


Figure 3.11: poly(A) FISH with double mutants. All strains were shifted to the *dhp1-154* restrictive temperature (37°C) 2 h before the experiment. Alexa488-conjugated oligo-d(T)$_{20}$ probe and DAPI staining was used to detect the relative localization of polyadenylated RNA in *S. pombe*. Images are representative of two independent replicates.

3.4 Discussion

Findings presented in this chapter underline differences between the two members of the CBC and allude to a role for Cbc1 that is separate to Cbc2. The traditional view of CBC postulates that both proteins are required for each other's function. This is hardly surprising, as by far the most prominent interactor for Cbc1 is Cbc2, and vice versa. This; however, does not exclude the possibility that these proteins also have roles mutually exclusive from each other. This chapter shows that Cbc1 depletion could revert the readthrough observed in Dhp1 ts mutant, alluding to an anti-cleavage role for Cbc1 in Pol II genes. Cbc2 depletion, on the other hand, results in a drastic downregulation of almost all Pol II genes, thereby causing a strong growth defect.

Expectedly, depletion of both CBC genes resulted in slower growth (Figure 3.2); but the cells were still viable unlike what was reported in fission yeast (Hayles *et al.*, 2013). This could be due to the fact that ~20% of the initial protein levels still remained after depletion (Figure 3.1B), and this might be enough to sustain life. It was also important that the depletion of one CBC protein did not affect the levels of the other protein (Figure 3.3). This finding was against what was reported in higher eukaryotes and budding yeast, where the stability of Cbc2 protein was dependent on Cbc1 (Fortes *et al.*, 1999; Pabis *et al.*, 2013). In fission yeast, CBC proteins might have a unique independence of stability, which may imply that the proteins have additional roles aside from CBC. This independence makes *S. pombe* a good model to study the effect of individual depletions of CBC proteins.

One interesting finding of this study came from the transcriptomics changes upon depletion of Cbc1 or Cbc2 (Figure 3.4). Cbc2 depletion resulted in a downregulation of many pc genes, which can be expected from a protein that binds the 7mG cap, stabilizing the Pol II transcript (Izaurralde *et al.*, 1994; Flaherty *et al.*, 1997). In contrast, Cbc1 depletion caused some genes to be downregulated, while others were upregulated. This may suggest that the major effect of losing Cbc2 is on the stability of the mRNA, while Cbc1 depletion results in problems in downstream functions, causing mis-regulation. This idea is supported by the inclusion of intron-containing genes in the upregulated group (Figure 3.5B). 56% of the genes upregulated only in Cbc1 depletion contained introns, as

opposed to the 32% observed in genes upregulated only in Cbc2 depletion. The opposite trend could be seen in genes downregulated only in one strain; as downregulated genes in Cbc1 depletion only had 15% intron-containing genes, while the overlap with Cbc2 downregulated genes was 55%. Briefly, intron-containing genes tended to be upregulated upon Cbc1 depletion.

This trend could be due to; a) an overall increase in transcription of the gene, b) an increase in intronic reads, or c) escape from RNA surveillance. Northerns presented in Figure 3.6 attest to first two possibilities, as for both genes tested there was an increase in transcriptional output and an increase in the unspliced product. Interestingly, while the unspliced band for *phb2* could be seen in Cbc2 depletion, *pre8* unspliced was not visible, affirming once more that depletion of either CBC protein has different downstream effects. On the other hand, Pol II ChIP-seq did not show much of a change in the Pol II occupancy in the depletion strains (Figure 3.7). This was also the case for *pre8* and *phb2* gene loci. Therefore, it can be speculated that upregulation of pc genes upon Cbc1 loss is not via a global increase of transcription, but rather a mis-regulation of RNA surveillance.

As mentioned, Pol II profiles did not change drastically in the two depletion strains, except for two aspects. First was a more pronounced bump at the TSS. This resembles promoter-proximal pausing of higher eukaryotes (Core and Lis, 2008); and while it can be readily detected in these organisms by ChIP-seq, in fission yeast the paused genes could only be detected recently by PRO-seq experiments (Booth *et al.*, 2016). Depleting CBC proteins might make pausing more pronounced due to defective recruitment of P-TEFb, stalling Pol II before the elongation, and thus making it easier to detect on a ChIP-seq experiment. Nevertheless, it might be worthwhile to assess the Pol II profiles with a GRO/PRO-seq after Cbc1/2 depletion to confirm this increase in “pausing”. Second difference was a shift of the peak post-pA site. This usually points to a termination defect and is illustrated by the readthrough in Pol II genes. However, neither RNA-seq profiles nor Northerns for individual transcripts showed any readthrough in depletion strains. This might be explained by the presence of a pro-active RNA surveillance that rapidly degrades the nascent RNA after pA cleavage.

Next, I set out to identify the underlying RNA surveillance pathway causing the differences in depletion strains. To this end, deletion or ts mutants of the nuclear exosome (Rrp6, Dis3) (Liu, Greimann and Lima, 2006), factors of the exosome (Mtl1, Red1, Pab2) (Lemay *et al.*, 2010; Marayati *et al.*, 2016), quality control SR protein Hrb1 (Krebber *et al.*, 2014), or the 3' exonuclease Dhp1 (Chalamcharla *et al.*, 2015) were chosen and combined with the depletion strains. The idea was to identify double mutants that had opposite effects on cell viability, which might in turn be due to differential regulation upon depletion of either CBC protein. Several aspects became apparent from this genetic screening. Firstly, even when the cell viability was affected the same way in two depletion strains (i.e. negative interaction in both cases), the potency of genetic interaction was different. For instance, the *dis3-54* mutation showed a stronger negative interaction with Cbc2 depletion as compared to Cbc1 depletion at 37°C. Conversely, *mtl1-1* yielded a stronger negative interaction with Cbc1 depletion at 37°C. Secondly, temperature played a key role in determining the genetic interaction. To illustrate, while *red1Δ* resulted in negative interaction with both depletion strains at most temperatures, a positive interaction with Cbc2 depletion could be observed at 37°C. This might be explained by the diverging metabolism and expression profiles of cells grown at heat-shock conditions. Thirdly, many double mutants (*pab2Δ*, *hrb1Δ*, *rrp6Δ* at 37°C) appeared to suppress the Cbc2 depletion growth defect. This can be explained by the increase in RNA stability conferred by the loss of RNA surveillance factors, attesting to the previous hypothesis that Cbc2 is mostly important for the stability of capped RNAs. Finally, there were indeed some double mutants (*pab2Δ*, *hrb1Δ* and *dhp1-154*) that showed opposite genetic interactions with the two depletion strains. Taken together, these findings imply that the effects of depleting Cbc1 or Cbc2 are not identical and CBC proteins could crosstalk with different surveillance factors to a differing extent.

The endo- and 3' exonuclease Dhp1 is known mostly for its role in transcription termination. Not surprisingly, the ts mutant *dhp1-154* causes readthrough from Pol II genes at restrictive temperature (Kim, Krogan, *et al.*, 2004). Interestingly, viability of *dhp1-154* deteriorated upon depletion of Cbc2 and ameliorated with Cbc1 depletion. The *dhp1-154* phenotype was easier to assess than the other exosome factors, which instead have many

indirect effects, making this double mutant a good candidate to investigate further. RNA-seq revealed that the readthrough phenotype of *dhp1-154* was reverted in a Cbc1 double mutant. Notably, Northern blot for two genes, *wis2* and *tom22*, showed that the transcripts were 3' extended in the *dhp1-154* single mutant, implying that not only the termination but also cleavage from the pA site was affected in this strain. The disappearance of the 3' extended transcript in Cbc1 double mutant was directly linked to depletion, as the extension was still present when the cells were grown in EMMG.

As explained in section 3.3.2, this rescue could be either because Cbc1 has an anti-cleavage role and depletion causes more efficient cleavage and termination, or because Cbc1 stabilizes the 3' extended transcripts in the *dhp1-154* mutant. ChIP-seq demonstrated a more pronounced dip at the pA site for Cbc1 depletion single and double mutants, implying that loss of Cbc1 might result in faster cleavage. The second possible explanation for the disappearance of 3' extended transcripts could not be assessed as part of this thesis due to time constraints. Here, a follow-up experiment would be to disrupt the exosome, as it is the main source of RNA surveillance in the nucleus, and check for the extended transcript in a *dhp1-154, rrp6Δ, NMT-Cbc1-FTP* triple mutant.

There are a number of pitfalls to the idea that Cbc1 might have an anti-cleavage role. Hallais et al. described a pro-cleavage role for CBC-Ars2 in HeLa cells, and noted that while the cleavage-stimulatory role of Cbc2 was dependent on the proximity of pA site to the cap, that of Cbc1 was not (Hallais *et al.*, 2013). They also underlined that CBC-Ars2 appeared to specifically stimulate premature pA signals, like the ones close to the TSS, within an intron, possibly to differentiate and degrade cryptic transcripts. It is therefore possible that the proximity of the cap to the pA site plays a key role in determining whether Cbc1 plays an anti- or pro-cleavage role. The changing protein composition of the transcription complex might also dictate what role Cbc1 will play in 3' end processing.

Another pitfall stems from the disappearance of the readthrough transcript in Cbc1 double mutant. Dhp1 is the only known nuclear 5' exonuclease that is capable of degrading nascent RNA during termination. If Cbc1 depletion indeed results in better cleavage of the 3' end, the double mutant should show an accumulation of nascent RNA, which does not

appear to be the case by RNA-seq. One possibility is that the reduced transcription turnover in the double mutant (Figure 3.10) allows for more efficient degradation of the readthrough RNA via other surveillance mechanisms. Probing for only the readthrough part of the transcript would shed light on this mystery.

The microscopy experiments revealed that not only the readthrough phenotype but also the cell morphology was rescued to an extent in the double mutant. On the other hand, a mild nuclear retention of poly(A)$^+$ RNA was visible. Since CBC-Ars2 can interact with TREX to promote export (Merz *et al.*, 2007), the lack of Cbc1 might affect nuclear export and even though the cleavage is rescued, the resulting mRNA may be retained in the nucleus. It is peculiar to note that such an effect was not observed in the Cbc2 depletion single mutant and was only weakly present in Cbc1 depletion. When combined with *dhp1-154*, this phenotype was elevated. This could be due to better cleavage and polyadenylation upon loss of Cbc1. Nascent RNA is cleaved more readily at the pA site and subsequently polyadenylated, allowing for detection with the FISH probe. Another possibility is that since Cbc1 depletion causes RNA processing defects such as intron retention (lane 2 in Figure 3.6; lane 4 in Figure 3.9), the mRNP composition of the resulting transcript does not pass as "export-ready" and is kept in the nucleus. However, due to a lack of CBC-mediated activation of RNA surveillance (Das, Butler and Sherman, 2003; Zhou *et al.*, 2017), they could not be degraded, and thus the accumulation observed by FISH.

In summary, I propose the following working model, illustrating the different roles of the two CBC proteins. Upon depletion of Cbc2, stability of almost all Pol II RNA is affected drastically, and this leads to a strong downregulation. On the other hand, Cbc1 depletion causes an upregulation of a subset of pc genes. Therefore, the main function of Cbc2 is to confer stability to the capped transcript, while Cbc1 might have additional roles in RNA processing/regulation (Figures 3.4, 3.5). These varying roles are evidenced by the following findings:

- Cbc1 depletion results in intron retention, indicating that Cbc2 alone is not sufficient to ensure proper RNA splicing. On the other hand, Cbc1 alone can ensure proper splicing for some genes, but not all (Figure 3.6).

- Differential regulation of genes upon Cbc1 vs Cbc2 depletion cannot be explained by transcriptional changes (Figure 3.7), and therefore should be mainly post-transcriptional. Indeed, two depletion strains have diverging genetic interactions with some of the RNA surveillance factors (Table 3.1). Deleting some of these factors suppressed Cbc2 depletion phenotype, attesting to the RNA-stabilizing role of Cbc2. In contrast, the same deletions deteriorated Cbc1 depletion phenotype.
- Cbc1 depletion, and not Cbc2 depletion, rescued the readthrough phenotype associated with the loss of Dhp1 function (Figures 3.8, 3.9). Since there is still Pol II transcribing beyond the pA site in the double mutant (Figure 3.10), I hypothesize that readthrough is rescued by a more active cleavage machinery. This is illustrated with a more pronounced drop of Pol II signal around pA site. RNA is cleaved and polyadenylated more efficiently in the absence of Cbc1; however, since the resulting transcript is still defective (i.e. not spliced properly), it is retained in the nucleus (Figure 3.11).

Chapter 4 Loss of Cbc3 affects RNA biogenesis

After checking the recruitment pattern and RNA dependence of Cbc3, it has become apparent that Cbc3 takes on a different role than CBC proper in *S. pombe*. Co-transcriptional recruitment profile for Cbc3 resembled those of some THO complex members; increasing over the gene body and peaking around the pA (Tuck and Tollervey, 2013; Fischl *et al.*, 2017). This raises the possibility that, similar to its role in higher eukaryotes (Gebhardt *et al.*, 2015), *S. pombe* Cbc3 might have a role in nuclear export.

Another intriguing finding was that Cbc3 was not recruited to all pc genes. What separated these genes from the others? What would happen to transcripts from these genes upon Cbc3 loss? Does Cbc3 bind these transcripts down the gene body or via the 7mG cap? Therefore, further investigation is needed to understand the function of *S. pombe* Cbc3.

This chapter begins with understanding the transcriptome-wide changes upon loss of Cbc3. Many pc genes are mildly downregulated, and interestingly, shorter genes are less affected by Cbc3 deletion. Pol II profiles reveal a drop of signal over the gene body, suggesting for Cbc3 a role in transcription elongation. Recruitment of elongation factor Paf1 to longer genes was affected upon Cbc3 loss, providing a possible explanation to this phenomenon. Moreover, like the human counterpart, deletion of *S. pombe* Cbc3 hindered export of poly(A)$^+$ RNA. Consequently, I propose a role for *S. pombe* Cbc3 in transcription elongation and nuclear export.

4.1 Pc genes are generally downregulated upon Cbc3 deletion

To check whether Cbc3 loss affects levels of Pol II transcripts, RNA-seq was carried out in a strain lacking Cbc3 gene *SPCC16C4.16c* versus wt strain in duplicates. *S. cerevisiae* spike-in was utilized for normalization. DESeq2 analysis revealed a general but mild

downregulation of pc genes. Only 6 pc genes out of 4104 appeared to be upregulated significantly, as opposed to 1918 downregulated (Figure 4.1A). In contrast, very few nc genes or CUTs were changed significantly (Figure 4.1B). This is in line with the Cbc3 recruitment profile presented earlier (section 2.1), showing Cbc3 signal mostly in pc genes. These imply that Cbc3 plays a role in regulating pc genes.

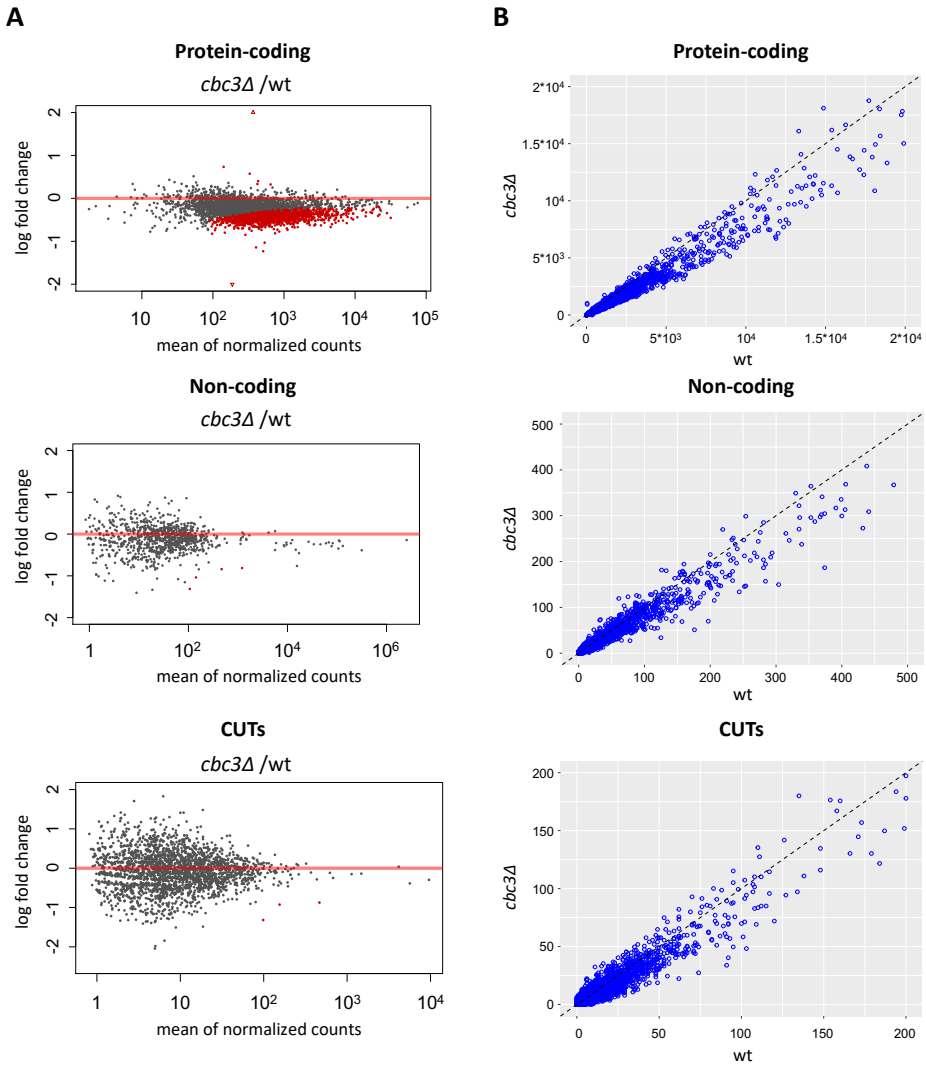


Figure 4.1: (A) Scatterplots showing changes in pc or nc gene transcripts, or CUTs after Cbc3 deletion with respect to wt. Red dots indicate significantly changing genes (adjusted p-value <

0.005). (B) X-Y scatter of read counts of individual genes in wt (x axis) and *cbc3*Δ (y axis). y = x is denoted by a dashed line.

When the overlap between Cbc3-recruiting genes by ChIP-seq and Cbc3-responsive genes by RNA-seq was assessed, it was revealed that most Cbc3-recruiting genes were indeed affected by Cbc3 loss (819 out of 1078; 76%); however, there were also additional genes downregulated in RNA-seq (1099 more). Thus, I started searching for a common property among genes affected by Cbc3 deletion in RNA-seq. Genes were grouped with respect to their GC-content, intron presence, number of introns; but no significant overlap could be found (data not shown). On the other hand, when the genes were grouped by their length, Cbc3 loss was shown to be affecting longer genes more. Figure 4.2 shows the percentage of genes in each gene length group overlapping with significantly changing genes in Cbc3 deletion. Comparing to the same overlap analysis for genes changing upon Cbc1 depletion, it can be observed that shorter genes were much less affected by Cbc3 deletion. As the gene length increased, overlap percentage went up to ~50%, indicating that genes longer than ~1000 bp tend to be affected more by the loss of Cbc3, as compared to shorter genes.

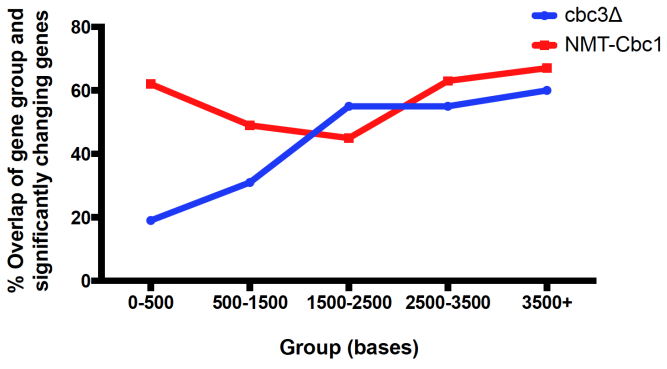


Figure 4.2: Shorter genes are less affected by Cbc3 loss. Pc genes are grouped with respect to their length and overlap of these gene groups with significantly changing genes upon Cbc3 deletion (blue line) or Cbc1 depletion (red line) was assessed.

4.2 Cbc3 promotes transcription elongation in longer genes by facilitating Paf1 recruitment

The gene length-dependent effect of Cbc3 brings about two possible means of regulation. First, transcription elongation might be affected, and while shorter genes could compensate for the elongation defect, longer genes might not. Second, Cbc3 might be playing a role in mRNA export similar to that hnRNP C plays in higher eukaryotes (McCloskey *et al.*, 2012), measuring the mRNP by selectively binding unstructured RNA regions longer than 200-300 bases and facilitating export of mRNAs longer than a certain threshold.

The Pol II occupancy profile in Cbc3 deletion was assessed by ChIP-seq. Metagene profiles exhibit a drop of Pol II signal over the gene body (Figure 4.3A). When the pc genes were grouped into those that were significantly downregulated in RNA-seq versus those that were not, Pol II reduction could only be seen in the former group. Next, processivity of the enzyme was assessed with a custom-made script (more information in section 6.5.4). Briefly, it calculates the mean of ratio of Pol II signal from two "windows" of 10 bins with a 50 bin-offset between them. In other words, it is the Pol II signal proportion of two 100-nt windows with 500 nt in between. The calculation is repeated until the end of the gene is reached, and windows are plotted as an x-y scatter plot. This measurement showed that the processivity of Pol II was reduced in *cbc3Δ* (Figure 4.3B), pointing to an elongation defect.

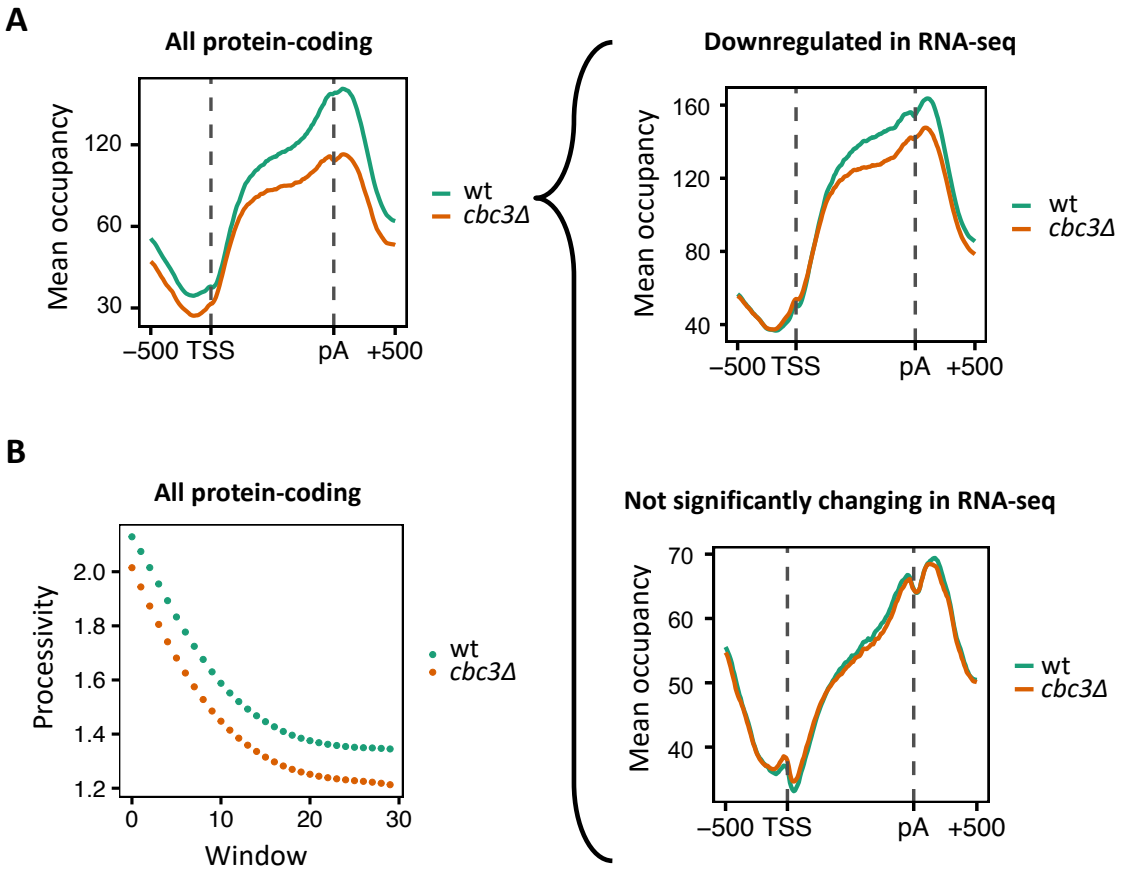


Figure 4.3: (A) Pol II occupancy metagene profiles in wt (green) or *cbc3*Δ (orange) for all pc genes (upper left), genes downregulated (upper right); or not significantly changing genes upon Cbc3 deletion (lower right). ChIP-seq alignments were binned to 0.5 kb before TSS, 1 kb for the gene body, and 0.5 kb after pA site. (B) Pol II processivity measurements in the two strains. Processivity is defined by the mean of ratio of Pol II signal from two windows of 10 bins with a 50 bin-offset between them. Data are represented as the average of two independent replicates.

To affirm that the drop of Pol II occupancy is due to defective elongation and not a consequence of problematic initiation, Ser5 and Ser2 phosphorylation levels were assessed. Whole cell extracts showed equal levels of Pol II (total CTD) and Ser5-phosphorylated CTD, and a mild increase in Ser2-phosphorylated CTD in *cbc3Δ* mutant (Figure 4.4). By ChIP-seq, it was confirmed that S5P remained unchanged while S2P had

increased over the gene body, pointing to an elongation problem rather than an initiation defect.

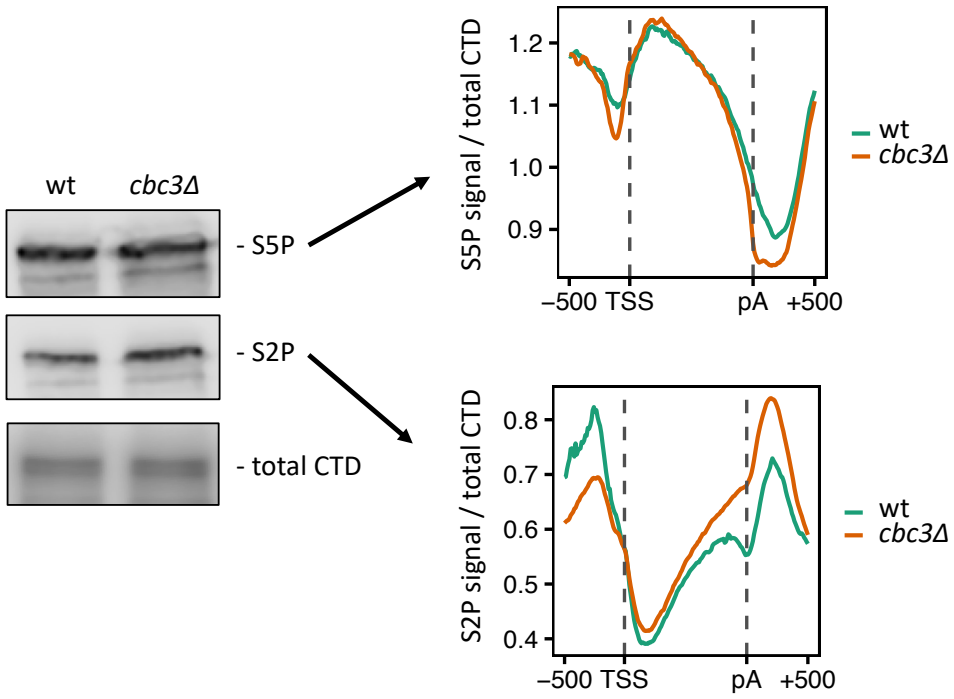


Figure 4.4: (Left) Representative Western blot images of Ser5-, Ser2-phosphorylated, or total CTD from wt or *cbc3*Δ whole cell extracts. Images are representative of three independent replicates. (Right) ChIP-seq metagene profiles for Ser5-phosphorylated (upper) or Ser2-phosphorylated (lower) Pol II in wt (green) or *cbc3*Δ (orange), normalized to the respective total CTD signal from each strain.

What connects Cbc3 to transcription elongation? It can be hypothesized that Cbc3 might facilitate the recruitment of an elongation factor. Indeed, Cbc3 deletion started to show growth defects when combined with the deletion of Paf1 (Figure 4.5A-B). To check whether Cbc3 deletion led to a defect in Paf1 recruitment, Paf1 ChIP was performed in either wt or *cbc3Δ* background. Loss of Cbc3 affected Paf1 recruitment to pc genes, especially further down the gene body (Figure 4.5C). Since differing Pol II levels might

affect Paf1 recruitment to chromatin (Kim, Ahn, *et al.*, 2004), I normalized Paf1 ChIP signal to the respective Pol II signal observed in both strains. Even then, Paf1 signal was reduced over the gene body for *cbc3Δ* (Figure 4.5C, right panel).

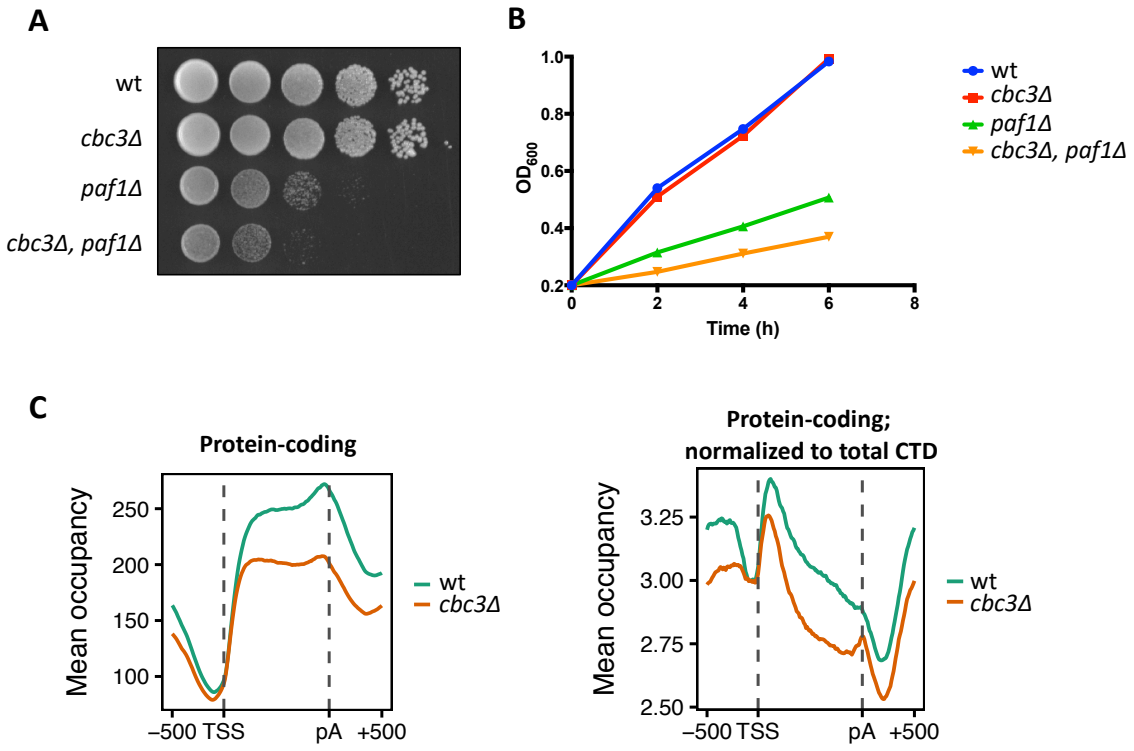


Figure 4.5: (A) Spot test with wt, *cbc3Δ*, *paf1Δ* and *paf1, cbc3Δ* strains. Serial dilutions were spotted on a YES plate and grown at 30°C for 5 days. Image is a representative of three independent replicates. (B) Growth curves of the strains in (A) in YES. Inoculations were grown overnight and diluted to OD_{600} = 0.2; and OD_{600} was measured for the consecutive 6 h. (C) Left, ChIP-seq metagene profile for Paf1 recruitment to pc genes in wt (green) or *cbc3Δ* (orange). Alignments were binned to 0.5 kb before TSS, 1 kb for the gene body, and 0.5 kb after pA site. Right-hand side profile is the same as left, but signal is normalized to total Pol II levels in the respective strains.

If the reduction in Paf1 recruitment underlies the downregulation of longer genes, the logic then follows that Paf1 signal should be more drastically reduced in this group. To this end, I used the same gene length groups as in Figure 4.2 and plotted Paf1 metagene and X-Y

scatters (Figure 4.6). This analysis revealed that as the gene length increased, reduction of Paf1 recruitment became more prominent. Interestingly, as the gene length increases, two distinct peaks of Paf1 become visible: one after the TSS and one closer to pA site. While both of them were reduced in longer genes in Cbc3 deletion, the latter peak was affected more. It is possible that the Paf1 recruitment close to the TSS was realized via a Cbc3-independent mechanism, while the recruitment during elongation relied on Cbc3. Taken together, these data point to a role in transcription elongation for *S. pombe* Cbc3, via facilitating recruitment of Paf1 to longer pc genes.

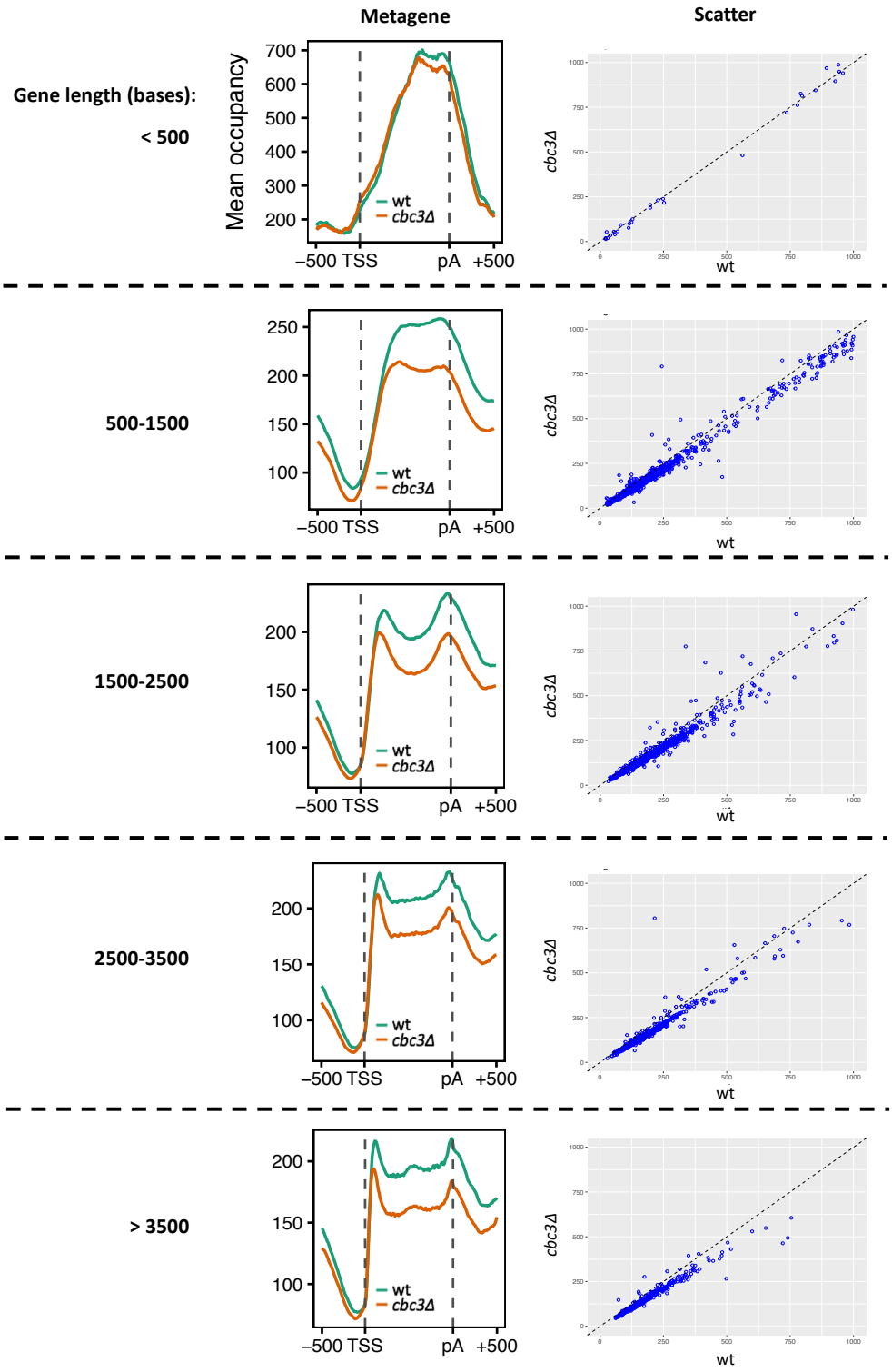


Figure 4.6: Paf1 recruitment to longer genes were more affected upon Cbc3 deletion. Pc genes of *S. pombe* were grouped by length into five subsets and recruitment of Paf1 was assessed in all. For metagenes, wt (green) and *cbc3Δ* (orange) signals were binned to 0.5 kb before TSS, 1 kb for the

gene body, and 0.5 kb after pA site. For scatter plots, read counts of each gene were plotted as wt (x axis) vs *cbc3Δ* (y axis). y = x is denoted by a dashed line.

4.3 Cbc3 facilitates nuclear export of poly(A)$^+$ RNA localization

The other potential role for *S. pombe* Cbc3 was in nuclear export, alluded to both by the THO/TREX component-like co-transcriptional recruitment profile and the general but mild downregulation of pc genes; as well as findings reported by Gebhardt et al. on human Cbc3. Furthermore, Stuart Wilson's lab had identified Cbc3 for the first time as a potential interactor of THO/TREX complex (Dufu *et al.*, 2010). To entertain this possibility, I tried combining *cbc3Δ* with deletion of Tex1. A member of the THO/TREX complex, its human orthologue (THOC3) was found to be interacting with Cbc3 (Gebhardt *et al.*, 2015). Moreover, Tex1 is non-essential in fission yeast, allowing for easier genetic interaction studies. Consequently, the double mutant showed a growth defect both on agar and in liquid media (Figure 4.7).

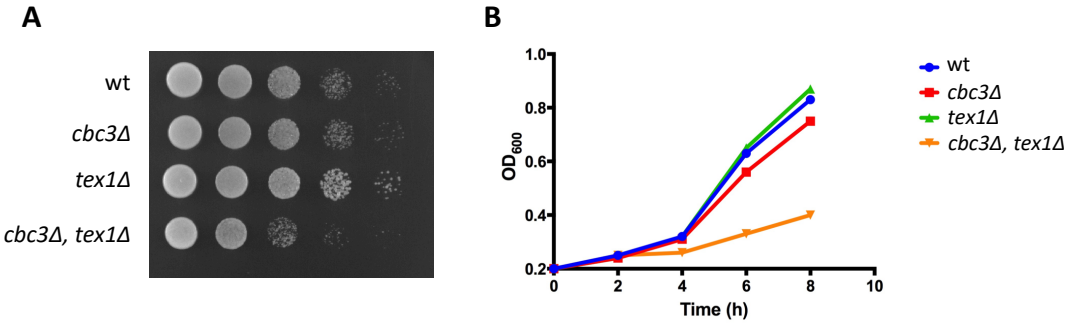


Figure 4.7: (A) Spot test with wt, *cbc3Δ*, *tex1Δ* and *cbc3Δ, tex1Δ* strains. Serial dilutions were spotted on a YES plate and grown at 30°C for 5 days. Image is a representative of two independent

replicates. (B) Growth curves of the strains in (A) in YES. Inoculations were grown overnight and diluted to OD_{600} = 0.2; and OD_{600} was measured for the consecutive 8 h.

Then, I checked the subcellular localization of poly(A)$^+$ RNA by FISH (Figure 4.8), as done in section 3.3.2. Compared to wt, Cbc3 deletion strain showed some nuclear retention alongside a more "granulated" signal. In comparison, Tex1 deletion by itself did not have a drastic effect on RNA localization, possibly due to a compensatory export mechanism that comes into play in this mutant. On the other hand, the double mutant exhibited a deformation of cell shape followed by strong nuclear retention. It can also be observed that this strain contained two subsets of phenotypes: one small, globular cell with strong nuclear retention; and one swollen cell where DAPI signal is dispersed. It is likely that when the former cell attempts to undergo mitosis, the latter phenotype presents itself. Therefore, loss of Cbc3 causes nuclear retention of poly(A)$^+$ RNA, which is further compounded upon deletion of Tex1.

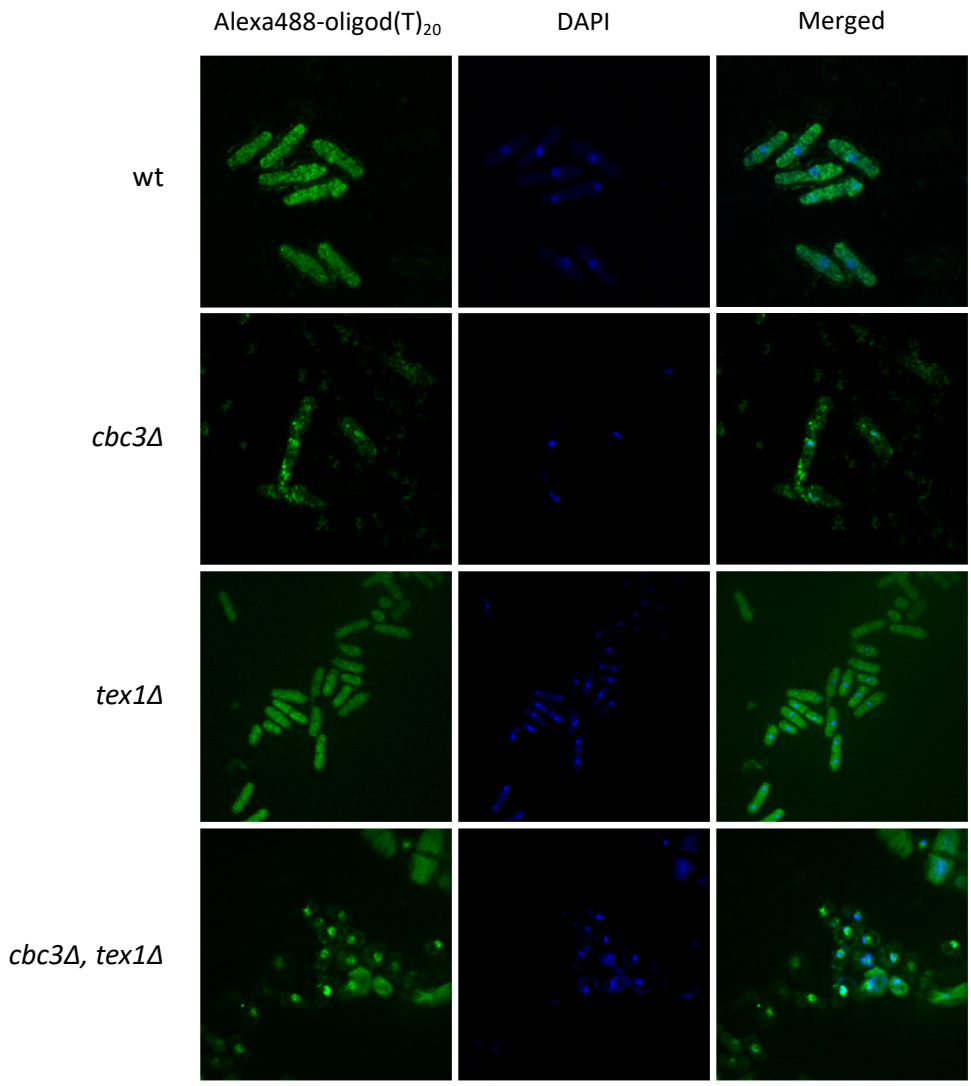


Figure 4.8: poly(A) FISH with wt, *cbc3Δ*, *tex1Δ* and *cbc3Δ, tex1Δ*. Alexa488-conjugated oligo-d(T)$_{20}$ probe and DAPI staining was used to detect the relative localization of polyadenylated RNA in *S. pombe*. Images are representative of three independent replicates.

4.4 Discussion

In this chapter, I reported my findings on the S. pombe orthologue of *NCBP3*, *SPCC16C4.16c* (referred to as Cbc3) and propose potential roles for its protein in transcription elongation and nuclear export. There are a number of points I would like to

discuss regarding the experiments, as well as some aspects of Cbc3 I have not touched upon in this study.

SPCC16C4.16c being a non-essential gene in *S. pombe*, it was somewhat surprising to see the general downregulation of pc genes upon Cbc3 deletion. This could not be attributed to a library size difference, thanks to the *S. cerevisiae* spike-in normalization. It is possible that the mild downregulation observed could be compensated for by other mechanisms to ensure cell survival. It is also noteworthy that Cbc3 deletion specifically affected pc genes, with almost no nc genes or CUTs significantly dysregulated. This corroborated the pc gene-specific recruitment profile of Cbc3 observed in ChIP-seq (section 2.1). Thus, Cbc3 binding might provide yet another means by which pc genes are distinguished from other Pol II transcripts. Combined with the finding that longer genes were more affected by Cbc3 deletion, similarities can be drawn between this mechanism and that of hnRNP C tetramer in mammals, which acts as a "molecular ruler" to bind only mRNAs and directing them to a different export mechanism by excluding PHAX binding (McCloskey *et al.*, 2012). To date, no such length-distinguishing factor was found in fission yeast; and it is possible that Cbc3 takes on this role in this organism. Furthermore, this would also explain why nc genes and CUTs were much less affected, as these transcripts tend to be on the shorter side of the length spectrum.

What brings Cbc3 to longer transcripts remained elusive. My RNase-treated ChIP-seq pointed to an RNA-independent means of recruitment, and I failed to find a sequence motif that favoured Cbc3 binding. However, ChIP-seq usually lacks this level of resolution, and instead, iCLIP or CRAC would be a better technique to ask this question. These techniques would also reveal whether Cbc3 is directly recruited to the RNA or somewhere else on the transcription machinery.

What led me to investigate the common properties among genes downregulated upon Cbc3 deletion was the fact that there were an additional 1099 genes that were downregulated by RNA-seq, which did not have Cbc3 signal by ChIP-seq. This is most likely due to the gene length-dependent effect on Paf1 recruitment upon Cbc3 loss. Though the molecular mechanism of Cbc3-mediated Paf1 recruitment remains to be elucidated, it is possible that

Paf1 reduction in longer genes result in slower elongation, and thus downregulation of the transcript. Alternatively, this might be due to sub-optimal pulldown of Cbc3 in ChIP-seq, resulting in loss of binding information. I cannot rule out the possibility that those additional 1099 genes still can bind Cbc3, albeit to a lower/more transient extent, which could not be detected by ChIP-seq.

By ChIP-seq, it was confirmed that there was a decrease in Pol II occupancy over the gene body after Cbc3 deletion, and this effect was the most prominent in genes significantly downregulated by RNA-seq. Processivity of the gene was also found to be less than that in wt. However, it is important to note that the processivity measurements utilized here is, by definition, is only a proxy of the Pol II metagene. If the slope of the metagene profile is steeper, the Pol II would be regarded as "more processive", and vice versa. As a control, I calculated and compared the processivity scores for two different total Pol II pulldowns from wt strain. Processivity differences observed after Cbc3 deletion was ~5-fold more than differences between two wt pulldowns. Furthermore, as compared to genes not significantly changed by RNA-seq, processivity difference in downregulated genes was 3.5-fold higher. Therefore, while the processivity calculation gives an idea as to "how much of the Pol II at any given point on a gene proceeds forward", it should not be taken as an absolute measurement. *In vivo*, techniques such as PRO- or GRO-seq could be utilized in a time course to precisely determine the position of Pol II, which could then be compared between wt and Cbc3 deletion. Additionally, *in vitro* studies would be required to determine precisely how much slower Pol II elongation complex is in *cbc3Δ*.

An interesting detail was that the not significantly changing genes appeared to have a small peak right before the TSS, resembling either an initiation problem or promoter-proximal pausing (Figure 4.3A). However, Ser5 phosphorylation levels of CTD remained unchanged in the Cbc3 deletion, making it unlikely to be initiation-related. To see if there is a negative correlation between pausing and Cbc3 dependency, overlap between paused genes in fission yeast reported by Lis lab (Booth *et al.*, 2016) and genes not changing upon Cbc3 deletion could be assessed. In mammals, Paf1 depletion resulted in increased release from promoter-proximal pausing and a subsequent increase in nascent and mature transcripts (Smith *et al.*, 2015). This pause release was associated with liberated enhancer

regions previously occupied by Paf1 (Chen *et al.*, 2017). On the other hand, Paf1 is also integral for elongation (Fischl *et al.*, 2017). Then, it is possible that, in Cbc3 deletion, genes prone to pausing have more transcription due to pause release via Paf1 clearance; however, since Paf1 recruitment during elongation is also affected, overall transcriptional output from said genes does not change significantly.

Figure 4.4 shows that while the overall total Pol II levels did not change, there was an increase of Ser2 phosphorylation over the gene body, peaking at pA site. Admittedly, it is hard to interpret a change in signal intensity in a ChIP-seq experiment. Naturally, it could be interpreted as an accumulation of the Ser2 phospho-mark; but it could also be interpreted as slower Pol II allowing for more efficient pulldown. For instance, Bachand lab reported increased signal for total Pol II ChIP-seq in *dhp1-154* mutant, possibly the result of a slow transcription cycle (Larochelle *et al.*, 2018), which was the opposite of what I observed in the same mutant (section 3.3.2; Figure 3.10). Thus, when interpreting these signal changes, one needs to take into account the context. Considering the general downregulation of pc genes observed upon Cbc3 deletion, there might be an increase in the phospho-mark in an attempt to encourage Pol II to accelerate, to compensate for the slow elongation. Yet, despite the presence of extensive S2P, Paf1 recruitment, and possibly the recruitment of other elongation factors as well, is impaired in Cbc3 deletion and thus Pol II processivity is affected. Furthermore, that the total S2P levels in the whole cell extract have increased also implies that the increase in signal intensity is real and not a result of slower Pol II being more “ChIPable”.

In human cells, elongation rate of Pol II positively correlates with gene length (Veloso *et al.*, 2014). This would explain why, upon Cbc3 deletion, downregulation is observed mostly on longer genes. The elongation defect is negligible, or can be compensated for, if the gene is shorter; however, longer genes need a faster Pol II, and thus are more affected. Paf1 recruitment patterns for the different gene length groups (Figure 4.6) present a mechanistic insight to this phenomenon. For all of the gene groups, the TSS-proximal Paf1 signal remains fairly the same. It can be argued that a short gene only consists of this TSS-proximal area, where Paf1 recruitment, and thereby elongation, is unaffected upon Cbc3

deletion. As the gene length increases, Paf1 signal difference between wt and mutant strain elevates, resulting in elongation defects and subsequent downregulation.

One possibility I did not cover in this study was the effect of Cbc3 deletion on Paf1 protein levels. That said, recruitment of Paf1 was affected differently in different gene groups (Figure 4.6) and for shorter genes, no significant change was observed in Paf1 levels. Furthermore, as stated earlier, all gene groups showed a TSS-proximal Paf1 peak, which stayed unchanged in Cbc3 deletion; the expectation would be that such a peak would also be affected by a global reduction in Paf1 protein levels. Nevertheless, while it is unlikely that Paf1 recruitment differences in Cbc3 deletion are due to a global reduction of its protein level, it might be worthwhile investigating.

Another way to support the hypothesis that Cbc3 deletion affects elongation would be to check its growth on plates containing elongation inhibitors like 6-azauracil or α-amanitin (Exinger and Lacroute, 1992; Brueckner and Cramer, 2008). Like the slow polymerase mutant in *S. cerevisiae* (De La Mata *et al.*, 2003), an elongation mutant should exhibit growth defects when grown on 6-azauracil.

Not only did Cbc3 deletion affect Pol II elongation, but it also resulted in nuclear retention of poly(A)$^+$ RNA. Gebhardt's findings in HeLa cells pointed to a role for human Cbc3 in mRNA export, and fission yeast orthologue appears to play a similar role. Supporting this, in my attempts to pull down Cbc3 interactors, RNA helicase Uap56 (human DDX39B) was found among the proteins interacting with *S. pombe* Cbc3. Surprisingly, no THO/TREX component was present in the Cbc3 eluate, despite the earlier findings in human cells (Dufu *et al.*, 2010; Gebhardt *et al.*, 2015). It is possible that this stemmed from the highly stringent purification conditions utilized, which broke apart the weaker and/or less transient interactions between Cbc3 and THO/TREX components.

When combined with the deletion of Tex1 (THOC3 in humans), Cbc3 deletion showed growth defect and aberrant cell morphology. In Figure 4.7, it can be seen that Tex1 single mutant appeared to grow somewhat faster than wt. This is most likely due to the small cell size of this mutant, interfering with the OD_{600} calculations. However, despite having a smaller, globular shape, which would also interfere with calculations, double mutant cells

grew slower, implying that the growth defect might be more pronounced than observed in the spot test. In fact, in liquid media, the growth defect of the double mutant was indeed more pronounced, with a doubling time of almost 8 h. This could be due to a problem in mitosis programme in these cells, whereby a cell undergoing mitosis looks swollen (Figure 4.8; upper right corner in *cbc3Δ, tex1Δ* image). Presumably, these cells might burst and die, considerably lengthening the doubling time for the strain.

Poly(A)$^+$ FISH experiments revealed that Cbc3 deletion indeed caused a mild nuclear retention phenotype (Figure 4.8). The retention was further compounded in the *cbc3Δ, tex1Δ* double mutant. Moreover, the RNA signal in Cbc3 deletion appeared more "granulated", especially in the cytoplasm, instead of the ubiquitous distribution in wt cells. Superficially, this resembles P-body staining in *S. pombe* (Wang, Chen and Wang, 2013). In budding yeast, mRNAs are targeted to P-bodies under stress conditions (Wang *et al.*, 2018), and since human Cbc3 was reported to have a role in mitigating viral stress (Gebhardt *et al.*, 2015), it is possible that fission yeast Cbc3 also has a stress-relieving role; not necessarily viral but more general, absence of which causes mRNAs to be localized in P-bodies. If this holds true, it would also explain the general downregulation of mRNAs in this mutant. One idea would be to stain a P-body marker, such as Dcp2, in poly(A)$^+$ FISH to confirm co-localization in Cbc3 deletion. Alternatively, a mutant where P-body function is compromised might be combined with Cbc3 deletion to assess whether the downregulation of mRNAs is rescued. However, it is important to note that my current data on Cbc3 do not point to any role/interactors in the cytoplasm; and thus, even if P-body localization is confirmed, this effect might be indirect.

Is the gene length difference observed in transcriptome changes reflected in the nuclear retention? My expectation would be that the longer mRNAs are retained more in the nucleus upon Cbc3 deletion. This could be checked by utilizing probes specific for only longer vs. shorter genes, or by fractionating nuclear and cytoplasmic RNA pools and check the levels of few chosen RNAs. It is also possible that the longer genes are not retained in the nucleus, but rather are targeted to P-bodies in the cytoplasm. Hence, the true fate that befalls on longer gene transcripts in Cbc3 deletion remains to be elucidated.

In conclusion, the findings reported in this chapter can be summarized as follows:

- Cbc3 deletion causes pc genes longer than ~500-700 bases in size to be downregulated. The effect is exclusive to pc genes, in line with the co-transcriptional recruitment profile for Cbc3 reported in Chapter 2.
- Upon Cbc3 deletion, both the occupancy and the processivity of Pol II in those genes significantly downregulated are reduced *in vivo*. While the Ser5 phospho-mark in CTD remained unchanged, there was an increase in Ser2 phosphorylation, interpreted as a compensatory action for slower elongation. Paf1 recruitment to longer genes is affected in Cbc3 deletion; thereby hinting at a mechanistic explanation for the slow enzyme.
- Cbc3 deletion also causes nuclear retention of polyadenylated RNAs, as well as more speckle-like distribution in the cytoplasm, which may be due to RNA targeted to P-bodies. Retention phenotype is further worsened when combined with deletion of a THO component, Tex1.

Chapter 5 General Discussion and Conclusions

In this thesis, fission yeast CBC was put under spotlight; striving to answer two main questions:

1. Is there a possibility of an alternative CBC?
2. Can CBC components also have CBC-independent roles alongside?

To elaborate on the former question, I studied the fission yeast orthologue of *NCBP3*, Cbc3; testing its recruitment, interacting partners, and biological function; comparing them with those of the canonical CBC. From my findings, I concluded that Cbc3 is a protein affecting RNA biogenesis both transcriptionally and post-transcriptionally (will be discussed in section 5.2); but does so in a mutually exclusive manner from the CBC (section 5.1).

For the second question, I tested the effect of depleting Cbc1 or Cbc2 individually on RNA levels and processing; and their genetic interactions with different RNA surveillance factors; and focused on one such factor (Dhp1) to further explore the differences observed with Cbc1 vs. Cbc2 depletion (section 5.3). It is apparent that Cbc1 and Cbc2 influence RNA levels and processing at varying levels, as well as interacting with RNA surveillance factors to differing extents. This difference is best exemplified in Cbc1 depletion rescuing the readthrough phenotype associated with Dhp1 (section 5.4). I hence deduced that even when the CBC is compromised, Cbc1 and Cbc2 can carry out separate functions. Therefore, caution should be exercised when working on CBC by depleting one component vs. the other, as they would bring about diverging changes downstream.

5.1 CBC: Is there room for another?

Gebhardt et al. proposed that Cbc3 could form an alternative CBC with Cbc1, substituting for the cap-binding feature of Cbc2. It contained an RRM domain similar to that of polyA-associated ribonuclease (PARN) (21.4% sequence identity) and of Cbc2 (16.4%), and was able to selectively bind 7mGTP than GTP (Gebhardt *et al.*, 2015). Since then, however, it was shown that Cbc3 can be in a ternary complex with Cbc1, Cbc2 and Ars2 (Schulze *et al.*, 2018), shedding doubt on the possibility of a mutually exclusive complex. To begin addressing such question, we should first know if fission yeast Cbc3 could bind RNA and/or the 7mG cap. An HHpred run (Söding, Biegert and Lupas, 2005) for the orthologue showed that, like its human orthologue, fission yeast Cbc3 shared the highest sequence similarity to PARN (36%). Depending on sequence similarity, HHpred hinted at a possible RRM domain between amino acids 34-91 and a 7mG binding domain between 24-34. This prediction was corroborated by the protein modelling tool Phyre2 (Kelley *et al.*, 2015), carried out by Dr. Krzysztof Kus. Intriguingly, *S. pombe* Cbc3 did not appear to have significant structural similarity with Cbc2 (5.6%, with an E-value of 0.017). One similarity was that DisMeta (Huang, Acton and Montelione, 2014) predicted disordered regions for both Cbc2 and Cbc3 after their respective RRM domains. Disordered regions have been found in conjunction with RRM domains in numerous proteins such as U2AF2, SRSF1, SRSF2 of human and Tat of human immunodeficiency virus (HIV), alluding to a central role in RNA binding (reviewed in (Järvelin *et al.*, 2016)). RNA binding by Cbc3 was confirmed in an ongoing mRNA interactome study from our lab (Kilchert et al., unpublished), as well as my data showing a reduction of Cbc3 signal upon RNase A treatment. ChIP-seq showed a more 3' recruitment for Cbc3, and while it is still possible that the protein is recruited to the 7mG cap at later stages of transcription, it is more likely that Cbc3 binds RNA further downstream, closer to the pA site.

Another finding pointing against an alternative CBC was the absence of Cbc3 in Cbc1 pulldown. This was surprising considering the co-IP experiments in mammalian cells (Gebhardt *et al.*, 2015). It is, of course, possible that such an interaction went unnoticed due to transient nature of a potential interaction between Cbc1 and Cbc3. Notwithstanding,

as was seen for canonical CBC pulldowns, if Cbc3 was an abundant interactor for Cbc1, it would have come up in our purifications.

While not included in this thesis, I also tested the genetic interaction between Cbc3 deletion and CBC depletion strains. There, combination of Cbc1 depletion and Cbc3 deletion showed a growth defect, while Cbc2 depletion did not have an additive effect. A hypothetical alternative CBC formed between Cbc1 and Cbc3 would have alleviated the loss of Cbc2; and conversely, depleting both Cbc2 and Cbc3 would have resulted in a worse phenotype, which was not observed. On the other hand, there is a possibility that Cbc1 and Cbc3 share some redundant functions, loss of both resulting in further growth defects.

Another set of data not included, but worthy of discussion, was the biochemical characterization of RNA and 7mG affinity of Cbc3, by Dr. Krzysztof Kus and Ms. Anna Wiecek. A Cbc3-RRM recombinant construct containing the putative RRM domain (1-112 aa) was unable to bind the cap analog in isothermal titration calorimetry. In line with this, m7GTP-conjugated beads pulled down full length Cbc3, but not Cbc3-RRM. Finally, fluorescence anisotropy studies showed moderate RNA affinity for full length, but not for Cbc3-RRM. Unfortunately, capped RNA was not tested in this experiment. As a result, it is hard to assess whether the Cbc3-RRM construct is indeed not able to bind the RNA, or non-functional. If the former holds, it could be argued that the C-terminal stretch of disordered region is important for RNA binding. A native purification with m7GTP beads with or without nucleases would also reveal if Cbc3 is able to bind the cap *in vivo*.

5.2 Function of Cbc3 in *S. pombe*

In light of the findings presented in Chapter 4, a working model for the function of Cbc3 in fission yeast is proposed (Figure 5.1). For pc genes longer than ~500 bases, Cbc3 takes on roles both in elongation and export. Upon co-transcriptional binding close to the pA site, Cbc3 facilitates Paf1 recruitment to the elongating Pol II, increasing the processivity of the enzyme. After cleavage and polyadenylation, Cbc3 helps with nuclear export alongside THO/TREX complex and possibly via its interacting partner Uap56. Due to the

size-specific recruitment of Cbc3, I postulate that this is a means by which longer pc genes are recognized and exported via the Yra1-Uap56 pathway instead of PHAX, similar to the "molecular ruler" hnRNP C tetramer in higher eukaryotes (McCloskey *et al.*, 2012). To further test the size specificity of Cbc3 recruitment, a Cbc3-recruiting non-essential gene could be replaced with a truncated version and its levels could be assessed in the Cbc3 deletion mutant, where the short gene should not be affected by loss of Cbc3 anymore.

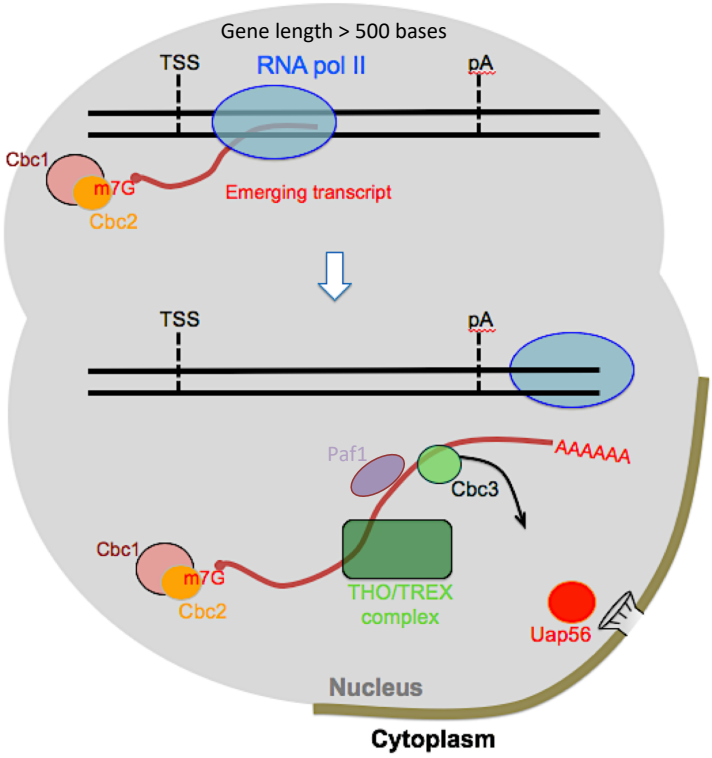


Figure 5.1: Working model for the role of Cbc3 in *S. pombe*. For genes longer than ~500 bases, Cbc3 is co-transcriptionally recruited toward the 3' end of the gene. This recruitment in turn facilitates Paf1 recruitment, increasing the elongation rate. Once the transcript is cleaved and polyadenylated, Cbc3 works in conjunction with THO/TREX complex and Uap56 to mediate RNA export.

There are a number of potential pitfalls to this model. First, the absolute elongation rate of Pol II was not assessed and only inferred from ChIP-seq, processivity measurements,

changes in the phospho-CTD and the subsequent changes in RNA levels. Though the slow elongation hypothesis is also backed up by Paf1 recruitment changes in longer genes, the elongation rates should still be assessed with nascent RNA techniques such as global run-on sequencing (Gardini, 2017) or *in vitro* transcription studies.

A second possible shortcoming is the link between Cbc3 and export. This hypothesis is backed up by the retention phenotype observed by FISH, Uap56 being an interactor of Cbc3, and the aberrant growth of *cbc3Δ, tex1Δ* double mutant. On the other hand, we still do not know whether the retention is specific to longer genes. Furthermore, from the granulated look of the cytoplasmic RNA signal in the mutant, it is possible that Cbc3 can direct RNA to P-bodies in the cytoplasm. Nuclear/cytoplasmic fractionation of RNA would reveal if there is a size selection in nucleus-retained RNAs. A potential role for Cbc3 in the cytoplasm also requires further investigation.

5.3 Cbc1/2 having CBC-independent roles

Previously, several publications have entertained the idea that Cbc1 and Cbc2 might also have CBC-independent functions. For instance, eIF4E-bound transcripts were found to contain Cbc1 only (Rufener and Mühlemann, 2013). Conversely, an alternatively spliced isoform of Cbc2, not containing the Cbc1-interacting domain or the 7mG binding pocket, is still able to bind RNA and is distributed to active transcription site (Pabis *et al.*, 2010). In *S. cerevisiae*, where CBC genes are non-essential, gene expression profiling of Cbc1 vs. Cbc2 deletion revealed a mere 15% of transcripts commonly regulated (Baron-Benhamou et al. 2003). Results presented in Chapter 3 confirm that Cbc1 and Cbc2 might have separate roles from each other and affect RNA biogenesis differently. In *S. pombe*, where both components are essential and where depleting one does not affect the protein stability of the other, Cbc1 and Cbc2 depletions did not affect gene expression profiles the same way; with many more genes upregulated in the former mutant. Moreover, splicing was affected more in Cbc1 depletion, whereas it could still be carried out to some extent in Cbc2 depletion, suggesting that Cbc1 can still mediate splicing in a Cbc2-independent manner. The depletion strains also interacted differently with deletions/ts mutants of RNA

surveillance factors, most pronounced differences observed with Pab2, Hrb1 and Dhp1. If the sole functions of Cbc1 and Cbc2 were those associated with CBC, disruption of either would have resulted in the same phenotype in surveillance factor double mutants.

On the other hand, it needs to be stressed out that in the purifications, Cbc1 was the most prominent interactor of Cbc2, and vice versa. The complex is stoichiometric and shades all other interactors in terms of abundance. Hence, it is safe to assume that, at any given point, most of Cbc2 is in complex with Cbc1, and the prominent function of the two proteins is that in CBC. Furthermore, in the ChIP-seq experiments, no gene was found where only one of the CBC proteins was recruited. Thus, it can be speculated that at least for co-transcriptional recruitment, both proteins are required. The Pol II ChIP-seq profiles in the two depletion strains showing identical changes, especially close to the TSS, highlight that for some of the transcriptional functions (i.e. P-TEFb recruitment), both proteins are required.

The core functional differences must, then, arise later on, where one protein is still able to recruit downstream effectors while the other cannot. Structural studies on CBC point out that while Cbc2 is responsible for recruitment to the 7mG cap, Cbc1 can act as an adapter protein, providing scaffolding for U1 snRNP, Ars2 and NEXT components in mammals (Lewis *et al.*, 1996; Mazza *et al.*, 2001; Andersen *et al.*, 2013; Hallais *et al.*, 2013). This also provides basis as to why, for instance, splicing was not affected in Cbc2 depletion as drastically as in Cbc1 depletion; Cbc1 is probably still able to interact and direct splicing machinery to nascent RNA, and RNA processing occurs, albeit to a lesser extent. On the other hand, the main function of Cbc2 appears to be related to RNA stability, as the depletion results in drastic downregulation of almost all Pol II transcripts. This could be either due to the free 7mG cap being targeted more easily by decapping enzymes, followed by degradation, or due to Cbc1 still actively targeting RNA to the exosome. In fact, the finding that deletion of certain surveillance factors alleviated Cbc2 depletion suggest that nuclear degradation is still active in Cbc2 depletion, likely via the activity of Cbc1 alone.

As a result, it needs to be stressed out that Cbc1 activity might not necessarily require Cbc2 presence. For this reason, care must be taken when working on CBC, as depletion of one

component would not be physiologically identical to depletion of the other. When possible, both Cbc1 and Cbc2 should be targeted to completely exhaust CBC function. Most of the research on CBC is carried out by targeting Cbc1 only, and while this is logical considering that Cbc1 is the adapter protein and the more abundant among them, one should take into account the possibility that function of Cbc2 is not entirely compromised.

However, the means of Cbc1 recruitment to the nascent RNA in the absence of Cbc2 remains elusive. Lacking an RRM domain or a 7mG binding pocket, Cbc1 is not expected to localize to the nascent RNA; however, my data on persistent splicing and RNA surveillance in Cbc2 depletion hints at a co-transcriptional recruitment for Cbc1, even in the absence of its partner. It might thus be worthwhile to check co-transcriptional recruitment of Cbc1 in the absence of Cbc2. If Cbc1 can still be recruited, it is possible that it is brought upon the RNA via other means, potentially by another factor containing an MIF4G-like domain.

A shortcoming of this study was that, for decent depletion of the gene via NMT promoter, the strains were grown in thiamine-containing media for 16 h. This is a long time to put cells under stress and can introduce artefactual results. However, with the emphasis of the study being the differences between two depletion strains, both of which are grown for 16 h in thiamine media, the presumed stress-related changes should be happening in both strains; and yet, the strains show differences. Therefore, while a certain level of stress caused by 16 h depletion could not be ruled out, it is unlikely that the differences observed in two depletion strains could be attributed to stress, solely. Nevertheless, a more rapid depletion technique, such as anchor-away (Haruki, Nishikawa and Laemmli, 2008) or the auxin-inducible degron system (Nishimura *et al.*, 2009) can be utilized to confirm the results.

5.4 Cbc1 and readthrough phenotype rescue: Anti-cleavage or anti-degradation?

The most striking difference between the CBC proteins was found when the depletion strains were combined with ts allele *dhp1-154*. While the Cbc1 double mutant grew faster, a synthetic phenotype was observed with Cbc2 double mutant. Interestingly, Cbc1 depletion, and not Cbc2 depletion, was found to rescue the readthrough phenotype of *dhp1-154* (Kim, Krogan, *et al.*, 2004). While being a clear indication that the absence of either CBC protein does not result in the same phenotype, this difference also justifies the divergent genetic interactions between Dhp1 and the two depletion strains.

Currently, two working hypotheses can be put forward. The first postulates that Cbc1 has an anti-cleavage role, absence of which causes more efficient cleavage in the double mutant, and though Pol II termination and turnover are still affected, readthrough is not included in the pre-mRNA. It was reported that CBC has a role in stimulating premature pA signal-mediated cleavage, such as one residing in a cap-proximal intron (Hallais *et al.*, 2013). It can be speculated that, for further downstream pA sites, Cbc1 might act as a checkpoint, slowing down the cleavage process to ensure RNA integrity (i.e. association with THO/TREX components, completion of splicing). In the absence of Cbc1, this checkpoint might be lifted, resulting in faster, but "sloppy", cleavage. This in turn results in not export-ready RNAs, which should be retained in the nucleus and degraded. However, due to compromised RNA surveillance in the Cbc1 depletion, these pre-mRNAs accumulate, as observed by poly(A)$^+$ FISH. Laitem et al. describes a Cdk9-mediated checkpoint at the pA site, where the inhibition of Cdk9 hinders Pol II elongation to the normal end of the transcription unit via early recruitment of Pcf11 (Laitem *et al.*, 2015). It can be speculated that Cbc1 promotes Cdk9 recruitment at the pA site, similar to its role proximal to TSS, absence of which causes premature cleavage/termination. To further cement this hypothesis, cleavage dynamics in Cbc1 depletion can be assessed. Moreover, Pol II occupancy around the pA site can be tested more directly with a nascent RNA-based technique such as NET-seq, as this would provide a better understanding of Pol II movement as compared to ChIP-seq. Another test would be deletion of a non-essential

component of the cleavage and polyadenylation machinery alongside Cbc1 depletion, as this would be expected to suppress the readthrough phenotype.

A second possible explanation for this phenomenon is that Cbc1 stabilizes the 3' extended transcripts that were the result of Dhp1 activity loss. This hypothesis then posits that Cbc1 depletion results in destabilization of these extended transcripts, and thus the reason why the readthrough phenotype is "rescued" in the double mutant. However, the fact that the extended transcript is present in Cbc2 depletion undermines this hypothesis to an extent, as so far, the most prominent role for Cbc2 appears to be RNA stabilization. Thus, it is unlikely that Cbc1 depletion only causes extended transcript degradation. In fact, the link between CBC and RNA surveillance is usually attributed to Ars2, which in turn binds Cbc1; therefore, the expectation would be that surveillance is compromised upon Cbc1 depletion, and hence, 3' extended transcripts should be even more stable at this state. Nevertheless, combining the Cbc1/Dhp1 double mutant with deletions of several surveillance factors would be a means to test this possibility. Thus, while this explanation is less likely in light of all my findings, further experiments are still necessary to challenge this hypothesis.

Chapter 6 Materials and Methods

6.1 Lists of oligonucleotides, vectors, *S. pombe* strains and antibodies used in this study

6.1.1 Oligonucleotides

Table 6.1: List of oligonucleotides used in this study. Note that the L4 and L5 primers contain a 5' homology arm to the vector to be used. These homology arms are indicated at the end of the table.

Name	Sequence	Purpose
Cbc1-Ctermtag L3	ACATTTTTCTCGAAAATCATTAG	Cbc1 tagging
Cbc1-FTP L4	TTTTTCCTCATTAGTTGCATCAG	Cbc1 tagging
Cbc1-FTP L5	AGAAAGTAGTCTATAACAAAAATG	Cbc1 tagging
Cbc1-Ctermtag L6	GTAGCTGTGCAAATGTCCCTTAA	Cbc1 tagging
Cbc2-Ctermtag L3	CTGTTTATTTGCTCTTCAAGTTC	Cbc2 tagging
Cbc2-FTP L4	ATTTTTTTTCCAACGATTGTATC	Cbc2 tagging
Cbc2-FTP L5	AAAGTTAGTTATATAATTCGGAC	Cbc2 tagging
Cbc2-Ctermtag L6	AATGTTAACGGATTTGTTCGTCC	Cbc2 tagging
Cbc3-Ctermtag L3	GATGATGATGTTTCACGAAATGC	Cbc3 tagging
Cbc3-FTP L4	ACGCTTGGAAACCCTGGAACGAC	Cbc3 tagging
Cbc3-FTP L5	TTGTGTTGTAGAGAAATATTAAATAC	Cbc3 tagging
Cbc3-Ctermtag L6	CTGTCTTCATGACTTTTACATTTC	Cbc3 tagging
Cbc1-HA L4	TTTTTCCTCATTAGTTGCATCAG	Cbc1 tagging
Cbc1-HA L5	AGAAAGTAGTCTATAACAAAAATG	Cbc1 tagging
Cbc2-HA L4	ATTTTTTTTCCAACGATTGTATC	Cbc2 tagging
Cbc2-HA L5	AAAGTTAGTTATATAATTCGGAC	Cbc2 tagging
Cbc1-NMT L3	CATCTTTTAAGTTCCTTCCTACC	NMT promoter integration

Cbc1-NMT L4	TTTGTGAGGAATGTTAGTGTCTT	NMT promoter integration
Cbc1-NMT L5	ATGTCTTCTTATCGGGGATCAAC	NMT promoter integration
Cbc1-NMT L6	CTTTTGAGATAGCGTCAGCCAG	NMT promoter integration
Cbc2-NMT L3	ATGATGGTTGATGCCATTTTGTT	NMT promoter integration
Cbc2-NMT L4	TTTTCCTAGGCAGTGGTTGTAGG	NMT promoter integration
Cbc2-NMT L5	ATGGCATCAATAACAAGGTTAG	NMT promoter integration
Cbc2-NMT L6	GAATCCGACTACATACCACCACG	NMT promoter integration
cbc3-delcas fwd	GTTGCTATTAATGGCATTCGAAG	Cbc3 deletion
cbc3-delcas rev	CGAGGTACAGCTATGAATCCATG	Cbc3 deletion
paf1-delcas L3	GATTAGGCAAGCATTCATTAATC	Paf1 deletion
paf1-delcas L4	TGATGTGTTTTCCACAAATACC	Paf1 deletion
paf1-delcas L5	TGTCTTTCATGCAGCATACCCT	Paf1 deletion
paf1-delcas L6	AAGCCATCTCCAAGGCCATC	Paf1 deletion
tbp1 fwd	CGTTTTGCCGCTGTTATTATGC	Northern probe
tbp1 rev	AGATGAGAAAGTACCGTGGGAG	Northern probe
wos2 fwd	CGGCCGCTACATCTTTTTCG	Northern probe
wos2 rev	AGTCCATACCGCCCATACCT	Northern probe
phb2 fwd	GATGAAGCGGGGACTTCC	Northern probe
phb2 rev	CTTTCGTTCCGGTTAGAG	Northern probe
pre8 fwd	TGGTGTTCGTCCGTTTGGTG	Northern probe
pre8 rev	CGAATCTCCGACTGGGACAATC	Northern probe
wis2 fwd	GGCGAGAAGTTTGAGGACGA	Northern probe
wis2 rev	GTCATCGTTGGGAACAGGGT	Northern probe
wos2_2 fwd	GTAGGTGACTTCAAGAACCTG	qPCR primer
wos2_2 rev	GTAACATACCTTCAGGAATTTG	qPCR primer
wos2_6 fwd	TTAGTCCATTTGGCGCAGGT	qPCR primer
wos2_6 rev	ACTGCCAACTTCCTCTTCCTC	qPCR primer
wos2_8 fwd	TGTACTTCAGAACCGTCGAAC	qPCR primer
wos2_8 rev	ACAGAAAAGCATATGTTGCCG	qPCR primer
tbp1_2 fwd	CCCTAACCCTGAAGATCGGG	qPCR primer

tbp1_2 rev	GAGCCGTGGTGGGTAAAG	qPCR primer
tbp1_6 fwd	TCGTTTGGAAGGTTTGGCTT	qPCR primer
tbp1_6 rev	CCATGCGATAAATCAAACCGGG	qPCR primer
tbp1_9 fwd	ACTGAGTATCTGCTGTTCCTGT	qPCR primer
tbp1_9 rev	TGGCTGCTTCTGCTTGCTAT	qPCR primer
phb2_2 fwd	GGATAAGGTATACTGAAGTGCCA	qPCR primer
phb2_2 rev	TTCCTTTAGGAAGTCCCCGC	qPCR primer
phb2_5 fwd	GCCCAGCAAGATGCTCAAAG	qPCR primer
phb2_5 rev	AGTTGCGCTGCTCTACCTTC	qPCR primer
phb2_7 fwd	ACCGCTAGATCCAAAAACCATC	qPCR primer
phb2_7 rev	GGGAAAGCGGCAAGGGAAT	qPCR primer
pFA6a L4 homology arm	TAGTCTCCACCCCCGCCTCCCCC	5' homology arm
pFA6a L5 homology arm	CCAGTTTAAACGAGCTCGAATTC	5' homology arm
pAG25 L4 homology arm	TTAATTAACCCGGGGATCCGTCGACC	5' homology arm
pAG25 L5 homology arm	CGAGCTCGAATTCATCGAT	5' homology arm
pSLF173 L4 homology arm	GTCATAGCTGTTTCCTGTGTGAA	5' homology arm
pSLF173 L5 homology arm	ACTTATAGTCGCTTTGTTAACTCGA	5' homology arm

6.1.2 Vectors

Table 6.2: List of vectors used in this study.

Number	**Description**	**Used for**	**Constructed in**
3973	pFA6a-3xFLAG-TEV-ProteinA-KanMX	Endogenous tagging	Vasilieva Lab
4027	pAG25-HA-FLAG-NAT	Endogenous tagging	Vasilieva Lab
2499	pSLF173-ura4-NMT (medium)	NMT promoter integration	Forsburg Lab (Forsburg and Sherman, 1997)

6.1.3 *S. pombe* strains

Table 6.3: List of *S. pombe* strains used in this study.

Number	Genotype	Reference/ Constructed by
YP4	h-, ura4-Δ18, leu1-32, ade6-M216	Dom Helmlinger
YP67	h-, dis3-54, ura4-Δ18, leu1	(Ohkura *et al*., 1988)
YP112	h-, dhp1-154	(Hanyu *et al*., 2009)
YP144	h+, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	(Bühler *et al*., 2008)
YP411	h-, rrp6Δ, NAT, ura4-Δ18, leu1-32	Cornelia Kilchert (Vasilieva Lab)
YP697	h+, Cbc1-FTP, KanMX, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP691	h+, Cbc2-FTP, KanMX, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP770	h+, cbc3Δ:KanMX, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	(Bitton *et al*., 2015)
YP788	h+, ura4-NMT-Cbc1-FTP, KanMX, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP789	h+, ura4-NMT-Cbc2-FTP, KanMX, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP832	h+, Cbc1-HA-FLAG, NAT, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP833	h+, Cbc2-HA-FLAG, NAT, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP834	h+, ura4-NMT-Cbc2, Cbc1-FTP, KanMX, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP842	h+, ura4-NMT-Cbc1, Cbc2-FTP, KanMX, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP899	h+, tex1Δ, ura4, cbc3Δ, KanMX, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP944	h+, paf1Δ, NAT, cbc3Δ, KanMX, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP1113	h+, Paf1-HTP, NAT, cbc3Δ, KanMX, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP1208	h?, ura4-NMT-Cbc1-FTP, KanMX, red1Δ, NAT, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study

YP1209	h?, ura4-NMT-Cbc2-FTP, KanMX, red1Δ, NAT, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP1224	h?, ura4-NMT-Cbc1-FTP, KanMX, pab2Δ, NAT, ura4-Δ18, leu1-32, ade6-M216	This study
YP1225	h?, ura4-NMT-Cbc2-FTP, KanMX, pab2Δ, NAT, ura4-Δ18, leu1-32, ade6-M216	This study
YP1266	h?, ura4-NMT-Cbc2-FTP, KanMX, dhp1-154, ura4-Δ18, leu1-32	This study
YP1344	h+, ura4-NMT-Cbc1-HA-FLAG, NAT, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP1345	h+, ura4-NMT-Cbc2-HA-FLAG, NAT, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	This study
YP1346	h?, ura4-NMT-Cbc2-FTP, KanMX, srp1Δ, NAT, ura4-Δ18, leu1-32	This study
YP1347	h?, ura4-NMT-Cbc1-FTP, KanMX, dhp1-154, ura4-Δ18, leu1-32	This study
YP1366	h?, ura4-NMT-Cbc1-FTP, KanMX, srp1Δ, NAT, ura4-Δ18, leu1-32	This study
YP1461	h?, ura4-NMT-Cbc1-HA-FLAG, NAT, dhp1-154, ura4-Δ18, leu1-32	This study
YP1462	h?, ura4-NMT-Cbc2-HA-FLAG, NAT, dhp1-154, ura4-Δ18, leu1-32	This study
YP151	h+, prp10-1	(Bayne *et al*., 2008)

6.1.4 Antibodies

Table 6.4: List of antibodies used in this study.

Antibody	**Origin**	**Manufacturer**	**Application**
α-CTD MABI0601	mouse	MBL International	ChIP (15 µl per sample) Western blot (1:1000)
α-S5P (3E8)	rat	Millipore	ChIP (15 µl per sample) Western blot (1:1000)
α-S2P (3E10)	rat	Millipore	ChIP (15 µl per sample) Western blot (1:1000)
α-Y1P (3D12)	rat	Active Motif	ChIP (15 µl per sample) Western blot (1:1000)

α-FLAG M2-HRP	mouse	Sigma	Western blot (1:5000)
Peroxidase-α-peroxidase (PAP)	rabbit	Sigma	Western blot (1:2000)
α-mouse	goat	Sigma	Western blot (1:10000)
α-rabbit	goat	Sigma	Western blot (1:10000)
α-rat	goat	Calbiochem	Western blot (1:10000)

6.2 *S. pombe* growth conditions and genetic manipulations

6.2.1 Media and growth conditions

Yeast strains that do not contain *no message in thiamine 1* (NMT) promoter were growing in Yeast Extract with Supplements (YES) media (5 g/L yeast extract, 30 g/L glucose, 0.225 g/L adenine, histidine, leucine, uracil and lysine) at 30°C.

Yeast strains with NMT promoter were grown in Edinburgh minimal medium with glutamate (EMMG) (3 g/L potassium hydrogen phthalate, 2.2 g/L Na_2HPO_4, 5 g/L NH_4Cl, 20 g/L glucose, 3.7 g/L glutamic acid monosodium salt hydrate, 1.05 g/L $MgCl_2.6H_2O$, 14.7 mg/L $CaCl_2.2H_2O$, 1 g/L KCl, 40 mg/L Na_2SO_4, 1 mg/L pantothenic acid, 10 mg/L nicotinic acid, 10 m/L inositol, 10 μg/L biotin) at 30°C and transferred to YES, letting them grow for 17 h at 30°C before harvesting.

Strains carrying the temperature sensitive *dhp1-154* allele (H229R) were grown at 30°C and shifted to 37°C 2 h before harvesting.

Strains carrying cold-sensitive allele *dis3-54* (P509L) were grown at 30°C and shifted to 25°C 2 h before harvesting.

prp10-1 temperature-sensitive splicing mutant were grown at 25°C and shifted to 37°C 2 h before harvesting.

6.2.2 Strain construction and transformations

Endogenous tagging/deletion of proteins, or integration of NMT promoter, was achieved by the two-step PCR method described by Bähler et al. (Bähler *et al.*, 1998). For the initial amplifications to generate homology regions, 250 ng of *S. pombe* genomic DNA was used as a template, with primers L3-L4 and L5-L6 (please refer to Table 6.3). Products were run on 1% agarose gel at 160V, and the fragments with the desired length were cut from the gel. ZymoClean Gel DNA Recovery kit (Zymo Research) was used to gel-extract the fragments. For the second PCR to obtain the tagging cassette, 30 ng of the isolated L3-4 and L5-6 fragments along with 50 ng of pFA6-KanMX vector were mixed with the PCR reaction containing L3 and L6 primers. The product was run on a 1% agarose gel, and the transformation cassette of the correct length was extracted from the gel.

4 ml-culture of the strain to be transformed was grown to OD_{600} = 0.8 and pelleted. Cells were washed twice in LiTE buffer (0.1 M LiAcetate, 1X Tris-EDTA). The transformation cassettes were mixed with single-stranded DNA and 100 µl of cells in LiTE buffer was added onto the mix. Following a 10 min-incubation at room temperature, 260 µl of PLATE buffer (40% PEG, 0.1 M LiAcetate, 1X Tris-EDTA) was added onto the cells, and the mix was incubated at 30°C for 1 h. DMSO was added to a final concentration of 10%. Heat shock was performed at 42°C for 5 min followed by an incubation on ice for 1.5 min. Cells were pelleted, resuspended in YES medium, grown overnight at 30°C shaker and streaked onto the appropriate selection plate.

6.2.3 Crosses, tetrad dissections and random spore analyses

h+ and h- homothallic strains to be crossed were streaked on EMMG plates. They were then mixed together on a fresh EMMG plate and let grown at 30°C incubator for two to three days. Once tetrads started to appear, a streak was made on a YES plate and tetrads were transferred to individual spots using a dissection microscope (Singer Instruments). After 5-6 h, burst-out spores were picked and dissected. After a week in 30°C incubator, spores were replica-plated to select for antibiotics resistance or auxotrophies.

For random spore analyses, a little amount from the mixed strains was added to 100 µl of autoclaved water and 2 µl of 20 mg/ml cytohelicase (Sigma) was added. Following day, 1% SDS was added in a 1:1 volume ratio. Mix was vortexed rigorously and spores were counted using a Thoma slide. Dilutions were made such that 300 spores were plated on each EMMG or YES plate. Once the colonies became visible, plates were replica-plated to select for antibiotics resistance or auxotrophies.

6.2.4 Spot tests and growth curves

For spot tests, strains were inoculated in 10 ml of EMMG or YES until they reach OD_{600} = 0.6-0.7. In the case of NMT-containing strains, two inoculations were grown in EMMG (non-depletion condition) and YES (depletion condition). They were then pelleted at 3500 rpm for 4 min and washed once with autoclaved water. Then, the pellets were diluted such that the undiluted spot would contain 1.5 x 10^5 cells. Further dilutions of 1:10, 1:100, 1:1000 and 1:10000 were made for each strain. They were then spotted on either EMMG or YES plates as 5 µl per spot and were grown at varying temperatures, as indicated in the figures. The plates were imaged each day until the 1:10000 spot from the wild type strain grew fully.

For growth curves, overnight inoculations either in EMMG or YES were diluted in the morning to OD_{600} = 0.2. OD_{600} measurements were taken every 2 h until they reached OD_{600}= 0.8.

6.3 Biochemistry techniques

6.3.1 RNA extraction

50 ml cultures of cells were grown to log phase (OD_{600}=0.7) and pelleted at 3500 rpm for 4 min. The cells were washed once with autoclaved water, and twice in 1X AE buffer (50 mM sodium acetate pH 5.0, 10 mM EDTA, 1% w/v SDS). They were resuspended in 500 µl of ice-cold 1X AE buffer, and SDS was added to a final concentration of 1.2%. 500 µl of phenol:chloroform:isoamyl alcohol pH 6.0 was added. The samples were vortexed for

30 sec and incubated at 65°C for 4 min followed by a 4 min incubation on ice. The mixture was then centrifuged down at 14000 rpm for 10 min in a cold (4°C) centrifuge to separate the aqueous phase containing RNA. The mentioned phase was then taken to a new 1.5 ml tube, and another equal volume of phenol:chloroform:isoamyl alcohol pH 6.0 was added. After vortexing for 20 sec, the mixture was centrifuged as before and the aqueous phase was taken.

To concentrate the RNA, ethanol precipitation method was used. Briefly, 1:10 volume of 3 M sodium acetate pH 5.2 and 2.5-fold volume 100% ice-cold ethanol were added. The tubes were inverted several times and incubated at -20°C for 1 h. Samples were then centrifuged at max speed for 15 min to pellet the RNA precipitate, washed once with 70% ice-cold ethanol and air-dried for 4 min. The pellets were resuspended in ultrapure water with a volume depending on their size.

For accurate determination of RNA concentrations, two dilutions (usually a 1:10 and a 1:20) were prepared from each sample and each dilution were measured twice using NanoDrop (Thermo Scientific). Concentrations were determined by taking the average of all measurements.

6.3.2 Protein extraction (TCA method)

For small-scale protein isolation, 5 ml-cultures grown overnight at 30°C shaker were pelleted and lysed in 1.85 N NaOH (Sigma) containing 1% β-mercaptoethanol (Sigma). 50% trichloroacetic acid (Sigma) was added to precipitate proteins. Samples were incubated on ice for 30 min and centrifuged down. Supernatant was discarded completely. The pellet was then dissolved in 2X laemmli buffer (Laemmli, Beguin and Gujer-Kellenberger, 1970) (24 mM Tris-Cl pH6.8, 14% glycerol, 0.004% bromophenol blue, 0.8% SDS, 2% β-mercaptoethanol) and 100 mM Tris (pH not adjusted) and boiled at 95°C for 10 min before loading onto an SDS gel as described in section 6.3.3, or alternatively, storing at -20°C.

6.3.3 SDS polyacrylamide gel

10-12% SDS PAA gels were prepared with a preceding 4% stacking gel. Proteins were denatured at 95°C for 10 min with 2X laemmli buffer before loading. The Mini-PROTEAN system (Bio-Rad) was used to run the gels in SDS running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS). The gels were generally run at 85 V for 15 min to let the proteins get into the gel, then the voltage was increased to 115 V until the proteins reached the end of the stacking gel, followed by electrophoresis at 165 V until the bromophenol blue dye reached the end of the gel.

6.3.4 Coomassie and silver staining

After the proteins were run on an SDS gel, as described in section 6.3.3, Coomassie staining was carried out using InstantBlue (Expedeon) according to manufacturer's instructions.

For silver staining, denatured proteins and marker were loaded into NuPAGE Bis-Tris 15% gels (Invitrogen). Gels were run in 1xMOPS running buffer (50mM MOPS, 50 mM Tris-HCl pH 7.7, 3.5mM SDS, 1mM EDTA) at 160 V for 1h. Instructions for the SilverQuest Silver Staining Kit (Invitrogen) followed for the silver staining. Depending on the strength of the bands, incubation time for the developing step was adjusted, from 4 min to 8 min. Reaction was subsequently stopped using the provided stop solution.

6.3.5 Western blotting

After the proteins were run on an SDS gel, as described in section 6.3.3, wet transfer was performed to blot the proteins onto Immobilon-P polyvinylidene fluoride (PVDF) membrane (Millipore), with constant current at 380 mA for 1 hr and 30 min, in 1X transfer buffer (25 mM Tris, 192 mM glycine, 1 mM EDTA, 20% methanol). Membrane was blocked in 5% skim milk (Sigma) in PBS-T (137 mM NaCl, 12 mM sodium phosphate pH 7.4, 2.7 mM KCl, 0.1% Tween-20) for 45 min at room temperature on a shaker. Overnight incubation with the primary antibody (Table 6.4) at 4°C was followed by horseradish

peroxidase-conjugated secondary antibody incubation for 2 h at room temperature, on the shaker. Visualization was performed using Clarity ECL Substrate (Bio-Rad).

6.3.6 Northern blotting

15 µg of total RNA extracted as described in section 6.3.1 was denatured at 65°C for 20 min and run on a 1.2% w/v agarose formaldehyde gel at 120 V for 3 h. The gel was then washed in 10 mM NaOH for 20 min on a shaker, rinsed with ultrapure water and washed with 20X SSC (3 M NaCl, 0.3 M sodium citrate pH 7.0) for 5 min and 10X SSC for 40 min. Capillary transfer of RNA from the gel to Hybond N+ membrane (Amersham) was carried out overnight. The next day, the RNA-blotted membrane was subjected to 120 mJ of UV irradiation using the Stratagene UV crosslinker. The membrane was then incubated in 3% glacial acetic acid (Sigma) for 10 min and stained with 0.03% methylene blue in 0.5 M sodium acetate pH 5.2 for 1-2 min, until the rRNA bands became visible. The background was washed away using MilliQ water; positions of the rRNA bands were marked with a pencil, and the membrane was imaged in a GelDoc system (Bio-rad). The membrane was rolled and placed in a hybridization tube and incubated in Pre-hybridization buffer (0.3 M sodium phosphate buffer pH 7.2, 1 g/ml BSA, 7% SDS, 1 mM EDTA) for 2 h at 42°C hybridization oven.

Probes that were previously amplified and gel-extracted as described in 2.4.1 using primers listed in Table 6.1, were primed using Prime-It II Random Primer Labeling Kit according to manufacturer's instructions. Briefly, 50 ng of probe was incubated with 10 µl of random hexamer primers at 95°C for 5 min, mixed with dATP buffer and 1 µl Klenow enzyme. In the hot room, 3 µl of α-P^{32}-dATP was added and the mixture was incubated in 37°C heat block for 30 min. The unincorporated radioactive nucleotides were washed away using ProbeQuant G-50 micro columns (GE Life Sciences).

The pre-hybridization buffer in the tube was replaced with 20 ml of the same buffer, 400 µl of ssDNA and the radioactively-labelled probe was added. The probe hybridization was performed in the 42°C hybridization oven overnight.

The membrane was then washed several times with 2X SSC and 0.1% SDS. Signal from the membrane was checked using a Geiger counter and the washes were continued until the signal was localized over a specific area on the membrane. The initial visualization was performed on an FLA-7000 Phosphorimager (Fujifilm). Longer exposures at -80°C were taken using Amersham Hyperfilm MP films and developed in the SRX-101A system (Konica Minolta).

6.3.7 Protein purification from yeast and mass spectrometry

6.3.7.1 Yeast grinding

Per pulldown, 2 L of yeast cultures were grown to the log phase (OD_{600}=0.7). After washing with water twice, cells were pelleted and resuspended in TMN buffer (20mM Tris-HCl pH 8.0, 5mM $MgCl_2$, 150mM NaCl, 10% glycerol) supplemented with protease and phosphatase inhibitors. Dropwise addition of the completely-resuspended yeast cells into a 50 ml-falcon tube submerged in liquid nitrogen was performed to obtain yeast "beads". Beads were then ground for 1 h using a mortar and pestle, with the regular cooling by liquid nitrogen. Once finely ground, the powder was transferred to a 50 ml-falcon and kept at -80°C until the time of purification.

6.3.7.2 Protein A purification and TEV cleavage

Yeast powders were let to melt, and volume was increased to 10 ml by the addition of TMN buffer with protease and phosphatase inhibitors. After centrifugation at 20000 g for 30 min, supernatant residing between the chromatin pellet and the fat layer was taken into a new tube. A 1:1 mixture of IgG sepharose (GE Healthcare) and rabbit IgG agarose slurry (Sigma) was added and the samples were incubated overnight to facilitate Protein A binding to the IgG beads. The next morning, samples were centrifuged down, washed twice with TMN buffer followed by another two washes with TEV buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM DTT, 0.05% NP-40). After the last wash, the samples were resuspended in the same volume of TEV buffer as the bead slurry volume used.

AcTEV protease (Thermo Fisher) was added and TEV cleavage carried out overnight at 4°C on a rotator.

6.3.7.3 FLAG purification

Samples were loaded onto micro bio-spin columns (Bio-rad) and centrifuged down to separate the supernatant from the bead slurry. Anti-FLAG M2 affinity gel resin (Sigma) was washed twice with TEV buffer. 60 µl of resin was used per 2 L of starting culture. Samples were incubated with the resin on a rotator for 2 h at 4°C. After extensive washes with the FLAG Base Buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% glycerol), elution was performed in the Elution Buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, 0.5 mM DTT, 0.5 mM $MgCl_2$, 0.5 mM $Mg(OAc)_2$) supplemented with 3X FLAG peptide (final concentration: 1 mg/ml) with rotation at 4°C for 30 min.

The elutions were subsequently subjected to tryptic digestion and 1 h LC-MS/MS Orbitrap by the Central Proteomics Facility at the Department of Biochemistry, University of Oxford.

6.4 Molecular biology techniques

6.4.1 Polymerase Chain Reaction (PCR)

PCR mixtures were prepared in a volume of 50 µl with 0.5 µM forward and reverse primers (Table 6.1), 0.2 mM dNTP mix, 1.5 mM $MgCl_2$, and 1 U Phire Hot Start II (Thermo Fisher) or Phusion HF (NEB), into which 250 ng of wt genomic DNA was added. Initial denaturation (5 min) and denaturation (30 sec) steps were carried out at 95°C (Phire) or 98°C (Phusion), and extension (30 sec) and final extension (7 min) steps were done at 72°C. The annealing temperatures were chosen with respect to the melt temperatures of the primers (Table 6.1). The reactions were performed in a thermal cycler (Bio-Rad). 50 µl of PCR product was mixed with 10 µl of 6X loading dye and run on a 1-1.5% agarose gel at 160V; images were taken under UV light.

6.4.2 Quantitative Real-time PCR

2 µg of yeast total RNA, prepared as described in section 6.3.1, was subjected to DNase RQ1 (Promega) treatment for 1 h at 37°C and the reaction was terminated using the provided stop solution. iScript Reverse Transcription kit (Bio-rad) was used to convert 100 ng of DNase-treated RNA to cDNA using random priming. 15 µl-reactions were prepared in duplicates for each assayed cDNA with Sensimix SYBR (No-ROX) Kit (Bioline). For each primer set, a standard curve was set up alongside using *S. pombe* wt cDNA with four dilutions to estimate the cDNA amount in experimental samples. Samples were run on Rotor-Gene Q thermal cycler (Qiagen) and cycle threshold values were calculated in Rotor- Gene Q software.

6.4.3 Chromatin Immunoprecipitation (ChIP)

The protocol used for ChIP experiments and summarized here was adapted from (Keogh, Podolny and Buratowski, 2003). Yeast cultures were grown to log phase in 200 ml and crosslinked with 11% (v/v) formaldehyde for 20 min, swirling the flasks every 2 min to ensure equal distribution of the crosslinker. 30 ml of Tris-Glycine buffer (20 mM Tris, 3 M glycine) was added to quench the reaction and cells were centrifuged down at 3500 rpm for 4 min at 4°C. Pellets were washed once with ice-cold 1X TBS (50 mM Tris-Cl pH 7.5, 150 mM NaCl) and FA lysis buffer (50 mM HEPES, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate) supplemented with 0.1% SDS and 1 mM PMSF.

The following day, pellets were resuspended in FA lysis buffer with 0.5% SDS and 1.5 ml of acid-washed glass beads (Sigma) was added. Cells were then vortexed for 30 cycles of 30 sec vortexing and 30 sec on ice. Supernatant was transferred to an ultracentrifuge tube and spun down at 120000 g for 20 min. The pellet was crushed using a wooden stick and resuspended in FA lysis buffer + 0.1% SDS. After another ultracentrifugation and crushing of the pellet, 1.6 ml of FA lysis buffer + 0.1% SDS was added and the samples were sonicated for 1 h and 20 min with a cycle of 15 sec on / 45 sec off using a Bioruptor (Diagenode). Sonicated samples were centrifuged at 16400 g for 20 min at 4°C and the

chromatin fraction in the supernatant was taken to a new tube. Concentrations were measured using NanoDrop, and the final concentration of NaCl was upped to 275 mM. 50 µl was taken as input from each sample.

Dynabeads Protein G (Thermo Fisher) for pol II ChIPs (see Table 6.4) or rabbit IgG agarose (Sigma) for pulldowns of proteins containing Protein A, were washed with 1X TE (10 mM Tris-Cl pH 8.0, 1 mM EDTA). Antibodies were conjugated in amounts stated in Table 6.4 for 2 h at 4°C rotator, and 10 µl of beads was added per 1 ml of sheared chromatin. After overnight incubation on a 4°C rotator, beads were washed extensively. Immunoprecipitated DNA fragments were eluted and de-crosslinked by incubation with 20 µg of pronase for 1 h at 42°C followed by overnight incubation at 65°C.

If the ChIP will be followed up by qPCRs, eluates were extracted using phenol:chloroform:isoamyl alcohol (25:24:1) and ethanol-precipitated. If the next step would be next-generation sequencing, the protocol was followed up as described in section 6.4.5.1.

6.4.4 polyA$^+$ RNA Fluorescence *in situ* Hybridization (FISH)

10 ml inoculations of strains were grown to log phase and crosslinked with 4% v/v formaldehyde for 30 min at RT. The cells were then centrifuged at 3500 rpm for 5 min and washed twice with ice-cold Buffer B (1.2 M sorbitol, 100 mM $KHPO_4$ pH 7.5). Cells were resuspended in spheroplasting buffer (1.2 M sorbitol, 100 mM $KHPO_4$ pH 7.5, 20 mM ribonucleoside vanadyl complex (RVC, Sigma), 0.5 mg/ml zymolyase 100T (MP Biomedicals), 0.2% (v/v) β-mercaptoethanol) and incubated at 37°C heat block for 1 h. After making sure that a spheroplasting efficiency of ~80% was reached, the reaction was stopped by washing with Buffer B. Spheroplasts were incubated in 1X PBS containing 0.01% Triton X-100 for 20 min. A small aliquot of cells was taken for RNase-treated controls. This aliquot was resuspended in 2X SSC and RNase A was added to 100 µg/ml final concentration. Treatment was carried out at 37°C for 1 h, and cells were subsequently washed with 2X SSC with 10% formamide.

25 ng of Alexa488-conjugated Oligo-d(T)$_{20}$ probe (made by Invitrogen) was mixed with 10 mg/ml yeast tRNA and 5 mg/ml ssDNA in a 1:2 ratio and dried in a SpeedVac vacuum concentrator (Thermo Fisher). The lyophilized mixture was resuspended in Hybridization Buffer F (10 mM $NaHPO_4$ pH 7.0, 20% formamide) and incubated for 3 min at 95°C. Equal volume of Buffer H (4X SSC, 4 mg/ml BSA, 20 mM RVC) was added. After the last wash, the cells were resuspended in the probe mix and incubated at 30°C shaker overnight.

The next day, samples were successively washed with 2X SSC + 10% formamide, 2X SSC + 0.1% Triton X-100 and 1X PBS. Nuclei were stained with DAPI (1:1000 in 1X PBS) for 10 min and excess DAPI was washed off with 1X PBS. Cells were subsequently mounted on poly-L-lysine-coated No: 1.5 coverslips (Deckgläser Precision Cover Slips, Marienfeld) and imaged in UltraView spinning-disc confocal microscope.

6.4.5 Library preparation

6.4.5.1 ChIP-seq DNA libraries

ChIP inputs and eluates, prepared as described in section 6.4.3, were concentrated using ChIP DNA Clean & Concentrator (Zymo Research) and eluted in 10 µl MilliQ water for inputs and 20 µl for eluates. Concentrations were determined with Qubit (Thermo Fisher) and dilutions were made such that starting material for library prep is 10-25 ng of DNA.

DNA libraries were prepared using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). NEBNext Multiplex Oligos for Illumina Index Primers Sets 1-4 (NEB) were utilized to barcode the libraries. Finished libraries were run on a DNA High Sensitivity chip (Agilent) in Bioanalyzer to assess size distribution and concentrations.

6.4.5.2 RNA-seq cDNA libraries

4 µg of RNA extracted as described in section 6.3.1 was first treated with DNase RQ1 (Promega) treatment for 1 h at 37°C in 200 µl final volume, followed by ethanol precipitation in the presence of 1 µl glycogen for 2 h at -20°C. After resuspension in MilliQ

water, the concentration was measured and 2.5 μg of RNA was subjected to ribodepletion using Ribo-Zero Yeast Gold Kit (Illumina), according to manufacturer's instructions. The samples before and after DNase RQ1 treatment and ribodepletion were run on an Bioanalyzer RNA chip (Agilent) to check the RNA integrity number (RIN) and to confirm the depletion of rRNA peaks.

cDNA libraries were prepared using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB). NEBNext Multiplex Oligos for Illumina Index Primers Sets 1-4 (NEB) were utilized to barcode the libraries. Finished libraries were run on a DNA High Sensitivity chip (Agilent) in Bioanalyzer to assess size distribution and concentrations.

6.5 Bioinformatics analysis

6.5.1 Sequence alignment

Paired-end reads from each of the four lanes in the NextSeq 500 (Illumina) flow cell were concatenated with respect to the read number. Low-quality and shorter reads were trimmed using Trimmomatic 0.36 (Bolger, Lohse and Usadel, 2014) using the following command:

```
java -jar trimmomatic.jar PE -phred33 fwd.fastq.gz rev.fastq.gz
output_forward_paired.fq.gz output_forward_unpaired.fq.gz
output_reverse_paired.fq.gz output_reverse_unpaired.fq.gz
SLIDINGWINDOW:4:20 MINLEN:36
```

FastQC (Andrews and Babraham Bioinformatics, 2010) was utilized to assess the quality of reads before and after trimming. The sequence alignment of the trimmed reads were done either using Bowtie2 (Langmead and Steven L Salzberg, 2013) or HISAT2 (Kim, Langmead and Salzberg, 2015) for ChIP-seq or RNA-seq libraries, respectively.

For Bowtie2 alignments, a concatenated index including *S. pombe* and *S. cerevisiae* genomes was used. Alignment was performed using standard settings and not allowing any discordant alignments ("--no-discordant").

HISAT2 alignments were carried out on Galaxy Server (Afgan *et al.*, 2016) using the standard options with adjusted maximum intron length ("-max-intronlen") to 5000 bases. For these alignments, separate index files were used for *S. pombe* (Schizosaccharomyces_pombe_1.1) and the spike-in genome (sacCer3).

6.5.2 ChIP-seq analysis

DeepTools package (Ramírez *et al.*, 2014) was utilized to analyse ChIP-seq alignments. bamCoverage function was used to tidy the data with bin size ("-bs") of 10, and the normalization factor calculated using spike-in reads was entered using the "--scaleFactor" option. Subsequently, computeMatrix scale-regions function was applied with the following options: "-m 1000 -b 500 -a 500 -bs 10". Finally, plotProfile function was used with the option "--outFileNameData" to obtain a tab-delimited file of the values, which was then plotted using ggplot2 (Ginestet, 2011) in R.

6.5.3 Differential expression analysis of RNA-seq alignments

Feature counts were obtained for coding, non-coding genes, cryptic unstable transcripts or antisense transcripts using HTseq (Anders, Pyl and Huber, 2015) with the S. pombe annotation provided by Cramer Lab (Eser *et al.*, 2016). Differential expression analyses of RNA-seq data was carried out with the DESeq2 package (Love, Huber and Anders, 2014) in R. Libraries were normalized to spike-in reads using the sizeFactors() function. After running the DESeq analysis to obtain the significantly changing genes, annotation names for the genes were extracted by backreferencing the gene identifiers to Eser annotation.

6.5.4 RNA polymerase II processivity measurement

To infer the speed by which RNA pol II moves along a transcription unit, the following script was written in R. Briefly, it calculates the mean signal in two bin windows ("leadwindow" and "lagwindow") separated by a given number of bins ("offset"), obtains the ratio between these mean values until it reaches the end of the transcription unit and plots the ratios. By definition, the script takes the derivative of the metagene function of a

total RNA pol II ChIP-seq experiment; therefore, the slopes of the metagene infer a relative "processivity speed", which can be compared to the processivity speed of pol II in the wt strain. The R script for the processivity measurements is given below:

```
## Dogukan Ulgen, February 2018
## Calculates polII processivity over the gene body from polII ChIP-seq
data.
## To start, you need to run computeMatrix scale-regions with your wt
and mutant (in that order) with -b 500, -a 500, -m 1000 and -bs 10.


library("reshape2")
library("plyr")
library("ggplot2")

setwd("/Users/dogukanulgen/Documents/Oxford/Lab/ChIP-seq/2017-10-05 wt
vs cbc3delta polII ChIP 2nd/Analysis/")

file <- "Cbc32nd.tab"
Rep2 <- read.table(file, header = TRUE, sep="\t")  # Read from the .tab
file obtained from plotHeatmap --outFileNameMatrix.
Rep <- cbind(Rep1, Rep2)


slidelength <- 10 ## Number of bins each window has.
offset <- 50 ## Number of bins between windows.

## Using the slidelength and offset values, this bit calculates the
numbers to be used in processivity measurements.
leadwindowstart <- 50 + offset + slidelength
leadwindowend <- leadwindowstart + slidelength
lagwindowstart <- 50
lagwindowend <- 50 + slidelength
totalwindows <- 150 - leadwindowend

list <- mat.or.vec(2,totalwindows) ## An empty matrix is generated, to
be filled with measurements later on.

for(i in 0:totalwindows) {  ## This is where the measurements are made
and written to the list.
  list[1,i] <-
mean(colMeans(Rep[,(leadwindowstart+i):(leadwindowend+i)])) /
mean(colMeans(Rep[,(lagwindowstart+i):(lagwindowend+i)]))
  list[2,i] <-
mean(colMeans(Rep[,(leadwindowstart+200+i):(leadwindowend+200+i)])) /
mean(colMeans(Rep[,(lagwindowstart+200+i):(lagwindowend+200+i)]))
}

## ------------------------
# If you want to combine replicates, use this bit!
# listRep1 <- t(list)
# listRep2 <- t(list)
# listAvg <- cbind(listRep1, listRep2)
# listAvg <- as.data.frame(listAvg)
```

```
#
# for(i in 1:totalwindows) {
#   listAvg[i,5] <- (listAvg[i,1] + listAvg[i,3]) / 2
#   listAvg[i,6] <- (listAvg[i,2] + listAvg[i,4]) / 2
# }
#
# listAvg <- listAvg[,5:6]
#
# list <- listAvg
# list <- as.data.frame(list)
# colnames(list) <- c("wt","cbc3d")
#
# list$x <- c(0:(totalwindows-1))
# meltme <- melt(list, id = "x")
#
# g <- ggplot(meltme,aes(x = x, y = as.numeric(value), color =
variable, group = variable)) +geom_point(size=2, shape=19) +
ggtitle(paste("polII processivity, wt vs. ssu72,
Slidelength=",slidelength ,", Offset=", offset)) + labs(x = "Window", y
= "Processivity")
# g + theme_linedraw(base_size = 24) + theme(panel.grid =
element_blank(), panel.border = element_rect(size = 1.0))  +
theme(axis.ticks = element_line(size = 1.2), axis.title =
element_text(size = 20)) +theme(plot.title = element_text(hjust = 0.5,
size=12)) + scale_colour_brewer(type = "qual", palette="Set1") +
theme(legend.title = element_blank(), legend.key.size = unit(1,"line"),
legend.text=element_text(size=16))

## ------------------------


## Finally, tidy up the data and plot using ggplot2.
list <- t(list)
list <- as.data.frame(list)
colnames(list) <- c("wt","cbc3delta")

list$x <- c(0:(totalwindows-1))
meltme <- melt(list, id = "x")

g <- ggplot(meltme,aes(x = x, y = as.numeric(value), color = variable,
group = variable)) +geom_point(size=2, shape=19) + ggtitle(paste("polII
processivity, polII sig. down, Slidelength=",slidelength ,", Offset=",
offset)) + labs(x = "Window", y = "Processivity")
g + theme_linedraw(base_size = 24) + theme(panel.grid =
element_blank(), panel.border = element_rect(size = 1.0))  +
theme(axis.ticks = element_line(size = 1.2), axis.title =
element_text(size = 20)) +theme(plot.title = element_text(hjust = 0.5,
size=12)) + scale_colour_brewer(type = "qual", palette="Set1") +
theme(legend.title = element_blank(), legend.key.size = unit(1,"line"),
legend.text=element_text(size=16))
```

Chapter 7 References

Afgan, E. *et al.* (2016) 'The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update', *Nucleic acids research*. doi: 10.1093/nar/gkw343.

Alexander, R. D. *et al.* (2010) 'Splicing-Dependent RNA polymerase pausing in yeast', *Molecular Cell*. doi: 10.1016/j.molcel.2010.11.005.

Allmang, C. *et al.* (1999) 'Functions of the exosome in rRNA, snoRNA and snRNA synthesis', *EMBO Journal*. doi: 10.1093/emboj/18.19.5399.

Anders, S., Pyl, P. T. and Huber, W. (2015) 'HTSeq-A Python framework to work with high-throughput sequencing data', *Bioinformatics*. doi: 10.1093/bioinformatics/btu638.

Andersen, P. R. *et al.* (2013) 'The human cap-binding complex is functionally connected to the nuclear RNA exosome', *Nature Structural and Molecular Biology*. Nature Publishing Group, 20(12), pp. 1367–1376. doi: 10.1038/nsmb.2703.

Andrews, S. and Babraham Bioinformatics (2010) 'FastQC: A quality control tool for high throughput sequence data', *Manual*. doi: citeulike-article-id:11583827.

Baejen, C. *et al.* (2014) 'Transcriptome Maps of mRNP Biogenesis Factors Define Pre-mRNA Recognition', *Molecular Cell*, 55(5), pp. 745–757. doi: 10.1016/j.molcel.2014.08.005.

Bähler, J. *et al.* (1998) 'Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe', *Yeast*, 14(10), pp. 943–951. doi: 10.1002/(SICI)1097-0061(199807)14:10<943::AID-YEA292>3.0.CO;2-Y.

Balatsos, N. A. A. *et al.* (2006) 'Inhibition of mRNA deadenylation by the nuclear cap binding complex (CBC)', *Journal of Biological Chemistry*, 281(7), pp. 4517–4522. doi:

10.1074/jbc.M508590200.

Bayne, E. H. *et al.* (2008) 'Splicing factors facilitate RNAi-directed silencing in fission yeast', *Science*. doi: 10.1126/science.1164029.

Bitton, D. A. *et al.* (2015) ' Identification of New Players in Cell Division, DNA Damage Response, and Morphogenesis Through Construction of Schizosaccharomyces pombe Deletion Strains ', *G3: Genes|Genomes|Genetics*, 5(3), pp. 361–370. doi: 10.1534/g3.114.015701.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*. doi: 10.1093/bioinformatics/btu170.

Booth, G. T. *et al.* (2016) 'Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast', *Genome Research*, 26(6), pp. 799–811. doi: 10.1101/gr.204578.116.

Booth, G. T. *et al.* (2018) 'Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast', *Nature Communications*. doi: 10.1038/s41467-018-03006-4.

Br??s, V., Yoh, S. M. and Jones, K. A. (2008) 'The multi-tasking P-TEFb complex', *Current Opinion in Cell Biology*, pp. 334–340. doi: 10.1016/j.ceb.2008.04.008.

Brannan, K. *et al.* (2012) 'mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription', *Molecular Cell*. doi: 10.1016/j.molcel.2012.03.006.

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

Brueckner, F. and Cramer, P. (2008) 'Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase II translocation', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.1458.

Bühler, M. *et al.* (2008) 'TRAMP-mediated RNA surveillance prevents spurious entry of

RNAs into the Schizosaccharomyces pombe siRNA pathway', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.1481.

Carpy, A. *et al.* (2014) ' Absolute Proteome and Phosphoproteome Dynamics during the Cell Cycle of Schizosaccharomyces pombe (Fission Yeast) ', *Molecular & Cellular Proteomics*. doi: 10.1074/mcp.m113.035824.

Carrillo Oesterreich, F., Preibisch, S. and Neugebauer, K. M. (2010) 'Global analysis of nascent RNA reveals transcriptional pausing in terminal exons', *Molecular Cell*. doi: 10.1016/j.molcel.2010.11.004.

Castello, A. *et al.* (2012) 'Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins', *Cell*. doi: 10.1016/j.cell.2012.04.031.

Chalamcharla, V. R. *et al.* (2015) 'Conserved factor Dhp1/Rat1/Xrn2 triggers premature transcription termination and nucleates heterochromatin to promote gene silencing', *Proceedings of the National Academy of Sciences*, 112(51), p. 201522127. doi: 10.1073/pnas.1522127112.

Chapman, R. D. *et al.* (2008) 'Molecular evolution of the RNA polymerase II CTD', *Trends in Genetics*. doi: 10.1016/j.tig.2008.03.010.

Chathoth, K. T. *et al.* (2014) 'A Splicing-Dependent Transcriptional Checkpoint Associated with Prespliceosome Formation', *Molecular Cell*. doi: 10.1016/j.molcel.2014.01.017.

Chen, F. X. *et al.* (2017) 'PAF1 regulation of promoter-proximal pause release via enhancer activation', *Science*. doi: 10.1126/science.aan3269.

Chen, Y. *et al.* (2009) 'DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation', *Genes and Development*. doi: 10.1101/gad.1834709.

Cheng, H. *et al.* (2006) 'Human mRNA Export Machinery Recruited to the 5′ End of mRNA', *Cell*. doi: 10.1016/j.cell.2006.10.044.

Chlebowski, A. *et al.* (2013) 'RNA decay machines: The exosome', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. doi: 10.1016/j.bbagrm.2013.01.006.

Cloutier, S. C. *et al.* (2012) 'The DEAD-box RNA helicase Dbp2 connects RNA quality control with repression of aberrant transcription', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M112.383075.

Connelly, S. and Manley, J. L. (1988) 'A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II.', *Genes & development*. doi: 10.1101/gad.2.4.440.

Core, L. J. and Lis, J. T. (2008) 'Transcription regulation through promoter-proximal pausing of RNA polymerase II', *Science*. doi: 10.1126/science.1150843.

Daneholt, B. (1997) 'A look at messenger RNP moving through the nuclear pore', *Cell*. doi: 10.1016/S0092-8674(00)81900-5.

Das, B., Butler, J. S. and Sherman, F. (2003) 'Degradation of Normal mRNA in the Nucleus of Saccharomyces cerevisiae', *Molecular and Cellular Biology*, 23(16), pp. 5502–5515. doi: 10.1128/mcb.23.16.5502-5515.2003.

Davis, C. A. and Ares, M. (2006) 'Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0507783103.

Dengl, S. and Cramer, P. (2009) 'Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in Vitro', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M109.013847.

Dufu, K. *et al.* (2010) 'ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex', *Genes and Development*, 24(18), pp. 2043–2053. doi: 10.1101/gad.1898610.

Edery, I. and Sonenberg, N. (1985) 'Cap-dependent RNA splicing in a HeLa nuclear extract.', *Proceedings of the National Academy of Sciences of the United States of America*, 82(22), pp. 7590–7594. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=3865180&ret

mode=ref&cmd=prlinks%5Cnpapers2://publication/uuid/426102BD-CE38-4687-99AB-B07F56C8A9C1.

Egan, E. D. *et al.* (2014) 'Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex', *Rna*, 20(6), pp. 867–881. doi: 10.1261/rna.044479.114.

Epshtein, V. *et al.* (2010) 'An allosteric mechanism of Rho-dependent transcription termination', *Nature*. doi: 10.1038/nature08669.

Eser, P. *et al.* (2016) 'Determinants of RNA metabolism in the Schizosaccharomyces pombe genome', *Molecular Systems Biology*. doi: 10.15252/msb.20156526.

Exinger, F. and Lacroute, F. (1992) '6-Azauracil inhibition of GTP biosynthesis in Saccharomyces cerevisiae', *Current Genetics*. doi: 10.1007/BF00351735.

Fischl, H. *et al.* (2017) 'Paf1 Has Distinct Roles in Transcription Elongation and Differential Transcript Fate', *Molecular Cell*, 65(4), pp. 685-698.e8. doi: 10.1016/j.molcel.2017.01.006.

Flaherty, S. M. *et al.* (1997) 'Participation of the nuclear cap binding complex in pre-mRNA 3' processing.', *Proceedings of the National Academy of Sciences of the United States of America*.

Fong, N. *et al.* (2015) 'Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition', *Molecular Cell*. doi: 10.1016/j.molcel.2015.09.026.

Forsburg, S. L. and Sherman, D. A. (1997) 'General purpose tagging vectors for fission yeast', *Gene*. doi: 10.1016/S0378-1119(97)00058-9.

Fortes, P. *et al.* (1999) 'Genetic and physical interactions involving the yeast nuclear cap-binding complex.', *Molecular and cellular biology*, 19(10), pp. 6543–6553.

Fresco, L. D. and Buratowski, S. (1996) 'Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing', *RNA society*.

Gardini, A. (2017) 'Global run-on sequencing (GRO-Seq)', in *Methods in Molecular*

Biology. doi: 10.1007/978-1-4939-4035-6_9.

Gebhardt, A. *et al.* (2015) ‘mRNA export through an additional cap-binding complex consisting of NCBP1 and NCBP3.’, *Nature communications*, 6, p. 8192. doi: 10.1038/ncomms9192.

Giacometti, S. *et al.* (2017) ‘Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate’, *Cell Reports*, 18(11), pp. 2635–2650. doi: 10.1016/j.celrep.2017.02.046.

Ginestet, C. (2011) ‘ggplot2: Elegant Graphics for Data Analysis’, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. doi: 10.1111/j.1467-985x.2010.00676_9.x.

Gonatopoulos-Pournatzis, T. and Cowling, V. H. H. (2014) ‘Cap-binding complex (CBC)’, *Biochemical Journal*, 457(2), pp. 231–242. doi: 10.1042/BJ20131214.

Görnemann, J. *et al.* (2005) ‘Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex’, *Molecular Cell*, 19(1), pp. 53–63. doi: 10.1016/j.molcel.2005.05.007.

Grenier St-Sauveur, V. *et al.* (2013) ‘Poly(A) Tail-Mediated Gene Regulation by Opposing Roles of Nab2 and Pab2 Nuclear Poly(A)-Binding Proteins in Pre-mRNA Decay’, *Molecular and Cellular Biology*. doi: 10.1128/mcb.00887-13.

Gu, W., Wind, M. and Reines, D. (2002) ‘Increased accommodation of nascent RNA in a product site on RNA polymerase II during arrest.’, *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.93.14.6935.

Gudipati, R. K. *et al.* (2012) ‘Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells’, *Molecular Cell*. doi: 10.1016/j.molcel.2012.08.018.

Hallais, M. *et al.* (2013) ‘CBC-ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing’, *Nature Structural and Molecular Biology*. Nature Publishing Group, 20(12), pp. 1358–1366. doi: 10.1038/nsmb.2720.

Hamm, J. and Mattaj, I. W. (1990) ‘Monomethylated cap structures facilitate RNA export

from the nucleus', *Cell*. doi: 10.1016/0092-8674(90)90292-M.

Hanyu, Y. *et al.* (2009) 'Genetic control of cellular quiescence in S. pombe', *Journal of Cell Science*. doi: 10.1242/jcs.046466.

Hardwick, S. W. and Luisi, B. F. (2013) 'Rarely at rest: RNA helicases and their busy contributions to RNA degradation, regulation and quality control', *RNA Biology*. doi: 10.4161/rna.22270.

Harigaya, Y. *et al.* (2006) 'Selective elimination of messenger RNA prevents an incidence of untimely meiosis', *Nature*. doi: 10.1038/nature04881.

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

Hartzog, G. A. and Fu, J. (2013) 'The Spt4-Spt5 complex: A multi-faceted regulator of transcription elongation', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. doi: 10.1016/j.bbagrm.2012.08.007.

Haruki, H., Nishikawa, J. and Laemmli, U. K. (2008) 'The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes', *Molecular Cell*. doi: 10.1016/j.molcel.2008.07.020.

Hayles, J. *et al.* (2013) 'A genome-wide resource of cell cycle and cell shape genes of fission yeast.', *Open biology*, 3(5), p. 130053. doi: 10.1098/rsob.130053.

Ho, C. K. and Shuman, S. (1999) 'Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme', *Molecular Cell*. doi: 10.1016/S1097-2765(00)80468-2.

Hoffman, C. S., Wood, V. and Fantes, P. A. (2015) 'An ancient yeast for young geneticists: A primer on the Schizosaccharomyces pombe model system', *Genetics*. doi: 10.1534/genetics.115.181503.

Hrossova, D. *et al.* (2015) 'RBM7 subunit of the NEXT complex binds U-rich sequences

and targets 3′-end extended forms of snRNAs', *Nucleic Acids Research*. doi: 10.1093/nar/gkv240.

Hsin, J. P. and Manley, J. L. (2012) 'The RNA polymerase II CTD coordinates transcription and RNA processing', *Genes and Development*. doi: 10.1101/gad.200303.112.

Huang, Y. J., Acton, T. B. and Montelione, G. T. (2014) 'DisMeta: A meta server for construct design and optimization', *Methods in Molecular Biology*, 1091, pp. 3–16. doi: 10.1007/978-1-62703-691-7_1.

Hujer, K. *et al.* (2016) 'Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II', *Cell*. Elsevier Inc., 165(2), pp. 372–381. doi: 10.1016/j.cell.2016.02.045.

Izaurralde, E. *et al.* (1994) 'A nuclear cap binding protein complex involved in pre-mRNA splicing', *Cell*, 78(4), pp. 657–668. doi: 0092-8674(94)90530-4 [pii].

Izaurralde, E. *et al.* (1995) 'A cap-binding protein complex mediating U snRNA export.', *Nature*, 376(6542), pp. 709–12. doi: 10.1038/376709a0.

J., B.-B. *et al.* (2003) 'The interaction of the cap-binding complex (CBC) with eIF4G is dispensable for translation in yeast', *Rna*, 9(6), pp. 654–662. doi: 10.1261/rna.5100903.

Järvelin, A. I. *et al.* (2016) 'The new (dis)order in RNA regulation', *Cell Communication and Signaling*, 14(1), p. 9. doi: 10.1186/s12964-016-0132-3.

Jeronimo, C. and Robert, F. (2014) 'Kin28 regulates the transient association of Mediator with core promoters', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.2810.

Jia, H. *et al.* (2011) 'The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex', *Cell*. doi: 10.1016/j.cell.2011.05.010.

Jiao, X. *et al.* (2013a) 'A Mammalian Pre-mRNA 5' End Capping Quality Control Mechanism and an Unexpected Link of Capping to Pre-mRNA Processing', *Molecular Cell*. Elsevier Inc., 50(1), pp. 104–115. doi: 10.1016/j.molcel.2013.02.017.

Jiao, X. *et al.* (2013b) 'A Mammalian Pre-mRNA 5' End Capping Quality Control

Mechanism and an Unexpected Link of Capping to Pre-mRNA Processing', *Molecular Cell*. doi: 10.1016/j.molcel.2013.02.017.

Kainov, Y. A. and Makeyev, E. V. (2020) 'A transcriptome-wide antitermination mechanism sustaining identity of embryonic stem cells', *Nature Communications*. doi: 10.1038/s41467-019-14204-z.

Katahira, J. (2015) 'Nuclear export of messenger RNA', *Genes*, 6(2), pp. 163–184. doi: 10.3390/genes6020163.

Kelley, L. A. *et al.* (2015) 'The Phyre2 web portal for protein modeling, prediction and analysis.', *Nature Protocols*. doi: 10.1038/nprot.2015.053.

Keogh, M.-C., Podolny, V. and Buratowski, S. (2003) 'Bur1 Kinase Is Required for Efficient Transcription Elongation by RNA Polymerase II', *Molecular and Cellular Biology*. doi: 10.1128/mcb.23.19.7005-7018.2003.

Kierzkowski, D. *et al.* (2009) 'The Arabidopsis CBP20 targets the cap-binding complex to the nucleus, and is stabilized by CBP80.', *The Plant journal : for cell and molecular biology*, 59(5), pp. 814–825. doi: 10.1111/j.1365-313X.2009.03915.x.

Kilchert, C. *et al.* (2015) 'Regulation of mRNA Levels by Decay-Promoting Introns that Recruit the Exosome Specificity Factor Mmi1', *Cell Reports*. Elsevier Ltd, 13(11), pp. 2504–2515. doi: 10.1016/j.celrep.2015.11.026.

Kim, D., Langmead, B. and Salzberg, S. L. (2015) 'Hisat2', *Nature methods*. doi: 10.1038/nmeth.3317.

Kim, J., Guermah, M. and Roeder, R. G. (2010) 'The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS', *Cell*. doi: 10.1016/j.cell.2009.12.050.

Kim, M., Krogan, N. J., *et al.* (2004) 'The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II', *Nature*. doi: 10.1038/nature03041.

Kim, M., Ahn, S. H., *et al.* (2004) 'Transitions in RNA polymerase II elongation

complexes at the 3′ ends of genes', *EMBO Journal*. doi: 10.1038/sj.emboj.7600053.

Komarnitsky, P., Cho, E. J. and Buratowski, S. (2000) 'Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription', *Genes and Development*. doi: 10.1101/gad.824700.

Krebber, H. *et al.* (2014) 'Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1', *Nature Communications*. Nature Publishing Group, 5(1), pp. 1–14. doi: 10.1038/ncomms4123.

De La Mata, M. *et al.* (2003) 'A slow RNA polymerase II affects alternative splicing in vivo', *Molecular Cell*. doi: 10.1016/j.molcel.2003.08.001.

LaCava, J. *et al.* (2005) 'RNA degradation by the exosome is promoted by a nuclear polyadenylation complex', *Cell*. doi: 10.1016/j.cell.2005.04.029.

Laemmli, U. K., Beguin, F. and Gujer-Kellenberger, G. (1970) 'A factor preventing the major head protein of bacteriophage T4 from random aggregation', *Journal of Molecular Biology*. doi: 10.1016/0022-2836(70)90402-X.

Lahudkar, S. *et al.* (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*. doi: 10.1093/nar/gkq1029.

Laitem, C. *et al.* (2015) 'CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.3000.

Langmead and Steven L Salzberg (2013) 'Bowtie2', *Nature methods*. doi: 10.1038/nmeth.1923.Fast.

Larochelle, M. *et al.* (2018) 'Common mechanism of transcription termination at coding and noncoding RNA genes in fission yeast', *Nature Communications*. Springer US, 9(1). doi: 10.1038/s41467-018-06546-x.

Lemay, J. F. *et al.* (2010) 'The Nuclear Poly(A)-Binding Protein Interacts with the

Exosome to Promote Synthesis of Noncoding Small Nucleolar RNAs', *Molecular Cell*. doi: 10.1016/j.molcel.2009.12.019.

Lemay, J. F. *et al.* (2014) 'The RNA exosome promotes transcription termination of backtracked RNA polymerase II', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.2893.

Lenasi, T., Peterlin, B. M. and 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), pp. 22758–22768. doi: 10.1074/jbc.M111.235077.

Lewis, J. D. *et al.* (1996) 'A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site', *Genes and Development*, 10(13), pp. 1683–1698. doi: 10.1101/gad.10.13.1683.

Lewis, J. D., Görlich, D. and Mattaj, I. W. (1996) 'A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing.', *Nucleic acids research*, 24(17), pp. 3332–6. doi: 10.1093/nar/24.17.3332.

Lidschreiber, M., Leike, K. and 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*. doi: 10.1128/mcb.00361-13.

Liu, Q., Greimann, J. C. and Lima, C. D. (2006) 'Reconstitution, Activities, and Structure of the Eukaryotic RNA Exosome', *Cell*. doi: 10.1016/j.cell.2006.10.037.

Logan, J. *et al.* (2006) 'A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene.', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.84.23.8306.

Love, M. I., Huber, W. and Anders, S. (2014) 'DESeq2', *Genome Biol*. doi: 10.1186/s13059-014-0550-8.

Lu, H. *et al.* (2006) 'The nonphosphorylated form of RNA polymerase II preferentially

associates with the preinitiation complex.', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.88.22.10004.

Lubas, M. *et al.* (2011) 'Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex', *Molecular Cell*. doi: 10.1016/j.molcel.2011.06.028.

Lubas, M. *et al.* (2015) 'The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis', *Cell Reports*. doi: 10.1016/j.celrep.2014.12.026.

Luo, M. J. *et al.* (2001) 'Pre-mrna splicing and mRNA export linked by direct interactions between UAP56 and aly', *Nature*. doi: 10.1038/35098106.

Luo, W., Johnson, A. W. and 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*. doi: 10.1101/gad.1409106.

Ma, W. K., Cloutier, S. C. and Tran, E. J. (2013) 'The DEAD-box protein Dbp2 functions with the RNA-binding protein Yra1 to promote mRNP assembly', *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2013.05.016.

Makino, D. L., Baumgärtner, M. and Conti, E. (2013) 'Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex', *Nature*. doi: 10.1038/nature11870.

Marayati, B. F. *et al.* (2016) 'The fission yeast MTREC and EJC orthologs ensure the maturation of meiotic transcripts during meiosis', *Rna*, 22(9), pp. 1349–1359. doi: 10.1261/rna.055608.115.

Matera, A. G. and Wang, Z. (2014) 'A day in the life of the spliceosome.', *Nature reviews. Molecular cell biology*, 15(2), pp. 108–21. doi: 10.1038/nrm3742.

Mayes, A. E. *et al.* (1999) 'Characterization of Sm-like proteins in yeast and their association with U6 snRNA', *EMBO Journal*, 18(15), pp. 4321–4331. doi: 10.1093/emboj/18.15.4321.

Mazza, C. *et al.* (2001) 'Crystal structure of the human nuclear cap binding complex',

Molecular Cell, 8(2), pp. 383–396. doi: 10.1016/S1097-2765(01)00299-4.

Mazza, C. *et al.* (2002) 'Large-scale induced fit recognition of an m7GpppG cap analogue by the human nuclear cap-binding complex', *EMBO Journal*, 21(20), pp. 5548–5557. doi: 10.1093/emboj/cdf538.

McCloskey, A. *et al.* (2012) 'hnRNP C tetramer measures RNA length to classify RNA polymerase II transcripts for export', *Science*. doi: 10.1126/science.1218469.

Meinel, D. M. *et al.* (2013) 'Recruitment of TREX to the Transcription Machinery by Its Direct Binding to the Phospho-CTD of RNA Polymerase II', *PLoS Genetics*, 9(11). doi: 10.1371/journal.pgen.1003914.

Meola, N. *et al.* (2016) 'Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts', *Molecular Cell*. doi: 10.1016/j.molcel.2016.09.025.

Merz, C. *et al.* (2007) 'Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment', *Rna*, 13(1), pp. 116–128. doi: 10.1261/rna.336807.

Meyer, S., Temme, C. and Wahle, E. (2004) 'Messenger RNA turnover in eukaryotes: Pathways and enzymes', *Critical Reviews in Biochemistry and Molecular Biology*. doi: 10.1080/10409230490513991.

Millevoi, S. and Vagner, S. (2009) 'Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation', *Nucleic Acids Research*. doi: 10.1093/nar/gkp1176.

Myers, L. C. *et al.* (1998) 'The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain', *Genes and Development*. doi: 10.1101/gad.12.1.45.

Ni, Z. *et al.* (2008) 'P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo.', *Molecular and cellular biology*, 28(3), pp. 1161–1170. doi: 10.1128/MCB.01859-07.

Nishimura, K. *et al.* (2009) 'An auxin-based degron system for the rapid depletion of

proteins in nonplant cells', *Nature Methods*. doi: 10.1038/nmeth.1401.

Nojima, T. *et al.* (2007) 'The interaction between cap-binding complex and RNA export factor is required for intronless mRNA export', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M700629200.

Nojima, T. *et al.* (2015) 'Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing', *Cell*. doi: 10.1016/j.cell.2015.03.027.

Nudler, E. (2012) 'RNA polymerase backtracking in gene regulation and genome instability', *Cell*. doi: 10.1016/j.cell.2012.06.003.

Ohkura, H. *et al.* (1988) 'Cold-sensitive and caffeine-supersensitive mutants of the Schizosaccharomyces pombe dis genes implicated in sister chromatid separation during mitosis', *The EMBO Journal*.

Ohno, M. *et al.* (2000) 'PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation', *Cell*. doi: 10.1016/S0092-8674(00)80829-6.

Osheim, Y. N., Proudfoot, N. J. and Beyer, A. L. (1999) 'EM visualization of transcription by RNA polymerase II: Downstream termination requires a poly(A) signal but not transcript cleavage', *Molecular Cell*. doi: 10.1016/S1097-2765(00)80465-7.

Pabis, M. *et al.* (2010) 'Binding properties and dynamic localization of an alternative isoform of the cap-binding complex subunit CBP20.', *Nucleus (Austin, Tex.)*, 1(5), pp. 412–21. doi: 10.4161/nucl.1.5.12839.

Pabis, M. *et al.* (2013) 'The nuclear cap-binding complex interacts with the U4/U6·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells.', *RNA (New York, N.Y.)*, 19(8), pp. 1054–63. doi: 10.1261/rna.037069.112.

Peterlin, B. M. and Price, D. H. (2006) 'Controlling the Elongation Phase of Transcription with P-TEFb', *Molecular Cell*, pp. 297–305. doi: 10.1016/j.molcel.2006.06.014.

Ramanathan, A., Robb, G. B. and Chan, S. (2016) 'mRNA capping : biological functions and applications', *Nucleic Acids Research*, pp. 1–16. doi: 10.1093/nar/gkw551.

Ramírez, F. *et al.* (2014) 'DeepTools: A flexible platform for exploring deep-sequencing data', *Nucleic Acids Research*. doi: 10.1093/nar/gku365.

Rougvie, A. E. and Lis, J. T. (1988) 'The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged', *Cell*. doi: 10.1016/S0092-8674(88)91087-2.

Rufener, S. C. and Mühlemann, O. (2013) 'eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells.', *Nature structural & molecular biology*, 20(6), pp. 710–7. doi: 10.1038/nsmb.2576.

Saha, N., Schwer, B. and Shuman, S. (1999) 'Characterization of human, Schizosaccharomyces pombe, and Candida albicans mRNA cap methyltransferases and complete replacement of the yeast capping apparatus by mammalian enzymes', *Journal of Biological Chemistry*. doi: 10.1074/jbc.274.23.16553.

Saldi, T. *et al.* (2016) 'Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing', *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2016.04.017.

Santos-Rosa, H. *et al.* (2015) 'Nuclear mRNA Export Requires Complex Formation between Mex67p and Mtr2p at the Nuclear Pores', *Molecular and Cellular Biology*. doi: 10.1128/mcb.18.11.6826.

Schneider, C. *et al.* (2012) 'Transcriptome-wide Analysis of Exosome Targets', *Molecular Cell*. doi: 10.1016/j.molcel.2012.08.013.

Schulze, W. M. *et al.* (2018) 'Structural analysis of human ARS2 as a platform for co-transcriptional RNA sorting', *Nature Communications*. Springer US, 9(1). doi: 10.1038/s41467-018-04142-7.

Schwer, B. and Shuman, S. (2011) 'Deciphering the RNA Polymerase II CTD Code in Fission Yeast', *Molecular Cell*. Elsevier Inc., 43(2), pp. 311–318. doi: 10.1016/j.molcel.2011.05.024.

Segref, A. *et al.* (1997) 'Mex67p, a novel factor for nuclear mRNA export. Binds to both poly(A)+ RNA and nuclear pores', *EMBO Journal*. doi: 10.1093/emboj/16.11.3256.

Segref, A., Mattaj, I. W. and Ohno, M. (2001) 'The evolutionarily conserved region of the U snRNA export mediator PHAX is a novel RNA-binding domain that is essential for U snRNA export', *RNA*. doi: 10.1017/S1355838201002278.

Sikorski, T. W. and Buratowski, S. (2009) 'The basal initiation machinery: beyond the general transcription factors', *Current Opinion in Cell Biology*. doi: 10.1016/j.ceb.2009.03.006.

Singh, G. *et al.* (2012) 'The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus', *Cell*. doi: 10.1016/j.cell.2012.10.007.

Smith, E. R. *et al.* (2015) 'PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II', *Cell*. Elsevier Inc., 162(5), pp. 1003–1015. doi: 10.1016/j.cell.2015.07.042.

Söding, J., Biegert, A. and Lupas, A. N. (2005) 'The HHpred interactive server for protein homology detection and structure prediction', *Nucleic Acids Research*, 33(SUPPL. 2). doi: 10.1093/nar/gki408.

Søgaard, T. M. M. and Svejstrup, J. Q. (2007) 'Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M701345200.

Stäßer, K. *et al.* (2002) 'TREX is a conserved complex coupling transcription with messenger RNA export', *Nature*. doi: 10.1038/nature746.

Stevens, A. and Poole, T. L. (1995) '5'-Exonuclease-2 of Saccharomyces cerevisiae. Purification and features of ribonuclease activity with comparison to 5'-exonuclease-1', *Journal of Biological Chemistry*. doi: 10.1074/jbc.270.27.16063.

Stolarski, R. *et al.* (1996) '1 H-NMR studies on association of mRNA cap-analogues with tryptophan-containing peptides', *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*. doi: 10.1016/0167-4838(95)00232-4.

Sträßer, K. (2002) 'Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export', *The EMBO Journal*. doi:

10.1093/emboj/19.3.410.

Strambio-De-Castillia, C., Niepel, M. and Rout, M. P. (2010) 'The nuclear pore complex: Bridging nuclear transport and gene regulation', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm2928.

Sträßer, K. and Hurt, E. (2001) 'Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p', *Nature*. doi: 10.1038/35098113.

Sugiyama, T. *et al.* (2015) 'The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome', *Nature Communications*, 6(1). doi: 10.1038/ncomms8050.

Taniguchi, I. and Ohno, M. (2007) 'ATP-Dependent Recruitment of Export Factor Aly/REF onto Intronless mRNAs by RNA Helicase UAP56', *Molecular and Cellular Biology*. doi: 10.1128/mcb.01341-07.

Tewari, R. *et al.* (2010) 'Armadillo-repeat protein functions: Questions for little creatures', *Trends in Cell Biology*. doi: 10.1016/j.tcb.2010.05.003.

Tieg, B. and Krebber, H. (2013) 'Dbp5 - From nuclear export to translation', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. Elsevier B.V., 1829(8), pp. 791–798. doi: 10.1016/j.bbagrm.2012.10.010.

Tuck, A. C. and Tollervey, D. (2013) 'A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs', *Cell*, 154(5), pp. 996–1009. doi: 10.1016/j.cell.2013.07.047.

Vanácová, S. *et al.* (2005) 'A new yeast poly(A) polymerase complex involved in RNA quality control.', *PLoS biology*. doi: 10.1371/journal.pbio.0030189.

Vasiljeva, L. *et al.* (2008) 'The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.1468.

Veloso, A. *et al.* (2014) 'Rate of elongation by RNA polymerase II is associated with

specific gene features and epigenetic modifications', *Genome Research*. doi: 10.1101/gr.171405.113.

Verdel, A. *et al.* (2004) 'RNAi-Mediated Targeting of Heterochromatin by the RITS Complex', *Science*. doi: 10.1126/science.1093686.

Vos, S. M. *et al.* (2018) 'Structure of paused transcription complex Pol II–DSIF–NELF', *Nature*. doi: 10.1038/s41586-018-0442-2.

Wang, C.-Y., Chen, W.-L. and Wang, S.-W. (2013) 'Pdc1 Functions in the Assembly of P Bodies in Schizosaccharomyces pombe', *Molecular and Cellular Biology*. doi: 10.1128/mcb.01583-12.

Wang, C. *et al.* (2018) 'Context-dependent deposition and regulation of mRNAs in P-bodies', *eLife*. doi: 10.7554/elife.29815.

Wasmuth, E. V., Januszyk, K. and Lima, C. D. (2014) 'Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA', *Nature*. doi: 10.1038/nature13406.

Wasmuth, E. V. and Lima, C. D. (2012) 'Exo- and Endoribonucleolytic Activities of Yeast Cytoplasmic and Nuclear RNA Exosomes Are Dependent on the Noncatalytic Core and Central Channel', *Molecular Cell*. doi: 10.1016/j.molcel.2012.07.012.

West, S. and Proudfoot, N. J. (2008) 'Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination', *Nucleic Acids Research*. doi: 10.1093/nar/gkm1112.

Will, C. L. and Lührmann, R. (2011) 'Spliceosome structure and function', *Cold Spring Harbor Perspectives in Biology*, 3(7), pp. 1–2. doi: 10.1101/cshperspect.a003707.

Worch, R. *et al.* (2005) 'Specificity of recognition of mRNA 5′ cap by human nuclear cap-binding complex', *RNA*. doi: 10.1261/rna.2850705.

Wyers, F. *et al.* (2005) 'Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase', *Cell*. doi: 10.1016/j.cell.2005.04.030.

Xiang, S. *et al.* (2009) 'Structure and function of the 5′3′ exoribonuclease Rat1 and its

activating partner Rai1', *Nature*. doi: 10.1038/nature07731.

Yamaguchi, Y. *et al.* (1999) 'NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation', *Cell*. doi: 10.1016/S0092-8674(00)80713-8.

Yamanaka, S. *et al.* (2010) 'Importance of polyadenylation in the selective elimination of meiotic mRNAs in growing S. pombe cells', *EMBO Journal*. doi: 10.1038/emboj.2010.108.

Zhang, H., Rigo, F. and Martinson, H. G. (2015) 'Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site', *Molecular Cell*. doi: 10.1016/j.molcel.2015.06.008.

Zhang, Z., Fu, J. and Gilmour, D. S. (2005) 'CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor, Pcf11', *Genes and Development*. doi: 10.1101/gad.1296305.

Zhou, Z. *et al.* (2017) 'Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export', *The EMBO Journal*, 36(19), pp. 2870–2886. doi: 10.15252/embj.201696139.

Appendix A Genetic screening of CBC depletion strains and RNA surveillance factors

This appendix includes all the genetic screens I carried out to reveal interactions between Cbc1 and Cbc2 depletion strains and different RNA surveillance factors. These results are summarized in Table 3.1 in section 3.3. Unless indicated otherwise, images are representative of two independent replicates. For ts strains, the restrictive temperature is indicated in red.

A.1 Rrp6

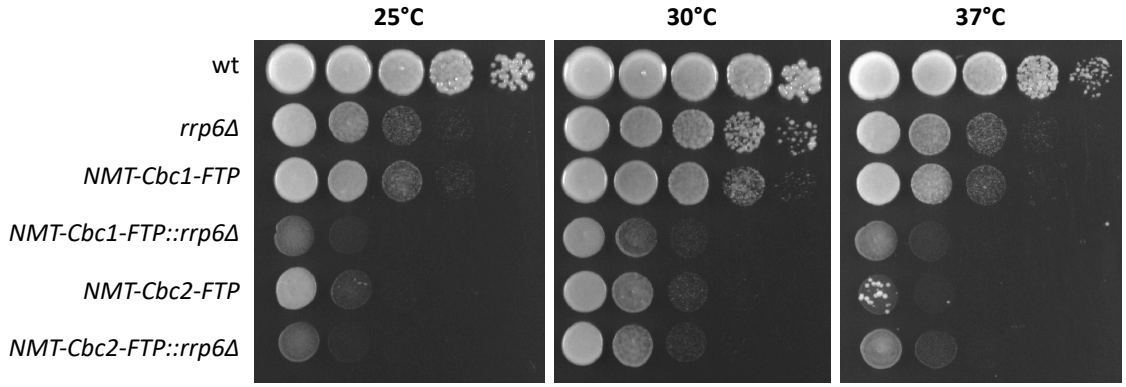


Figure A.1: Screening of genetic interaction between *rrp6Δ* and the depletion strains, grown at 25°C, 30°C or 37°C.

A.2 Dis3

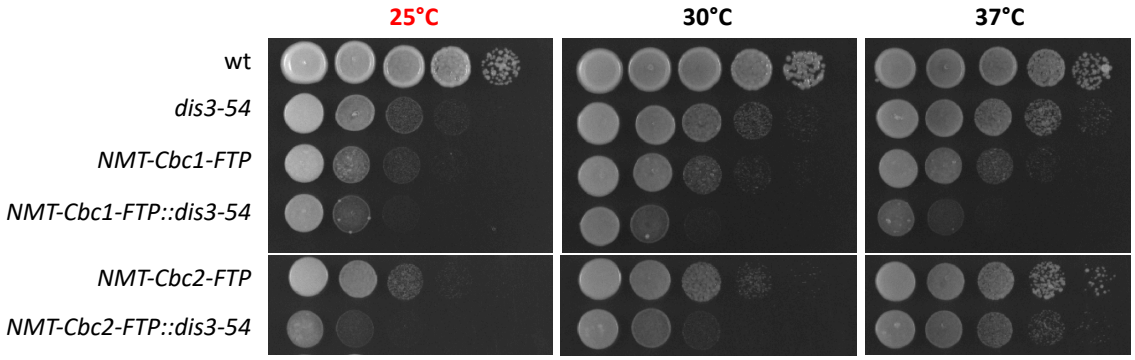


Figure A.2: Screening of genetic interaction between *dis3-54* and the depletion strains, grown at 25°C, 30°C or 37°C. Restrictive temperature for *dis3-54* is 25°C.

A.3 Mtl1

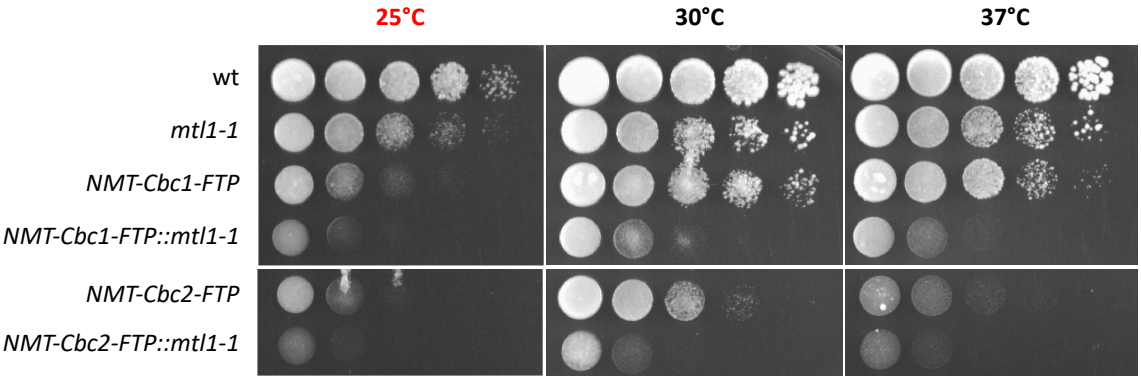


Figure A.3: Screening of genetic interaction between *mtl1-1* and the depletion strains, grown at 25°C, 30°C or 37°C. Restrictive temperature for *mtl1-1* is 25°C.

A.4 Red1

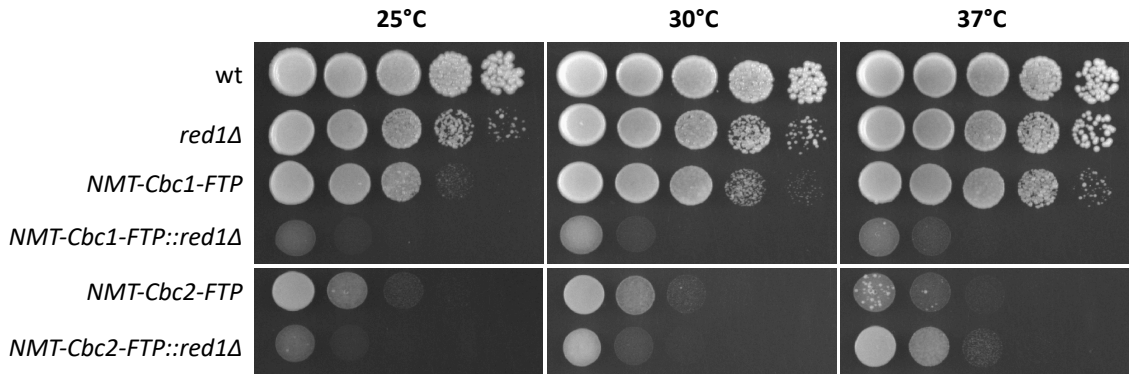


Figure A.4: Screening of genetic interaction between *red1Δ* and the depletion strains, grown at 25°C, 30°C or 37°C.

A.5 Pab2

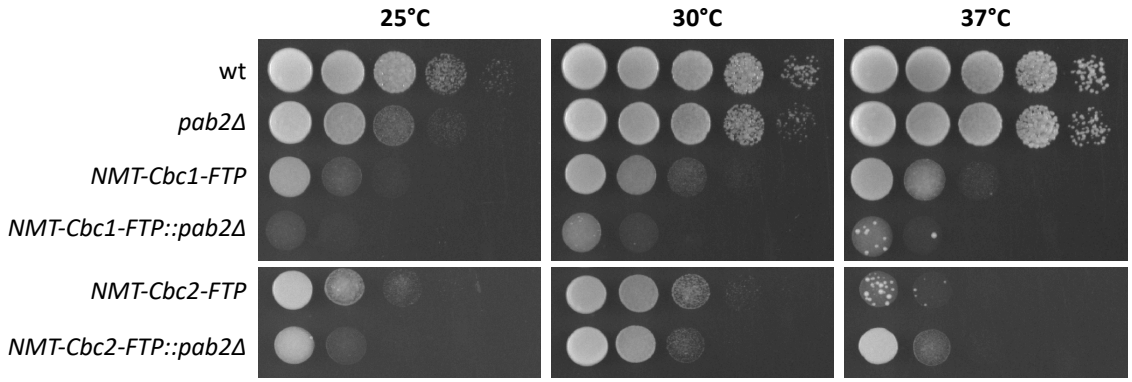


Figure A.5: Screening of genetic interaction between *pab2Δ* and the depletion strains, grown at 25°C, 30°C or 37°C.

A.6 Hrb1

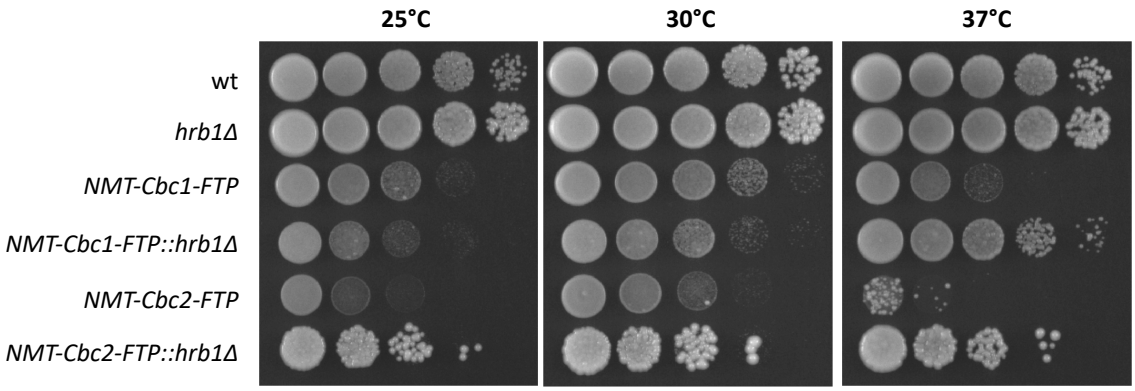


Figure A.6: Screening of genetic interaction between *hrb1Δ* and the depletion strains, grown at 25°C, 30°C or 37°C.

A.7 Dhp1

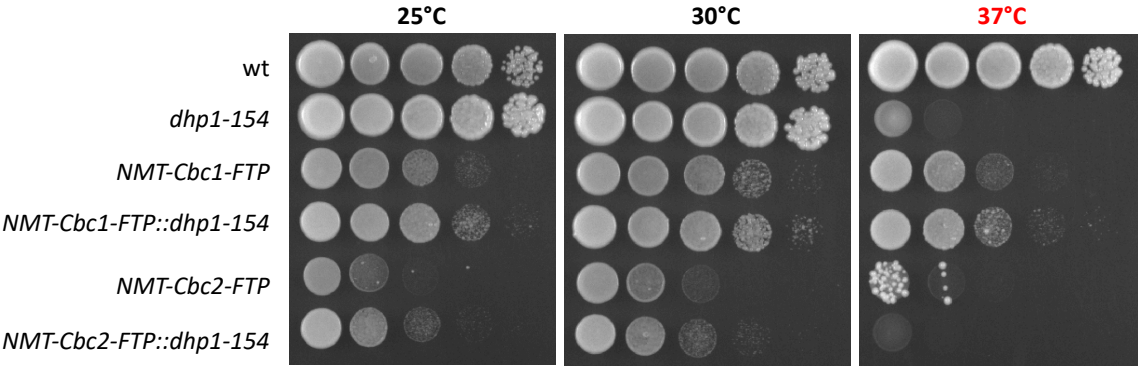


Figure A.7: Screening of genetic interaction between *dhp1-154* and the depletion strains, grown at 25°C, 30°C or 37°C. Restrictive temperature for *dhp1-154* is 25°C.