

Dynamics of RNA Transcription and Transcript Stabilisation in Wild-type and Mutant *Saccharomyces cerevisiae*



Ronja Woloszczuk

Department of Biochemistry

St Catherine's College

Trinity Term 2016

Thesis submitted in partial Fulfilment of Requirements for Degree of
Doctor of Philosophy

Abstract

Ribonucleic acids (RNAs) are a diverse and highly dynamic class of molecules. Coding RNAs facilitate the production of proteins from the heritable information stored in deoxyribonucleic acid (DNA), while pervasively transcribed non-coding RNAs have catalytic, scaffolding and regulatory functions.

Since cellular requirements vary with time, cell type and environmental condition, RNA pools are highly dynamic. Steady-state RNA levels are balanced through the production of RNAs via transcription and the removal of RNA via degradation. Both RNA transcription and degradation rates are highly regulated and can vary over several orders of magnitude even among closely related or overlapping RNAs. What is more, these rates frequently correlate with transcript translatability, *i.e.* how efficiently individual transcripts are being translated into proteins.

This means that the cell fine-tunes protein levels indirectly via RNA pools through different processes to coordinate cell growth and the response to various stimuli using interconnected feed-back loops. RNA transcription is the most studied aspect of the regulatory machinery that determines cellular RNA concentrations. However, the absence of important components of RNA degradation pathways also lead to a global deregulation of steady-state RNA levels.

The aim of this thesis is to link different aspects of RNA metabolism in the budding yeast *Saccharomyces cerevisiae*. This is achieved by characterising steady state RNA concentrations, transcription rates and nucleosome occupancy, that is DNA packaging, which is inherently linked to transcription, in

wild-types and two degradation pathway mutants.

The study of wild-type yeast conducted in this thesis shows that 3' end formation of stable RNAs is linked to a slowdown of RNAP II and a depletion in nucleosomes. Moreover, the analysis of 3' ends reveals a novel class of intragenic transcripts that are constitutively transcribed and terminated, at least in part, via non-canonical termination pathways.

The absence of Rrp6p, a nuclear exoribonuclease, causes the stabilisation of unstable RNAs in the nucleus. In spite of previous reports linking Rrp6p to transcriptional read-through, RNA stabilisation does not appear to be linked to wide-spread changes in transcription or nucleosome occupancy. Stabilised RNAs have diverse 3' ends that primarily map to the gene body of non-coding genes and are likely to be terminated by the Nrd1p-Nab3p-Sen1p termination pathway. The observed 3' ends are associated with distinct sequences and generally map to the end of nucleosome depleted regions.

In *xrn1Δ* defects in RNA degradation lead to a global stabilisation of all RNAs and specific secondary changes in chromatin and RNA transcription. These changes are likely to be driven by altered concentrations of transcription factors and chromatin remodellers. Binding of these factors to DNA leads to nucleosome depletion at promoters in the absence of an external stimulus, ultimately causing an altered responsiveness to acute environmental stress and a distinct growth phenotype.

Hence, RNA degradation pathways are a major determinant of levels of specific RNA isoforms. Changes in the activity of RNA degradation pathways result in altered cellular signalling and modified transcriptional responses. RNA

stabilisation is, in part, hard-wired into the sequence of the RNA and in part controlled by exoribonucleases in the cytoplasm and nucleus. The absence of these exoribonucleases has detrimental effects on the cell and affects its ability to respond to the environment.

Acknowledgements

I would like to thank my supervisors Prof Jane Mellor and Dr Benjamin Schuster-Böckler for their inputs, helpful suggestions, patience and consistent support throughout this thesis. In addition, I would like to acknowledge Dr Harry Fischl, who produced, with the exception of the *rrp6Δ* NET-Seq data, all the NET-Seq data discussed in this thesis, Dr Françoise Howe, who measured the cell sizes of wild-type and *xrn1Δ* yeast, Anitha Nair, who carried out the *xrn1Δ* life span experiments and helped with many experiments throughout my thesis, and Philip Lorenz and Ellie Maitland, who collected the GAL10-GFP FACS data. Moreover, I would like to thank Naomi Petala for her help with the MNase sequencing, Rhadika Patel for her instructions of RNA sequencing and Dr Struan Murray for his support with my initial bioinformatics analyses. I would also like to thank the whole of Mellor lab for their help with various protocols, their useful insight and for providing a nice environment, as well as my funding bodies, the EPSRC and the Leatherseller's company, for providing the funds that made this DPhil possible. Last, but not least, I would like to extend a big thank you to my friends and family, who made Oxford the experience that it was, listened to my complaining and supported me anyway.

1.4.1.4	3'->5' degradation	28
1.4.2	mRNA surveillance	29
1.4.2.1	Non-sense mediated decay	29
1.4.2.2	No-Go decay	31
1.4.2.3	Non-Stop decay	32
1.4.3	Degradation of stable non-coding RNAs	33
1.4.4	Nuclear decay	34
1.4.4.1	The nuclear exosome	34
1.4.4.2	RNA surveillance by the 5'->3' degradation machinery	35
1.4.4.3	Splicing	35
1.4.5	Summary of RNA degradation pathways.....	35
1.5	The dynamic nature of the RNA transcriptome.....	36
1.5.1	Transcriptional induction and suppression	36
1.5.1.1	The environmental stress response.....	37
1.5.1.2	The glucose to galactose Shift.....	39
1.5.2	Degradation rate dynamics.....	42
1.6	Aims and outline	44
2	Materials and Methods	48
2.1	Strains and transformations	48
2.2	Yeast media and growth conditions.....	49
2.3	Glucose to galactose Shifts	50
2.4	Drop plates	50
2.5	Western blotting.....	51
2.6	3' end poly(A)+ RNA sequencing.....	52
2.6.1	Library generation	52
2.6.2	Data processing (general)	53
2.6.3	Differential expression analysis	54
2.6.4	Data processing (polyadenylation sites)	55
2.6.5	Motif analysis	55

2.6.6	Structure analysis	55
2.7	Micrococcal nuclease sequencing (MNase-Seq)	56
2.7.1	Sample collection.....	56
2.7.2	Library generation and sequencing	57
2.7.3	Data processing	58
2.8	Nascent Elongating Transcript Sequencing (NET-Seq)	59
2.8.1	Sample collection.....	59
2.8.2	Immunoprecipitation	59
2.8.3	Linker ligation and fragmentation.....	60
2.8.4	Size selection.....	60
2.8.5	Reverse transcription	61
2.8.6	Circularisation	61
2.8.7	PCR amplification	61
2.8.8	Sequencing	63
2.8.9	Sequence alignment and processing.....	63
2.8.10	Annotations.....	63
2.8.11	Additional data sets	63
2.8.12	Data analysis and visualisation	65
2.9	Fluorescence-Activated Cell Sorting (FACS)	65
3	Roadblock Termination is a Widespread Phenomenon in Wild-type Yeast.....	67
3.1	Introduction	67
3.1	Results	69
3.1.1	Nucleosome depletion is a feature of most promoters and terminators of coding and non-coding genes	69
3.1.2	NET-Seq reads accumulate in distinct regions of the gene	72
3.1.3	Spike-ins improve RNA-Seq experiment accuracy	76
3.1.4	3' end-Seq data can be used to map PAS.....	77
3.1.5	The majority of PAS maps to the 3' end of genes	80
3.1.6	Terminators are inherently bi-directional	81
3.1.7	PAS are not used with equal frequencies	84

3.1.8	Intragenic PAS are a common feature of coding genes	84
3.1.9	Intragenic and 3' end PAS are associated with distinct motifs	87
3.1.10	Intragenic PAS are terminated by NNS but not CPFC	92
3.1.11	PAS increase prior to TF binding sites	94
3.1	Discussion	96
4	Transcription Termination Occurs Via Fail-Safe Mechanisms	102
4.1	Introduction	102
4.2	Results	105
4.2.1	<i>rrp6Δ</i> strains contain increased concentrations of unstable non-coding RNAs ..	105
4.2.2	The number of 3' end position at non-coding genes is elevated in <i>rrp6Δ</i> mutants	108
4.2.3	<i>rrp6Δ</i> -specific PAS occur across the gene and map antisense to coding genes	109
4.2.4	The deletion of <i>RRP6</i> has little effect on coding genes	111
4.2.5	<i>rrp6Δ</i> mutants accumulate more antisense PAS	112
4.2.6	Rrp6p targets the majority of tRNAs and snRNAs	114
4.2.7	<i>rrp6Δ</i> -specific PAS are used less frequently than conserved PAS	117
4.2.8	Positions of <i>rrp6Δ</i> -specific PAS correlate with Rrp6p and Rrp44p binding	118
4.2.9	<i>rrp6Δ</i> -specific PAS occur downstream of highly transcribed regions	119
4.2.10	<i>rrp6Δ</i> -specific sites occur downstream of NDRs	123
4.2.11	<i>rrp6Δ</i> -specific sites occur at defined sequences	126
4.2.12	The CFP binding sites are found at both <i>rrp6Δ</i> -specific PAS and conserved PAS	129
4.2.13	Nab3p and Nrd1p binding sites are enriched proximal to PAS	131
4.2.14	Pcf11p and Nrd1p are bound to terminators of both most PAS	135
4.2.15	CPF subunits are depleted from <i>rrp6Δ</i> -specific PAS non-coding PAS	137
4.2.16	<i>rrp6Δ</i> is more susceptible to environmental stresses	139
4.3	Discussion	140
5	Xrn1p is a major indirect repressor of untimely transcription	145
5.1	Introduction	145

5.2	Results	152
5.2.1	The absence of Xrn1p results in elevated RNA concentrations in most RNAP II-transcribed RNAs.....	152
5.2.2	Increases in RNA concentrations correlate with an elevated number of RNA isoforms	154
5.2.3	Increases in RNA concentrations do not correlate with Xrn1p binding	155
5.2.4	The absence of Xrn1p affects RNA concentrations of specific groups of genes	157
5.2.5	The deletion of <i>XRN1</i> leads to the accumulation of additional transcripts.....	163
5.2.6	RNAs with increased concentrations have increased transcription rates	165
5.2.7	<i>xrn1Δ</i> are characterised by decreases in nucleosome occupancy at transcription factor binding sites	168
5.2.8	Glucose to galactose shifts result in a global decrease in RNA concentrations .	175
5.2.9	Changes in RNA concentrations and transcription rates in during a glucose-to-galactose shift are indicative of a stress response	178
5.2.10	The transcription of ribosomal protein genes is less dynamic in <i>xrn1Δ</i>	183
5.2.11	Transcription of galactose genes is induced more quickly in <i>xrn1Δ</i> mutants .	185
5.2.12	Lower degradation rates can lead to increases in protein concentrations.....	188
5.2.13	<i>xrn1Δ</i> are more resistant to some forms of stress	189
5.3	Discussion	190
6	Conclusion and Future Work.....	196
7	Bibliography	199

Abbreviations

AMP	Adenosine Monophosphate
ATG	Start Codon
ATP	Adenosine Triphosphate
CF	Cleavage Factor
ChIP	Chromatin Immunoprecipitation
CPFC	Cleavage and Polyadenylation Factor Complex
CUT	Cryptic Unstable Transcript
DNA	Deoxyribonucleic Acid
ESR	Environmental Stress Response
FACT	Facilitates Chromatin Transcription
GO	Gene Ontology
GTF	General Transcription Factor
GRO-Seq	Genomic Run-On
GRR	Growth Rate Response
lncRNA	Long Non-Coding RNA
mRNA	Messenger RNA
m7G cap	7-methylguanylate cap
ncRNA	Non-Coding RNA
NDR	Nucleosome Depleted Region
NET-Seq	Nascent Elongating Transcript Sequencing

NGD	No-Go Decay
NGS	Next-Generation Sequencing
NMD	Nonsense-mediated decay
NNS	Nrd1p-Nab1p-Sen1p
NSD	Non-Stop Decay
NUT	Nrd1p-unstable transcripts
ORF	Open Reading Frame
pAp	3'-Phospho-Adenosine-5'-Phosphate
PAS	Polyadenylation Site
PIC	Pre-Initiation Complex
3' Seq	Poly(A)+ 3' end Sequencing
PRO-Seq	Precision Nuclear Run-On
PTC	Pre-Mature STOP Codon
PTM	Post-Translational Modification
RNA	Ribonucleic Acid
RNA-Seq	Ribominus RNA Sequencing
RiBi	Ribosomal Biogenesis Protein
RBP	RNA Binding Protein
RPG	Ribosomal Protein Gene
RNAP	RNA Polymerase
rRNA	Ribosomal RNA

SAGA	Spt-Ada-Gcn5p acetyltransferase
snRNA	Small Nucleolar RNA
snRNA	Small Nuclear RNA
SUT	Stable Uncharacterised Transcripts
TEF	Transcription Elongation Factors
TF	Transcription Factor
TES	Transcription End Site
tRNA	Transfer RNA
TSS	Transcription Start Site
TU	Transcription Unit
UAS	Upstream Activator Sequence
uORF	Upstream Open Reading Frame
UTR	untranslated region
XRN1	Exoribonuclease 1
XUT	Xrn1p-Unstable Transcripts
YMC	Yeast Metabolic Cycle

List of Figures

Figure 1.1 Types of RNAs.	2
Figure 1.2. Orientations of non-coding transcripts.	Error! Bookmark not defined.
Figure 1.3 Models of regulation of gene expression by lncRNAs.....	7
Figure 1.4. PIC assembly.	10
Figure 1.5. Nucleosomes and NDR formation.....	13
Figure 1.6. Phosphorylation patterns of RNAPII.	15
Figure 1.7. Termination complexes.	18
Figure 1.8. Models of PAS-dependent and PAS-independent transcriptional termination.....	20
Figure 1.9. Canonical cytoplasmic decay pathways.....	24
Figure 1.10. Nonsense mediated decay.....	30
Figure 1.11. No-Go decay.	32
Figure 1.12. Nuclear degradation machinery.	35
Figure 1.13. Glucose repression and <i>GAL</i> gene induction.....	40
Figure 2.1. Normalisation of drop plate measurements.	51
Figure 2.2. 3' end poly(A)+ RNA-Seq workflow.....	Error! Bookmark not defined.
Figure 2.3. FACS measurements.....	Error! Bookmark not defined.
Figure 3.1. Sequencing techniques.....	69
Figure 3.2. MNase-Seq read distributions.....	71
Figure 3.3. NET-Seq read distributions.	74
Figure 3.4. RNA-Seq experiments correlations.....	77

Figure 3.5. Mapping of PAS from 3' RNA-Seq data	79
Figure 3.6. Positions of PAS at the 3' end of genes.	83
Figure 3.7. Intragenic PAS and 3' end PAS are characterised by the same features.	86
Figure 3.8. PAS-associated motifs.	90
Figure 3.9. Binding of CPFC and NNS subunits to PAS.	93
Figure 3.10. PAS Numbers are increased in the vicinity of transcription Factor binding sites. ..	95
Figure 3.11. Models of transcriptional termination.	98
Figure 4.1. Changes in RNA abundance between WT and <i>rrp6Δ</i> mutants.....	107
Figure 4.2. Changes in the number of PAS in <i>rrp6Δ</i> mutants compared to wildtypes.....	109
Figure 4.3. Distribution of PAS in wildtypes and <i>rrp6Δ</i> strains.	110
Figure 4.4. Examples of <i>rrp6Δ</i> -specific and conserved PAS.....	113
Figure 4.5. Transcription and transcript stabilisation of tRNAs and snoRNAs.	107
Figure 4.6. Frequency of RNA-Seq reads at conserved and <i>rrp6Δ</i> -specific sites.	118
Figure 4.7. Rrp6 and Rrp44 binding to RNAs.....	119
Figure 4.8. Nascent transcription rates in <i>rrp6Δ</i>	121
Figure 4.9. Nascent RNA transcription at <i>rrp6Δ</i> -specific and conserved PAS.	123
Figure 4.10. Nucleosome depletion at canonical and <i>rrp6Δ</i> -specific PAS.	125
Figure 4.11. Figure: Nucleotide distributions upstream and downstream of mapped PAS.	127
Figure 4.12. Positioning of CPF binding motifs at PAS.	131
Figure 4.13. Distribution of Nab3p and Nrd1p binding sites.	132
Figure 4.14. Nrd1p and Nab3p binding sites at different PAS.....	134

Figure 5.1. Models of Xrn1p-mediated regulation of gene expression.	149
Figure 5.2. Changes in poly(A) + RNA abundance in <i>xrn1Δ</i> compared to wild-type under optimal growth conditions.	153
Figure 5.3. Distribution of 3' end isoforms and associated secondary structures.	155
Figure 5.4. Xrn1p binding to RNA and DNA.	156
Figure 5.5. Comparison of RNA concentrations in <i>xrn1Δ</i> with <i>rrp6Δ</i> and <i>sua7-1</i> mutants.	158
Figure 5.6. Changes in RNA concentration in selective groups of genes.	162
Figure 5.7. NET-Seq profiles in wild-type and <i>xrn1Δ</i> yeast.	166
Figure 5.8. Changes in RNA transcription rates.	168
Figure 5.9. Nucleosome occupancy in <i>xrn1Δ</i>	171
Figure 5.10. Nucleosome occupancy at transcription factor binding sites in <i>xrn1Δ</i> and wild-type strains.	174
Figure 5.11. Differences and similarities in nucleosome occupancy, RNA transcription rates and stable RNA concentrations in <i>xrn1Δ</i> and wild-type yeast at the <i>GAL1/GAL10</i> locus.	176
Figure 5.12. Changes in RNA concentrations and transcription rates after a glucose-to-galactose-shift.	177
Figure 5.13. Regulation of ribosomal protein genes during a glucose-to-galactose shift.	184
Figure 5.14. Regulation of <i>GAL</i> genes during a glucose-to-galactose shift.	186
Figure 5.15. <i>xrn1Δ</i> phenotype.	189
Figure 5.16. Model of Xrn1p transcriptional regulation.	194

List of Tables

Table 1.1. Types of non-coding transcripts in <i>S. cerevisiae</i>	4
Table 2.1. Yeast strains.....	49
Table 2.2. Primary Antibodies for Western Blots.....	52
Table 2.3. Oligonucleotides used for NET-Seq library generation.	62
Table 2.4. Data sets used in this thesis. *TheTEF-Seq method is briefly outlined in Chapter 4.	64
Table 3.1. GO enrichment for genes with intragenic PAS.....	85
Table 3.2. Motifs associated with intragenic and 3' end PAS.....	88
Table 4.1. Identified tRNAs in the wild-type and <i>rrp6Δ</i>	115
Table 4.2. Motifs identified within 100 nucleotides of conserved and <i>rrp6Δ</i> -specific PAS.....	128
Table 5.1 Fold-changes in RNA concentrations in <i>xrn1Δ</i> , <i>rrp6Δ</i> and <i>sua7-1</i> mutants compared to wild-type yeast.....	153
Table 5.2. Gene Ontology terms associated with coding factions of <i>xrn1Δ</i> -stabilised transcripts.	160
Table 5.3. Gene Ontology terms associated with transcripts that were only detected in <i>xrn1Δ</i> yeast.....	165
Table 5.4. Gene Ontology terms associated with up- and downregulated transcripts after a glucose to galactose shift.....	180
Table 5.5. Gene Ontology terms associated with changes in transcription rates in wild-types minutes after a glucose to galactose shift.....	181
Table 5.6. Gene Ontology terms associated with changes in transcription rates in <i>xrn1Δ</i> minutes after a glucose to galactose shift.....	182

Table 5.7. Gene Ontology terms associated with changes in transcription rates in *xrn1Δ* and wild-types 60 minutes after a glucose to galactose shift..... 183

1 Introduction

1.1 Overview

Ribonucleic acids (RNAs) are a diverse group of molecules with a variety of functions. More specifically, RNAs act as temporary storage units of heritable information during protein translation, as scaffolds and as enzymatic and regulatory components.

The regulation of cellular RNA concentrations is complex, dynamic and tailored towards the current needs of the cell. This introduction outlines the types of RNAs present in cells and the interconnected regulatory processes that determine intracellular RNA concentrations.

1.2 Types of RNAs

By definition, ribonucleic acids are strings of nucleotides that are synthesised by copying a template DNA strand in a reaction catalysed by an RNA polymerase (RNAP). In principle, any transcribed region that forms a single RNA transcript and has a fixed transcription start site (TSS) and a transcription end site (TES), can form a transcription unit (TU). TUs generally contain one or more genes with varying degrees of conservation. However, conservation is not a prerequisite and ultimately an RNA function is largely dependent on its RNA's protein coding and catalytic potential. Therefore, RNAs have been categorised based on these characteristics.

1.2.1 Coding RNAs

A large fraction of the RNAs transcribed in a cell contains protein coding sequences, which can be translated by ribosomes. Due to this, they are referred to as messenger RNAs (mRNAs) and typically have evolutionarily conserved open reading frames (ORFs), *i.e.* ORFs which code for orthologous proteins in different organisms, flanked by untranslated regions (UTRs). All eukaryotic mRNAs are synthesised by RNAP II, 5' capped and 3' polyadenylated and exported to the cytoplasm, where they are translated. This is distinct from bacterial RNAs, which have a triphosphate at their 5' end and are not polyadenylated. Some mRNAs are constitutively transcribed, while other genes are being switched on and off in responses to particular stimuli.

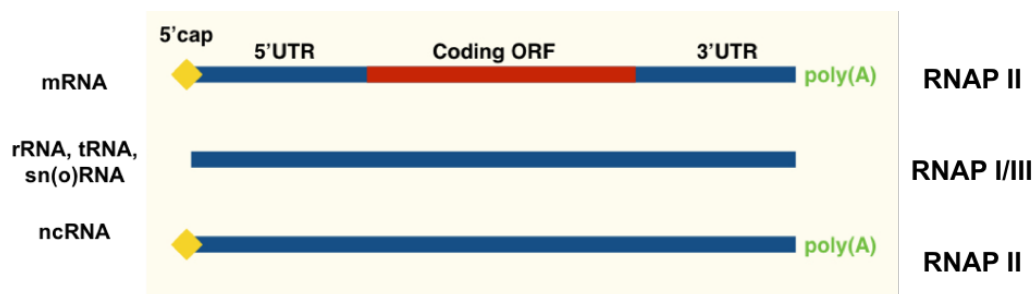


Figure 1.1 Types of RNAs.

1.2.2 RNAP I and RNAP III transcribed non-coding RNAs

Ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs) and small nuclear RNAs (snRNAs) are RNAs without protein coding functions that are transcribed by RNAP I (rRNAs), RNAP III (tRNAs, snRNAs and snoRNAs) and sometimes RNAP II (snRNAs and snoRNAs). Rather than storing a message, these RNAs have catalytic and scaffolding functions of their own. Unlike mRNAs, they are not capped or polyadenylated, although they can be chemically modified at specific positions.

1.2.3 RNAP II transcribed non-coding RNAs

Long non-coding RNAs (lncRNAs) are pervasively transcribed by RNAP II throughout the genome (David et al. 2006; Xu et al. 2009; van Dijk et al. 2011; Schulz et al. 2013). lncRNAs are generally shorter than mRNAs but, by definition, longer than 200 nucleotides. Like other RNAP II transcribed RNAs, stable lncRNAs are 5' capped and 3' polyadenylated. Unlike mRNAs, however, lncRNAs lack any apparent protein-coding function, although many lncRNAs contain cryptic ORFs that are translated into short, unstable peptides (Smith et al. 2014).

lncRNAs are a common feature of transcriptomes across all branches of life (Morris & Vogt 2010; David et al. 2006; Djebali et al. 2012; Derrien et al. 2011; Raghavan et al. 2012). As lncRNAs are ten-fold less abundant, on average, than coding mRNAs (Ozsolak et al. 2010) and comparatively poorly conserved between species (Guttman et al. 2009), non-coding transcripts were initially thought to originate from noisy transcription (Struhl 2007). However, lncRNAs have independent promoters resulting in condition and tissue-specific expression patterns (Rhee & Pugh 2012; Granovskaia et al. 2010; Nguyen et al. 2014; Murray et al. 2012). While this does not necessarily mean that all non-coding RNAs perform or have specific functions in the cell, it suggests that non-coding RNA transcription can be regulated independently. Moreover, while the primary sequence of non-coding RNAs is often poorly conserved, promoters and secondary structure evolution is constrained across species (Derrien et al. 2012; Eddy 2014), suggesting that lncRNA transcription is regulated.

1.2.3.1 lncRNA types

In yeast, lncRNAs are characterised by their stability and the circumstances under which they can be observed. In fact, very little lncRNA transcription leads to the accumulation of stable transcripts (Table 1.1). Only a subgroup of non-coding transcripts, called stable uncharacterised transcripts (SUTs), are readily detected in wild-type yeast (Xu et al. 2009). These RNAs, are, like mRNAs, frequently exported to the cytoplasm and degraded by the 5'->3' exoribonuclease 1 (Xrn1p) (Marquardt et al. 2011; van Dijk et al. 2011).

Type	Full Name	Stability	Degradation Pathway	Mutants	Reason for Accumulation	References
CUT	Cryptic Unstable Transcript	Unstable	Mostly NNS and nuclear exosome, some Xrn1p dependent	<i>RRP6</i> , <i>TRF4</i> , <i>TRF5</i> , <i>AIR1</i> , <i>AIR2</i> , <i>MTR4</i>	Failed termination or degradation	Wyers et al. 2005; Neil et al. 2009
NUT	Nrd1p-Unstable Transcript	Unstable	NNS and nuclear exosome	<i>NRD1</i>	Failed termination	Schulz et al. 2013
SUT	Stable Uncharacterised Transcripts	Stable	Mostly Xrn1p dependent, some nuclear exosome dependence	-	-	Xu et al. 2009
XUT	Xrn1p-Unstable Transcript	Unstable	Xrn1p dependent	<i>XRN1</i>	Failed degradation	van Dijk et al. 2011
SRT	Ssu72-Restricted Transcripts	Unstable	Mostly NNS and nuclear exosome	<i>SSU72</i>	Decreased promoter directionality, increased aberrant initiation	Tan-Wong et al. 2012

Table 1.1. Types of non-coding transcripts in *S. cerevisiae*.

Unstable lncRNAs accumulate in a variety of mutants. For instance, deletion of *XRN1* results in the accumulation of Xrn1-unstable transcripts (XUTs) and the stabilisation of most SUTs (Marquardt et al. 2011; van Dijk et al. 2011).

In addition, short-lived transcripts referred to as Cryptic Unstable Transcripts (CUTs) are retained in the nucleus where they are targeted for degradation immediately after transcription has ceased (Thiebaut et al. 2006). CUTs primarily result from spurious transcription of bidirectional promoters (Helen Neil et al. 2009; Xu et al. 2009) and are targeted for degradation by the nuclear exosome via an alternative Nrd1p-dependent termination pathway discussed in section 1.3.3 (Schulz et al. 2013; Jamonnak et al. 2011; Creamer et al. 2011; Marquardt et al. 2011). Hence, deletion of the nuclear exosome subunit *RRP6* (Xu et al. 2009; Davis & Ares 2006), *NRD1* (Schulz et al. 2013), and components of the TRAMP complex (Wyers et al. 2005) as well as of *UPF1* (Malabat et al. 2015) all lead to the accumulation of CUTs.

Non-canonical RNAs can also accumulate due to the initiation of aberrant transcription. This has, in particular, been observed in mutants that are unable to form gene loops, *i.e.* do not form links between the 5' and 3' ends of genes, and are unable to control the directionality of transcriptional initiation from bidirectional promoters (Tan-Wong et al. 2012; Grzechnik et al. 2014).

1.2.3.2 lncRNA function

In spite of the pervasiveness of lncRNA transcription, their function remains controversial. Some lncRNAs have been linked to the regulation of sense transcription, although regulatory mechanisms vary between loci. Broadly speaking, the regulation of coding transcript concentrations by non-coding

RNAs can occur via transcriptional effects in *cis* or can be mediated by the non-coding transcript in *cis* or *trans* once transcription has been completed.

In the first case, the act of transcription of the non-coding gene itself changes the surrounding transcriptional landscape, ultimately suppressing or enhancing the transcription of neighbouring genes. Non-coding TUs frequently either overlap with or are transcribed from the opposite strand as, that is antisense to, coding TUs (Figure 1.2).

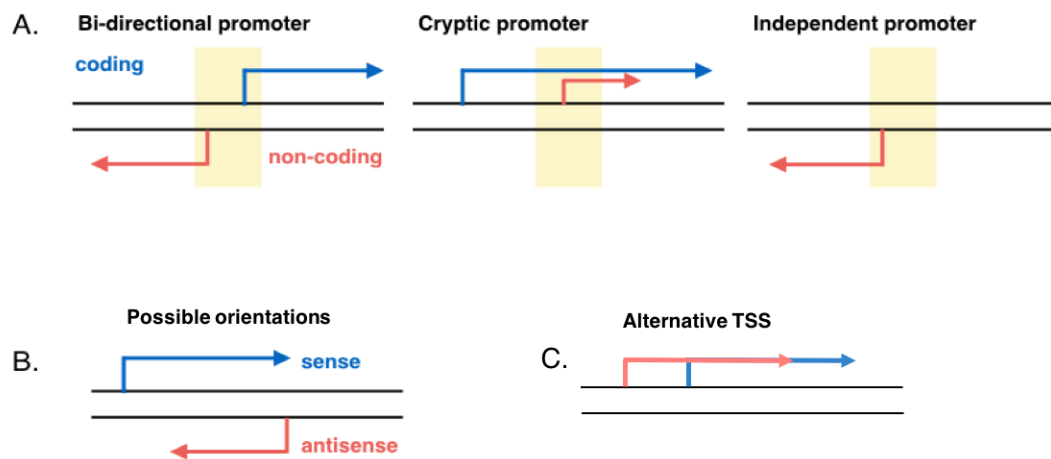


Figure 1.2. Orientations of non-coding transcripts.

A. Types of promoters. B. Antisense orientation. C. Alternative promoters resulting in distinct transcripts.

In yeast, lncRNAs are most commonly transcribed from shared bidirectional promoters, but can also originate from cryptic promoters, 3' regions of genes, intron-exon boundaries, alternative TSS, or, less frequently, independent promoters either within gene bodies or intergenic regions. This means that antisense TUs can overlap with important regulatory features of the sense TU including its promoter and terminator (Murray et al. 2012; Pelechano et al. 2013; Churchman & Weissman 2011; Kopcewicz et al. 2007). Hence, non-coding

transcription, especially in the antisense direction, can result in transcriptional interference (Figure 1.3).

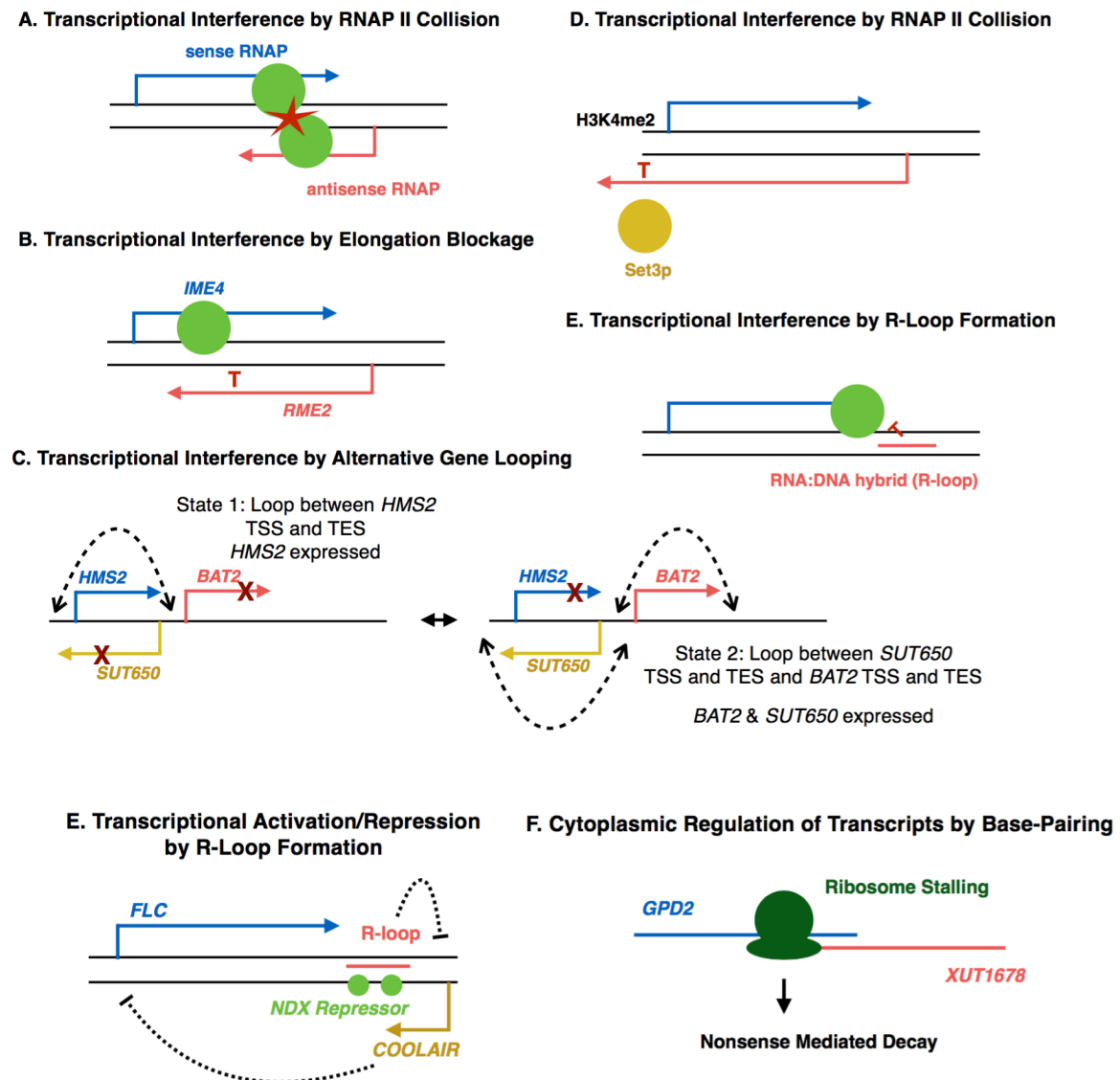


Figure 1.3 Models of regulation of gene expression by lncRNAs.

Transcriptional interference can be achieved through the collision of opposing RNAP II complexes (Figure 1.3 (A), Prescott & Proudfoot 2002; Hobson et al. 2012), the blocking the elongation of RNAPII (Figure 1.3 (B), Hongay et al. 2006), gene-loop mediated transcriptional insulation (Figure 1.3 (C), Nguyen et

al. 2014) or the modification of the local chromatin environment (Figure 1.3 (D), Hainer & Martens 2011; Kim et al. 2012). However, transcription rates of sense and antisense genes are not correlated genome-wide (Murray et al. 2015).

Alternatively, lncRNAs control coding RNA transcript concentrations post-transcriptionally. This can occur via a variety of mechanisms. For instance, the formation of RNA:DNA hybrid structures (R-loops, Figure 1.3 (E)) can cause RNA polymerase pausing and transcriptional termination (Skourti-Stathaki et al. 2011). In addition, the presence of R-loops can result in a depletion of nucleosomes that is associated with gene activation or silencing (Sun et al. 2013).

Moreover, base-pairing between sense and antisense transcripts in the cytoplasm leads to changes in translation rates or RNA decay of sense transcripts (Wery et al. 2016, Figure 1.3 (F)).

1.3 RNAP II RNA transcription

During the process of RNA transcription, a template DNA strand is copied by RNAPs into RNAs, thereby making the genetic information stored in the form of DNA accessible to the cell. For transcription to occur, RNAP II must be released from the promoter and progress from the transcription start site (TSS) to the transcription end site (TES) while synthesizing DNA-templated RNA. Productive transcription requires the recruitment of various co-factors to the RNAP II that affect the local chromatin, facilitate DNA melting and co-transcriptionally modify and process the nascent RNA (Sainsbury et al. 2015; Porrua & Libri 2015; Zhou et al. 2012). As such, RNA transcription is a highly regulated process that is defined by three phases: initiation, elongation and termination.

1.3.1 Transcription initiation

For transcription to be initiated, the RNAP II must bind and melt DNA at the gene promoter. However, the binding of the RNAP II complex to the promoter is not spontaneous, but is dependent on the bending of DNA, remodelling of chromatin and nucleosome depletion, as well as the assembly of the pre-initiation complex (PIC), a multimeric complex consisting of general transcription factors (GTFs).

1.3.1.1 RNAPII PIC assembly

PIC assembly occurs in several steps and involves the consecutive binding of the GTFs (TFIIB, TFIID, TFIIIE, TFIIF and TFIIH) together with RNAP II and requires the remodeling of the promoter (Sainsbury et al. 2015). This process is schematically outlined in Figure 1.4.

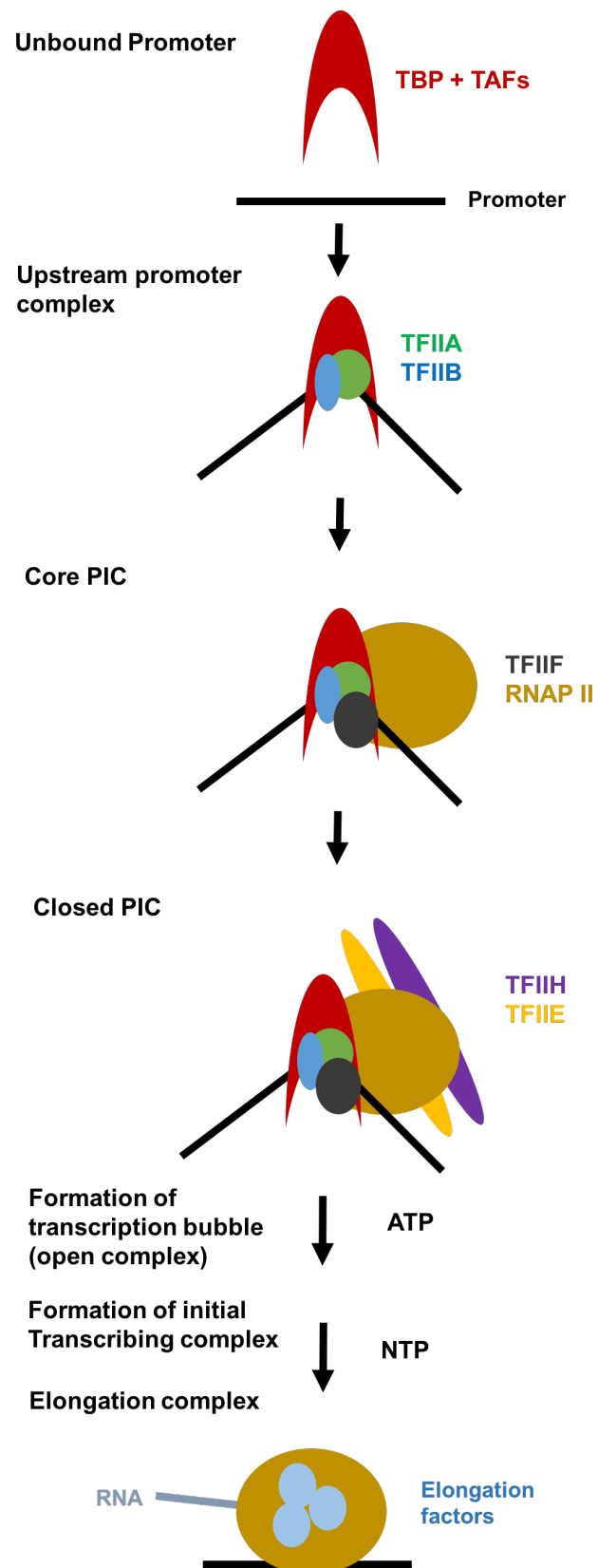


Figure 1.4. PIC assembly.

In brief, the TATA binding protein (TBP) and variable TBP-associated factors (TAFs) form a promoter-specific TFIID complex (Kuras et al. 2000; Huisinga & Pugh 2004). TBP, as part of this complex, then binds to the TATAAWAR (TATA-box, W-A or T, R-A or G) consensus sequence or non-canonical motifs approximately 30-140 base pairs upstream of the TSS (Rhee & Pugh 2012). Next, TFIIA and TFIIB (Sua7p in yeast) are recruited, which stabilises the interaction of TBP with the DNA. The formation of this upstream promoter complex aids the bending of the DNA by TBP and facilitates the binding of RNAP II (Imbalzano et al. 1994; Pinto et al. 1992).

Once RNAP II has been recruited the complex is referred to as the core PIC. Subsequent binding of TFIIIE and TFIIH triggers promoter opening by melting the DNA, followed by the insertion of the template strand into the RNAP II cleft. This process constitutes the transition from a closed complex to a transcription-competent open complex (Kim 2000; Treutlein et al. 2012). In principle, PICs can assemble on both strands of the DNA resulting in bidirectional transcription initiation (Rhee & Pugh 2012; Venters & Pugh 2009). The PIC's orientation on the DNA is determined by the Mediator complex and TFIIIF (Bernecky & Taatjes 2012), which is recruited together with RNAP II (Rani et al. 2004). Therefore, bidirectional transcription from a single promoter is, at least in part, regulated independently, resulting in condition and tissue-specific expression of mRNAs and lncRNAs (Nguyen et al. 2014; Granovskaia et al. 2010).

1.3.1.2 Nucleosome depletion and histone modifications

Most eukaryotic DNA is tightly wrapped around nucleosomes, *i.e.* octameric

histone complexes, that are precisely positioned on the DNA (Figure 1.5). PICs cannot assemble in nucleosome occupied regions of the DNA, due to the occlusion of transcription-factor binding sites and sterical hindrances that prevent GTF binding. Therefore, transcription-competent PIC assembly requires the removal nucleosomes from the DNA. Nucleosome depleted regions (NDRs) are a characteristic of actively transcribed promoters and terminators (Fan et al. 2010).

Nucleosome depletion is partially hard-wired into the DNA through the occurrence of poly(A/T) stretches at the promoter and terminator (Segal & Widom 2009). These sequences are less bendable than GC rich sequences, making nucleosome binding more fragile. In addition, specific histone chaperones and chromatin remodelling complexes, such as, but not exclusively, Asf1p, RSC and SWI/SNF, can be recruited to the promoter to create or maintain NDRs (Figure 1.5 (D)). These complexes act by either shifting nucleosomes along the DNA, resulting in phased nucleosome assemblies, or by evicting nucleosomes from the DNA altogether (Venkatesh & Workman 2015). Nucleosome depletion occurs upstream of the TSS, resulting in stringently positioned -1 and +1 nucleosomes upstream and downstream of the TSS respectively (Mavrich et al. 2008). In spite of their exact positioning, the -1 and +1 nucleosomes at the promoter are turned over rapidly (Dion et al. 2007). As genes are being switched on and off, nucleosome positioning is subject to change and, similarly to transcription, dependent on the current state of the cell (Huebert et al. 2012; Teif et al. 2012).

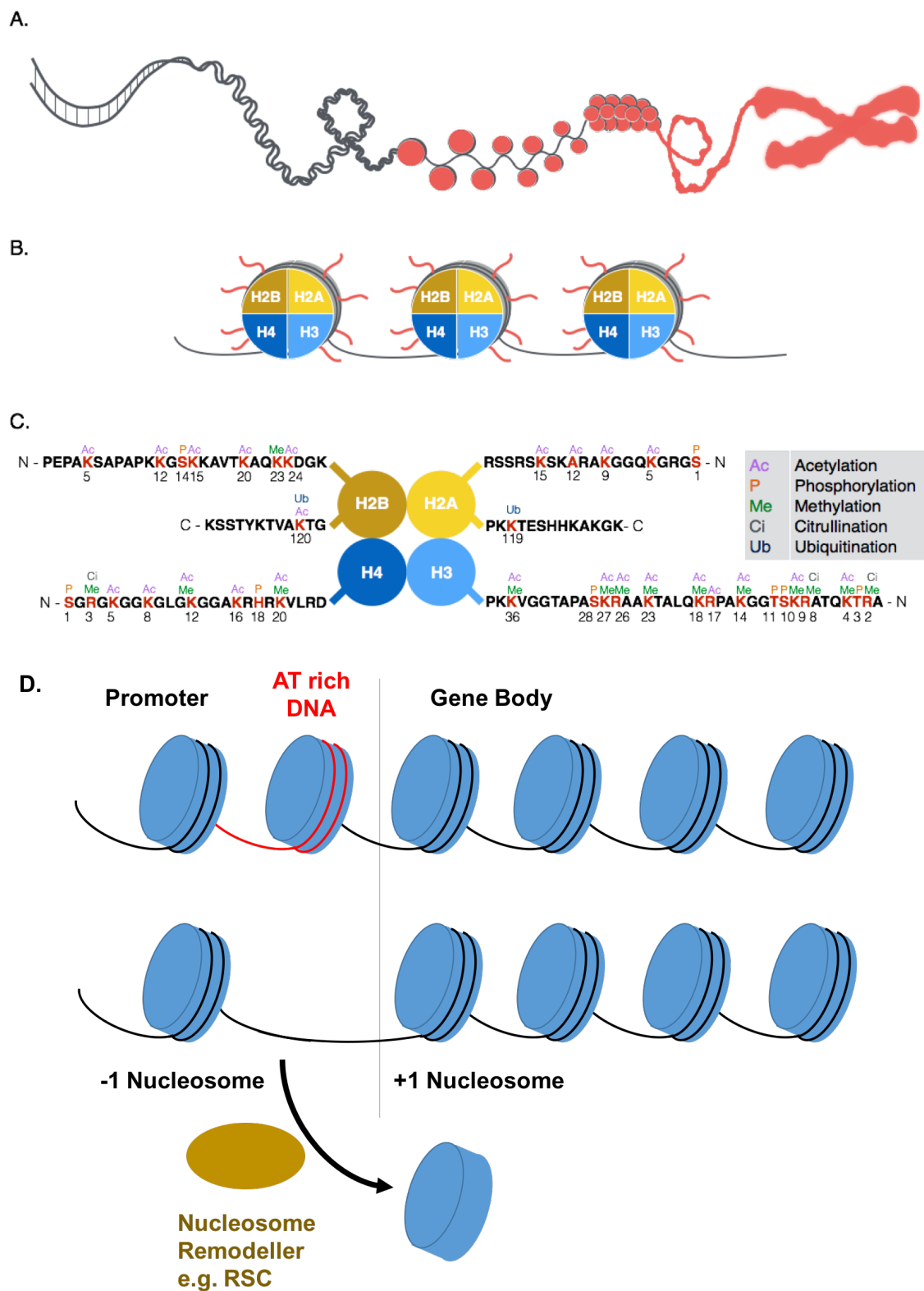


Figure 1.5. Nucleosomes and NDR formation.

A. Chromatin structure. B. Nucleosomes and their histone subunits. C. Known histone modifications. D. Nucleosome depletion from the promoter.

However, nucleosome remodelling at the TSS can occur independently of the initial recruitment of PIC components and is not inherently dependent on transcription (Grimaldi et al. 2014; Huebert et al. 2012). Instead, NDR formation occurs prior to the initial recruitment of RNAP II (Venters & Pugh 2009). The process is aided by transcription factors, histone chaperones and chromatin remodellers that are thought to act cooperatively and compete with nucleosomes for DNA binding (Floer et al. 2010; Ransom et al. 2009; Kharerin et al. 2016). One current model suggests that transcription factors bind transiently to the DNA and recruit remodelling complexes such as SWI/SNF. These complexes then remodel the region around the transcription factor binding site, thereby giving the general transcription machinery a window of time in which to associate with the promoter (Voss et al. 2011).

NDR formation facilitates the incorporation of alternative histone variant H2A.Z, a characteristic of active promoters (Yen et al. 2013; Ranjan et al. 2013). Incorporation of H2A.Z, together with H3K56 acetylation, destabilises the nucleosome (Williams et al. 2008), likely facilitating RNAP II escape from the promoter. In addition, H2A.Z contributes to the recruitment of chromatin remodellers and transcription factors, further stimulating transcription (Draker et al. 2012). Chromatin remodellers such as RSC are also recruited to the NDR in a histone acetylation-dependent manner, which is required for NDR maintenance (Hartley & Madhani 2009).

1.3.2 Transcription elongation

Once the RNAP II is poised upstream of the TSS, transcription can be initiated. The presence of RNAP II alone, however, does not necessarily lead to

transcription initiation. In particular in mammals, RNAP II frequently pauses proximal to the promoter (Adelman & Lis 2012).

After successful initiation and promoter escape, RNAP II progresses over the TU. This requires the exchange of GTFs for transcription elongation factors (TEFs). The sequential recruitment of these factors is mediated by the differential phosphorylation of heptameric repeats in the C-terminal domain (CTD) of RNAP II subunit Rpb1p (Figure 1.6; Buratowski 2003; A Mayer et al. 2010; Mayer et al. 2012; Harlen et al. 2016), in particular at Ser5 and Ser2 (Corden 2016). In a first step, heptameric CTD repeats are phosphorylated at Ser5 by Kin28p, stimulating the recruitment of the capping complex, which catalyses the co-transcriptional formation of the 7-methylguanylate (m7G) 5'cap during the early stages of the transcription cycle (Mandal et al. 2004). Next, TEFs replace the GTFs and stimulate transcription elongation (Hirtreiter et al. 2010; Grohmann et al. 2011). Over the course of elongation, Ser5p is replaced by Ser2p, a step which correlates with the exchange of TEFs for termination factors, in particular Pcf11p (Lunde et al. 2010). RNAP II can also be phosphorylated at Ser7 and Tyr1. Moreover, the separation of phosphorylation marks is not always strict and overlaps frequently occur (Kim et al. 2010).

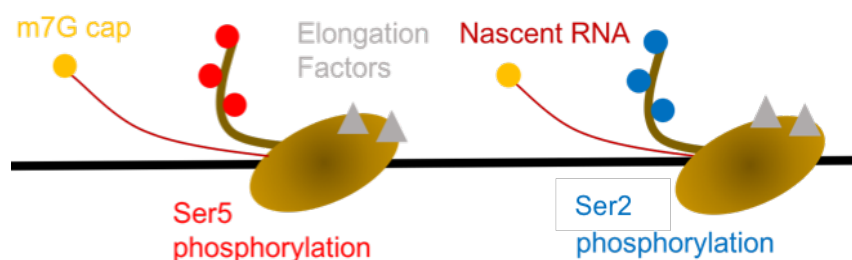


Figure 1.6. Phosphorylation patterns of RNAPII.

Progression of RNAP II also requires the action of several multi-subunit complexes that remove the nucleosome barrier by transiently displacing H2A/H2B (Belotserkovskaya et al. 2003; Pavri et al. 2006). This process is supported by several histone chaperones. Active transcription correlates with the accumulation of histone modifications, including methylation of H3K4 at the 5' end and H3K36 methylation at the 3' end of the gene among others, which are deposited co-transcriptionally by histone-modifying enzymes (Tanny 2015).

1.3.3 Transcript processing, polyadenylation and transcription termination

Transcription termination is the final step of the transcription process and can occur via two main pathways in *S. cerevisiae*. Mammalian termination mechanisms are not discussed in detail in this section. One of the pathways present in yeast, called the cleavage and polyadenylation factor complex dependent termination pathway, mainly targets nascent mRNAs, which are destined to be exported from the nucleus. The second one, called the Nrd1p-Nab3p-Sen1p (NNS) pathway, specialises in the termination of snoRNAs and snRNAs as well as pervasively transcribed and defective transcripts.

1.3.3.1 Cleavage factor and polyadenylation complex dependent termination

The Cleavage Factor and Polyadenylation Complex (CFPC)-dependent termination pathway involves the recruitment of several multimeric complexes to the 3' UTRs of protein-coding transcripts. These complexes are called the cleavage factors (CF) I A/I B, cleavage and polyadenylation factor (CPF) and

the Rat1p. Ultimately, their recruitment to the nascent RNA and RNAP II leads to the cleavage of the transcript at a polyadenylation site (PAS), followed by the polyadenylation of the nascent transcript and transcriptional termination.

Two events aid the recruitment of the CPF and CF complexes. First, Pcf11p, a component of the cleavage factor I A complex (CF I A), binds to the Ser2 phosphorylated CTD at the end of the gene (Lunde et al. 2010). Pcf11p's partner Clp1p then directly interacts with the CPF complex, causing its recruitment to the 3' end of the RNA (Ghazy et al. 2012).

Second, RNA-binding proteins in the CPF and CF complexes itself recognise motifs in the 3'UTR resulting in the association of the CPF-CF complex with the RNA (Figure 1.7). The *S. cerevisiae* termination signals consist of four distinct elements, namely an efficiency element reminiscent of the TATA box (EE, TAYRTA Y; can be any pyrimidine, R any purine), a positioning element (PE, AAWAAA; W can be an adenine (A) or thymine (T)), U-rich regions upstream and downstream of the cleavage site and the cleavage site (Guo & Sherman 1996). Unlike human cells which have a clearly defined polyadenylation sequence, yeast can tolerate considerable variation at all of these elements. While the presence of these elements as well as structural elements enhances cleavage efficiency (Moqtaderi et al. 2013), yeast does not strictly require all elements to be present for cleavage and polyadenylation to occur (Dichtl & Keller 2001). EE and PE located approximately 60-30 nucleotides upstream of the cleavage site are bound by CF IB and CF IA, while U-rich elements are bound by CPF (Figure 1.7 (A, B), Mischo & Proudfoot 2013).

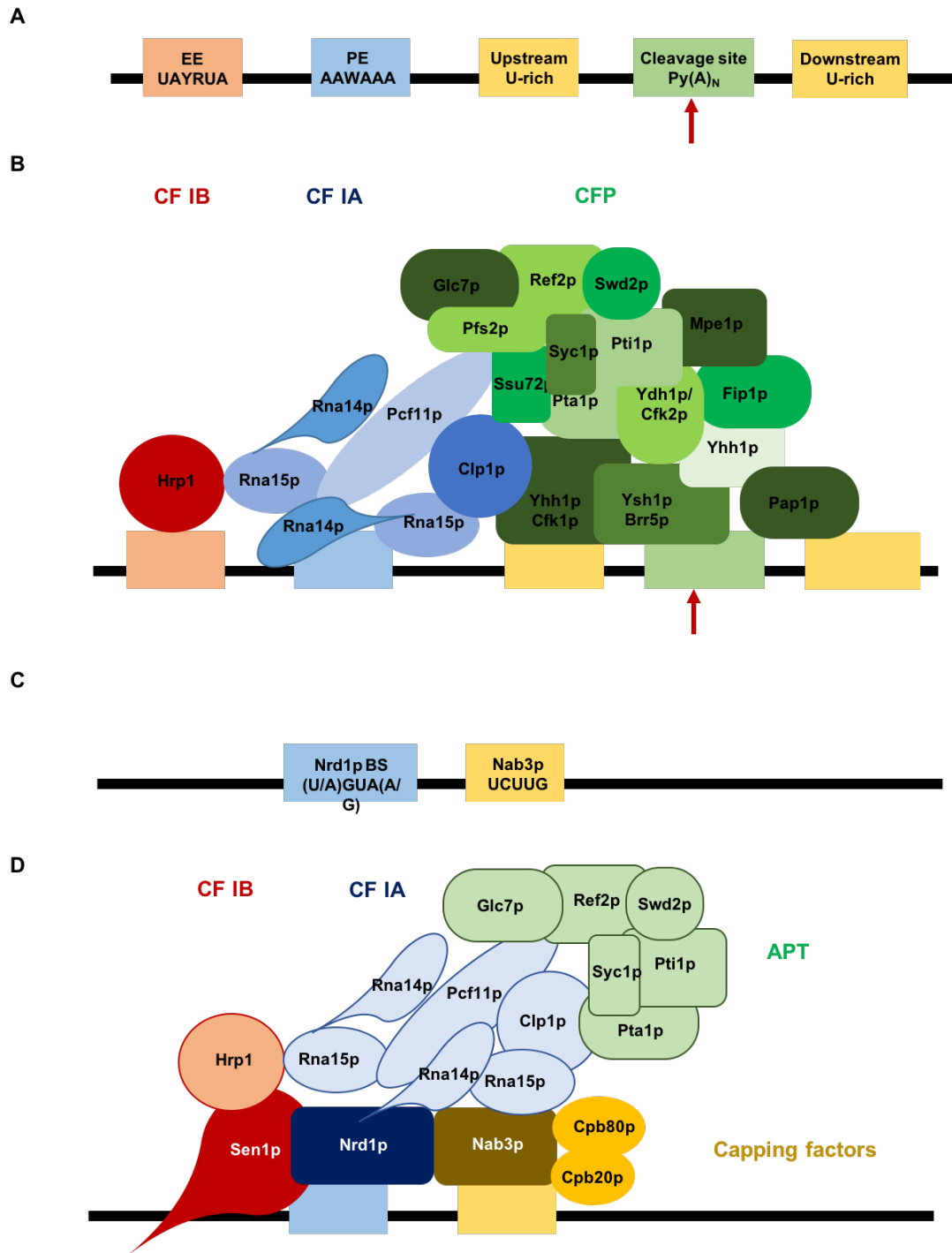


Figure 1.7. Termination complexes.

(A) Sequence elements associated with polyadenylation. EE- efficiency motif. PE- positioning motif. (B) CPFC complex components. (C) NNS binding sites (BS). (D) NRD-CPF-APT (expected composition). Adapted from Mischo & Proudfoot 2013.

The binding of the CF complexes is the primary determinant of the choice of cleavage site. By contrast, the cleavage and polyadenylation activities themselves are contained in the CPF complex, whereby the endonuclease Ysh1p ultimately cleaves the nascent transcript at the PAS after CPF association with the RNA (Mandel et al. 2006). After cleavage, the CPF-associated polymerase Pap1p catalyses the addition of adenosines to the end of the transcript (Ezeokonkwo et al. 2012). The newly formed poly(A) tail is then bound by Nab2p or Pab1p in order to stabilise the transcript and to target it for nuclear export (St-Sauveur et al. 2013; Schmid et al. 2012).

The exact position of the cleavage site is only minimally defined in yeast. In fact, rather than having one polyadenylation site, most genes have multiple polyadenylation sites, which occur in a 300bp window after the STOP codon (Ozsolak et al. 2010; Moqtaderi et al. 2013). PAS choice is regulated and has implications for RNA stability (Geisberg et al. 2014).

1.3.3.2 PAS-dependent termination

Cleavage of the transcript does not automatically lead to RNAP II transcription termination. In fact, RNAP II continues to transcribe RNAs downstream of the PAS.

Two main models have been proposed for transcription termination after cleavage (Figure 1.8). The allosteric model, which is mainly based on *in vitro* data, suggests that PAS-dependent termination is mediated by direct binding of CPFC components, such as Pcf11p, to the RNAP II. This causes a conformational change that dislodges the elongation complex due to the absence of anti-termination factors (Richard & Manley 2009).

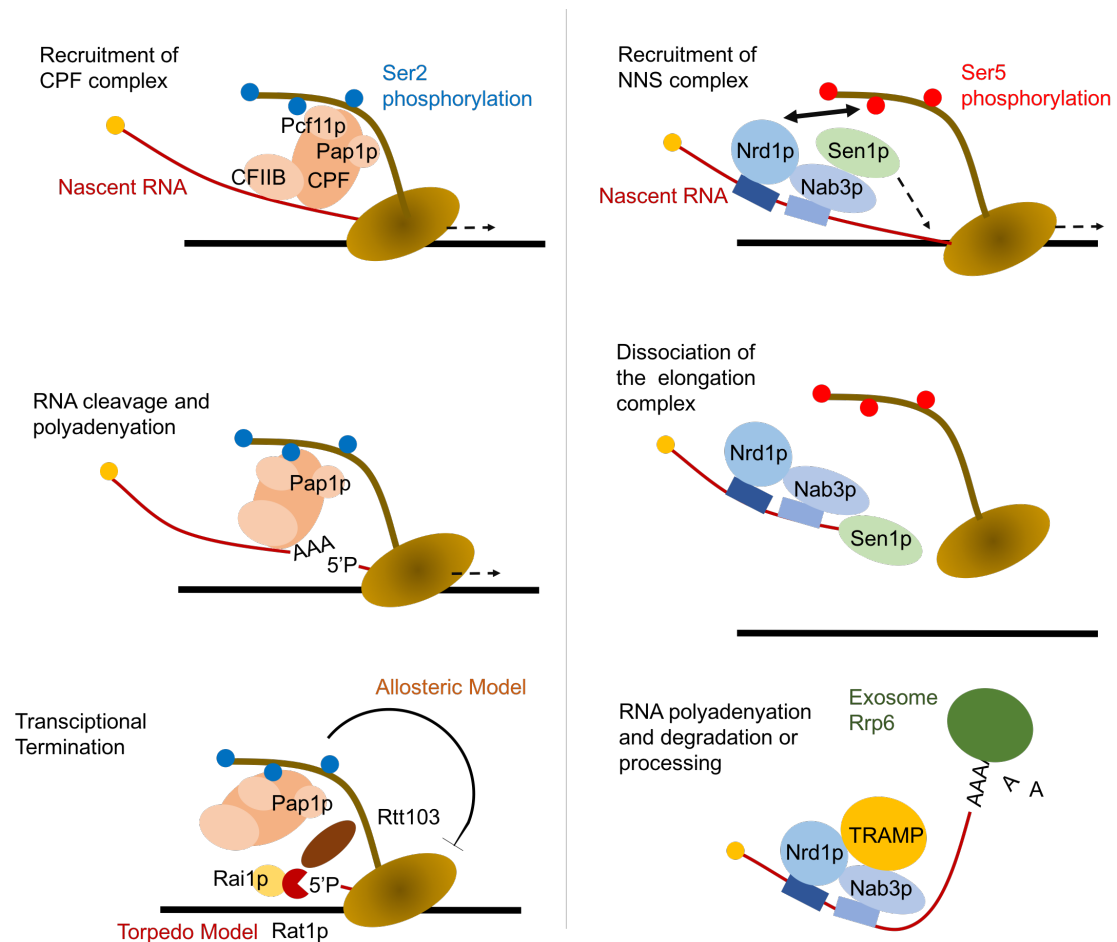


Figure 1.8. Models of PAS-dependent and PAS-independent transcriptional termination.

By contrast, the torpedo model proposes that Rat1p, the nuclear homologue of the cytoplasmic 5'→3' exoribonuclease Xrn1p, is required for efficient transcriptional termination. After the 5' phosphate becomes exposed owing to the cleavage of RNA at the PAS, the nascent RNA becomes a substrate for Rat1p. Rat1p degrades the nascent transcript co-transcriptionally and eventually catches up with RNA Pol II, triggering termination (Kim et al. 2004; West et al. 2004). The degradation of the nascent RNA does not appear to be the main cause of termination in this scenario. Instead, it allows Rat1p to associate with the transcribing polymerase and, together with Rtt103, to dismantle the elongation complex bound to RNAP II (Pearson & Moore 2013).

As both models are not mutually exclusive, it is possible that the Rat1p and CPF-CF act synergistically to trigger termination and degradation of the nascent transcript (Luo et al. 2006).

1.3.3.3 The NNS pathway

Transcription can also be terminated via the NNS (Nrd1p-Nab3p-Sen1p) pathway. Unlike the CPFC-dependent pathway, the NNS pathway mainly targets cryptic and short stable transcripts including snRNAs and snoRNAs (Jamonnak et al. 2011; Schulz et al. 2013; Creamer et al. 2011). The effector proteins in this pathway are not conserved in higher eukaryotes, but an equivalent mechanism appears to have been evolved (reviewed in Porrua and Libri 2015).

The NNS complex consists of RBPs Nrd1p and Nab3p and the RNA helicase Sen1p. Nrd1p and Nab3p bind to a (U/A)GUAA/G and UCUUG motif respectively (Carroll et al. 2007). These motives are abundant in unstable transcripts, but depleted from coding transcripts, leading to the preferential early termination of CUTs. Nrd1p associates with the RNAP II via Ser5 phosphorylated residues in the CTD (Vasiljeva et al. 2008), allowing the NNS complex to scan for RNA binding motifs in the nascent RNA. Sen1p is thought to release RNAP II from the transcript, potentially via a Rho-like mechanism, that is by unwinding the RNA in an ATP-dependent manner and causing the dismantlement of the RNAP elongation complex (Figure 1.7 (C, D); Porrua and Libri 2013).

Released transcripts are targeted for processing (snoRNAs) or degradation (CUTs) by the nuclear exosome exoribonucleases Rrp6p and Rrp44p and, in

the case of snoRNAs, the endonuclease Rnt1 after polyadenylation by the TRAMP (Trf/Air/Mtr4p Polyadenylation) complex. The NNS, exosome and TRAMP complex are able to interact physically with each other (Vasiljeva et al. 2008; Fasken et al. 2015). Nrd1p is able to recruit TRAMP after the RNAP II has dissociated leading to oligoadenylation of the RNA, which is then directly fed into the nuclear exosome complex (Wyers et al. 2005; Tudek et al. 2014). Although Rrp6p and the nuclear exosome primarily act as one complex, they have been shown to target some transcripts independently (Gudipati et al. 2012). Rrp6p has also recently been shown to also to directly influence the termination of a subgroup of NNS transcripts (Fox et al. 2015).

There is some overlap between the NSS and PAS-dependent pathways, as Pcf11p, for instance, is also able to contribute to termination via the NNS pathway (Grzechnik et al. 2015) and Nrd1p-Nab3p targets approximately a quarter of protein-coding transcripts (Webb et al. 2014). In addition, NNS components are able to bind CF IA/IB and a minimal version of the CPF, called the APT (associated with Pta1p) complex (Figure 1.7), which is thought to have functions in the 3' end processing of non-coding RNAs (Nedea et al. 2003).

1.4 RNA degradation

All RNAs must be degraded and, similarly to transcription, RNA degradation is a tuneable, highly dynamic and regulated process.

All known degradation pathways and their essential components are conserved from yeast to human, although some additional proteins fine-tune the RNA degradation machinery in mammals (Houseley & Tollervey 2009). Only pathways occurring in *S. cerevisiae* will be discussed here, although most of these pathways have been conserved in higher eukaryotes.

While the bulk of RNA processing and mRNA surveillance takes place in the nucleus, most functional RNAs are degraded in the cytoplasm, whereby degradation can be initiated via a variety of pathways (Parker 2012).

1.4.1 Canonical cytoplasmic RNA turnover

The majority of RNA turnover in the cytoplasm is initiated via a translation-independent mechanism (Figure 1.9). This means that degradation is initiated independently of translational activity, i.e. in the absence of, for instance, ribosome stalling on defective transcripts. RNA degradation of coding and non-coding cytoplasmic RNAs can be initiated at both the 5' and the 3' ends of transcripts by the 5'->3' and 3'->5' degradation machineries. However, it is thought that degradation in the 5'->3' direction is more prevalent (Hsu & Stevens 1993).

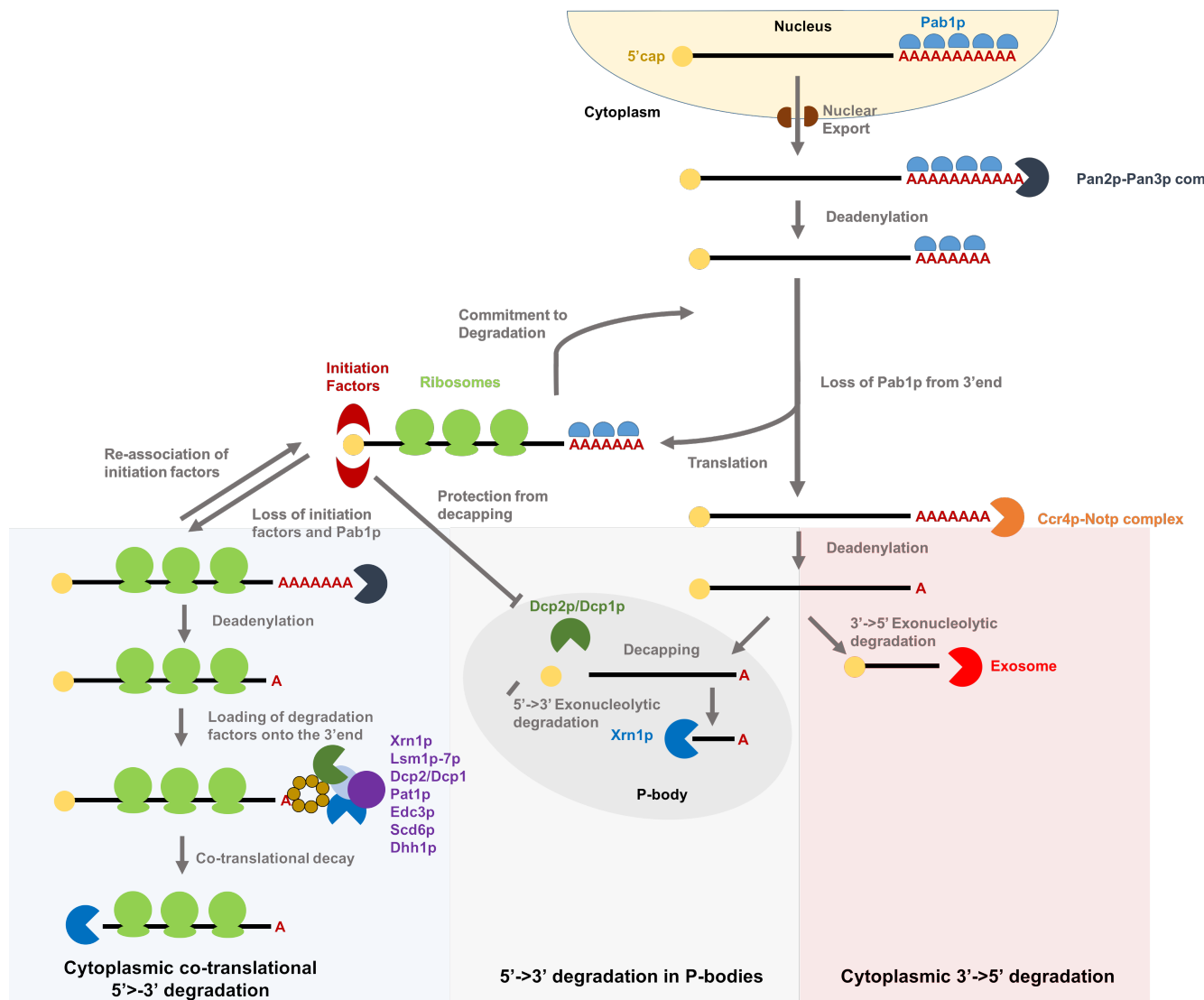


Figure 1.9. Canonical cytoplasmic decay pathways.

1.4.1.1 Deadenylation

For transcripts to become substrates for the RNA degradation machineries, they must be deadenylated and, at least in the case of the 5'→3' degradation pathway, decapped (Figure 1.9). Deadenylation along with decay itself are thought to be the rate-limiting steps in this process (Decker & Parker 1993; Muhlrad & Parker 1992).

Deadenylation is catalysed by two complexes, the Ccr4p/Pop2p/Notp and Pan2p/Pan3p deadenylation complexes, whereby Ccr4p and Pan2p are the respective catalytic subunits (Figure 1.9). These complexes target the same transcript consecutively in the wild-type. First, Pan2p, stimulated by Pan3p and polyadenylation binding protein (Pab1p), shortens the poly(A) tail from ~90 to ~65 nucleotides (Brown & Sachs 1998; Boeck et al. 1996). This step occurs co-translationally after cytoplasmic import and before the transcript has been committed to decay. Further deadenylation by Ccr4p/Pop2p/Notp is inhibited by Pab1p at this stage.

The second stage of deadenylation is only initiated once Pab1p has started dissociating from the poly(A) tail (Tucker et al. 2001). Deadenylation rates are regulated through interactions with RNA binding proteins and the deadenylation complexes. For instance, ~10% of mRNAs are targeted by Pufp proteins (Jackson et al. 2004; Ulbricht & Olivas 2008), a family of proteins that promotes deadenylation through direct interactions with Pop2p (Goldstrohm et al. 2006).

1.4.1.2 Decapping

After deadenylation, decapping is initiated once the cap-binding complex, consisting of eIF4E (the cap-binding protein), eIF4A (a DEAD-box RNA helicase) and eIF4G (the cap-binding complex scaffolding protein), bound to the 7-methylguanylate (m7G) cap at the 5' end of the RNA is exchanged for the decapping complex. (Tharun & Parker 2001).

On intact RNAs, the cap-binding complex is stabilised through direct interactions with Pab1p bound to the 3' end of the RNA, as Pab1p binding indirectly increases the cap-binding complex's affinity for m7G (Safaei et al.

2016; O'Leary et al. 2013). Once the poly(A) tail length reaches 10-12 residues, however, the poly(A) tail is too short to support Pab1p binding. Loss of Pab1p from the 3' end of the transcript results in a destabilisation of the cap-binding complex. The inhibition of translation, for instance by preventing the formation of the 48S translation pre-initiation complex, also accelerates the loss of the capping complex from the 5' end of the transcript (Coller & Parker 2005; Nissan et al. 2010).

Once the m7G is exposed, the RNA becomes a substrate for the Dcp2p/Dcp1p decapping complex. Dcp2p cleaves the 5' cap and releases 7-methylGDP and a RNA with an exposed 5' monophosphate. Exposed 5' phosphates are recognised by 5'->3' exoribonuclease Xrn1p resulting in the degradation of RNA (Floor et al. 2008).

Although Dcp2p and Dcp1p are sufficient for decapping (Dunckley 1999), rates are further stimulated by the binding of various decapping activators. These co-factors increase decapping rates by inducing a conformational change that increases the complex' affinity for RNA (Edc1p, Edc2p and Xrn1; Borja et al. 2011; Braun et al. 2012; Wurm et al. 2016) or by preventing Dcp2p from adopting an inactive confirmation (Smc6p and Edc3p; Harigaya & Parker 2010; Fromm et al. 2012)). Decapping is also stimulated by the Pat1/Lsm1-7 complex, a multi-subunit complex that assembles at the 3' end of transcripts after Pab1p has dissociated from the poly(A) tail (Wu et al. 2014). As both the Pat1p/Lsm1-7p complex and Xrn1p are integral components of the 5'->3' degradation machinery, this suggests that components of this decay pathway directly regulate decapping rates.

1.4.1.3 5'->3' degradation

Decapped RNAs are primarily hydrolysed by the exoribonuclease Xrn1p, which is recruited to the RNA via direct interactions with the Pat1p/Lsm1-7p complex and Dcp1p (Wurm et al. 2016; Nissan et al. 2010; Bouveret et al. 2000).

Hence, Pat1p acts as a scaffolding protein that, together with the Lsm1p-7p RNA helicase (Chowdhury et al. 2007; Chowdhury et al. 2014), acts as a platform in the assembly of the degradation machinery and connects the 5' decapping machinery with the 3' end of the RNA and the proteins bound there, thus connecting the RNA ends in a circular manner (Nissan et al. 2010; Sharif & Conti 2013).

Since Xrn1p primarily localises to cytoplasmic mRNA-processing bodies (P-bodies; Sheth and Parker 2003; Kedersha et al. 2005), Xrn1p-catalysed degradation has been suggested to primarily occurs in P-bodies, which are self-associating aggregates of translationally-repressed RNAs bound to components of the 5'->3' degradation, decapping, non-sense mediated decay and occasionally deadenylation machineries (Eulalio et al. 2007; Parker & Sheth 2007).

However, while P-bodies are major sites of decapping and deadenylation (reviewed in Decker and Parker 2012), whether mRNAs are primarily degraded in P-bodies remains controversial. Recent studies indicate that RNAs are degraded co-translationally by Xrn1p (Hu et al. 2009; Pelechano et al. 2015), suggesting that sequestration of RNAs into ribosome-free P-bodies is not a pre-requisite for the initiation of RNA decay.

Moreover, some capped mRNAs are recycled from P-bodies to stress granules,

where they associate with translation initiation factors and form stalled translation pre-initiation complexes, from which they can re-enter the translation cycle. Both the abundance of P-bodies and stress granules increase during stress conditions, when both RNA degradation and translation is repressed on a globally and RNAs are temporarily stowed away, suggesting that these aggregate structures act as storage units for RNAs (Balagopal & Parker 2009; Buchan & Parker 2009).

1.4.1.4 3'->5' degradation

RNAs can also be degraded co-translationally in the 3'->5' direction by the cytoplasmic exosome complex (Anderson & Parker 1998). The exosome consists of nine catalytically inert subunits, which form the exosomal ring and cap structures, and Rrp44p, an RNase with both exonucleolytic and endonucleolytic activities (Mitchell et al. 2013; Dziembowski et al. 2007; Bonneau et al. 2016). Deadenylated RNA is targeted to the exosome via the Ski RNA helicase complex and the Ski7p cofactor, which mediates interactions between the exosome and the Ski complex (Anderson & Parker 1998; Brown et al. 2000).

The RNA enters the exosome via its cap domain in an unwound conformation and then traverses the internal channel in a similar fashion to the one observed in the proteasome. At the end of the channel the RNA is degraded by Rrp44p (Schaeffer et al. 2009; Makino et al. 2013). Since combined knockouts of the Rrp44p exoribonuclease and endoribonuclease activities are synthetically lethal (Schaeffer et al. 2009), the exact targeting of RNAs to the cytoplasmic exosome remains relatively poorly understood. Current data suggests that the concentrations of fewer than 10% of mRNAs are affected by the mutation of

either the endonucleolytic or the exonucleolytic enzymatic sites of Rrp44p (Gudipati et al. 2012).

1.4.2 mRNA surveillance

Not all RNAs, which are exported to the cytoplasm, are fully functional. Some RNA transcript contain truncated ORFs; others have premature stop codons (PTCs) or chemical modifications that cause ribosomes stalling. All of these cases are picked up by the mRNA surveillance machinery in a translation-dependent manner.

1.4.2.1 Non-sense mediated decay

Nonsense-mediated decay or NMD is an integral part of the mRNA surveillance machinery (Figure 1.10). It is triggered through the recognition of sub-optimally positioned, premature stop codons (PTCs). NMD is primarily responsible for the degradation of mRNAs containing truncated ORFs, although mRNAs with full-length ORFs and upstream ORF (uORF)-containing lncRNAs can also be targeted by NMD (Wery et al. 2016; Durand & Lykke-Andersen 2013; Lykke-Andersen & Jensen 2015; He & Jacobson 2015).

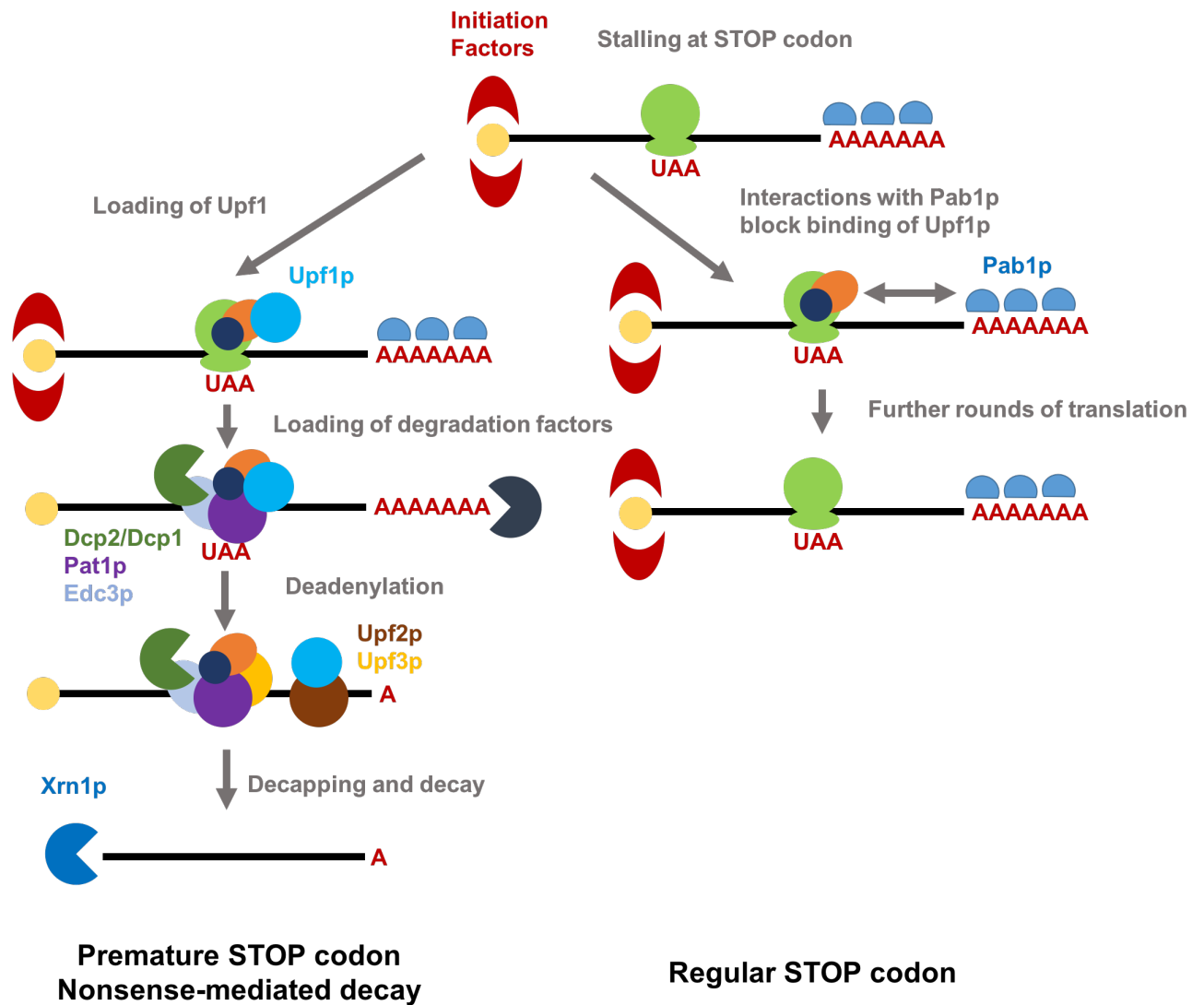


Figure 1.10. Nonsense mediated decay.

During the first step of NMD, the translation termination complex consisting of eukaryotic release factors eRF1 and eRF3 are recruited to the PTC. Once bound, eRF3 can directly interact with up-frameshift factor 1 (Upf1p) (Czapinski et al. 1995; Wang et al. 2001), which associates non-specifically with the 3'UTR in the absence of the termination complex. Upf1p together with Upf2p and Upf3p forms the UPF complex required for the recruitment of RNA degradation factors. Remodelling of the UPF complex, triggered by GTP hydrolysis, results in the release of the nascent peptide and eRF3, as well as an increased affinity

for decay factors and the Ccr4p/Pop2p/Notp complex. Binding of these factors increases deadenylation and decapping rates and ultimately causes the degradation of the transcript, predominately by Xrn1p (Shoemaker & Green 2012).

Interactions between eRF3 and the Pab1p block degradation via the NMD pathway (Amrani et al. 2004). For this reason, PTCs are recognised based on their distance from the poly(A) tail (Cao & Parker 2003; Singh et al. 2008) and, in addition to PTCs, long 3'UTRs and upstream ORFs can also trigger NMD, contributing to isoform specific decay rates (Smith et al. 2014; Wery et al. 2016). Since NMD is inactivated during stress, cells use this mechanism to control RNA and protein concentrations under these conditions (Gardner 2008).

1.4.2.2 No-Go decay

No-Go decay (NGD) targets mRNAs, which are occupied by stalled ribosomes (Figure 1.10). Stalls can occur when (a) the RNA contains stem loops or pseudoknots, (b) the RNA is chemically damaged, for instance, due to UV and γ -irradiation or (c) the RNA contains several consecutive non-optimal codons or amino-acids with the same charge. For a transcript to become a substrate for NGD, ribosomes must stall in the translation elongation phase, resulting in an empty aminoacyl-tRNA acceptor (A) site (Lykke-Andersen & Bennett 2014). Dom34p and Hbs1p, two proteins structurally similar to eRF1 and eRF3, bind to the A site and recruit an endonuclease via an unknown mechanism (Figure 1.11; Doma & Parker 2006). Endonucleolytic cleavage releases two RNA fragments that are substrates for Xrn1p and the exosome. The cleavage also triggers recycling of the ribosome by Dom34p and Hbs1p and proteosomal degradation of the nascent peptide (Tsuboi et al. 2012). NGD is particularly

important in the presence of oxidative stress, which damages RNAs (Simms et al. 2014).

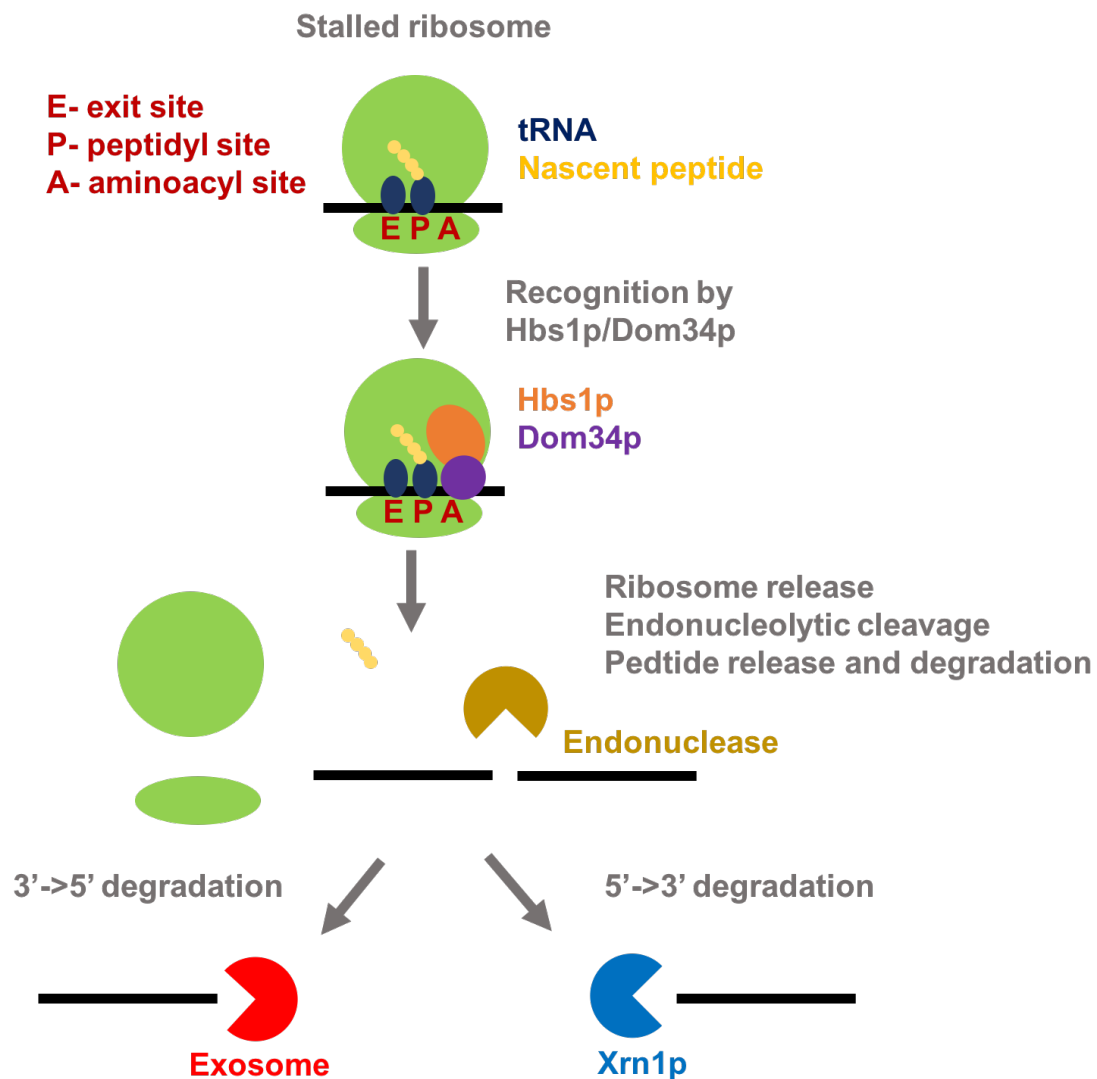


Figure 1.11. No-Go decay.

1.4.2.3 Non-Stop decay

Non-Stop decay (NSD) is a pathway related to NGD that removes RNAs with polyadenylated truncated ORFs without any STOP codons (Frischmeyer et al. 2002). The absence of a STOP codon causes ribosome stalling in the poly(A) tail, as interactions of polylysines in the nascent peptide with the positively

charged ribosome exit channel cause the ribosome to slow down. Alternatively, on short poly(A) tails, ribosomes are unable to progress, when they reach the end of the transcript (Klauer & van Hoof 2012). Stalling of the ribosome leads to the recruitment of Ski7p, a protein with a similar structure as Hsb1p, probably in conjunction with an unknown binding partner (van Hoof et al. 2002; Schaeffer & van Hoof 2011). Ski7p directly interacts with the exosome and targets the transcript to 5'→3' degradation. The nascent peptide is readily recognised by different E3 ligases, polyubiquitylated and degraded by the proteasome (Dimitrova et al. 2009).

1.4.3 Degradation of stable non-coding RNAs

Stable non-coding RNAs are protected from degradation through interactions with RNA binding proteins (RBPs) and chemical modifications that alter RNA structure. For instance, non-functional rRNAs cause ribosome stalling and are picked up by the NGD pathway or are degraded via a vacuolar pathway by autophagy (Cole et al. 2009). Moreover, mature tRNAs are only targeted for degradation by Xrn1p and its nuclear homolog Rat1p after becoming hypoacetylated resulting in an altered secondary structure (Whipple et al. 2011; Chernyakov et al. 2008). Similarly, pre-tRNAs with defects in their maturation or modification are readily degraded by the nuclear exosome (Vanáčová et al. 2005; Vanacova & Stefl 2007) as are pre-mature snoRNAs and rRNAs (Kadaba et al. 2006).

1.4.4 Nuclear decay

In addition to the processing of various RNAs, one main function of nuclear decay is to remove cryptic and defective RNAs before they can be exported to the cytoplasm.

There are three main classes of RNAs that are targeted for degradation in the nucleus, namely defective RNAs, spurious non-coding transcripts and RNAs that are by-products from nuclear processing after RNA transcription termination and intron splicing.

RNAs are considered defective when they fail to acquire a 5' cap or poly(A) tail or are not properly spliced or packaged. The formation of the 3' end is linked to nuclear export rates and RNAs lacking proper poly(A) tails are also retained in the nucleus (Hilleren & Parker 2001).

1.4.4.1 The nuclear exosome

Degradation of spurious transcripts, for instance CUTs, is primarily carried out by the nuclear exosome, in which Rrp6p binds to the core exosome complex and acts as the main catalytic subunit. In addition, Rrp6p can dissociate from the core exosome and form a complex with Rrp47p (Makino et al. 2015).

Rrp6p/exosomes, Rrp6p/Rrp47p and the core exosome complexes target different subgroups of nuclear transcripts, in particular during the RNA processing of snoRNAs, snRNAs, tRNAs, rRNAs and intron-containing genes (Schneider et al. 2012; Gudipati et al. 2016). In addition to its function in nuclear and processing decay, the exosome has been implicated in the nuclear retention of RNAs (Vodala et al. 2008), RNA surveillance, *i.e.* a quality control mechanism that prevents defective RNAs from being exported to the cytoplasm

(Figure 1.12 (A), Hilleren et al. 2001) and transcriptional termination, discussed in section 1.3.3.3 (Fox et al. 2015).

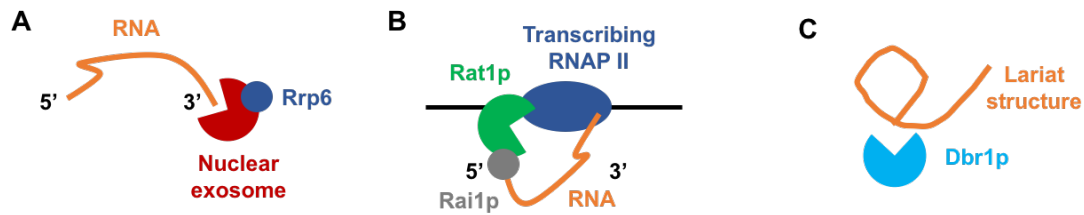


Figure 1.12. Nuclear degradation machinery.

1.4.4.2 RNA surveillance by the 5'->3' degradation machinery

RNAs that fail to acquire a 5' cap or poly(A) tail have an exposed 5' monophosphates are recognised by Rai1p and degraded by its binding partner, the Xrn1 homolog Rat1p. The Rai1p-Rat1p complex travels with the transcribing polymerase and removes, in particular, non-methylated cap structures before degradation (Figure 1.12 (B), Jiao et al. 2010).

1.4.4.3 Splicing

Unspliced pre-mRNAs are spliced by the spliceosome complex, leaving a functional mRNA and a branched intron lariat. These branched splicing by-products and mispliced RNAs are debranched by Dbr1p and exported to the cytoplasm, where they are degraded by Xrn1p and the cytoplasmic exosome (Figure 1.12 (C), Hilleren & Parker 2003).

1.4.5 Summary of RNA degradation pathways

A variety of specialised pathways degrade RNAs in the cytoplasm and nucleus. These pathways are responsible for canonical RNA degradation and RNA surveillance and quality control and can be regulated in a condition-dependent manner. Regulation of degradation rates frequently occurs upstream of the

degradation event itself, for instance, through the regulation of deadenylation. The final step in each pathway is the hydrolysis of RNAs by the 5'->3' and 3'->5' degradation machineries. The enzymes that catalyse degradation have partially complementary and overlapping functions, explaining why knockouts of some catalytic components of the RNA degradation machinery result in severe growth phenotypes but not lethality.

1.5 The dynamic nature of the RNA transcriptome

With very few exceptions, RNA is constantly being transcribed and degraded in cellular organisms. RNA levels change with growth conditions, cellular age, stress and differentiation. Moreover, under steady-state conditions, copy numbers of individual RNAs can vary over six orders of magnitude (Holland 2002). Maintenance of appropriate RNA copy numbers requires the coordination of RNA synthesis and degradation (Tavazoie et al. 1999; Cacace et al. 2012). Hence, both transcription and RNA degradation rates are regulated dynamically during cyclic transcription programs, for instance during the cell-cycle (Eser et al. 2014), and in response to changes in environmental conditions (Pelechano et al. 2010; Castells-Roca et al. 2011; Vasudevan & Peltz 2016). Some well-characterised systems are discussed below.

1.5.1 Transcriptional induction and suppression

Most experiments in yeast are carried out under optimal growth conditions, during which *S. cerevisiae* are able to grow in the exponential phase. However, when conditions become suboptimal, cells are able to respond through the

transcriptional induction of specific genes, for instance by switching on *GAL* genes in response to a change in carbon source availability, as well as by undergoing generic changes in the form of the environmental stress response.

1.5.1.1 The environmental stress response

The environmental stress response (ESR) refers to a global change in cellular transcript concentrations that is induced after exposure to a stimulus, which ultimately results in cellular stress, down-regulation of growth rates and changes in cellular morphology. This generic response affects the RNA abundance of approximately 27% of all yeast genes, largely independently of the precise stimulus (Brauer et al. 2008; Gasch et al. 2000). Stimuli include lack of nutrients, the presence of harmful chemicals as well as thermic heat. In addition, mutants with slow-growth phenotypes inherently have RNA transcriptome signatures reminiscent of the ESR (O'Duibhir et al. 2014). Relative RNA concentrations are remarkably similar under all of these circumstances, although the exact time course of the response can vary with the type of stress and its severity (Gasch et al. 2000).

It is thought that the occurrence of the ESR signature reflects changes in the average metabolic state and the relative distribution of cell cycle phases in the population (Slavov et al. 2012; Slavov & Botstein 2011). Metabolism also plays a role. In *S. cerevisiae*, the yeast metabolic cycle (YMC) is characterised by three distinct phases (oxidative, reductive building and reductive charging) with different rates of oxygen consumption, intra-cellular NADPH levels and transcriptional programmes (Tu et al. 2005; Kuang et al. 2014). The YMC can occur independently of the cell cycle. However, when the two cycles coincide, the onset of DNA replication is timed so that it does not temporally overlap with

oxidative growth to avoid DNA damage by reactive oxygen species (Slavov et al. 2011; Burnetti et al. 2016).

Stressed cells grow more slowly, as stress effectively causes cells to remain in the G0 phase of the cell cycle for a longer period of time. In addition, available energy sources are limiting so that oxidative growth becomes more prevalent, while energy-intensive translation is repressed. These features are characteristics of the reductive charging phase of the YMC. Overall, autophagy, vacuolar, and stress response genes are transcriptionally induced in response to stress and low growth rates. By contrast, ribosomal, translation, and mitochondrial genes are repressed (Brauer et al. 2008; Slavov & Botstein 2011).

Changes in RNA abundance occur, in part, due to the regulation of nascent transcription rates by cooperatively-acting transcription factors (Miller et al. 2011). On the other hand, RNA abundance during the ESR is also controlled by changes in RNA decay rates, as stress and growth rates lead to a general stabilisation of RNAs. The exception to this rule are Ribosomal Protein Genes (RPGs) and Ribosomal Biogenesis Genes (RiBis), which are degraded rapidly after the onset of stress (Munchel et al. 2011; García-Martínez et al. 2016).

To summarise, the ESR is a generic response that is activated in the presence of cellular stress due to changes in growth rates and cellular metabolism. It is defined by specific changes in RNA abundance and regulated both through the adjustment of nascent transcription and RNA decay rates.

1.5.1.2 The glucose to galactose Shift

The glucose to galactose shift in *S. cerevisiae* is one of the best-studied systems of transcriptional regulation across all model organisms. In essence, yeast responds to glucose depletion in the presence of galactose by switching on genes required for galactose metabolism, while simultaneously repressing glucose catabolism genes.

Since anaerobic glucose catabolism facilitates high growth rates, glucose is yeast's preferred carbon source and is exclusively metabolised when available at high concentrations in favour of alternative carbon sources (Schuller 2003).

To conserve energy, genes required for alternative carbon source catabolism are repressed in the presence of glucose (Figure 1.13 (A)). One important effector of glucose-dependent catabolite repression is Mig1p (Nehlin et al. 1991), a protein that is imported into the nucleus in the presence of high glucose concentrations (De Vit et al. 1997).

Yeast sense the availability of glucose through signalling via the non-transporting glucose carrier homologs Snf3p and Rtg2p and intracellular AMP/ATP ratios. Glucose depletion leads to a change in intracellular signalling and an increase in AMP/ATP ratios. These changes activate the AMP-dependent protein kinase Snf1p (Kim et al. 2013). One of Snf1p's targets is Mig1p, which is inactivated by phosphorylation, resulting in the derepression of respiratory genes and high-affinity hexose transporters (Figure 1.12 (A); Ostling & Ronne 1998). Furthermore, activatory transcription factors including Adr1p, Sip4p, Cat8p, Hsf1p and Msn2p, among others, are phosphorylated by the Snf1p-Snf4-Gal83p and thereby activated, leading to the transcriptional

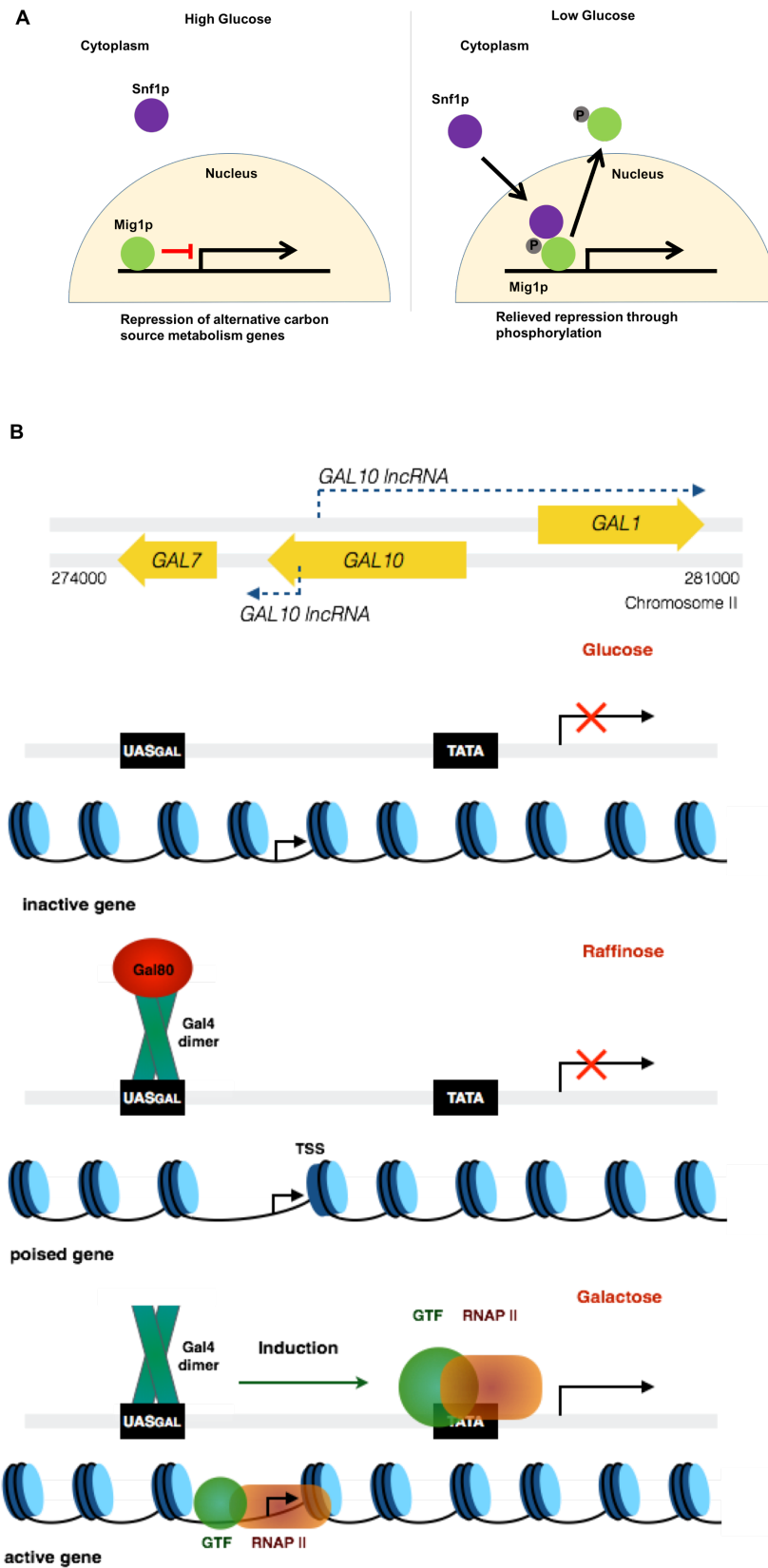


Figure 1.13. Glucose repression and *GAL* gene induction.

(A) Glucose repression by Mig1p. (B) Promoter architecture and transcription of galactose genes in glucose, raffinose and galactose.

induction of genes involved in alternative carbon source and fatty acid metabolism, gluconeogenesis and stress response genes. Snf1p also regulates RNA degradation rates of Mig1p-dependent RNAs in the cytoplasm by phosphorylating the exoribonuclease Xrn1p (Kayikci & Nielsen 2015).

Apart from generic changes that occur due to glucose depletion, changes in the glucose to galactose ratio lead to the transcriptional induction of the *GAL* regulon (Escalante-Chong et al. 2015), which consists of a total of 12 genes including four genes that are essential for the uptake of galactose (*GAL2*) and its metabolism (*GAL1*, *GAL7*, *GAL10*) (reviewed in Weinhandl et al. 2014). The *GAL* regulon is primarily under the control of the Gal4p (Griggs & Johnston 1991), a transcriptional activator that dimerizes and, with the help of the chromatin remodeller complex RSC (Floer et al. 2010), binds to upstream activator elements (UAS) in the presence of galactose (Lohr & Lopez 1995). Binding at the UAS leads to the recruitment of essential components of the transcriptional machinery including the histone acetyltransferase and deubiquitylase SAGA, TFIID and the transcriptional master-regulator Mediator (reviewed in Traven et al. 2006). Gal4p is transcriptionally repressed by Mig1p and transcribed at 5-fold lower levels in glucose than in any other carbon source (Trumbly 1992). Gal4p dimers are formed and bind the UAS_{*GAL*} in carbon sources other than glucose, for instance raffinose, but are unable to stimulate gene-expression in the absence of galactose, as the Gal4p is bound and inactivated by Gal80p (Figure 1.13 (B)). This repression is relieved upon galactose-dependent binding of Gal3p to Gal80p (Lavy et al. 2012), thus resulting in the transcription of *GAL* regulon genes. *GAL80* is also part of the *GAL* regulon so that *GAL* transcription ultimately reaches a steady state due to

this inhibition feedback loop. Moreover, the rate of induction of individual genes depends on the number of UAS_{GAL} at the promoter that can be bound by Gal4p (Lohr & Lopez 1995). Several other factors are also thought to contribute to the induction of *GAL* transcription. For instance, once translated, the galactokinase Gal1p, a homologue of *GAL3* can also impair Gal80p binding to Gal4p (Zacharioudakis et al. 2007). Gal1p is detectable for up to seven generations after exposure to galactose leading to a memory effect that results faster re-activation of transcription during a secondary exposure (Kundu et al. 2007; Brickner et al. 2007). Furthermore, the Reb1p-driven transcription of a lncRNA at the *GAL10/GAL* locus (*ncGAL10*) has been also been implicated in the regulation of *GAL1* and *GAL10*, although this view remains controversial (Houseley et al. 2008). Initial studies show *ncGAL10* contributing to a Rpd3S complex-dependent glucose repression in *cis* (Houseley et al. 2008; Geisler et al. 2012), while later studies suggest that the presence of the *ncGAL10* transcript leads to the formation of a RNA-DNA hybrid structure (R-loop) and subsequent transcriptional induction (Cloutier et al. 2013; Cloutier et al. 2016). It has also been suggested that the transcription of *ncGAL10* has no significant effect on the induction of *GAL1* and *GAL10* in galactose, but is important for preventing transcriptional leakage (Lenstra et al. 2016).

1.5.2 Degradation rate dynamics

Ultimately, the balance between transcription and degradation rates determines

RNA concentrations in the cell. Therefore, RNAs which are present at similar concentrations can have vastly different half-lives, which are often balanced with RNA transcription rates under steady-state conditions (Shalem et al. 2011; Dori-Bachash et al. 2012).

However, degradation rates are not static and can be adjusted, for instance in response to stress (Munchel et al. 2011; Castells-Roca et al. 2011). Another example of the dynamic adjustment of degradation rates occurs during the cell cycle, where periodic, non-contemporaneous changes in transcription and degradation rates are required to shape expression peaks (Eser et al. 2014).

In spite of this, how RNA stability is determined remains relatively poorly understood, although some general principles have been established. Generally speaking, RNA stability correlates negatively with transcript length (Geisberg et al. 2014; Feng & Niu 2007; Barrass et al. 2015) and, in the case of mRNAs, positively with codon-optimality (Presnyak et al. 2015). Overall this suggests that genes that are more readily translated are generally more stable. High rates of translation probably cause a reduction in co-translational decay by preventing access of the 5'->3' degradation machinery to the RNA (Hu et al. 2009; Pelechano et al. 2015), thus resulting in increased RNA stability.

Both 5' UTRs and promoters have also been shown to affect RNA half-lives, although translation-independent regulation of stability appears to occur via the 3' end (Trcek et al. 2011; Bregman et al. 2011). RNA stability is not *per se* linked to 3' UTR length; rather it is influenced by the binding of proteins and changes in RNA structure. 3' UTRs frequently contain one or multiple binding motifs for RBPs (Hogan et al. 2008) and, in some cases, differential loading of

RPBs onto alternative 3' isoforms can explain differences in their stabilities (Gupta et al. 2014).

10% of yeast transcripts also contain U-rich motifs that can stabilise RNAs by increasing the amount of secondary structure through base-pairing with the poly(A) tail (Geisberg et al. 2014), thereby blocking the access of deadenylases and components of the degradation machinery. In contrast to differential RBP binding, RNA secondary structure confirmations can be altered even by single nucleotide changes in the 3' UTR. This is important, as in yeast, most genes give rise to a multitude of highly similar RNA isoforms, some of which differ by as little as one nucleotide (Pelechano et al. 2013; Mata 2013; Moqtaderi et al. 2013). Despite these minute changes to their primary sequence, many of these isoforms have been shown to have significantly different half-lives (Geisberg et al. 2014).

In summary, this suggests that the selection of RNA stabilities is regulated dynamically and, at least in part, independently from RNA transcription rates. Up to date, only the regulation of approximately a third of yeast RNAs can be mechanistically explained, but it is evident that the stabilisation of individual RNAs is non-random.

1.6 Aims and outline

Broadly speaking, the aim of this thesis is to further characterise links and feedback mechanisms between chromatin structure, transcriptional processes and stable RNA concentrations. As outlined in this introduction, DNA

packaging, transcription and RNA degradation are all highly dynamic and interconnected processes that ultimately determine cellular RNA pools. Although these processes are frequently co-regulated, most studies focus on one particular aspect of these processes. Hence, in order to obtain a more complete picture of this co-regulation, it is important to examine these processes in parallel.

By comparing genome-wide data sets recording RNA nucleosome positioning, transcription rates, the concentration of RNAs and the positioning of 3' ends in wild-type yeast and deletion mutants of the exoribonucleases *XRN1* and *RRP6*, the studies conducted for this thesis are able to capture differences in RNA transcription and chromatin structure that are a direct or indirect result of RNA stabilisation.

The thesis has been structured into three results chapters.

The focus of Chapter 3 is the description of sequencing techniques used throughout this thesis as well as to establish links between RNA transcription, nucleosome depletion and 3' end formation. The data shows that formation of 3' ends of stable RNAs is linked to a slow-down in RNA polymerase progression and a depletion of nucleosomes. Moreover, wild-type yeast accumulate stable transcripts that are terminated within the gene body of coding genes, likely via non-canonical roadblock-like mechanisms.

Chapter 4 characterises transcriptional dynamics in an *rrp6Δ* strain. The absence of Rrp6p leads to a global stabilisation of transcripts that are terminated by the NNS complex. As the 3' ends of Rrp6p-targeted RNAs are clearly distinguishable from those observed in the wild-type, it is therefore

possible to directly compare dynamics of CPFC-termination and NNS termination at the level of nucleosome depletion, transcription protein binding and 3' end sequence diversity. The data supports a function for NNS/Rrp6p in the removal of products of early termination at both coding and non-coding genes, in a manner that is independent of nucleosome depletion. Moreover, some Rrp6-targets are terminated and cleaved at canonical cleavage sites, suggesting that even unstable RNAs can be either subjected to 3' end processing or are terminated via a fail-safe mechanism.

The aim of Chapter 5 was to determine whether and under which circumstances Xrn1p influences transcriptional rates and uncover the underlying mechanisms behind this regulation. Xrn1p has been linked to transcriptional repression and transcriptional induction (discussed in Chapter 5). However, the role of Xrn1p in transcriptional regulation remains controversial. Although the stabilisation of non-coding RNAs has been investigated most thoroughly in recent years, RNA-Seq data shows that the deletion of *XRN1* leads to the stabilisation of most coding transcripts including some that are likely to stem from leaky transcription. Moreover, the *xrn1Δ* mutants are characterised by specific changes in their nucleosome occupancies, in particular nucleosome depletion at transcription factor binding sites, linked to specific changes in transcription rates and early transcriptional induction during glucose to galactose shifts. Although nobody has investigated protein concentrations in *xrn1Δ* mutants up to this point, it is likely that these changes are secondary effects due to the accumulation of transcription factors and chromatin remodellers.

In summary, this thesis focuses on different aspects of RNA metabolism, by examining direct links between chromatin structure and transcription, in particular transcriptional termination. Moreover, it tries to address the question as to how and in what manner the stabilisation of RNAs in the cytoplasm and nucleus contributes to a dysregulation of transcriptional induction.

2 Materials and Methods

2.1 Strains and transformations

All *S. cerevisiae* strains used in this study were based on the BY4741 strain (Euroscarf) and are listed in Table 2.1. Deletion strains and tagged strains were constructed, as previously described, by inserting a PCR-amplified selection marker into an Open Reading Frame through homologous amplification (Longtine et al. 1998). In brief, for deletion strains 40bp primers with ends homologous to the flanking regions of the desired deletion site were used to amplify the selection marker from a pFA6a-*KanMX6*, pFA6a-*HIS3MX6*, pFA6a-*TRP1* or pFA6a-*LEU2* plasmid to confer resistance to kanamycin or complement histidine, tryptophan or leucine prototrophy respectively. Analogously, for tagged strains *FLAG*, *HA* or *GFP* sequences were amplified with primers that allowed in-frame insertion at the 3' end of the gene upstream of the stop codon using pFA6a-*GFP*, pFA6a-*HA* plasmids or pFA6a-3x*FLAG* (plasmid constructed by Fischl, H).

For transformation, yeast were grown to mid-log phase in a 50 mL culture, pelleted (3 minutes, 5000 rpm) and permeabilised by resuspension in 450 µl 0.1M LiAc/TE (>1 hr, 4°C). 100 µl of cells were mixed with 10 µl of purified PCR-product (Qiagen PCR purification kit), 10 µl calf thymus DNA (Sigma) and 700 µl 0.1 M LiAc/TE/40% PEG, incubated at 30°C for 30 minutes, heat-shocked for 20 minutes at 42°C, pelleted, resuspended in purified water and plated onto agar selection plates. In the case of kanamycin selection, strains were recovered for 4 hours at 30°C. Strains were screened by PCR and confirmed by sequencing.

Strain	Parent	Genotype	Origin
BY4741	S228C	<i>MATα;his3Δ1;leu2Δ0; met15Δ0;ura3Δ0;</i>	Euroscarf
<i>xrn1-Δ</i>	BY4741	<i>xrn1-Δ::kanMX</i>	Euroscarf
<i>rrp6-Δ</i>	BY4741	<i>rrp6-Δ::kanMX</i>	Euroscarf
<i>sua7-1</i>	BY4741	<i>sua7-1::kanMX</i>	Howe, F.
<i>bem1-Δ</i>	BY4741	<i>bem1-Δ::kanMX</i>	Euroscarf
<i>GAL10-GFP</i>	BY4741	<i>GAL10-GFP::kanMX6</i>	Nair, A.
<i>xrn1-Δ GAL10-GFP</i>	<i>xrn1-Δ</i>	<i>xrn1-Δ GAL10-GFP::kanMX6</i>	Maitland, E.
<i>XRN1-HA</i>	BY4741	<i>XRN1-HA::HIS3MX6</i>	Nair, A.
<i>RPB3-3xFLAG</i>	BY4741	<i>RPB3-3xFLAG::HIS3MX6</i>	Fischl, H.
<i>xrn1-Δ RPB3-3xFLAG</i>	<i>RPB3-3xFLAG</i>	<i>xrn1-Δ::LEU2 RPB3-3xFLAG::HIS3MX</i>	Fischl, H.
<i>rrp6-Δ RPB3-3xFLAG</i>	<i>RPB3-3xFLAG</i>	<i>rrp6-Δ::kanMX6 RPB3-3xFLAG::HIS3MX</i>	This study
<i>sua7-1 RPB3-3xFLAG</i>	<i>RPB3-3xFLAG</i>	<i>Sua7-1::kanMX6 RPB3-3xFLAG::HIS3MX</i>	Fischl, H.
<i>S. pombe</i>		<i>u⁺,ura4-Δ18,leu1-32,ade6-M216</i>	Wittman, S.

Table 2.1. Yeast strains.

2.2 Yeast media and growth conditions

S. cerevisiae strains were streaked out from glycerol stocks onto YPD (1% yeast extract (Difco), 2% bactopectone, 2% glucose) containing agar plates (YPDA, YPD, 2% agar (Difco)) and grown for a minimum of two days at 30°C. For liquid culture experiments, the cells were resuspended in YPD and grown overnight. The following day, a fresh YPD culture was inoculated at OD₆₀₀ 0.2 and grown to approximately OD₆₀₀ 0.6 (30°C, 170 rpm). *S. pombe* was treated in an equivalent manner, but grown in YES (0.5% yeast extract, 3% glucose, 0.0225% adenine,

0.0225% histidine, 0.0225% leucine, 0.0225% uracil, 0.0225% lysine hydrochloride).

2.3 Glucose to galactose Shifts

For glucose to galactose shifts, yeast were grown as described in 2.2, but pelleted (3 minutes, 5000rpm) at approximately OD₆₀₀ 0.55, transferred into YPG (1% yeast extract (Difco), 2% bactopectone, 2% galactose) and collected after 15, 30 or 60 minutes in galactose.

2.4 Drop plates

To assess growth rates, the yeast were grown as described in 2.2 (exponential growth phase) or left to grow for three days (stationary phase) and exposed to one of the following: YPDA (no stress), Complement Supplement Mixture Media (CSM, 0.025% yeast nitrogen base, 2% glucose, 2% agar), acute hydrogen peroxide stress (10mM for exponential phase cultures or 100mM for stationary phase cultures for 60 minutes), acute heat shock (50°C water bath for 10 minutes), acute heat shock followed by acute hydrogen peroxide stress, DMSO, 12mM caffeine in YPDA, 100ng/μl cyclohexamide in DMSO and YPDA, 150nM rapamycin in YPDA, 1 M NaCl or 100mM hydroxyurea in YPDA. Hydrogen peroxide and heat shock were applied to liquid media cultures. All other chemicals were mixed into the agar containing media. The yeast were serially 10-fold diluted six times from OD₆₀₀ 0.4 and drop plated onto the stress plates. The plates were scanned daily and analysed by calculating the cumulative intensity of dilution drops 1-3 over four days. Growth rates were assessed by measuring the intensity of these drops in ImageJ. Growth rates of mutants

strains were normalised to the wild-type growth rate as well as their own growth rate in optimal growth conditions (YPDA without stresses). An example of these calculations is shown in Figure 2.1.

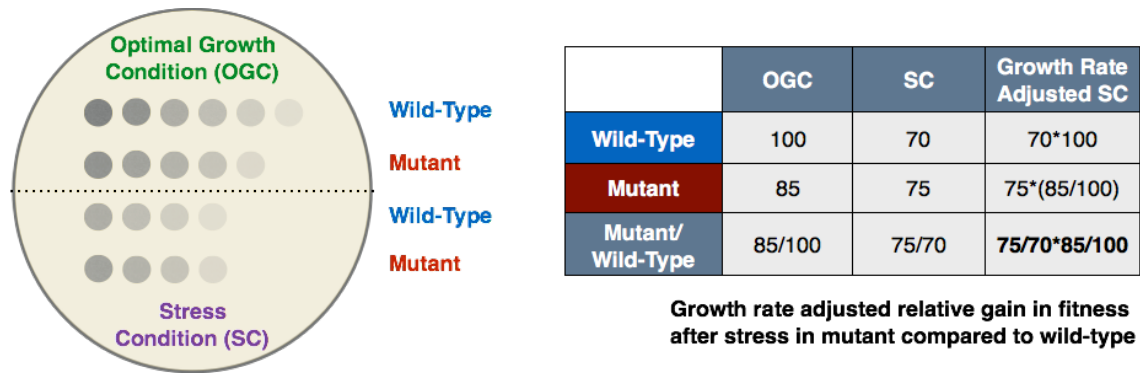


Figure 2.1. Normalisation of drop plate measurements.

The Figure shows an example of yeast growth under optimal growth conditions and after exposure to a stress as well as a hypothetical calculation after intensity measurements.

2.5 Western blotting

Protein samples were collected either as described in 2.8.2 or by creating whole cell extracts. For the latter a pellet of cells from a 25ml culture was resuspended in 60µl 1M DTT, 300µl 8M urea, 240µl 2x loading buffer (100mM Tris-HCl (pH 6.8), 20% glycerol, 4% SDS) and 100µl acid-washed glass beads. The cells were vortexed for 3 minutes and boiled for 5 minutes. Relative protein concentrations were measured with a 1:1 Bradford:water assay at OD₅₉₅. An appropriate amount of each sample was loaded and run on a 7.5%, 10% or 15% polyacrylamide gel and transferred onto nitrocellulose membranes using a semi-dry transfer method (15V, 45 minutes). The membranes were blocked in 10ml 5% BSA in TBST (20mM Tris-HCl (pH 7.5), 150mM NaCl, 0.1% Tween-10) for 1 hour. The primary antibody (Table 2.2) was added and the membrane

incubated for 1 hour, washed three times in TBST for 10 minutes, incubated with the 2.5µl HRP-conjugated secondary rat, mouse or rat antibodies in 10ml 2.5% BSA for 45 minutes and washed again three times in TBST for 10 minutes. The samples were visualised with a chemiluminescence assay (Pierce) on x-ray film.

Antibody	Epitope recognised	Species reactivity	Dilution
M2 (Sigma)	FLAG (DYKDDDK)	Mouse	1:5000
σ-80 (Santa Cruz)	RBP1, residues 1-80	Rabbit	7:1000
MAB1501 (Roche)	HA	Mouse	1:1000

Table 2.2. Primary Antibodies for Western Blots

2.6 3' end poly(A)+ RNA sequencing

2.6.1 Library generation

S. cerevisiae and *S. pombe* were grown in parallel at 30°C at 170rpm, mixed at a 2.5:1 ratio (normalised to OD₆₀₀) and pelleted (3 minutes, 5000rpm). The RNA was extracted using the miRNeasy kit (Qiagen) (Figure 2.2 (A)). Libraries were generated from 1µg RNA with a 3' end QuantSeq kit (Lexogen, see workflow in Figure 2.2 (B)), which relies on reverse transcription of polyadenylated RNAs, second strand synthesis using random barcoded primers before cDNA amplification. cDNA concentrations were measured on a 2100 bioanalyzer machine (Agilent) using a High Sensitivity DNA chip. Samples were diluted to a concentration of 350pmol/l. Up to 18 samples were multiplexed, loaded onto an Ion PITM Chip (ThermoFisher) using the Ion Torrent Chef (ThermoFisher) and the Ion Torrent Hi-Q Chef kit. The samples were sequenced on an Ion Proton Sequencer (ThermoFisher) with the Ion PI Hi-Q Sequencing 200 Kit using the Ion

P1B 3' Adapter (ATCACCGACTGCCATAGAGAGGCTGAGAC) as a template (520 flows, 200nt library length settings).

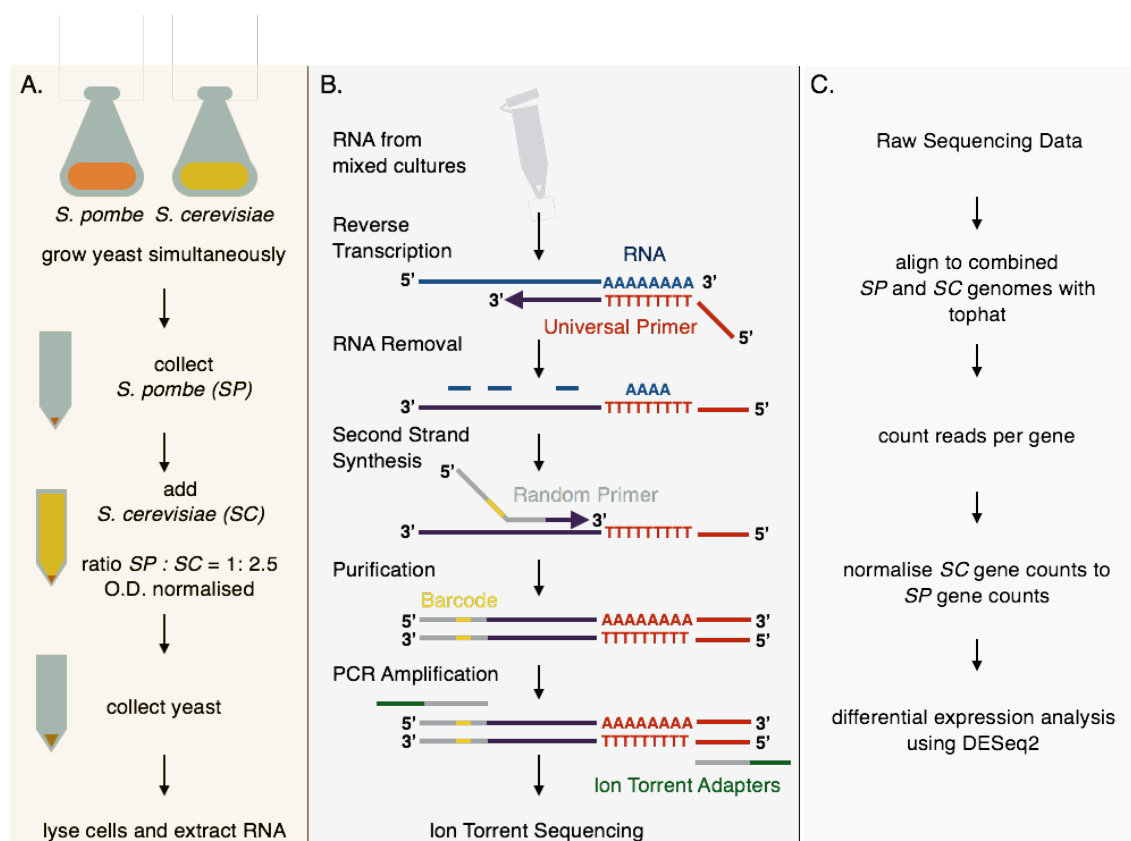


Figure 2.2. 3' end poly(A)+ RNA-Seq workflow.

(A) Growth conditions and sample collection for 3' end poly(A)+ RNA-Seq. (B) Workflow for Lexogen 3'end sequencing kit. (C) Outline of read/data processing after sequencing.

2.6.2 Data processing (general)

Pre-data processing (IonTorrent™ Suite Software) excluded non-barcoded reads and reads, in which the universal and barcoded IonTorrent adapters were not identifiable, and trimmed universal primer sequences and barcodes from the FastQ sequences. The remaining, trimmed FastQ sequences were trimmed using cutadapt (Martin 2011) to remove low-quality nucleotides from the 5' end (quality <20) and poly(A) sequences from the 3' end. Reads shorter than 20nt were removed and the remaining reads were aligned to a combined bowtie2

index of *S. cerevisiae* genome built R64-1-1 and (Ensembl, September 2011) and *S. pombe* (Ensembl, EF2) using Tophat (parameters: -N 2 --max-intron-length 1000 -p 4 -g 2 --read-edit-dist 2 --no-novel-juncs --transcriptome-index=transcriptome_data). The resulting was trimmed to contain only reads aligning to a single site using samtools (-bq 20). Coverage files were created using the bedtools genomecov function in a strand-specific manner and *S. cerevisiae* reads normalised to total *S. pombe* reads.

2.6.3 Differential expression analysis

Differential expression was assessed using the DESeq2 Bioconductor package R. All *S. cerevisiae* genes were normalised to *S. pombe* using the sizeestimate function, which calculates a scaling factor based on the total number and read distribution of *S. pombe* reads in each sample. After normalisation to *S. pombe*, the number of *S. cerevisiae* reads counts aligning to a region between the STOP codon to 300nt downstream of the STOP codon was calculated for protein coding genes. Similarly, for non-coding genes read counts were calculated for a 300nt window surrounding the main annotated PAS (150nt upstream of the annotated PAS to 150nt downstream of the annotated PAS). Genes were said to be differentially expressed if they showed more than a 2-fold difference in expression and a maximum Benjamini-adjusted p-value of 0.05. Overall fold-difference in expression estimates were based on calculated fold-changes in DESeq2. For heatmaps, expression changes relative to the wild type were plotted by ward clustering (R heatmap2 function).

2.6.4 Data processing (polyadenylation sites)

To find polyadenylation sites only pre-processed FastQ sequences with a minimum poly(A) length of 12nt were considered and processed as in 2.7.2. Uniquely aligned reads were trimmed to the 3' end using bedtools and combined into a bed coverage file. Occurrence of polyadenylation was assessed relative to the STOP codon as defined by the *Saccharomyces* Genome Database (SGD, www.yeastgenome.com), *i.e.* in the gene body of protein coding genes between start and stop codons and up to 300nt after the stop codon in a strand specific manner. The primary polyadenylation site was defined as the site with the most reads within a gene or the 300nt termination zone. The number of polyadenylation sites per gene was defined as the number of nucleotides between the start codon and 300nt after the stop codon that had a minimum number of 3' ends mapping to it. This cut-off was dependent on sequencing depth. Annotation of polyadenylation sites as mutant- specific occurred when PAS were present in one strain, but absent from another.

2.6.5 Motif analysis

Motifs were analysed using the MEME suite tool DREME and Centrimo functions (version 4.11.2) using standard settings (Machanick & Bailey 2011). Sequences of 130 nucleotides from 100 nucleotides upstream of the PAS to 30 nucleotides downstream of the PAS were submitted.

2.6.6 Structure analysis

Structures were predicted with the Vienna-RNA-tool package 2.1. (Gruber et al. 2008) using standard settings.

2.7 Micrococcal nuclease sequencing (MNase-Seq)

2.7.1 Sample collection

1 litre YPD was inoculated with *S. cerevisiae* and grown as described in 2.2. Once a density of OD₆₀₀ 0.6 was reached, cells were crosslinked by adding 27ml 37% formaldehyde (1% final concentration) for 10 minutes at room temperature, quenched with 50 ml 2.5M glycine, pelleted (5 minutes, 5000rpm), washed 3 times in 30ml ice-cold TBS (50mM Tris-Cl 150mM NaCl, pH 7.5) and resuspended in 10ml TBS. The cell density was determined using spectrophotometry (OD₆₀₀ of 1.0 corresponds to roughly 3×10^7 cells/ml) and 10^{10} cells selected for subsequent treatment. Cells were spheroblasted for 30 minutes at 35°C in 10ml 1M Sorbitol, 0.5mM 2-mercaptoethanol, 20mg/ml zymolase 20T (MP Biomedicals), pelleted (10 minutes, 5000rpm) and resuspended in SBD buffer (1M Sorbitol, 50 mM NaCl, 10 mM Tris pH 7.5, 5 mM MgCl₂, 1 mM CaCl₂, 0.075% NP-40, 1mM 2-mercaptoethanol). The cells were split into 8 samples and digested with 2.5U, 5U, 10U, 20U, 40U, 80U, 160U and 320U Nuclease S7 (Sigma) for 10 minutes at 37°C. The reaction was stopped by adding 100µl pre-warmed stop buffer (5% SDS, 50 mM EDTA) and incubated overnight with 100µl 20mg/ml Proteinase K (Roche). 100µl 5M NaCl were added and the sample incubated for 3 hours at 65°C, followed by a 1 hour incubation at 37°C after the addition of 1µl 10mg/mg RNase A (Sigma). The DNA was purified using a PCR purification kit (Qiagen). 1µg per sample was run on a 2% agarose gel to select of comparable titrations. The most comparable samples were run on a SizeSelect E-Gel (Invitrogen) and sizes around 150bp removed from the gel.

2.7.2 Library generation and sequencing

Libraries were generated using the Fast DNA Library Prep Set for Ion Torrent (NEB). 1µg DNA in 51µl H₂O was end repaired with 6µl end repair buffer and 3µl end repair mix at 25°C for 20 minutes followed by enzyme inactivation at 70°C for 10 minutes. The sample was mixed with 22µl H₂O, 10µl T4 DNA Ligase Buffer, 6µl T4 DNA Ligase, 0.5µl barcoded adapter (Ion Xpress Barcode Adapters 1-16 Kit (ThermoFisher)), 0.5µl universal adapter and 1µl Bst 2.0 WarmStart DNA polymerase and incubated for 15 minutes at 25°C followed by enzyme inactivation at 65°C for 5 minutes. The DNA was combined with 180µl AMPure XP beads (Beckman) and incubated for 5 minutes at room temperature. The beads were collected on a magnetic rack, the supernatant discarded and the beads washed twice with 200µl freshly prepared 80% ethanol, dried for 5 minutes at room temperature and resuspended in 25µl 0.1X TE. The beads were incubated for 2 minutes, the beads collected and 20µl run on an E-Gel, where fragment sizes between 250-300bp were selected. The DNA solution was topped up to 46µl and mixed with 4µl primers (Primer 1: CCATCTCATCCCTGGGTGTC, Primer 2: CCACTACGCCTCCGCTTTCCTCTCTATG) and 50µl NEBNext Q5 Hot Start HiFi PCR Master Mix. The mixture was subjected to 30 seconds at 98°C followed by 6 cycles of 10 seconds at 98°C, 30 seconds at 58°C and 30 seconds at 65°C and 5 minutes at 65°C. The DNA was mixed with 90µl AMPure XP beads, incubated for 5 minutes at room temperature, collected on a magnetic rack, washed twice with 200µl 80% ethanol, dried for 5 minutes at room temperature and resuspended in 35µl 0.1X TE. The beads were collected

and 30µl of the supernatant collected. DNA concentrations were assessed on a Bioanalyzer machine using a High Sensitivity DNA chip. Samples were diluted to 350 pmol/l and multiplexed (up to 6 samples per chip), loaded onto an Ion PI™ Chip (ThermoFisher) using the Ion Torrent Chef (ThermoFisher) and the Ion Torrent Hi-Q Chef kit. The samples were sequenced on an Ion Proton Sequencer (ThermoFisher) with the Ion PI Hi-Q Sequencing 200 Kit using the Ion P1B 3'Adapter (ATCACCGACTGCCATAGAGAGGCTGAGAC) as a template (520 flows, 236nt library length settings).

2.7.3 Data processing

FastQ sequences were trimmed using cutadapt (Martin 2011) to remove low-quality nucleotides. Reads shorter than 20nt were removed and the remaining reads were aligned to *S. cerevisiae* genome build R64-1-1 and (Ensembl, September 2011) and *S. pombe* (Ensembl, EF2) using Bowtie2 (parameters: --phred33 --end-to-end -q -S). Only unique reads were kept and coverage files created using bedtools. MNase positions and differences between strains and conditions were calculated using the danpos2 dpos function. Nucleosome positions were assigned relative to annotated TSS and PAS, whereby the +1 nucleosome was considered to be the first nucleosome downstream of the annotated TSS and the -1 nucleosome the first nucleosome upstream of the annotated TSS.

2.8 Nascent Elongating Transcript Sequencing (NET-Seq)

2.8.1 Sample collection

The protocol is based on a published protocol (Churchman & Weissman 2011). 2 litres of cells were grown from a pre-culture as described in 2.2 and harvested by filtration onto a 0.45µl pore size filter (Whatman). The cell mass was scraped off the filter, flash-frozen in liquid nitrogen and ground in a 50ml grinding jar with a 25mm stainless steel ball in a mixer mill (MM400 (Retsch)) in 6 cycles of 3 minutes each at a 15Hz shaking frequency, whereby the sample was chilled in liquid nitrogen between cycles. 1g of the yeast grindate was collected and stored at -80°C.

2.8.2 Immunoprecipitation

The grindate was resuspended in a lysis buffer (20mM HEPES (pH 7.4), 110 mM KOAc, 0.5% Triton-X-100, 0.1% Tween 20, 10 mM MnCl₂, complete proteinase inhibitors (Roche), 50 U/ml SUPERase.In RNase inhibitors (Invitrogen)) and incubated with 660U of DNase I (Promega) at 4°C for 20 minutes. The supernatant was collected after pelleting (16000g, 4°C, 10 minutes) and mixed with 0.5 ml prewashed (2 washes in 10ml lysis buffer) anti-FLAG M2 affinity agarose beads (Sigma). The solution was incubated on a nutator for 2.5 hours at 4°C. The beads were washed four times in 10ml wash buffer (20mM HEPES (pH 7.4), 110mM KOAc, 0.5% Triton-X-100, 0.1% Tween 20, 1mM EDTA) and incubated two times in 300µl elution buffer (lysis buffer with 1mg/ml 3xFLAG peptide) for 30 minutes at 4°C. RNA was extracted from the combined sample with the miRNeasy kit (Qiagen). Immunoprecipitation

efficiency was assessed by Western blotting of samples collected before immunoprecipitation, before and after elution.

2.8.3 Linker ligation and fragmentation

3µg RNA in 30 µl Tris-HCl from step 2.8.2. was denatured (2 minutes, 80°C) and 3' end ligated to 50 ng/µl cloning linker 1 (Integrated DNA technologies, Table 2.3) in a ligation buffer (12% PEG 8000, 1x T4 Rnl2 ligation buffer, 10 U/µl T4 truncated Rnl2 (NEB)) for 3 hours at 37°C. The ligated RNA was fragmented in an alkaline fragmentation buffer (100mM NaCO₃ (pH 9.2), 2mM EDTA) for 35 minutes at 95°C. The fragmented RNA was precipitated for 30 minutes in 0.75ml isopropanol, 0.5ml H₂O, 60µl 3M NaOAc (pH 5.5) and 2µl 15mg/ml GlycoBlue (Ambion), pelleted (16000g, 4°C, 30 minutes), washed in 0.75 ml 80% ethanol and resuspended in 10µl 10mM Tris-HCl (pH 7.0).

2.8.4 Size selection

The RNA from step 2.8.3. was mixed with 10µl 2x TBE-urea loading dye (89mM Tris-Cl, 89mM boric acid, 2mM EDTA, 12% Ficoll, 7M urea, 2.5mg/ml Orange G), denatured (2 minutes, 80°C and run at 200V for 32 minutes on a 10% TBE urea gel (Biorad). The gel was stained with 5µl SybrGold (Invitogen). RNA with lengths between 38-95 nt were cut out of the gel. The gel pieces were ground through centrifugation in microfuge tubes and melted in 200µl H₂O (70°C, 10 minutes). Gel debris was removed by centrifuging the gel mixture in Costar-Spin-X columns (Corning) and the RNA precipitated as in step 2.8.3.

2.8.5 Reverse transcription

The RNA from step 2.8.4. was mixed with 3.28µl 5x First-Strand Buffer, 0.82µl 10mM dNTPs and 0.5µl 100µM RT primer (Table 2.3), denatured (2 minutes, 80°C) and reverse transcribed for 30 minutes at 48°C after adding 0.5µl SUPERase.In, 0.82µl 0.1M DTT and 0.82µl Superscript III (ThermoScientific). The RNA was degraded with 1.8µl 1M NaOH (98°C, 20 minutes) and the solution neutralised with 1.8µl 1M HCl.

2.8.6 Circularisation

The cDNA was denatured (3 minutes, 95°C) after adding 20µl 2x TBE loading dye and run on a 10% TBE-urea gel (Biorad) at 200V for 50 minutes. The gel was stained and disrupted as in 2.8.4 after excision of the RT product, the cDNA precipitated as in 2.8.3 and resuspended in 15µl 10mM Tris-HCl (pH 8.0). The RT product was circularised by adding 2µl 10x CircLigase buffer 1µl 1mM ATP, 1 µl 50mM MnCl₂ and 1µl CircLigase (Epicentre) and incubating the mixture for 60 minutes at 60°C followed by heat-inactivation of the enzyme (10 minutes, 80°C).

2.8.7 PCR amplification

5µl circularised RNA was combined with 16.7 µl 5xHF Phusion buffer, 1.7µl 10mM dNTPs, 0.4µl 100µM barcoded primer (Table 2.3), 59.2µl H₂O and 0.8µl Phusion polymerase (NEB), split between five 0.2ml tubes and subjected to 30 seconds at 98°C followed by up to seven cycles of 98°C for 10 seconds, 60°C for 10 seconds and 72°C for 10 seconds. One tube was removed after each

cycle starting from cycle 3 and placed on ice. All samples were mixed with 3.4µl loading dye (15% Ficoll, 0.0025% Orange G) and run on an 8%TBE gel (Invitrogen) at 180V for 55 minutes and stained with 5µl SybrGold. The sample with the most unsaturated signal was selected, excised from the gel, ground and soaked in 0.67ml soaking buffer (0.3M NaCl, 10mM Tris-HCl (pH 8.0), 1mM EDTA) overnight at room temperature. The next day the DNA was precipitated and washed as described in 2.8.3. and resuspended in 10µl 10mM Tris-HCl (pH 8.0).

Oligonucleotides	Sequence
Linker-1	5AppCTGTAGGCACCATCAAT/3ddC
RT-primer	5Phos/ATCTCGTATGCCGTCTTCTGCTTG/iSp18/CACTCA /iSp18/TCCGACGATCATTGATGGTGCCTACAG
Universal Primer	CAAGCAGAAGACGGCATAACGA
Barcoding primer 1	AATGATACGGCGACCACCGAGATCTACACGATCGGAAG AGCACACGTCTGAACTCCAGTCACATGCCATCCGACGA TCATTGATGG
Barcoding primer 2	AATGATACGGCGACCACCGAGATCTACACGATCGGAAG AGCACACGTCTGAACTCCAGTCACTGCATCTCCGACGA TCATTGATGG
Barcoding primer 3	AATGATACGGCGACCACCGAGATCTACACGATCGGAAG AGCACACGTCTGAACTCCAGTCACGGCTACTCCGACGA TCATTGATGG
Barcoding primer 4	AATGATACGGCGACCACCGAGATCTACACGATCGGAAG AGCACACGTCTGAACTCCAGTCACTTAGGCTCCGACGA TCATTGATGG
Barcoding primer 5	AATGATACGGCGACCACCGAGATCTACACGATCGGAAG AGCACACGTCTGAACTCCAGTCACTGACCATCCGACGA TCATTGATGG
Sequencing primer	TCCGACGATCATTGATGGTGCCTACAG

Table 2.3. Oligonucleotides used for NET-Seq library generation.

5App – 5' adenylated, 3ddC- 3' blocked with dideoxy C, 5Phos- 5' phosphorylated end, iSp18- 18 carbon spacer, underlined- barcode.

2.8.8 Sequencing

Samples quality and quantity was assessed using a TapeStation (Agilent) and qPCR. The barcoded samples were multiplexed and 50nt sequenced with a Illumina HiSeq machine using a sequencing primer (Table 2.3).

2.8.9 Sequence alignment and processing

FastQ sequences were trimmed using cutadapt (Martin 2011) to remove adapter sequences. Reads shorter than 16nt were removed and the remaining reads were aligned to the *S. cerevisiae* genome built R64-1-1 (Ensembl, September 2011) using Bowtie (Langmead 2011) with the following parameters: -n 1, -e 70, -l 28, -k 1, -m1 -S. The resulting BAM file contained uniquely aligned reads that were narrowed to their 3' end and combined into bed files using samtools and bedtools.

2.8.10 Annotations

Transcription units were defined by their START and STOP codons for protein coding genes and most common TSS and polyadenylation sites (PAS) based for non-coding genes (Wery et al. 2016). Nucleosome positions were used as defined by the MNase-sequencing results (see section 2.7).

2.8.11 Additional data sets

Data sets from published and unpublished studies (Fischl, H.) were used to complement the analyses. They are listed in Table 2.4.

Data Set	Origin
NET-Seq (excluding <i>rrp6Δ</i>)	Fischl, H.
<i>rrp6Δ</i> NET-Seq	This study
Pcf11p-TEF-Seq*	Fischl, H.
Nrd1p-TEF-Seq*	Fischl, H.
Rrp6p-CRAC	Schneider et al. 2012
Rrp44p-CRAC	Schneider et al. 2012
Hrp1p-CRAC	Tuck and Tollervey 2013
Nab2p-CRAC	Tuck and Tollervey 2013
Pap1p-CRAC	Tuck and Tollervey 2013
Nrd1p-CRAC	Jamonnak et al 2011
Nab3p-CRAC	Jamonnak et al 2011
Polyadenylation sites	Pelechano et al 2013
Polyadenylation sites and 3' end poly(A)+ RNA concentrations (wildtype, <i>rrp6Δ</i> , <i>xrn1Δ</i> , <i>sua7-1</i>)	This study
Polyadenylation sites and RNA stability	Geisberg et al 2014
RNA-Seq, degradation rates and transcription rates	Sun et al 2013
Degradation rates and transcription rates	Miller et al 2011
Transcription rates	Pelechano et al 2010
NET-Seq data (used for a correlation with data collected by Fischl, H)	Churchman and Weissman 2011
Xrn1-ChipExo	Haimovich et al 2013
Transcription factor binding sites	Rhee and Pugh 2012
Transcription factor binding sites	Kasinathan et al 2014
Protein turnover	Christiano et al 2016
Protein concentrations	Ghaemmaghami et al 2003
Protein concentrations	Chong et al 2015
RNA concentration	Wery 2016
RNA concentration	Malabat 2015
MNase digestion data (wildtype, <i>rrp6Δ</i> , <i>xrn1Δ</i>)	This study

Table 2.4. Data sets used in this thesis.

***The TEF-Seq method is briefly outlined in Chapter 4.**

2.8.12 Data analysis and visualisation

The data was analysed and visualised using the integrative genome viewer (IGV) and custom scripts based on built-in functions in Matlab (2014 version) and the following R-packages (R version 3.2.2, Bioconductor version 3.2): DESeq2, Gviz, GenomicRanges, GenomicAlignments, BiocParallel, Rsamtools, GenomicFeatures, pheatmap, gplots, RColorBrewer, mclust, ggplot2, PoiClaClu.

DESeq2 was used to identify differentially transcribed genes, whereby significance was defined as a 2-fold difference in transcription rates and an adjusted p-value of smaller than 0.05 to account for multiple testing.

Correlations between data sets were calculated using Spearman correlations and significance between groups using Mann-Whitney tests (since the data was not normally distributed). Differences between data sets were determined using principal component analyses.

Metagene profiles were plotted by aligning genes by their TSS, PAS, START codons STOP codons or nucleosome positions and determining the median count values among all selected genes at each nucleotide in a window of a maximum of 1500nt surrounding the selected site.

2.9 Fluorescence-Activated Cell Sorting (FACS)

FACS was used to assess protein accumulation in GFP-tagged strains after a glucose to galactose shift or glucose shutdown and to measure DNA replication. For the former, 0.5 ml exponentially growing cells were harvested from liquid cultures (5 minutes, 7000 rpm), resuspended in 4%

paraformaldehyde in 1xPBS, light-protected and incubated for 45 minutes. The cells were washed twice (5 minutes, 7000 rpm) in 400 μ l 1xPBS and stored up to one month in 400 μ l 1xPBS at 4°C. For FACS, 50 μ l cells were diluted 1ml 1xPBS and fluorescence measured for 10000 cells per sample on a FACS Calibur machine (BD Biosciences) with a 488nm-laser and a 530/30nm band pass filter. Samples were collected in YPD and 2, 4, 6, 8 and 16 hours after a shift from YPD to YPG.

For cell cycle analysis, the cells were fixed by adding 2.4ml cold 100% ethanol to 1ml culture (final concentration 70%) and stored at 4°C. 100 μ l cells were washed (1 minute, 7000 rpm) in 1ml 50mM sodium citrate, resuspended in 0.5ml 50mM sodium citrate containing 0.1mg/ml RNase A and incubated at 37°C for 2 hours. 0.5 ml 4 μ g/ml propidium iodide in 50mM sodium citrate was added and fluorescence measured for 50000 cells per sample by FACS using a 488nm-laser and a 585/42nm band pass filter.

The data were analysed using a custom script based on the FlowClust R-package (Lo et al. 2009). A schematic is shown Figure 2.3.

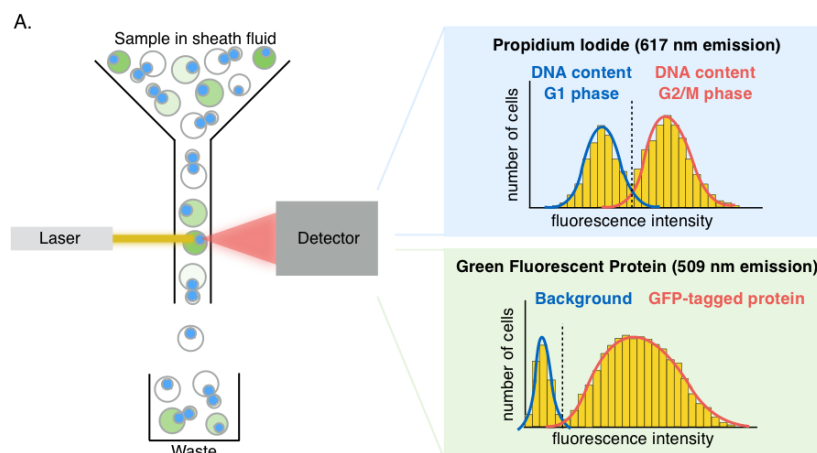


Figure 2.3. FACS measurements

3 Roadblock Termination is a Widespread Phenomenon in Wild-type Yeast

3.1 Introduction

Chromatin states, transcription rates and stable RNA pools in the cytoplasm are inherently linked. Nucleosome depletion and specific histone modifications are a characteristic of active RNA transcription, which itself is a prerequisite for the accumulation of RNA. While nucleosome depletion is an inherent feature of promoters and can occur prior to the onset of transcription (Struhl & Segal 2013), nucleosome depletion at terminators is linked to transcriptional termination and 3' end formation (Fan et al. 2010).

Termination can occur via the CPFC-dependent pathway and the NNS pathway, discussed in the Introduction, and is required for transcript release from the DNA and 3' end processing. In addition to the CPFC and NNS termination pathways, termination can occur via non-canonical mechanisms, for instance, through the occurrence of Reb1p roadblocks. In this scenario, the collision of RNAP II with a Reb1p transcription factor (Colin et al. 2014), bound to the DNA, leads to the pausing of RNAP II, followed by the ubiquitination of the stalled RNAP II and its dissociation from the DNA. Reb1p unstable transcripts (RUTs) are then rapidly degraded by the nuclear exosome. This alternative termination pathway has been proposed to be a fail-safe mechanism to prevent transcriptional read-through.

The aim of this chapter is to establish links between chromatin, active transcription and 3' end polyadenylation dynamics in wild-type yeast grown

under optimal steady-state growth conditions. The mechanism and patterns described in this chapter will serve as a reference for comparisons with *rrp6Δ* and *xrn1Δ* mutants and non-steady state growth conditions in discussed in Chapters 4 and 5.

Chromatin dynamics were recorded using MNase-Seq (Kaplan et al. 2009), nascent transcriptome rates using NET-Seq (Churchman & Weissman 2011) and RNA concentrations with normalised 3' end poly(A)+ sequencing (Figure 3.1). Furthermore, 3' end sequencing data was used to map of polyadenylation sites (PAS) with a simplified protocol. As the same sequencing techniques were employed throughout this thesis, the description of protocols in this chapter will contribute to the understanding of Chapters 4 and 5.

The data show that the vast majority of PAS map to nucleosome depleted regions (NDRs) and that the occurrence of PAS correlates with an increase in the density of nascent transcription reads. Intragenic PAS are a wide-spread phenomenon within coding genes and are likely to be, at least in part, terminated by alternative road-block termination pathways.

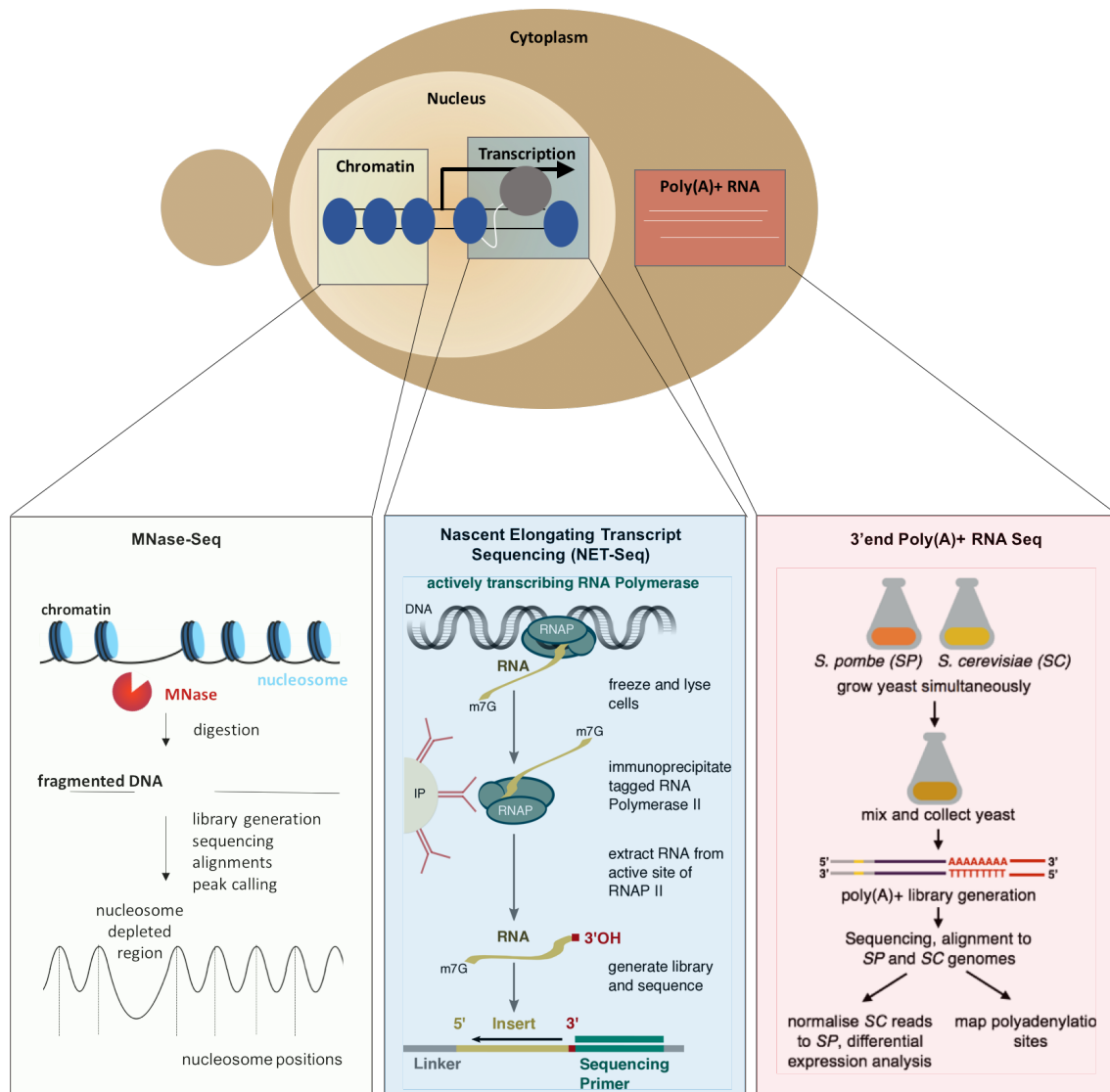


Figure 3.1. Sequencing techniques

3.1 Results

3.1.1 Nucleosome depletion is a feature of most promoters and terminators of coding and non-coding genes

MNase-Seq relies on the digestion of unbound DNA by MNase (Kaplan et al. 2009). The enzyme preferentially digests DNA in easily accessible regions, *i.e.* DNA sequences that are not wrapped around nucleosomes. Therefore,

stretches of DNA, which are sequenced disproportionately often, are generally bound to nucleosomes. By contrast, low read counts occur in regions in which nucleosomes binding is inherently unfavourable due to DNA sequence patterns or in which nucleosomes are evicted from the DNA by nucleosome remodelling enzymes. Stable nucleosome depleted regions (NDRs) are particularly common at active promoters and terminators (Kaplan et al. 2009; Fan et al. 2010).

In the data set presented here, NDRs could be detected upstream of the majority of annotated START codons and TSS of both coding and non-coding genes (Figure 3.2). As the position of TSS and PAS can vary with growth and environmental conditions, all the data presented in this thesis was mapped relative to the START and STOP codon for coding genes.

Nucleosome phasing was evident at the +1 to +4 nucleosomes in coding genes, whereby the +1 nucleosome was defined as the first nucleosome downstream of the NDR. By contrast, in non-coding genes nucleosome phasing was generally restricted to the +1 nucleosome, although this might reflect the relative uncertainty of current annotations. For some coding genes with large NDRs, some poorly bound nucleosomes were also distinguishable within the NDR.

When taking only genes with defined NDRs into account, the mean NDR size at the promoter was 221 nucleotides for coding and 248 nucleotides for non-coding genes. The larger size of non-coding gene NDRs can be explained by the fact that many non-coding RNAs are expressed from bi-directional promoters, which are inherently associated with larger NDR sizes (Neil et al. 2009; Xu et al. 2009; Rhee & Pugh 2012b).

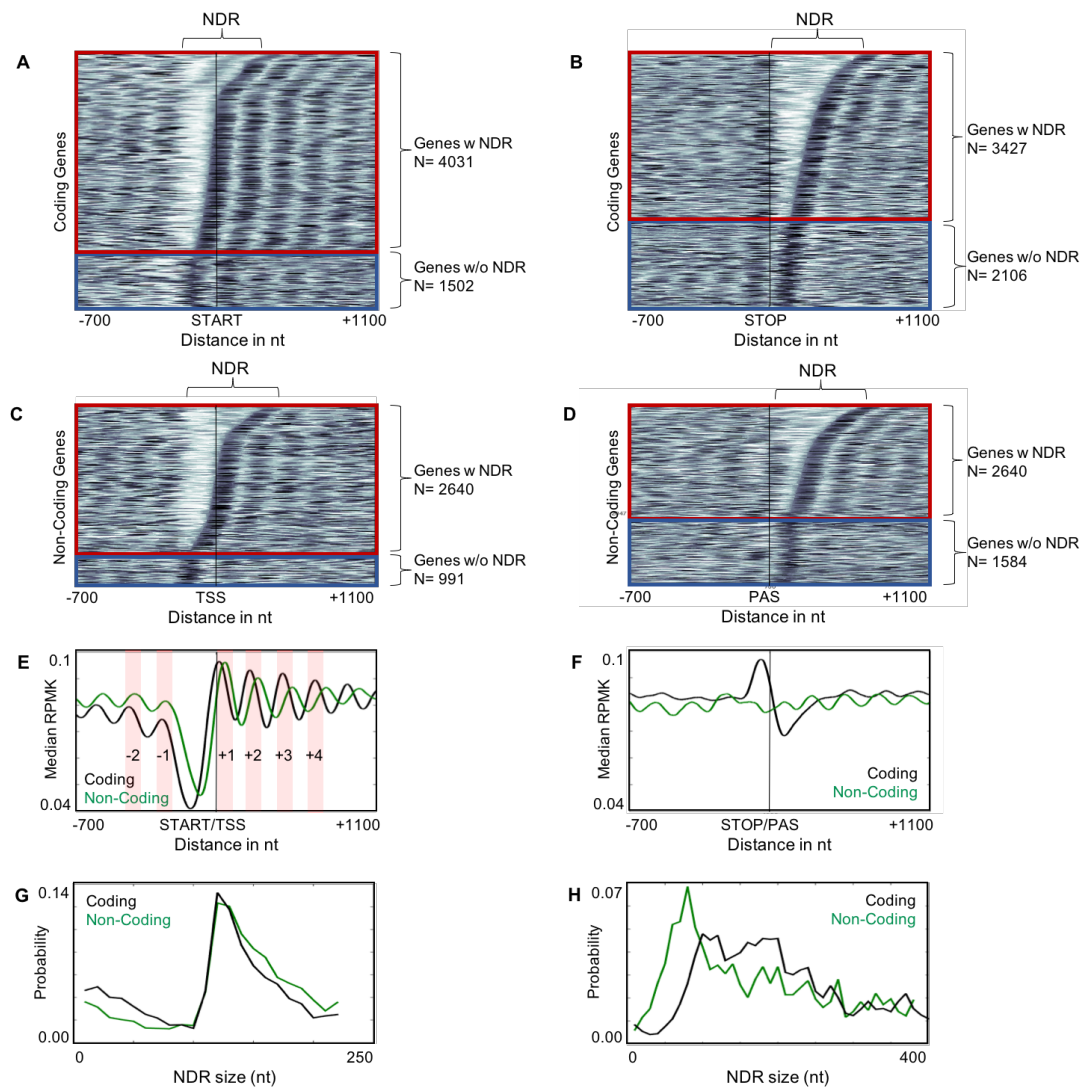


Figure 3.2. MNase-Seq read distributions.

(A-D) Heatmap of nucleosome occupancy surrounding the (A) START and (B) STOP codon of coding and the (C) TSS and (D) PAS of non-coding genes. Dark colours indicate high occupancy regions. The data was normalised for each gene individually. (E-F) Median nucleosome occupancy of coding and non-coding genes at the (E) START codon/TSS and (F) STOP codon/PAS. The red boxes indicate nucleosome positions. (G-H) Distribution of NDR sizes of coding and non-coding genes at the (G) START codon/TSS and (H) STOP codon/PAS. NDR- nucleosome depleted region.

Nucleosome depletion could also be detected downstream of the STOP codon at coding genes or the annotated polyadenylation site (PAS) at non-coding genes. However, the positioning of nucleosomes at the 3' end was less stringent and fewer genes had clearly defined NDRs. Although NDRs are visible in heatmaps, in which the data is normalised separately to the relative

maximum read count of each gene, nucleosome depletion is not detectable in the metagene profiles of non-coding genes. Non-coding NDR sizes at the terminator are also shorter on average, as coding and non-coding genes have a mean NDR size of 217 and 203 nucleotides respectively.

This suggests that fewer nucleosomes are evicted at non-coding gene PAS. Since nucleosome depletion at terminators is linked to 3' end formation during transcription rather than DNA sequence patterns and directly correlates with RNAP II occupancy (Fan et al. 2010), it is likely that nucleosome depletion at terminators of non-coding genes is more restricted because transcriptional rates are lower than in coding genes. In addition, annotated PAS often do not match the end of nascent transcription in NET-Seq data at non-coding gene (discussed below). Therefore, the position of 3' ends of non-coding transcripts might either be variable or badly annotated, which could in turn be reflected in the relative lack of nucleosome depletion.

3.1.2 NET-Seq reads accumulate in distinct regions of the gene

NET-Seq is a technique that measures the amount of immuno-precipitated RNA in the active site of RNAP II. As such, the number of measured reads per nucleotide can either reflect the amount of transcript that is being produced or the transcription rate itself.

The amount of transcript that is being produced can vary over the length of the gene body due to abortive RNAP II transcription or, more rarely, as a result of transcription initiation from intragenic TSS. Similarly, transcription rates can be adjusted through the binding of co-factors to the actively transcribing RNAP II and RNAP II pausing.

Hence, RNAP II density is dependent on the transcription initiation frequency, elongation rate and processivity, and mathematical modelling has shown that varying all three parameters can lead to indistinguishable next-generation sequencing reads profiles (Ehrensberger et al. 2013).

Therefore, it is important to bear in mind that NET-Seq signals are a semi-quantitative measure of active transcription. Nevertheless, measured relative transcriptions rates from NET-Seq experiments correlate reasonably well with data from published studies (Figure 3.3 (A), $R=0.8194$), which directly measure the incorporation of labelled nucleotides into the RNA, and are anti-correlated with RNA degradation rates (Figure 3.3).

While transcription is inherently noisy, the majority of NET-Seq reads map to annotated coding regions. An example of a typical profile over a genic region is shown in Figure 3.3 (B). In wild-type yeast, 5734 coding genes and 2640 non-coding genes are transcribed under optimal growth conditions. Although more than half of all coding genes have measurable antisense transcription ($N=2987$), NET-Seq transcription rates are not correlated in the sense and antisense directions (Figure 3.3 (C), Murray et al. 2012). However, nascent transcription rates correlate with concentrations of polyadenylated RNAs ($R=0.4944$) and are anti-correlated with nucleosome depletion at the promoter ($R= -0.2680$).

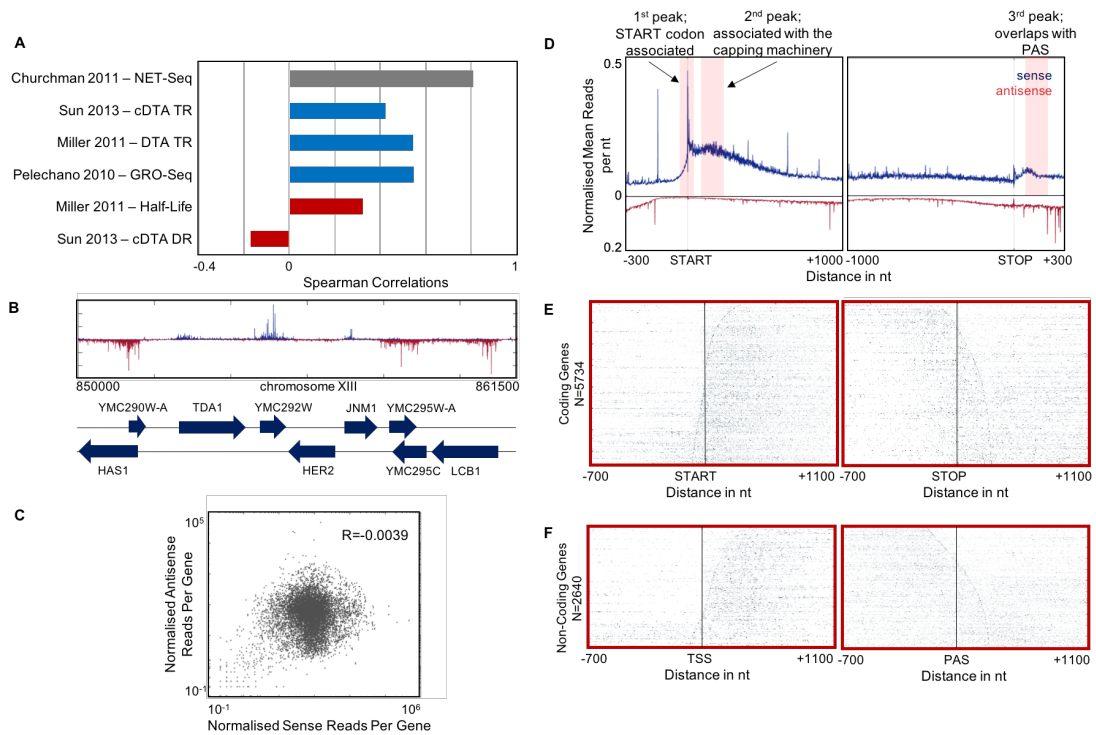


Figure 3.3. NET-Seq read distributions.

(A) Spearman correlations of NET-Seq data with measurements from previous studies. TR- transcription rate. DR- degradation rate. (B) Example of NET-Seq data. (C) Correlation of sense and antisense reads. (D) Metagene profile of NET-Seq reads at coding genes. (E-F) Heatmap of NET-Seq reads in (E) coding and (F) non-coding genes. The NET-Seq data for this Figure was collected by Fischl, H.

In metagene profiles, NET-Seq reads accumulate in the 5' region of both coding and non-coding genes and level off towards the end of the gene. In fact, three regions with increased read density can be distinguished (Figure 3.3. (D)).

The first peak overlaps with the START codon in coding genes, *i.e.* before the RNAP II has entered the coding region. It is possible that this peak reflects stalled polymerase, which has not yet fully committed to transcription.

The second 5' peak's maximum is located approximately +140-300 nucleotides downstream of the START codon. As initiation factors are thought to be exchanged for elongation factors in the same region of the gene (Mayer et al. 2010), the observed read accumulation might reflect a slowdown in RNAP II

speed to accommodate this process. Alternatively, increased read counts per nucleotide could be caused by abortive RNAP II transcription. Consistent with this, the peak overlaps with the binding of subunits of the TRAMP complex to the RNA (Tuck & Tollervey 2013), suggesting that defective RNAs, after a failure to cap the transcript, are being removed by the nuclear exosome. Alternatively, the transcription machinery could compete with the nuclear RNA degradation machinery, resulting in abortive transcription at the beginning of coding genes.

NET-Seq read counts are also increased over the first 300 nucleotides after the STOP codon. Again, this could reflect a slow-down of the RNAP II prior to RNA cleavage and polyadenylation.

Antisense transcription was primarily detected downstream of the STOP codon at the 3' end and upstream of the START codon relative to the sense coding transcript. The majority of these 3' end antisense transcription reads stem from the transcription of convergent genes located downstream of the sense gene. In addition, the reads can originate from the transcription of short, 3' end transcripts (Pelechano et al. 2013), which are transcribed from gene terminators that support the assembly of PICs (Murray et al. 2012). Antisense transcription proximal to the START codon is likely the product from bi-directional transcription from shared promoter regions (Venters & Pugh 2009; Neil et al. 2009; Xu et al. 2009).

In general, coding gene transcription is generally initiated at or just upstream of the START codon (Figure 3.3 (E)). By contrast, transcription of non-coding genes appears to frequently be initiated downstream of the annotated TSS

(Figure 3.3 (F)). Similarly, in coding genes transcription is terminated after the STOP codon, whereas non-coding genes are sometimes terminated prior to their annotated PAS. Hence, gene annotations for non-coding genes either do not reflect commonly used TSS and PAS or are highly variable between different laboratories, conditions and even within particular strains. It is likely that the precise locations of both TSS and PAS is inherently more variable for non-coding RNAs.

3.1.3 Spike-ins improve RNA-Seq experiment accuracy

Total intracellular RNA concentrations are regulated through RNA transcription and degradation rates and vary with environmental conditions and growth rates. Therefore, mutations that affect RNA metabolism or growth rates can result in a change in total intracellular RNA concentrations.

It is crucial to take these changes into account when comparing relative differences between mutants and growth conditions. Altered global RNA concentrations can be detected during sequencing by normalising *S. cerevisiae* specific reads to a *S. pombe* control. Rather than preparing the *S. cerevisiae* RNA exclusively, budding yeast strains were grown in parallel with *S. pombe* and mixed at a fixed ratio prior to RNA extraction. Libraries generated were then generated from the mixed RNA. As a constant number of *S. pombe* cells was added to each sample, total *S. pombe* read numbers were subsequently used to normalise between different strains.

This normalisation results in accurate and repeatable estimates of relative global RNA levels and minimizes differences between biological replicates ($R=0.9877$). Normalised read counts from the 3' ends of genes (STOP codon to

300nt downstream, a window which should capture the majority of 3' end reads (Oszolak et al. 2010)) partially correlate with published RNA-Seq data (Figure 3.4 (A)), whereby the extent of the correlation depends on whether a poly(A)+ RNA-Seq (better correlation) or a total RNA-Seq (poorer correlations) method was used for library preparation. Moreover, the 3' end read counts correlate moderately well with protein concentrations (Figure 3.4 (B); Wery et al. 2016; Malabat et al. 2015; Sun et al. 2013; Chong et al. 2015; Ghaemmaghami et al. 2003; Christiano et al. 2016).

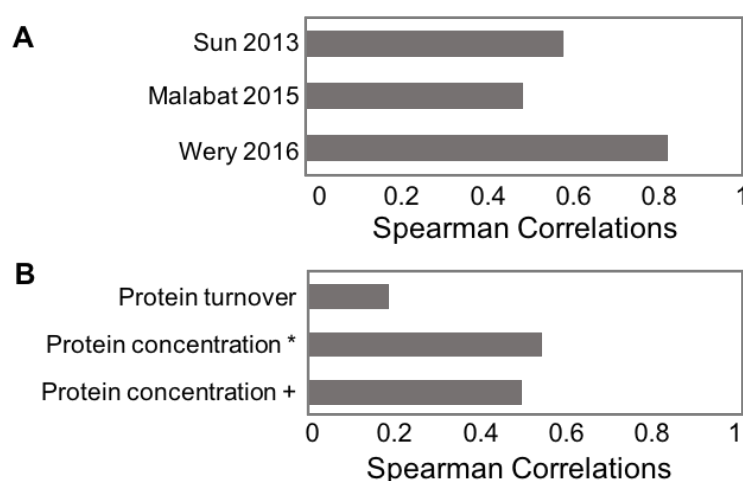


Figure 3.4. RNA-Seq experiments correlations.

Spearman correlations of 3' end poly(A)+ RNA-Seq reads with (A) total, ribosomal-depleted (Sun et al. 2013; Malabat et al. 2015) and poly(A)+ RNA-Seq data sets (Wery et al. 2016) and (B) protein turnover (Christiano et al. 2016) and concentrations (*Ghaemmaghami et al. 2003; + Chong et al. 2015).

3.1.4 3' end-Seq data can be used to map PAS

Poly(A)+ 3' end reads can also be used to map polyadenylation sites. During the library generation, poly(T) primers anneal to the poly(A) tails and serve as primers during the reverse transcription. Therefore, poly(A) tails are sequenced

together with the templated RNA sequence. The position of the polyadenylation site can be determined by trimming the poly(A) tail from the end 3' end of the read and mapping the terminal nucleotide after alignment to the reference genome (Figure 3.5 (A)).

Only reads with high quality scores and a minimum of 12 adenosines at their 3' ends were chosen. It is unlikely that these sequences are artefacts as long A-rich sequences are depleted from genic regions in yeast to prevent ribosome stalling after the translation of poly-lysines. Moreover, sequences downstream of the mapped PAS were examined for the presence of long poly(A) stretches and, while some A-rich regions could be detected, downstream sequences were not enriched in long stretches of consecutive adenosines. This suggests that the selected reads are likely to stem from true polyadenylated RNAs.

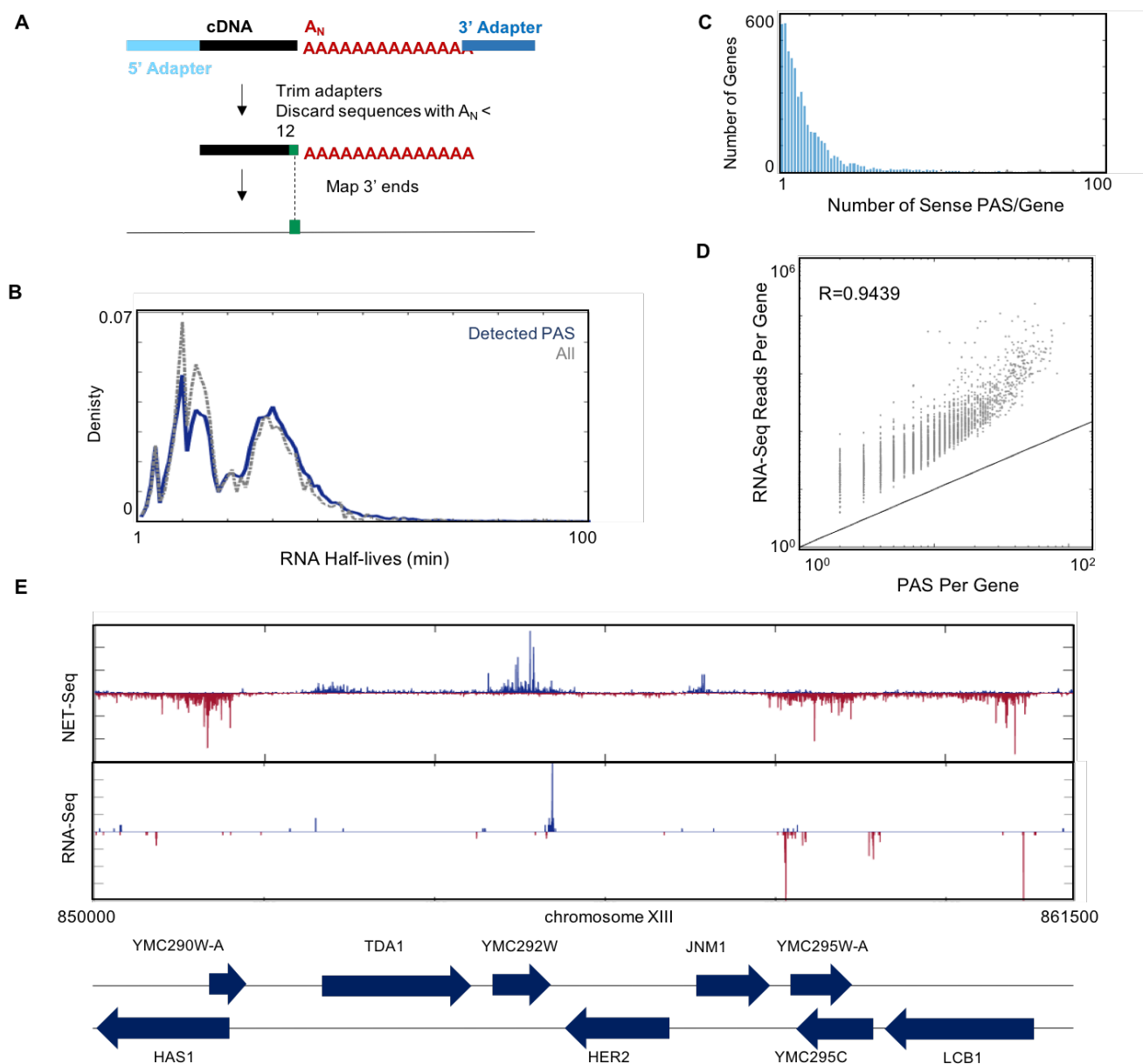


Figure 3.5. Mapping of PAS from 3' RNA-Seq data.

(A) Workflow of PAS mapping. (B) Distribution of RNA stabilities associated with detected RNA isoforms. The PAS detected in this study are shown in blue. (C) Distribution of the number PAS per coding gene. (D) Scatter plot and spearman correlation of the number of PAS and RNA-Seq reads per gene. (E) Example of NET-Seq reads compared to PAS distribution on chromosome XIII.

A total of 37061 PAS could be detected in the RNA sequencing data, whereby PAS were considered to be real if they had at least two 3' ends reads mapping

to the same nucleotide. 86% of the mapped positions (N= 31933) had previously been reported in a TIF-Seq data set (Pelechano et al. 2013). In addition, 2754 of the detected PAS were also reported in a separate, less comprehensive study (Geisberg et al. 2014), which also measured RNA stabilities of individual polyadenylated RNAs. While the majority of the PAS reported in the published data could be detected with the method described here, some PAS of less stable RNAs could not be reproduced. Hence, there might be a slight bias against the detection of unstable RNAs in the data set described here (Figure 3.5 (B)). Nevertheless, the mapped PAS broadly reproduce previously reported PAS detected by different techniques and laboratories. An example of the position of these mapped reads at a gene locus is shown in Figure 3.5 (E).

This means that PAS can be mapped accurately and directly from 3' end sequencing reads using a simplified protocol, which does not require any size selection steps or RNA fragmentation as in previous protocols (Pelechano et al. 2012). Moreover, library generation is fast, requires only approximately 500ng RNA per library and allows for multiplexing. Therefore, using standard 3' end library sequencing kits might be a cheap and quick alternative to traditional poly(A) mapping protocols.

3.1.5 The majority of PAS maps to the 3' end of genes

The majority of PAS mapped to coding genes. Overall, 32842 reads were distributed over 5000 genes, corresponding to an average of 6.57 PAS per coding gene (Figure 3.5 (C)). This is significantly higher than the number observed at non-coding genes, which had an average of 0.43 PAS per gene.

Overall, the number of PAS is correlated with the total number of RNA-Seq reads, suggesting that variation in PAS selection, or at least its detection, is a function of transcript abundance (Figure 3.5 (D)).

79.3 % of coding gene PAS mapped to the 3' end of the gene within a region from the STOP codon to 350 nucleotides downstream of the STOP codon, whereby the maximum number of PAS per nucleotide occurred at +73 nucleotides downstream of the STOP codon. Polyadenylation sites frequently occurred in close proximity to one another with a median distance of four nucleotides between sites, confirming the existence of a polyadenylation zone described in previous reports (Ozsolak et al. 2010; Moqtaderi et al. 2013).

The majority of non-coding PAS also mapped to the 3' end of the annotated non-coding genes, with the number of PAS peaking proximal to the annotated PAS in the metagene profile. However, the distribution of PAS varied widely and 60% of mapped PAS were located upstream of the annotated PAS, suggesting that either annotations are incorrect, or alternatively, that there is no fixed termination point of non-coding RNAs. This agrees with another study conducted in human cells, in which lincRNAs are also prone to showing variable termination sites (Schlachow et al. 2017).

3.1.6 Terminators are inherently bi-directional

Approximately 50% of expressed genes (N= 2866) also had PAS mapping to the opposite strand in the antisense direction over the gene body and the 3' end of the sense gene. Like sense and antisense transcription rates, the number of RNA-Seq reads ($R=-0.0013$) and the number of PAS in the sense and antisense direction are not correlated ($R=0.0169$).

On average, at genes with antisense reads 6.55 antisense PAS (aPAS) could be detected per gene. Most of these mapped to downstream of the sense STOP codon (Figure 3.6 (A, C)). In contrast to sense PAS, the number of detected aPAS increased with the proximity to the STOP codon, peaking one nucleotide upstream of it. Hence, it appears that the STOP codon, or regulatory elements in its vicinity, act as a barrier to transcriptional progression in the antisense direction. One explanation for this is the occurrence of high numbers of Nrd1p/Nab3p binding sites in the vicinity of the STOP codon in both the sense and antisense direction (see Figure 4.13 (A) in the following Chapter). Hence, it is likely that bi-directional transcriptional termination is encoded in the DNA to prevent transcriptional read-through into neighbouring coding genes.

A total of 1535 coding genes share a terminator with at least one coding or non-coding TU. For this reason, some of the antisense PAS are likely to stem from coding and non-coding genes that are transcribed from the opposite strand and share a terminator with the upstream sense gene. In agreement with this, the number of aPAS is significantly higher at genes with overlapping terminators ($p=7.6 \times 10^{-37}$). However, even when these overlapping terminators are excluded, an average of 2.60 aPAS was observed per gene. Therefore, and consistent with nascent transcription data, the region downstream of the STOP codon must be transcribed in the antisense direction likely resulting in short, polyadenylated transcripts.

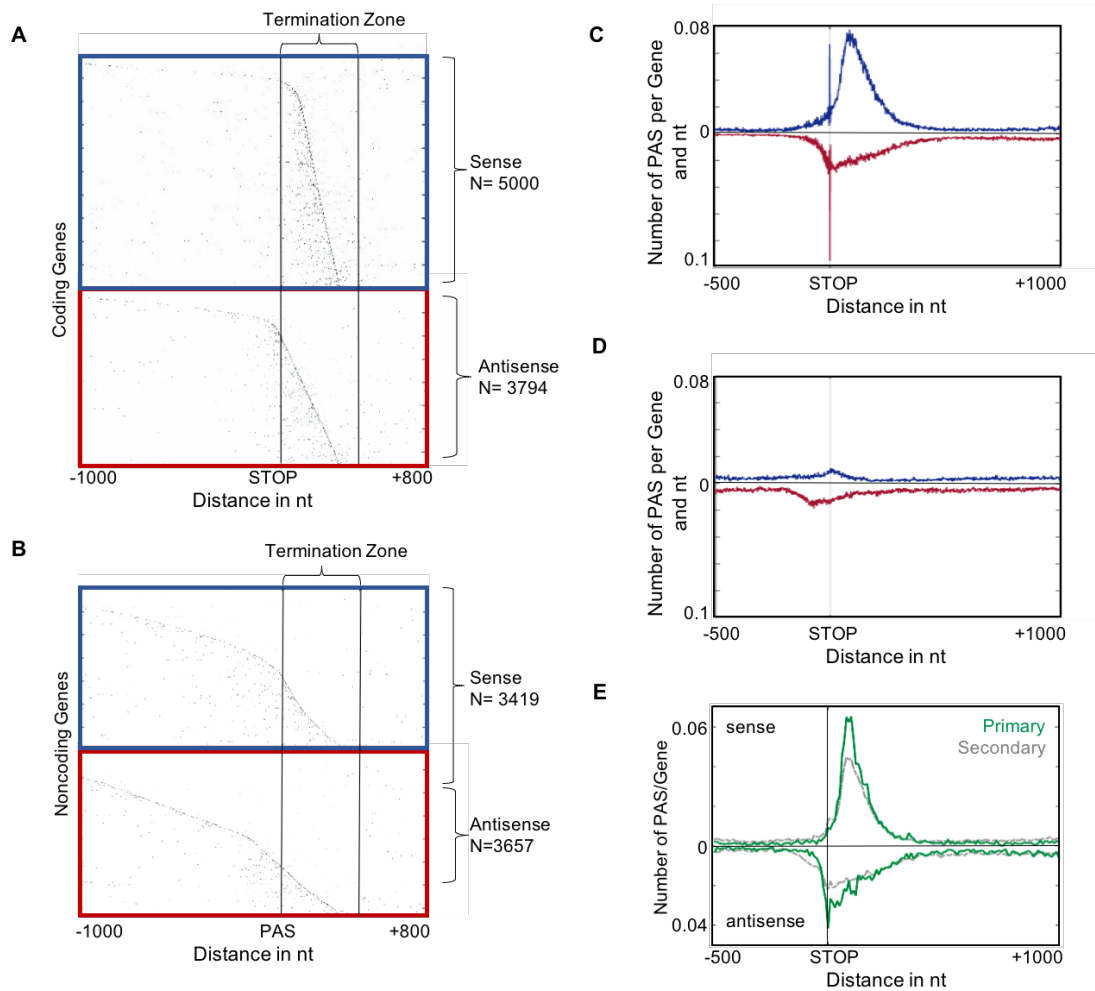


Figure 3.6. Positions of PAS at the 3' end of genes.

(A-B) Heatmap of PAS distributions of (A) coding and (B) non-coding genes. The blue box indicates reads in the sense direction, the red box reads in the antisense direction. Dark dots indicate the positions of mapped PAS. (C-D) Number of PAS per gene and nucleotide relative to (C) the STOP codon in coding genes and (D) the annotated PAS in non-coding genes in the sense (blue) and antisense (red) direction. (E) Distribution of primary (green) and secondary (grey) PAS in the vicinity of the STOP codon in coding genes.

In addition to convergent coding genes, many non-coding transcripts 3' ends map to a region downstream of the STOP codon of coding genes, which also contain PAS regulating element, suggesting these regions are inherently bi-directional. Consistent with this, many terminators contain a large number of

Hrp1p binding sites, a component of the CPFC that can terminate transcription independent of its orientation (de Boer et al. 2014). Bidirectional terminators might have a role in preventing transcription read-through into adjacent genes.

3.1.7 PAS are not used with equal frequencies

Most genes (N=4460) have one PAS that is used with a higher frequency than all others. This PAS will hereafter be referred to as the primary PAS. In the sense direction, 71% of primary PAS within the first 200 nucleotides after the STOP codon. Similarly, in the antisense direction most genes have their primary aPAS upstream of the STOP codon. By contrast, secondary PAS positions are slightly less restricted and occur more frequently outside these defined boundaries (Figure 3.6 (E)).

Primary PAS are mostly conserved between wild-type and mutant strains (discussed in Chapters 4 and 5), but can shift with environmental conditions. For instance, 1682 primary PAS are shifted 60 minutes after the yeast were shifted from glucose to galactose, although the observed positional changes were generally minor.

3.1.8 Intragenic PAS are a common feature of coding genes

20.75 % of coding gene PAS were detected in the intragenic region between the START and STOP codon (Figure 3.7 (B)). In total, intragenic PAS were detectable in 51.4% of all genes (N= 2568; 48.7% of these reads were reproducible in the TIF-Seq data). 248 of these genes only had detectable intragenic PAS, without having any detectable 3' end PAS. Therefore, although some reads might originate from errors during the library generation process,

the frequency in the occurrence of intragenic alternative polyadenylation sites suggests that intragenic alternative polyadenylation is likely to be a common feature of transcription in budding yeast.

Term	p-Value
cytoplasmic translation	1.36×10^{-41}
translation	2.18×10^{-32}
ribosome biogenesis	2.39×10^{-32}
rRNA processing	5.08×10^{-28}
Ribosome	1.97×10^{-27}
maturation of SSU-rRNA	4.99×10^{-15}
endonucleolytic cleavage in ITS1	7.16×10^{-10}
ribosomal large subunit biogenesis	2.24×10^{-9}
endonucleolytic cleavage in 5'-ETS	2.15×10^{-6}
maturation of LSU-rRNA	2.06×10^{-6}
endonucleolytic cleavage to generate mature 5'-end of SSU-rRNA	5.08×10^{-6}
ribosomal small subunit biogenesis	4.62×10^{-6}
ribosomal small subunit assembly	2.85×10^{-5}
rRNA export from nucleus	1.67×10^{-4}
DNA-templated transcription, termination	2.18×10^{-4}
ribosomal large subunit assembly	2.15×10^{-4}
formation of translation preinitiation complex	1.25×10^{-3}

Table 3.1. GO enrichment for genes with intragenic PAS.

Gene ontology (GO) enrichment analysis was conducted on gene with intragenic PAS using gprofiler (Reimand et al. 2007).

Intragenic PAS did not originate from annotated non-coding genes and were distributed across the whole of the gene body (Figure 3.7 (A)). No link to other features such as introns could be identified, although intragenic sites were highly enriched in genes associated with translation and ribosome biogenesis in gene ontology searches (Table 3.1).

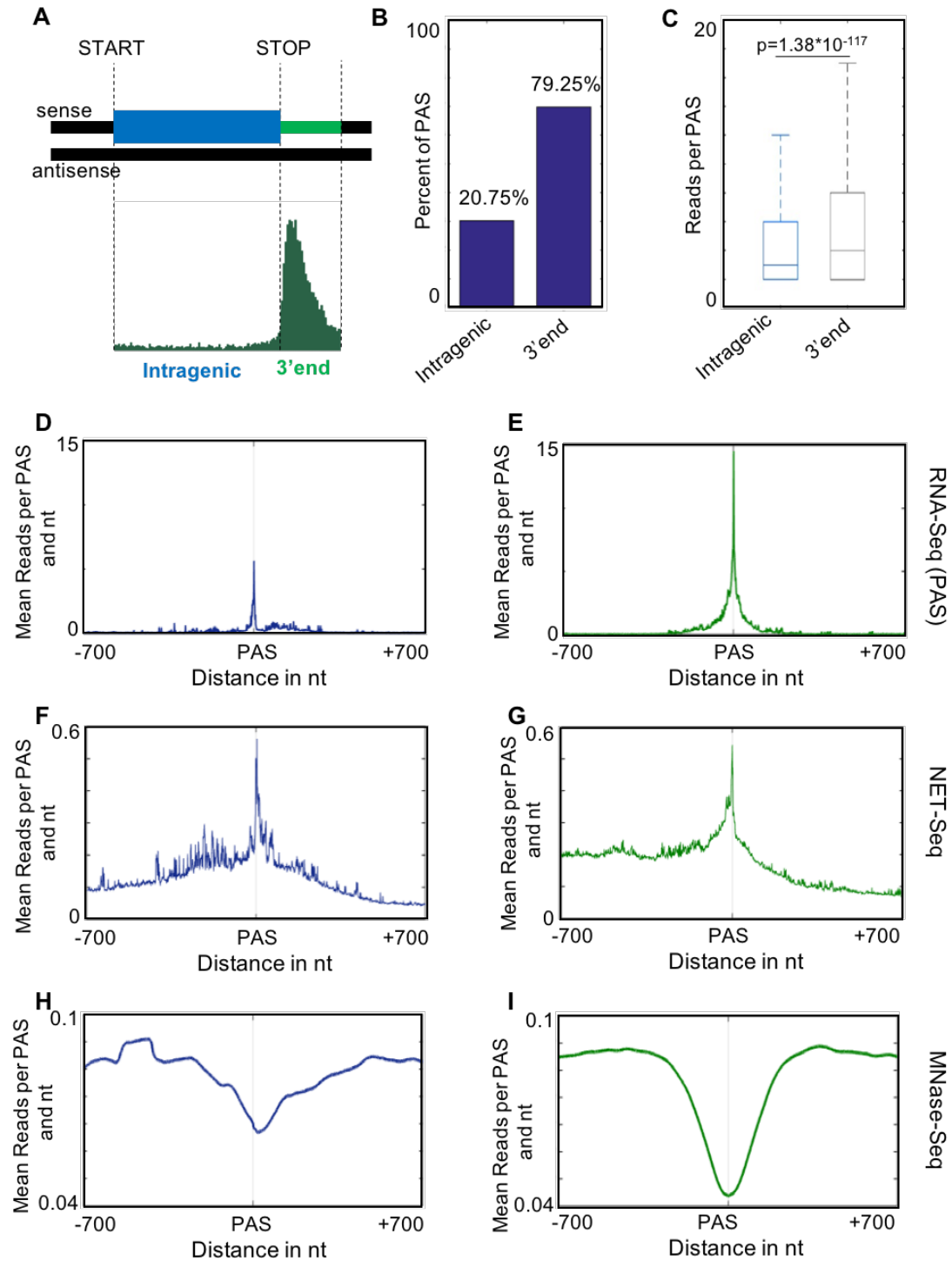


Figure 3.7. Intragenic PAS and 3' end PAS are characterised by the same features.

(A) Definition of intragenic and post 3' end PAS and distribution of PAS across the gene body and 3' end. (B) Percentage of intragenic PAS. (C) Number of Reads in intragenic (blue) and post STOP (grey) PAS. (D-E) PAS distribution surrounding (D) intragenic and (E) post STOP PAS. (F-G) NET-Seq read profiles surrounding (F) intragenic and (G) post STOP PAS. (H-I) Nucleosome occupancy surrounding (H) intragenic and (I) post STOP PAS. Significance between intragenic and post STOP PAS read counts was tested using a Wilcoxon ranksum test.

Overall, intragenic PAS accumulated fewer reads per PAS than 3' end PAS (Figure 3.7 (C, D, E)). However, both intragenic and 3' end PAS are associated with similar features.

Firstly, NET-Seq reads accumulate at the polyadenylation site (Figure 3.7 (F,G)), suggesting that RNAP II transcription speed is diminished at the PAS, probably to allow RNA cleavage to occur. Secondly, both intragenic and 3' end PAS show signs of nucleosome depletion compared to the surrounding region, a side-effect of transcription termination (Figure 3.7 (H, I); Fan et al. 2010). Lastly, both 3' end PAS and intragenic PAS are associated with similar nucleotide frequencies (Figure 3.8 (A, B)), which are a characteristic of cleavage sites (Ozsolak et al. 2010; Moqtaderi et al. 1996; Tuck & Tollervey 2013).

3.1.9 Intragenic and 3' end PAS are associated with distinct motifs

Sequences surrounding 3' end PAS and intragenic PAS were scanned for motifs enrichments (Table 3.2). Motifs matching canonical efficiency elements (EE, TAYRTA), positioning elements (PE, AAWAAA), U-rich motifs and cleavage sites were enriched in sequences flanking both intragenic and 3' end PAS. In addition, some distinct motifs could also be identified for intragenic and 3' end PAS.

Intragenic PAS (N=4509)			3'end PAS (N=19612)		
	Number of sites	E-value		Number of sites	E-value
	948	7.3*10 ⁻⁹⁹		13413	1.8*10 ⁻¹⁰⁵⁸
	1303	4.1*10 ⁻⁵¹		8488	5.8*10 ⁻²³⁵
	1069	5.4*10 ⁻⁴⁶		7906	1.4*10 ⁻²³⁰
	782	6.7*10 ⁻⁴⁴		4271	2.6*10 ⁻¹⁷⁸
	1052	4.1*10 ⁻⁴¹		3512	1.2e*10 ⁻¹¹²
	973	1.3*10 ⁻³⁸		7667	6.3*10 ⁻⁹⁰
	279	6.4*10 ⁻³⁴		3386	2.5*10 ⁻⁶³
	292	1.8*10 ⁻³⁰		5297	9.7*10 ⁻⁴⁶
	976	2.3*10 ⁻²⁶		1009	4.0*10 ⁻⁴²
	163	6.3*10 ⁻²²		2616	3.0*10 ⁻³³
	111	5.0*10 ⁻²⁰		226	2.5*10 ⁻²⁹
	107	8.6*10 ⁻²⁰		271	1.1*10 ⁻²⁰
	145	1.2*10 ⁻¹⁸		137	2.5*10 ⁻¹⁹
	123	5.2*10 ⁻¹⁸		353	3.0*10 ⁻²⁰
	1320	5.3*10 ⁻¹⁸		454	2.0*10 ⁻²¹
	169	8.5*10 ⁻¹⁷		833	2.9*10 ⁻¹⁸
	124	2.2*10 ⁻¹⁶		454	6.4*10 ⁻¹⁸
	113	1.8*10 ⁻¹⁵		169	9.2*10 ⁻¹⁸
	165	1.1*10 ⁻¹⁴		1667	2.0*10 ⁻¹⁷
	455	6.9*10 ⁻¹⁶		606	5.0*10 ⁻¹⁷

Table 3.2. Motifs associated with intragenic and 3' end PAS.

Motif enrichment analysis was performed in sequences 100 nucleotides upstream to 30 nucleotides downstream of the mapped PAS using the DREME tool (Machanick & Bailey 2011). E-values correspond to Benjamini-adjusted p-values.

However, while canonical CPFC binding motifs were detected in both data sets, the distribution of these motifs around the PAS was distinct at intragenic, *i.e.* PAS occurring between the START and STOP codons of coding genes, and 3' end PAS, occurring in a region between the STOP codon and +300 nucleotides downstream (Table 3.2).

For 3' end PAS, efficiency elements were enriched 30-60 nucleotides and positioning element are ~20 nucleotides upstream of the PAS respectively. In addition, U-rich elements flanked the cleavage site, which was located just downstream of the mapped PAS (Figure 3.8 (D)).

By comparison, only the cleavage site (Py(A)_N) was stringently positioned at intragenic PAS and, as a percentage, more intragenic PAS (70.9%) than 3' end PAS (29.3%) had optimal cleavage sites. All other CPFC-associated motifs were distributed randomly around the PAS (Figure 3.8 (C)). Therefore, only RNAs with poly(A) tails generated at the 3' ends of genes might be able to support binding of the CPF and CFs.

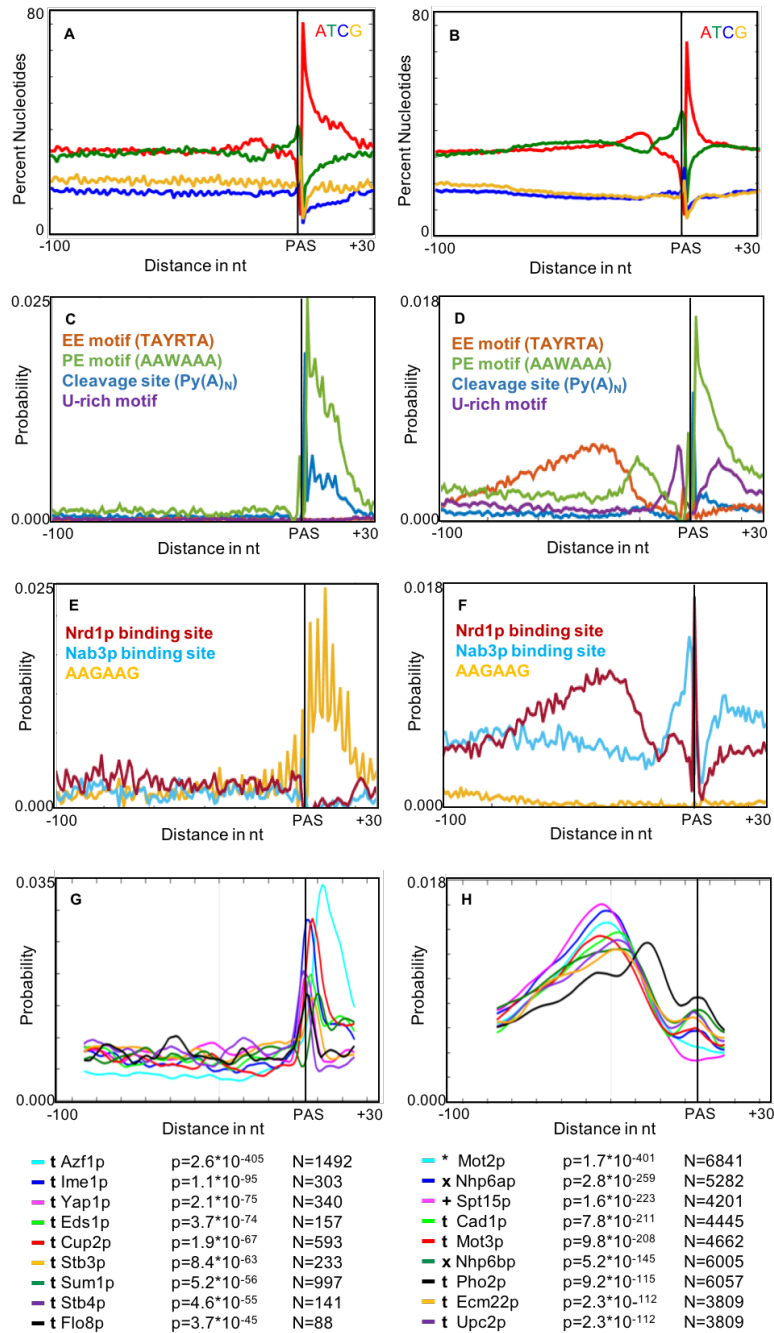


Figure 3.8. PAS-associated motifs.

(A-B) Nucleotide distribution in the vicinity of (A) intragenic (B) 3' end PAS. (C-D) CPFC motifs enriched in sequences surrounding (C) intragenic and (D) 3' end PAS. (E-F) NNS motifs and intragenic-enriched motif in sequences surrounding (E) intragenic and (F) 3' end PAS. Annotated transcription factor binding motif occurrence in sequences surrounding (G) intragenic and (H) 3' end PAS. * CCR4/NOT termination, x Nucleosome remodelling complex, + Transcription initiation factor, t Transcription factor. (G-H) were calculated with CENTRIMO (Machanick & Bailey 2011).

In addition, sequences upstream of 3' end PAS were enriched in Nab3p and Nrd1p binding motifs. Approximately 29.5% of all PAS at the 3' end had both Nab3p and Nrd1p binding sites within 50 nucleotides upstream of the mapped PAS. Both the frequency of Nrd1p and Nab3p binding sites spiked proximal to the PAS (Figure 3.8 (F)). Furthermore, Nrd1p binding sites frequency was increased 30-60 nucleotides upstream of mapped 3' end PAS. Hence, the enrichment of Nrd1p and Nab3p binding sites overlapped with EEs and U-rich elements, suggesting that RNA binding proteins from both termination pathways are positioned in a similar configuration on the RNA. Moreover, both Nrd1p/Nab3p and CPFC binding sites frequently occurred in the vicinity of canonical cleavage sites. Such a fail-safe termination mechanism, which is likely linked to the overlap between STOP codons and Nrd1p binding sequences, has previously been suggested by Webb et al. 2014 and it is likely that both termination pathways can support RNA cleavage and potentially 3' end processing, although no mechanism for NNS-directed RNA cleavage has been described to date.

Similar to CPFC motifs, Nrd1p and Nab3p binding motifs appeared to be randomly positioned in sequences upstream of intragenic PAS (Figure 3.8 (E)), although a slight depletion in the frequency of Nab3p/Nrd1p motif occurrence could be detected just downstream of the mapped PAS. However, sequences just downstream of intragenic PAS were enriched in transcription factors binding motifs (Figure 3.8 (G)). These sequences were only associated with intragenic PAS, as they could not be detected at either 3' end PAS or non-coding gene PAS (not shown).

By contrast, 3' end PAS were enriched in motifs that are bound by Nhp6ap (Figure 3.8 (H)), a nucleosome remodeller that interacts with the FACT complex to evict nucleosomes (Formosa et al. 2001), and Mot2p, a subunit of the Ccr4p-Not complex, which regulates transcriptional elongation rates (Kruk et al. 2011). As these proteins have roles in the establishment of NDRs and regulate transcription, they might be able to directly aid 3' end formation.

3.1.10 **Intragenic PAS are terminated by NNS but not CPFC**

Published data sets, which measure the relative amount of crosslinking of CF-CPF and NNS with RNA (Jamonnak et al. 2011; Baejen et al. 2014), were used to test whether 3' end PAS and intragenic PAS are targeted by the CF-CPF and NNS complexes (Figure 3.9). As only RNAs crosslinked to the protein of interest were sequenced, a high number of sequencing reads mapping to a particular region is indicative of high amounts of RBP binding. Sequences downstream of 3' end PAS are enriched in RNAs that crosslinked with components of the cleavage factors (Figure 3.8 (A-D)), the polyadenylation machinery and poly(A) binding proteins (Figure 3.9 (E, F)). In addition, sequences upstream of 3' end PAS are bound by Nrd1p and Nab3p (Figure 3.9 (G, H)).

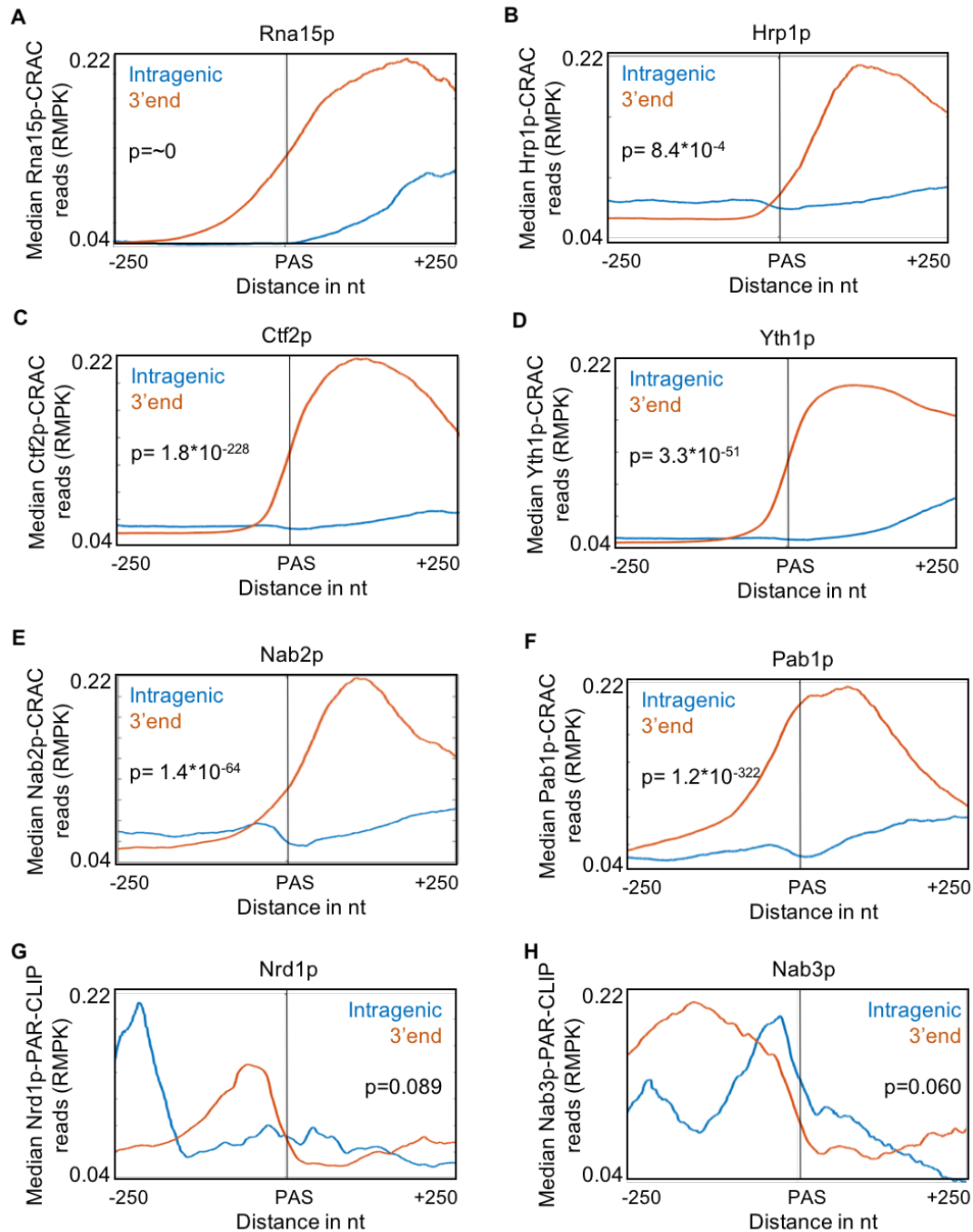


Figure 3.9. Binding of CPFC and NNS subunits to PAS.

(A-D) Binding of CPFC components (A) Rna15p (CFIA), (B) Hrp1p (CFIB), (C) Ctf2p (CPF) and (D) Yth1p (CPF), (E-F) poly(A) binding proteins (E) Pab1p and (F) Nab2p and (G-H) NNS subunits (G) Nrd1p and (D) Nab3p to RNAs at intragenic (blue) and 2'end PAS (red). P-values were calculated using Student's T-tests. Data was taken from (Jamonnak et al. 2011; Baejen et al. 2014) and represents the relative amount of protein crosslinked to RNA. All data was plotted relative to the mapped intragenic (mapped between a START to STOP codon) and 3'end (STOP to +300nt downstream of the STOP codon) PAS.

By contrast, no enrichment in binding of any CF or CPF components could be detected at intragenic PAS. Instead, intragenic PAS tended to be bound by NNS subunits. In total, 23.4% of all intragenic PAS had detectable levels of Nrd1p and Nab3p upstream of the mapped PAS. In addition, 36.5% of PAS were associated with either Nrd1p or Nab3p binding. Hence, approximately 60% of intragenic PAS are possibly targeted by the NNS termination pathways, suggesting that it is the primary termination pathway at these sites. However, an additional 40% do not appear to be targeted by either the NNS or CPFC termination pathway.

3.1.11 PAS increase prior to TF binding sites

A variety of transcription factor binding sites were identified just downstream of intragenic PAS. Unfortunately, these transcription factors have only been linked to particular sequences through *in vitro* assays and studies of deletion mutants. It is therefore impossible to know whether the mapped binding sites are actually targeted by transcription factors *in vivo* and, hence, to assess whether the increase in the number of PAS upstream of these sites is directly linked to transcription factor binding.

As a substitute, other transcription factors, whose binding sites were mapped using ChIP-Seq, ChIP-exo and ORGANIC techniques were analysed with respect to the occurrence of PAS upstream of their annotated binding sites (Rhee & Pugh 2012a; Kasinathan et al. 2013; Ostrow et al. 2014).

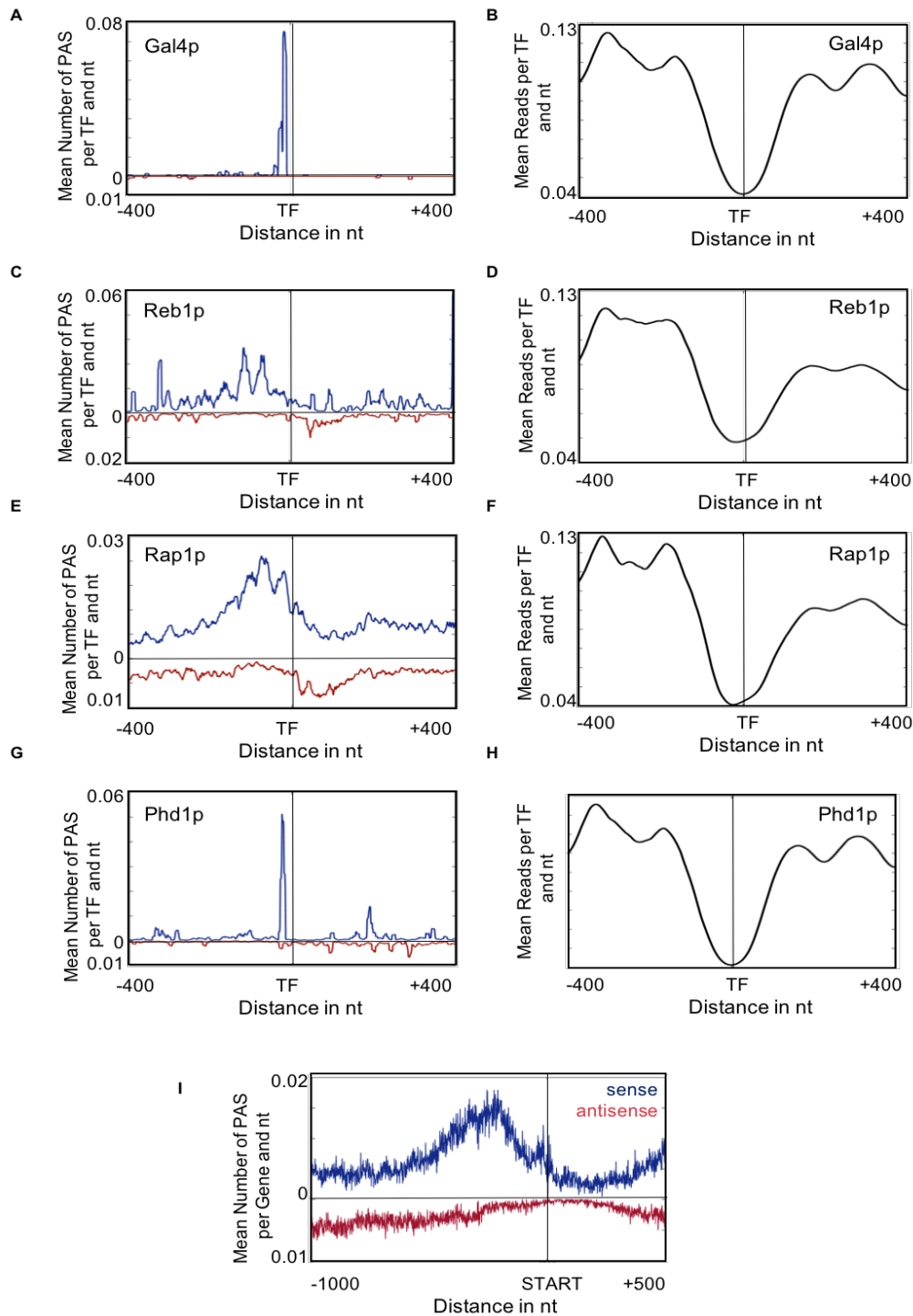


Figure 3.10. PAS Numbers are increased in the vicinity of transcription Factor binding sites.

PAS and nucleosome occupancy in the vicinity of (A,B) Gal4p*, (C,D) Reb1p, (E,F) Rap1p and (G,H) Phd1p TF binding sites. (I) Mean distribution of PAS surrounding the START codon.

*Nucleosome depletion and PAS at Gal4p binding sites were measured 60 minutes after cells were shifted to galactose.

While some factors, including Fkh1p, Fkh2p and Abf1p (not shown), did not show any accumulation of PAS upstream of the binding site, Reb1p (which was previously shown to act as a road-block termination factor; Colin et al. 2014), Rap1p, Gal4p and Phd1p were all associated with increased read-counts upstream of their binding sites (Figure 3.10 (A, C, E, G)). At the same time, nucleosome depletion was observed at the mapped binding sites, which might be occurring to accommodate the transcription factor (Figure 3.10 (B, D, F, H)). As many transcription factors bind at the promoter, the occurrence of PAS was also mapped relative to the START codon of coding genes. As the number of PAS is increased upstream of transcription start sites in the vicinity of the promoter, transcription factors binding to the promoter could prevent transcriptional read-through from upstream genes into the promoter of neighbouring genes.

3.1 Discussion

Transcription is an inherently noisy process. This is reflected by the fact that, even under optimal growth conditions, most genes have multiple transcription start and polyadenylation sites (Pelechano et al. 2013; Gupta et al. 2014; Geisberg et al. 2014; Moqtaderi et al. 2013; Ozsolak et al. 2010). However, in spite of this variation, the ultimate position of these TSS and PAS is, to a large part, determined by the underlying DNA sequences and supported by characteristic changes to the chromatin structure (Kaplan et al. 2009).

Most PAS are located within terminators that are enriched in A/T rich sequences (de Boer et al. 2014) and support the establishment of nucleosome depleted regions through the recruitment of chromatin remodellers (Alen et al.

2002). Due to the increase in nucleosome density in the flanking regions of the NDR, the process of nucleosome remodelling at the 3' ends of genes can in-and-of-itself contribute to transcriptional termination by causing a reduction in polymerase speed (Huang et al. 2013). Moreover, as RNAP II pausing is also associated with intron splicing (Harlen et al. 2016), a reduction in RNAP II elongation rates might be a requirement for RNA processing.

In agreement with this, the data presented in this Chapter supports a scenario in which 3'end formation, determined by the recruitment the CPFC and NNS complexes to the RNA (Ozsolak et al. 2010; Moqtaderi et al. 2013), is inherently associated with nucleosome depletion and a slowdown in RNAP II speed.

As sequences flanking 3' end PAS are bound by CPFC and Nrd1p and Nab3p, it seems that both termination pathways can support the formation of stable mRNAs. This confirms results from a previous study, which measured Nrd1p-Nab3p binding at between 20-30% of protein coding transcripts (Webb et al. 2014). NNS and CPFC subunits likely associate with the RNA via their binding motifs, which are enriched upstream of the cleavage site. Interestingly, both NNS and CPFC motifs were found in very similar regions relative to the cleavage site, implying that the relative positions of both types of binding motifs are major contributors to cleavage site selection. No evidence of NNS-mediated RNA cleavage and polyadenylation has been published to date. In fact, Ysh1p, the endoribonuclease that catalyses the cleavage reaction at polyadenylation sites does not associate with NNS subunits (reviewed in Mischo & Proudfoot 2013). Therefore, although NNS targeted and CPFC targeted PAS are both characterised by the occurrence of cleavage sites, it is unclear whether NNS

mediated cleavage occurs *in vivo* or whether NNS-targeted 3' ends are formed via alternative mechanisms.

Excluding non-coding genes, canonical CPFC as well as NNS termination motifs are mainly enriched at the 3' end of genes (Webb et al. 2014; Tuck & Tollervey 2013). Therefore, most genes' primary PAS map to the termination zone, *i.e.* within the first 300 nucleotides after the STOP codon (Ozsolak et al. 2010). More than one third of coding genes share their termination zones with another convergent gene, suggesting that, like promoters (Rhee & Pugh 2012; Neil et al. 2009; Xu et al. 2009), terminators are inherently bidirectional (Figure 3.11 (A)). As most transcripts are terminated before the start of the gene body of their convergent neighbour, bi-directional terminators might prevent transcriptional read-through into the gene body of convergent genes, an inherent danger in compact genomes such as *S. cerevisiae*'s.

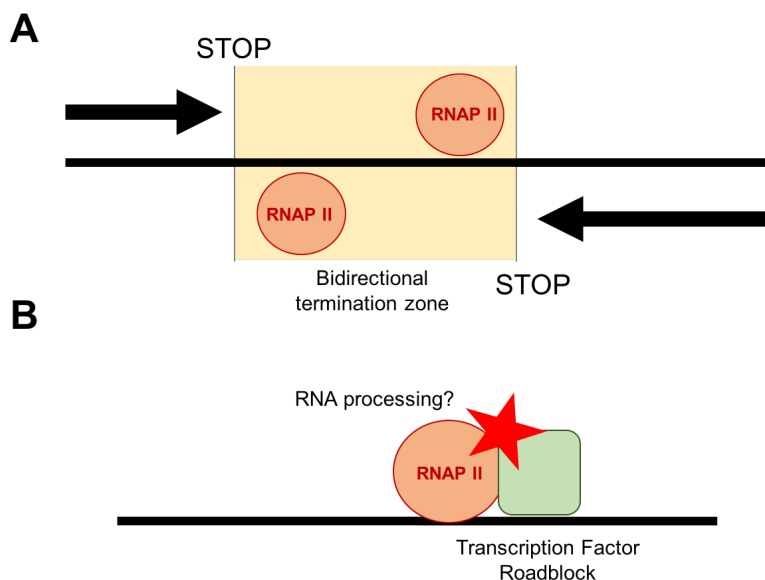


Figure 3.11. Models of transcriptional termination.

(A) Depiction of a bidirectional termination zone. STOP codons act as barriers to transcriptional progression. (B) Termination via a roadblock mechanism.

Since yeast can tolerate considerable variability in their CPFC and NNS binding motifs, most genes have multiple PAS in addition to the primary PAS (Pelechano et al. 2013; Geisberg et al. 2014; Gupta et al. 2014). Secondary PAS were present both at the 3' end and in the intragenic region of many genes. In fact, more than 20% of PAS mapped to the gene body of coding genes, suggesting that the occurrence of intragenic PAS is a wide-spread phenomenon.

In principle, these PAS could simply stem from RNAs that are terminated early on in the transcription process due to defects in RNA processing (Tuck & Tollervey 2013). However, there are several pieces of data that argue against this.

Firstly, early termination products are inherently unstable, have short oligo(A) tails and are targeted for degradation shortly after their transcription. Therefore, they are generally only detectable in yeast poly(A)⁺ RNA-Seq data sets from strains in which *RRP6*, the catalytic subunit of the exosome, is mutated. Indeed, the deletion of *RRP6*, leads to the stabilisation of a variety of RNAs, including early termination products, and a diversification of 3' ends (discussed in Chapter 5). Since intragenic PAS are observed in strains with catalytically active Rrp6p, however, this suggests that they either originate from stable RNAs or are abundant enough to be detectable even in wild-type yeast.

Secondly, if intragenic PAS were a by-product of early termination, one would expect intragenic PAS to map to the all genes irrespective of their molecular function. However, intragenic PAS accumulate primarily in genes coding for ribosomal subunits, genes required for ribosomal biogenesis or co-factors of

cytoplasmic translation. Moreover, the exact positions of the majority of these PAS are reproducible in different experimental set-ups and strains (data not shown).

Lastly, both 3' end PAS and intragenic PAS are associated with nucleosome depletion and spikes in NET-Seq data, both of which are inherently linked to 3' end formation (Fan et al. 2010; Huang et al. 2013). Moreover, intragenic PAS are positioned adjacent to canonical cleavage sites, suggesting that they are a product of 3' end processing. Therefore, it seems likely that transcripts that are terminated at intragenic PAS are generated through a regulated process.

As between 25-60% of intragenic sites appear to be targeted by the NNS termination pathway, a large part of this termination and processing is probably carried out by the NNS complex and its binding partners. Considering no significant binding of CF and CPF subunits could be detected at intragenic PAS, however, it is unlikely that the remaining 40% of intragenic transcripts are terminated via this pathway. Instead, transcription factors might contribute to termination of intragenic PAS in a roadblock-like mechanism, reminiscent of the one reported for Reb1p (Colin et al. 2014). However, unlike in the Reb1p-mediated roadblock mechanism, in the mechanism proposed here transcriptional termination would have to result in the formation of stable RNAs (Figure 3.11 (B)).

Consistent with this, predicted transcription factor binding sites are enriched downstream of mapped PAS and, conversely, the number of PAS is increased just upstream of some known transcription factor binding sites. Hence, termination by transcription factors might be a wide-spread phenomenon,

although the exact mechanism might be transcription factor specific and will have to be further investigated.

In summary, variation in 3' end formation is wide-spread in yeast.

The model presented here suggests that, at the 3' end, cleavage and polyadenylation of transcripts occurs within a transcription termination zone, which is frequently shared between genes. Termination can be initiated via either the CPFC or NNS termination pathway, both of which lead to cleavage and RNA processing.

By contrast, intragenic polyadenylation sites, which are a pervasive phenomenon, are formed after termination via either the NNS termination pathway or via a roadblock-like alternative termination mechanism (Figure 3.11 (B)).

4 Transcription Termination Occurs Via Fail-Safe Mechanisms

4.1 Introduction

At the end of the RNA transcription cycle, RNAP II transcription is terminated via the cleavage and polyadenylation factor (CPF) dependent or Nrd1p-Nab3p-Sen1p (NNS) termination pathways. The former, which involves RNA cleavage and polyadenylation by the cleavage factor IA/IB and cleavage and polyadenylation complex (CPFC), is primarily activated at protein coding genes, while the latter terminates the majority of RNAP III transcribed RNAs and RNAP II transcribed non-coding RNAs (Porrua & Libri 2015). In spite of these preferences, there is significant overlap between CF and NNS-targets and many protein coding transcripts are capable of being terminated by either the canonical CPF- or the NNS-pathway (Rondón et al. 2009; Webb et al. 2014).

One determinant of termination pathway choice is rooted in the phosphorylation patterns of the C-terminal domain (CTD) of the RNAP II. Ser2 phosphorylation of the CTD favours the binding of Pcf11p (Meinhart & Cramer 2004), an essential subunit of the cleavage factor IA (CFIA) complex, which is required for transcription termination and 3' end processing (Amrani et al. 1997). By contrast, Nrd1p preferentially binds to Ser5 phosphorylated CTDs (Vasiljeva et al. 2008). Since the RNAP II CTD is generally phosphorylated at Ser5 in the early gene body and at Ser2p in the late gene body, Nrd1p and Pcf11p tend to localise to the beginning and end of genes respectively (Mayer et al. 2010; Creamer et al. 2011). However, this is only a guideline as there is substantial

overlap between Nrd1p and Pcf11p in genome-wide data sets at individual genes (Kim et al. 2010).

In addition to CTD phosphorylation, the choice of termination pathway is also influenced by the occurrence of specific RNA binding motifs that facilitate the recruitment of the termination complexes. Cleavage and 3' end processing of the RNA by the CPF is directed by upstream efficiency elements (EE), positioning elements (PE), U-rich elements, as well as the cleavage site (CS) itself (Guo & Sherman 1996; Dichtl & Keller 2001). Although UAUUA and AA(U/A)AAA are optimal for the binding of CF IB and CF IA to the EEs and PEs respectively, *S. cerevisiae* can tolerate considerable variations in these motifs (Ozsolak et al. 2010). In fact, in the presence of U-rich motifs, yeast can dispense of the EE and PE all together at some transcripts, although this has negative consequences for cleavage efficiency (Dichtl & Keller 2001). Moreover, cleavage efficiency correlates with the occurrence of secondary structure elements in the form of stem loops (Moqtaderi et al. 2013).

Nrd1p and Nab3p bind to GUAA/G and UCUUG motifs on the RNA via their RRM domains. The occurrence of their binding motifs in the RNA is a crucial determinant of NNS-specificity (Carroll et al. 2004; Jamonnak et al. 2011; Creamer et al. 2011). While variations in the Nrd1p motif can be tolerated and still allow for RNA binding, they reduce the proteins' affinity for RNA *in vitro*. Nab3p and Nrd1p bind RNA cooperatively and, hence, the distance between their respective motifs influences their affinity for specific RNAs (Carroll et al. 2007; Porrua et al. 2012). For this reason, NNS targets frequently have multiple Nrd1p and Nab3p binding motifs in close proximity to another. However, Nrd1p

and Nab3p are also capable of binding RNA on their own with weak affinity and recruiting the TRAMP complex and nuclear exosome (Hobor et al. 2011).

The choice of termination pathway correlates with the stability of RNAP II transcribed RNAs. Broadly speaking, CF/CPF-terminated polyadenylated transcripts are stabilised and exported from the nucleus. Their poly(A) tails are recognised by Pab1p, which is loaded onto the poly(A) tail through direct interactions with the CFP immediately after polyadenylation (Amrani et al. 1997). Binding of Pab1p protects the RNA from degradation by the nuclear exosome and other exonucleases and facilitates the export of the RNA to the cytoplasm via the nuclear pore complexes (Brune et al. 2005).

By contrast, excluding snoRNAs and some mRNAs, most NNS-terminated RNAP II-transcribed RNAs are polyadenylated by the TRAMP complex and degraded by the nuclear exosome immediately after their transcription (Houseley & Tollervey 2009). As Rrp6p degrades oligo(A) tails generated by Trf4p, a TRAMP subunit (Schmid et al. 2012), poly(A) tails of NNS targets tend to be shorter, further reducing nuclear export rates. Moreover, non-coding transcripts are bound by Nab2p, a Pab1p homolog. Similar to Pab1p, Nab2p shuttles between the cytoplasm and nucleus and can facilitate nuclear export (Fasken et al. 2008). However, Rrp6p is able to displace Nab2p, resulting in the exposure of the RNA's poly(A) tail and its degradation by Rrp6p. As Rrp6p is the primary effector of RNA degradation downstream of NNS-mediated termination, a large fraction of transcripts is stabilised and hyperpolyadenylated in the absence of Rrp6p (Neil et al. 2009; Thiebaut et al. 2006; Thompson & Parker 2007; Schmid et al. 2012).

Using poly(A)+ 3' end RNA-Seq data it is possible to map 3' ends that are present in the wild-type and compare them those that are only detectable in the *rrp6Δ*. As Rrp6p targets transcripts that are terminated via the NNS termination pathway, 3' ends that occur exclusively in the *rrp6Δ* mutant should primarily stem from NNS-targeted RNAs. By contrast, PAS that are present in both the wild-type and *rrp6Δ* yeast presumably originate from CPF-targeted transcripts. Hence, it is possible to directly compare and contrast PAS of CPF- and NNS-targets and link them to specific features such as nucleotide content. Considering this premise, the aim of this Chapter is to further characterise the PAS of NNS-targeted transcripts through a direct of comparison of their 3' ends.

4.2 Results

4.2.1 *rrp6Δ* strains contain increased concentrations of unstable non-coding RNAs

Consistent with previous reports (Xu et al. 2009; Neil et al. 2009), the *rrp6Δ* mutant was characterised by high concentrations of polyadenylated non-coding RNAs (Figure 4.1. (A)). In total, concentrations of 481 polyadenylated non-coding RNAs increased more than two-fold in the *rrp6Δ* mutant compared to the wildtype (Figure 4.1. (C)). In addition, 1480 non-coding RNAs, which were absent in the wildtype RNA-Seq data, could be detected in the mutant.

Unstable CUTs and NUTs, *i.e.* known targets of Rrp6p-mediated degradation (Schulz et al. 2013; Xu et al. 2009), were particularly affected (Figure 4.1 (A)). Concentrations of CUTs and NUTs that were present in both the mutant and wildtype increased by 3.6-fold and 4.8-fold respectively. Moreover, 441 CUTs

(47.7%) and 508 NUTs (33.29%) were present in the *rrp6Δ* but absent in wildtype yeast.

Concentrations of stable SUTs and Xrn1p-targeted XUTs were less affected and only elevated by 1.8- and 1.6-fold respectively, although the total number of detectable SUTs and XUTs also rose by 193 (22.79%) and 338 (18.80%) genes respectively.

By contrast, overall concentrations of coding RNAs did not change in the absence of Rrp6p (Figure 4.1 (B)).

Using a Rrp6p cross-linking-and-cDNA-analysis (CRAC) dataset (Schneider et al. 2012), which measures Rrp6p binding to RNA, *rrp6Δ*-stabilised RNAs were compared with RNAs that did not show a significant change in concentration in the *rrp6Δ* mutant. *rrp6Δ*-stabilised RNAs were characterised by elevated levels of Rrp6p binding in the data set ($p=3.1 \times 10^{-11}$), confirming that these RNAs are targeted by Rrp6p (Figure 4.1 (C)).

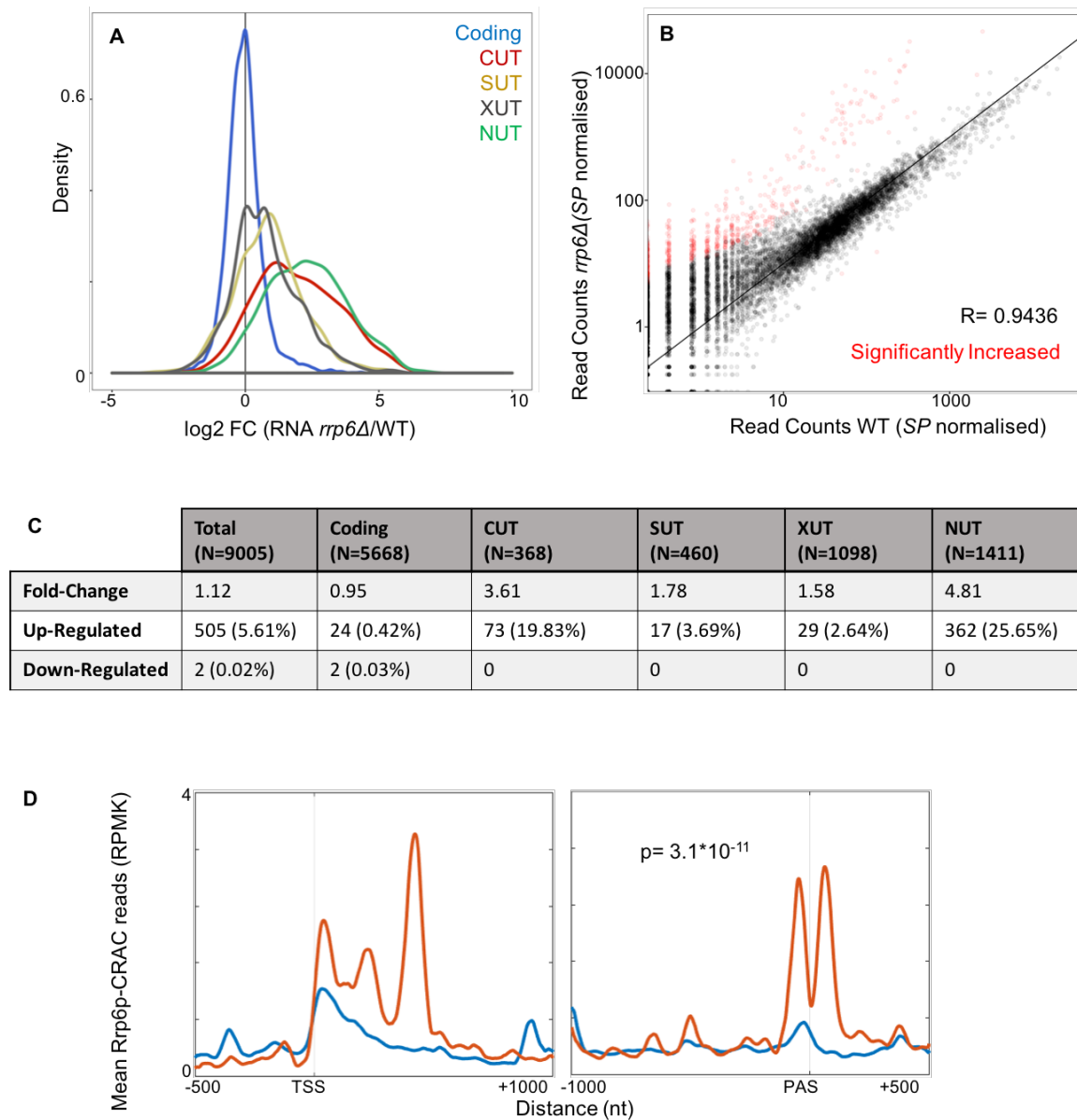


Figure 4.1. Changes in RNA abundance between WT and *rrp6Δ* mutants.

(A) Distribution of fold-changes in different gene classes. (B) Scatter plot of RNA-Seq read counts in *rrp6Δ* and wildtype yeast. (C) Summary of fold-changes and the number of genes affected in the *rrp6Δ* mutant. Fold change estimates were based on three replicates and calculated using DESeq2. Genes were defined as up-regulated or down-regulated if their RNAs increased or decreased more than two-fold and had a Benjamini adjusted p-value of less than 0.05. (D) Metagene profile of Rrp6-binding to RNAs of unaffected (blue) and *rrp6Δ*-stabilised RNAs (red). The Rrp6-CRAC data was taken from (Schneider et al. 2012)

4.2.2 Number of 3' end position at non-coding genes is elevated in *rrp6Δ* mutants

Correlating with the increase in RNA concentrations, RNA 3' end positions were more diverse in the *rrp6Δ* mutant. Overall, the number of distinct 3' ends increased 1.27-fold in the *RRP6* deletion mutant compared to the wildtype after normalisation to sequencing depth, resulting in a total number of 51761 PAS in the *rrp6Δ* strain.

29239 of these PAS were also detected at precisely the same position in the wildtype. Hence, approximately 78.89% of the canonical 3' ends present in the wildtype are conserved in the *rrp6Δ* mutant. These 3' ends will be referred to as conserved PAS in the remainder of the text, while remaining 23392 PAS overserved exclusively in the *rrp6Δ* strain will be called *rrp6Δ*-specific PAS. PAS in this context will refer to the positions of all observed 3' ends, independently of whether the 3' end was likely formed via the CPFC pathway or not. Examples of conserved and *rrp6Δ*-specific PAS are shown in Figure 4.4.

53.1% of *rrp6Δ*-specific PAS mapped to annotated non-coding, RNAP II-transcribed genes. By comparison, only 7.2% of conserved PAS mapped to the same class of genes. Hence, this corresponds to an 8.4-fold increase in the number of non-coding PAS. CUTs were particularly affected by the deletion of *RRP6*, which resulted in a 16.0-fold increase in the number of PAS. However, the number of PAS also increased significantly at NUTs, XUTs and SUTs, which were characterised by 6.7-, 5.5- and 4.3- fold increases in the number of PAS respectively (Figure 4.2 (A, B)). Therefore, the fold increase in the diversity

of PAS is greater than the increase in RNA concentrations, suggesting that the choice of termination sites is highly variable in non-coding genes

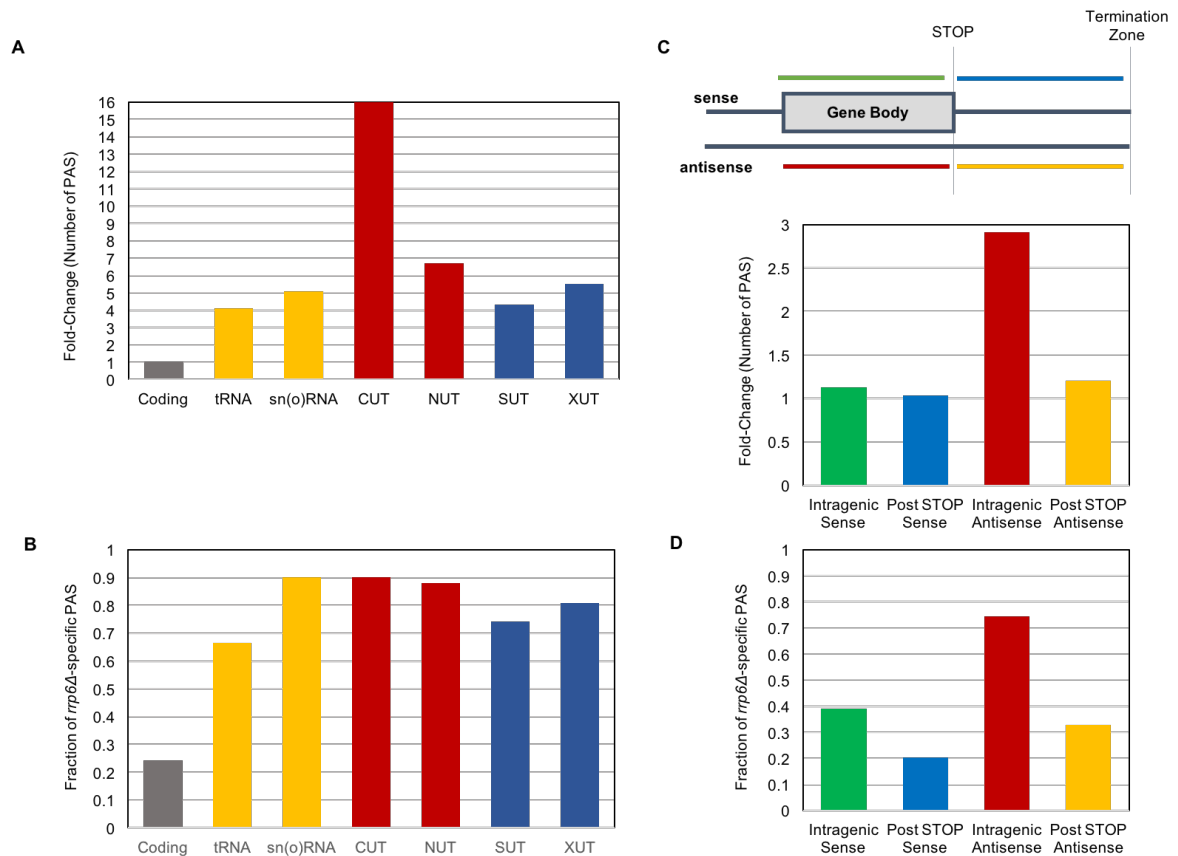


Figure 4.2. Changes in the number of PAS in *rrp6Δ* mutants compared to wildtypes.

(A) Fold changes in the number of PAS in *rrp6Δ* compared to wildtypes in different gene types. (B) Fraction of *rrp6Δ*-specific PAS as a function of the total number of PAS wildtypes in different gene types. (C) Fold changes in the number of PAS in *rrp6Δ* compared to wildtypes at different positions in coding genes. (B) Fraction of *rrp6Δ*-specific PAS as a function of the total number of PAS wildtypes at different positions in coding genes.

4.2.3 *rrp6Δ*-specific PAS occur across the gene and map antisense to coding genes

rrp6Δ-specific PAS did not specifically map to the 3' ends of non-coding genes. Instead, they were distributed over the whole of the gene body as well as the 3' end of the gene (Figure 4.2 (A, B), Figure 4.3 (A, B)). Interestingly, the number of PAS increased more substantially over the gene body than at the 3' end of

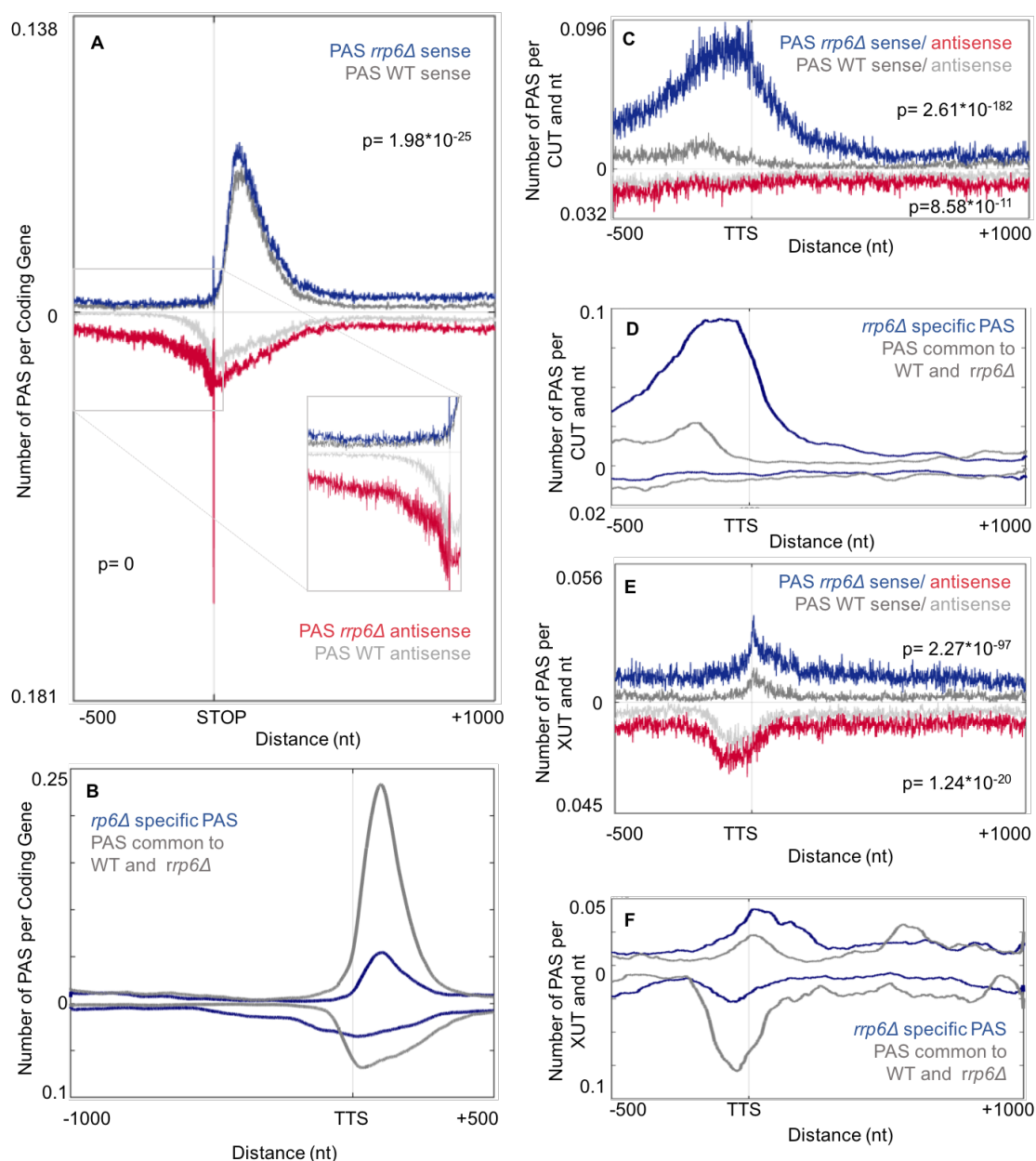


Figure 4.3. Distribution of PAS in wildtypes and *rrp6Δ* strains.

(A,C,E) Number of PAS per nucleotide at the 3' end of (A) coding genes (C) CUTs and (E) XUTs in wild-type and *rrp6Δ* yeast. (B,D,F) Distribution of conserved and *rrp6Δ*-specific PAS at the 3' end of (B) coding genes, (D) CUTs and (F) XUTs in wild-type and *rrp6Δ* yeast. TTS- transcription termination site. STOP- stop codon. nt-nucleotides.

non-coding genes. Hence, non-coding genes PAS are, unlike coding gene PAS, not restricted to a termination zone (Figure 4.3 (C-F)). Moreover, 40.4% of *rrp6Δ*-specific non-coding gene PAS map antisense to the gene body of coding genes, a region which is normally comparatively depleted of PAS in wildtype yeast.

4.2.4 The deletion of *RRP6* has little effect on coding genes

In contrast to non-coding genes, the total number of PAS at coding genes was virtually unaffected (1.05-fold change). In addition, 79.8% of all 3' ends mapped to exactly the same position in the wildtype and *rrp6Δ* mutants. Similarly, positions of 93.6% of primary PAS, *i.e.* the PAS to which the largest fraction of reads map at a particular gene, were unaffected by the deletion of *RRP6*.

The remaining 20.2% PAS, which mapped to the 3' end of coding genes, were *rrp6Δ*-specific. On average, these PAS were shifted by only 17 nucleotides from their next proximal PAS, likely to resulting in minor changes in, for instance, transcript stability.

This means that most PAS at the 3' end of coding genes were unaffected by the absence of Rrp6p. Furthermore, when changes were observed, these primarily affected secondary PAS and are unlikely to affect the coding potential of the transcript.

By comparison, the position of intragenic sites was more variable, as 37.5% of PAS in the *rrp6Δ* mutant did not have a corresponding match in the wildtype. However, it is possible that this is not an *rrp6Δ*-specific phenomenon, but reflects an inherently variability in the positioning of intragenic PAS.

4.2.5 *rrp6Δ* mutants accumulate more antisense PAS

In spite of the minor changes observed on the sense strand, significantly more PAS accumulated on the antisense strand of coding genes. Overall, antisense PAS increased 2.1-fold in the *rrp6Δ* mutant, even when annotated non-coding genes and convergent, overlapping coding genes were excluded from the analysis. Approximately half of these antisense PAS mapped opposite to the gene body of the coding gene. For clarification of where these PAS were located, refer to the Figure 4.2 (C), where the region is indicated in red. Antisense PAS mapping opposite to the gene body of the coding gene were shifted, on average, 341 downstream nucleotides from the preceding PAS. These large shifts are consistent with the stabilisation of additional, potentially unannotated, non-coding transcripts.

When taking into account PAS mapping to non-coding genes, 1308 coding genes had an increased number of PAS antisense to their gene body. Of these, 770 genes did not have any detectable antisense PAS prior to the deletion of *RRP6*. No association with any molecular pathway could be identified, suggesting that the appearance of these PAS is independent of the type of transcript on the sense strand. The distribution of PAS in the vicinity of the STOP codon is shown in Figure 4.3 (A, B). Moreover, examples of the

distribution of polyadenylation sites as well as NET-Seq reads and termination factor binding patterns in wild-types and *rrp6Δ* mutants are shown in Figure 4.4.

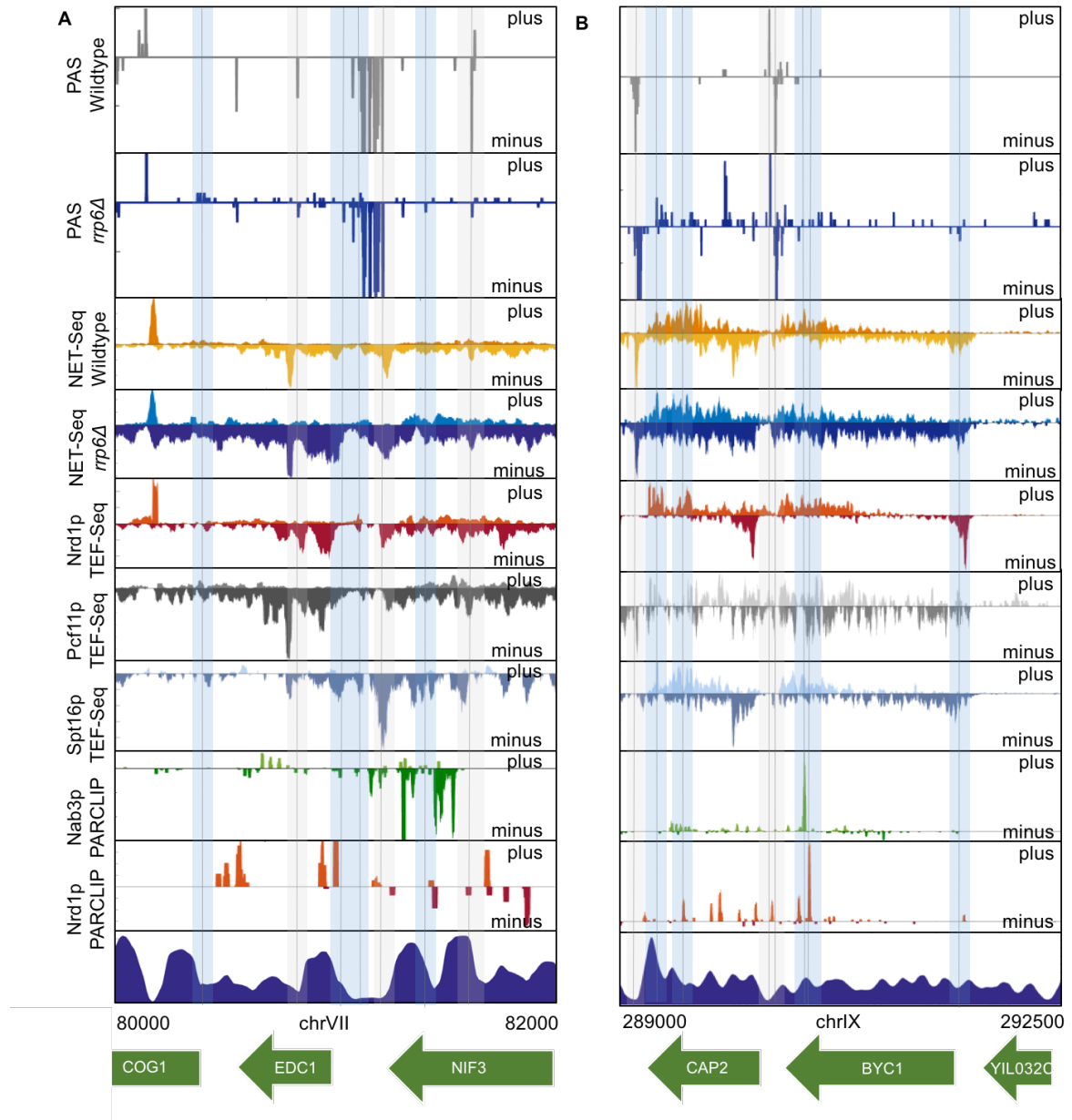


Figure 4.4. Examples of *rrp6Δ*-specific and conserved PAS.

Blue boxes indicate *rrp6Δ*-specific PAS, grey boxes conserved PAS. NET-Seq data (WT), Nrd1p, Pcf11p and Spt16p TEF-Seq data was collected by Fischl, H. Nrd1p and Nab3p PAR-CLIP data were taken from Jamonnak et al. 2011.

4.2.6 Rrp6p targets the majority of tRNAs and snRNAs

Additional polyadenylation sites were also identified in the vicinity of snoRNAs and tRNAs, whereby the number of PAS increased by 4.1- and 5.1-fold respectively.

Low concentrations of 40 polyadenylated tRNAs could be detected in the wild-type RNA-Seq data. However, polyadenylated transcripts were much more abundant in the *RRP6* deletion mutant, in which 126 tRNA genes (42.1% of all tRNAs) had detectable polyadenylated transcripts (Figure 4.5). Eight tRNA precursors, which were previously shown to be targets of Trf4p (Skowronek et al. 2014), were detected, as well as 70 additional tRNAs (see Table 4.1). 35.3% of PAS mapped to the gene body of tRNAs, while the remaining 64.7% mapped to the 3' end of the tRNAs. The data are consistent with a role for Rrp6p in 3' end processing at almost half of all tRNAs, as well as the removal of either defective splicing products or prematurely terminated transcripts. As Rrp6p is thought to play a minor role in tRNA 3' end maturation (Skowronek et al. 2014; Copela et al. 2008), these results suggest that Rrp6p targets a wider range of tRNAs than previously thought.

Wild-type	<i>rrp6Δ</i> -specific	
tE(CUC)D	tA(UGC)L	tP(UGG)M
tE(UUC)E2	tD(GUC)N	tP(UGG)O3
tE(UUC)G2	tE(CUC)D	tQ(UUG)D1
tE(UUC)G3	tE(UUC)E1	tQ(UUG)D2
tE(UUC)K	tE(UUC)E2	tQ(UUG)H
tG(CCC)D	tE(UUC)G3	tR(ACG)D
tG(GCC)D2	tE(UUC)J	tR(ACG)E
tG(GCC)P1	tF(GAA)M	tR(ACG)K
tH(GUG)G2	tG(CCC)D	tR(ACG)L
tK(CUU)F	tG(GCC)D2	tR(CCG)L
tK(CUU)M	tG(GCC)E	tR(CCU)J
tK(UUU)G2	tG(GCC)F1	tR(UCU)G1
tL(CAA)A	tG(GCC)J2	tR(UCU)M2
tL(CAA)D	tG(GCC)P1	tS(CGA)C
tL(CAA)G1	tH(GUG)E2	tS(GCU)O
tL(UAA)B2	tH(GUG)G2	tS(UGA)E
tM(CAU)E	tI(AAU)B	tS(UGA)I
tM(CAU)J3	tI(AAU)E2	tS(UGA)P
tM(CAU)O1	tI(UAU)D	tT(CGU)K
tM(CAU)P	tI(UAU)L	tT(UGU)G1
tP(AGG)C	tK(CUU)D1	tT(UGU)G2
tP(AGG)N	tK(CUU)F	tT(UGU)H
tQ(UUG)D3	tK(CUU)I	tV(AAC)M1
tR(ACG)D	tK(CUU)K	tV(AAC)M3
tR(ACG)E	tK(UUU)G2	tV(CAC)D
tR(ACG)K	tK(UUU)K	tV(CAC)H
tR(ACG)L	tL(CAA)A	tV(UAC)B
tR(CCG)L	tL(CAA)D	tW(CCA)G1
tS(CGA)C	tL(CAA)G1	tW(CCA)G2
tS(GCU)O	tL(CAA)M	tW(CCA)K
tS(UGA)E	tL(UAA)B2	tW(CCA)P
tS(UGA)I	tL(UAA)N	tX(XXX)D
tS(UGA)P	tL(UAG)L1	tY(GUA)D
tT(CGU)K	tM(CAU)E	tY(GUA)F2
tT(UGU)G2	tM(CAU)O1	tY(GUA)J1
tT(UGU)H	tM(CAU)O2	tY(GUA)O
tV(UAC)B	tM(CAU)P	
tW(CCA)G1	tP(AGG)C	
tW(CCA)P	tP(AGG)N	
tX(XXX)D	tP(UGG)H	

Table 4.1. Identified tRNAs in the wild-type and *rrp6Δ*.

Red indicates tRNAs that are known Rrp6p-targets (Copela et al. 2008).

A similar phenomenon could be observed at snoRNAs. snoRNAs are terminated at terminators I and II (Figure 4.5. (D); Richard & Manley 2009). The Nrd1p/Nab3p- and TRAMP-dependent terminator I is the principal termination site in yeast. By contrast, transcripts terminated at terminator II are TRAMP-independent, but processed by the nuclear exosome (Grzechnik & Kufel 2008). Transcripts terminated at both terminators are polyadenylated by Pap1p, whereby polyadenylation is a prerequisite for snoRNA maturation. Moreover, both pathways require Rrp6p for processing, as snoRNA precursors are stabilised in *rrp6Δ* mutants (Kim et al. 2006).

The absence of Rrp6p affected all snoRNAs. Both the concentration of polyadenylated snoRNAs transcripts and the number of PAS per gene was elevated at all 81 known snoRNAs. The majority of these PAS mapped either to the 3' end of the snoRNA or to 30 base pairs downstream of the annotated 3' end, *i.e.* positions corresponding to terminator I and II. While terminator I appeared to be almost exclusively used in the wildtype, in *rrp6Δ* yeast snoRNAs were frequently terminated at terminator II. In fact, more individual PAS were detected at terminator II than at terminator I in the *rrp6Δ* mutant.

Furthermore, intragenic PAS were more common in the *RRP6* deletion mutant. Hence Rrp6p appears to have a role in terminator selection, processing of snoRNAs and the removal of defective transcripts.

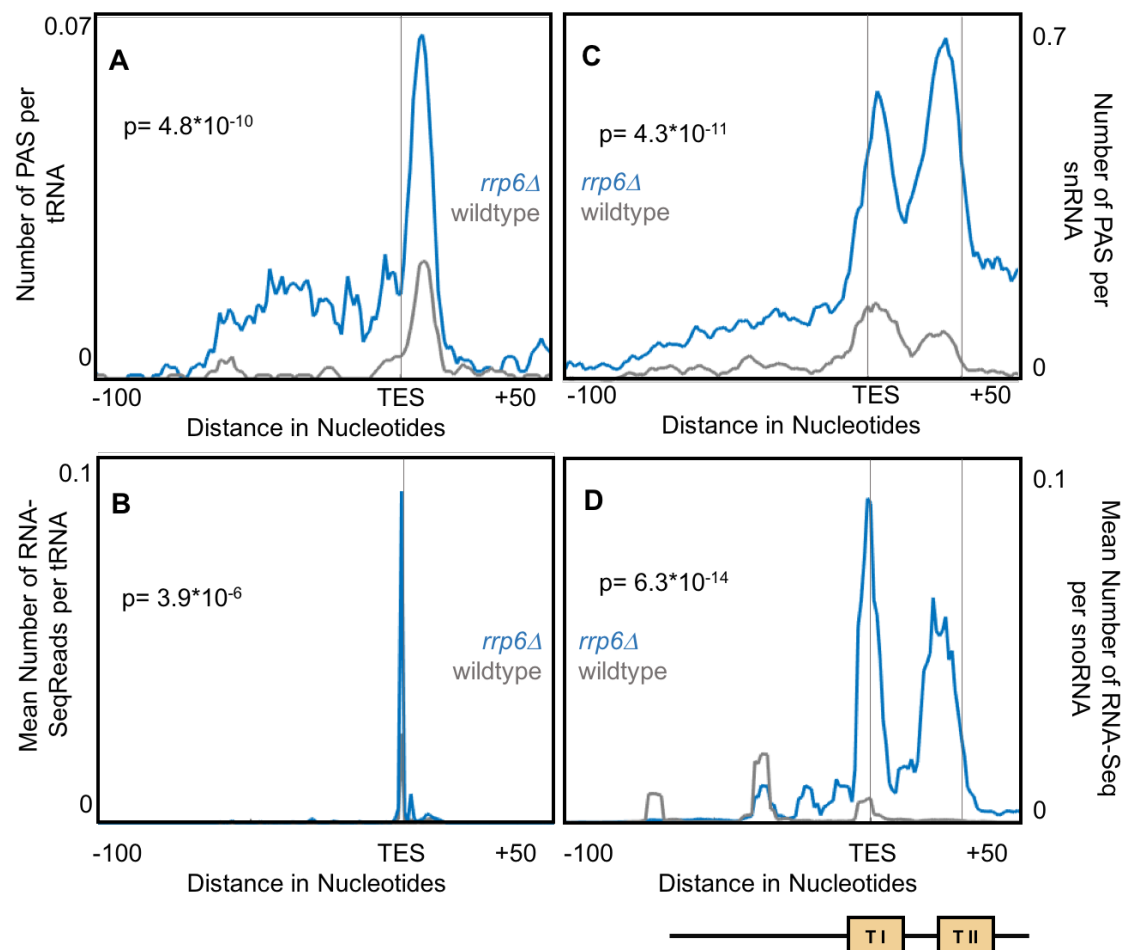


Figure 4.5. Transcription and transcript stabilisation of tRNAs and snoRNAs.

(A,C) Distribution of RNA-Seq reads at the 3' ends of (A) tRNAs and (C) snoRNAs. (B,D) Distribution of RNA-Seq reads at the 3' ends of (B) tRNAs and (D) snoRNAs. TI- snoRNAs terminator I. TII- snoRNAs terminator II. *rrp6Δ* mutants data is shown in blue, wild-type data in grey.

4.2.7 *rrp6Δ*-specific PAS are used less frequently than conserved

PAS

Similar to canonical PAS, *rrp6Δ*-specific PAS generally were associated with increased RNA-Seq read densities in the vicinity of the PAS (Figure 4.6).

However, the average number of reads per PAS was significantly lower than for canonical PAS, suggesting that *rrp6Δ*-specific PAS were used less frequently than canonical PAS even in the *rrp6Δ* mutant.

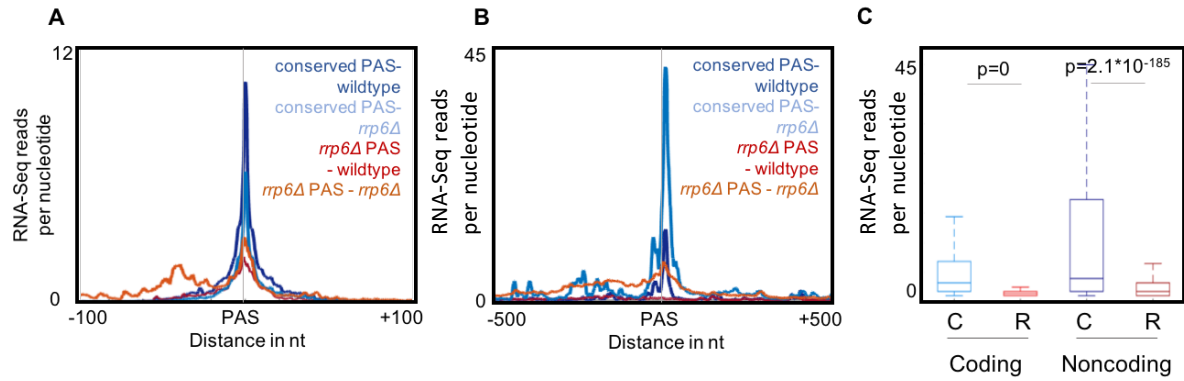


Figure 4.6. Frequency of RNA-Seq reads at conserved and *rrp6Δ*-specific sites.

RNA-Seq read densities at (A) coding and (B) non-coding genes. (C) Comparison of reads per PAS at conserved and *rrp6Δ*-specific PAS of coding and non-coding genes. C- conserved PAS. R- *rrp6Δ*-specific PAS.

4.2.8 Positions of *rrp6Δ*-specific PAS correlate with Rrp6p and Rrp44p binding

To confirm that sequences of *rrp6Δ*-specific PAS are bound by components of the nuclear exosome, binding of Rrp6p and Rrp44p to RNA was examined using published CRAC data sets (Schneider et al. 2012). At coding genes, *rrp6Δ*-specific PAS were associated with significantly higher read counts than conserved PAS, whereas at non-coding genes both conserved and *rrp6Δ*-specific PAS were enriched in Rrp44p- and Rrp6p-CRAC reads (Figure 4.7). This suggests that at some coding genes Rrp6p might have function in the degradation or processing of particular transcripts. Moreover, in agreement with previous reports (Schneider et al. 2012; Gudipati et al. 2012), it implies that both comparatively stable as well as unstable non-coding RNAs can be targeted by the nuclear exosome.

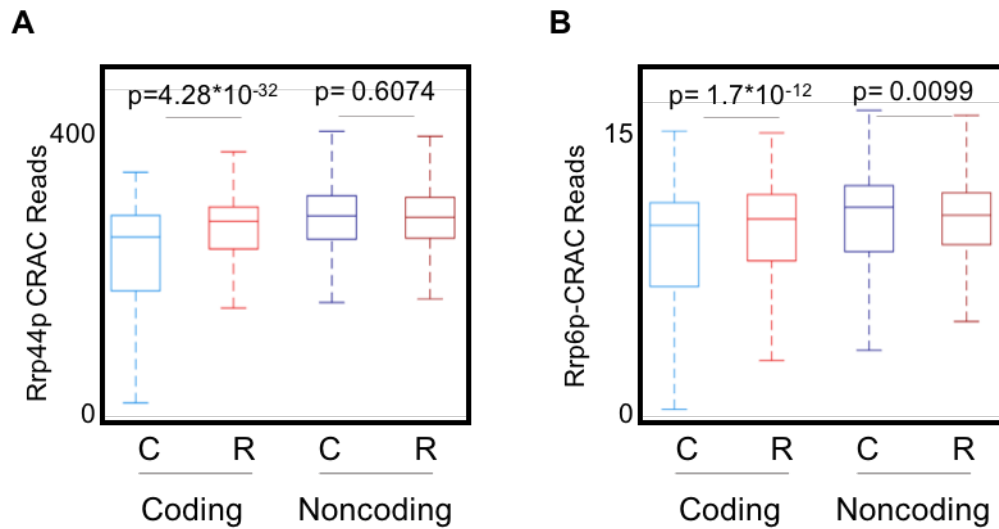


Figure 4.7. Rrp6 and Rrp44 binding to RNAs.

(A) Rrp44 binding to RNAs upstream of conserved and *rrp6Δ*-specific PAS. (B) Rrp6 binding to RNAs upstream of conserved and *rrp6Δ*-specific PAS. C- conserved PAS. R- *rrp6Δ*-specific PAS.

4.2.9 *rrp6Δ*-specific PAS occur downstream of highly transcribed regions

Nascent transcription rates were recorded using NET-Seq in both the *rrp6Δ* and wildtype strains. Although global read distributions were shifted slightly towards the 3' end of the gene in the *rrp6Δ* strain (Figure 4.8 (A-D)), transcription rates were highly correlated in the wildtype and *rrp6Δ* at both coding and non-coding genes (Figure 4.8 (E, F)).

However, nascent transcription rates per nucleotide were generally slightly elevated at coding genes in the *rrp6Δ* compared to the wildtype strain. By contrast, transcription rates at non-coding genes, whose RNAs are stabilised in the absence of Rrp6p, were not affected in the same way (Figure 4.8 (H)).

Hence it is likely that transcription rates of non-coding genes, whose RNAs are

stabilised in the absence of Rrp6p, are regulated by feedback mechanisms aimed at minimizing their accumulation.

Transcription rates also correlated with mature RNA concentrations. The magnitude of these correlations was very similar at coding genes in the *rrp6Δ* and wildtype strains, but was markedly increased in *rrp6Δ* yeast at non-coding genes (Figure 4.8. (G)). As non-coding gene transcription rates are correlated in *rrp6Δ* and wildtype strains, this suggests that non-coding genes are constitutively transcribed in the wild-type, but stabilised in *rrp6Δ* mutant. Similar conclusions were reached in previous reports (Neil et al. 2009; Pelechano et al. 2013; Murray et al. 2015; Churchman & Weissman 2011),

Consistent with the constitutive transcription of Rrp6p-targets, regions upstream of *rrp6Δ*-specific PAS were transcribed at high rates in the wildtype (Figure 4.9 (A-B)). In fact, as many *rrp6Δ*-specific PAS mapped to the gene body in the vicinity of the transcription start site of non-coding genes, they originated from regions that are associated with high read counts.

In spite of this, fewer NET-Seq reads mapped to the actual polyadenylation site in *rrp6Δ*-specific PAS than compared to conserved PAS. Since, the accumulation of NET-Seq reads at polyadenylation sites is likely associated with polymerase pausing, it appear that fewer RNAP II are stalled at *rrp6Δ*-specific PAS than conserved PAS (Figure 4.9 (C,E)).

In addition, the data suggest that termination at *rrp6Δ*-specific PAS is not the result of delayed termination after transcriptional read-through, as has recently

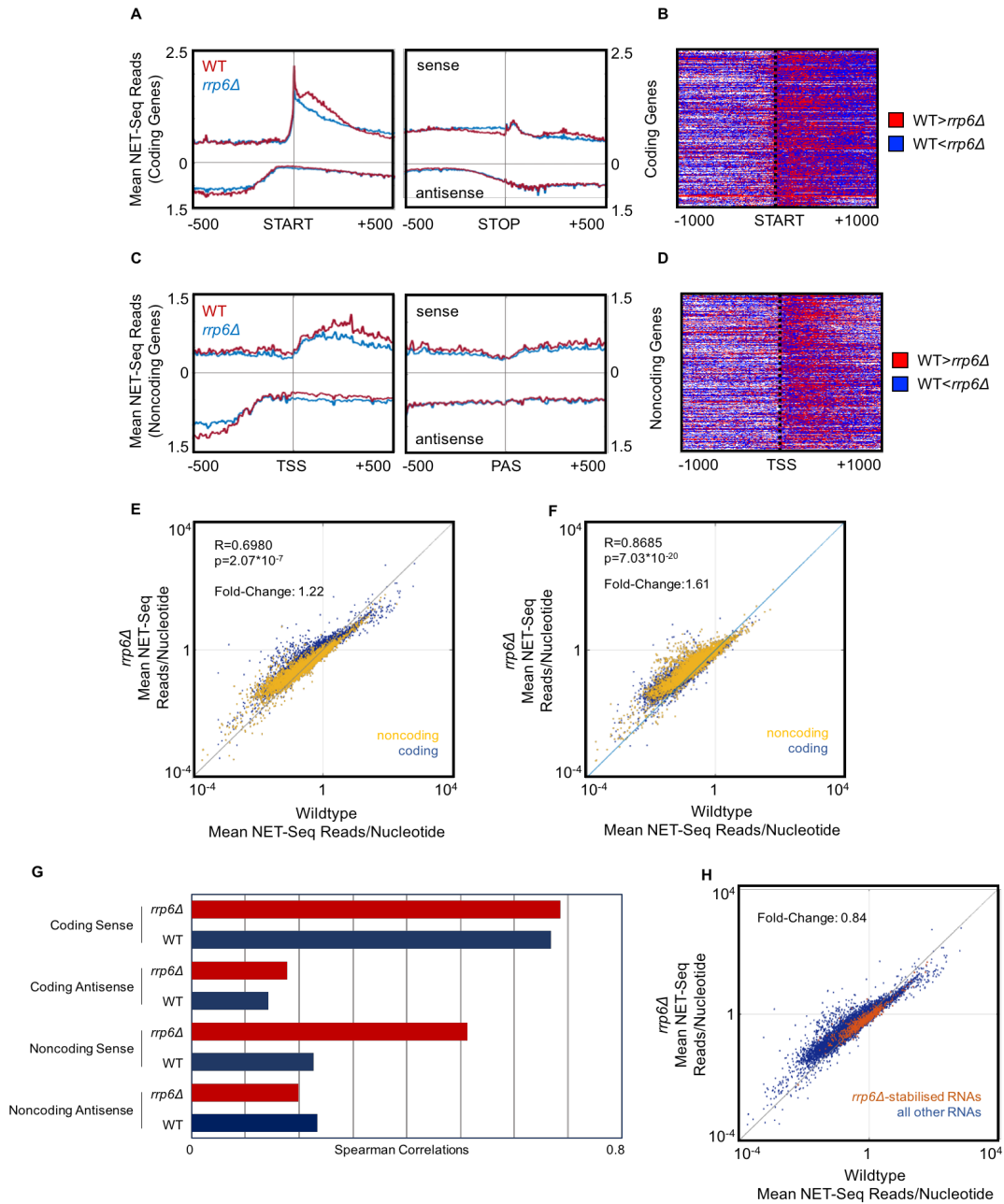


Figure 4.8. Nascent transcription rates in *rrp6Δ*.

(A,C) Metagene profiles of wild-type and *rrp6Δ* strain NET-Seq read distributions at (A) coding and (C) non-coding genes. (B,D) Heatmap of changes in read numbers at the 5' end of individual (B) coding and (D) non-coding genes. Positions with higher read numbers in the wild-type are shown in red, positions with higher read numbers in the *rrp6Δ* strain are shown in blue. (E-F) Scatter plots of NET-Seq read per nucleotide in wild-type and *rrp6Δ* strains in the (E) sense and (F) antisense direction of coding (blue) and non-coding (yellow) genes. R-values indicate the Spearman correlations, p-values the statistical significance in the difference between the number of reads in the wild-type and *rrp6Δ* strains. The fold-change value indicates the average fold-change in NET-Seq reads per gene between the two strains. (G) Spearman correlations of RNA and NET-Seq data at different gene classes. (H) Scatter plot of NET-Seq reads per nucleotide in wild-type and *rrp6Δ*. Genes whose transcripts are stabilised in *rrp6Δ* mutants are shown in red, genes whose transcripts are unaffected are shown in blue. START-START codon. STOP-STOP codon. TSS- transcription start site. PAS- annotated polyadenylation site.

been proposed (Lemay et al. 2014; Fox et al. 2015). If transcriptional read-through were the main cause for accumulation of *rrp6Δ*-specific isoforms, one would expect a higher number of reads to be detectable proximal to *rrp6Δ*-specific PAS in the *rrp6Δ* strain. To test this, the ratio of NET-Seq reads mapping proximal to *rrp6Δ*-specific PAS, *i.e.* a region that should only be transcribed in *rrp6Δ* mutant, was compared to a region further upstream, *i.e.* a region that should be transcribed in wild-types as well as *rrp6Δ*. Since there was no statistically significant difference in the ratio of reads between wild-types and *rrp6Δ* mutants, one can assume that the majority of *rrp6Δ*-specific PAS are not a by-product of transcriptional read-through (Figure 4.9 (F)). However, although the data suggest that transcriptional read-through is not a wide-spread mechanism, the possibility that Rrp6p prevents read-through at specific loci cannot be excluded.

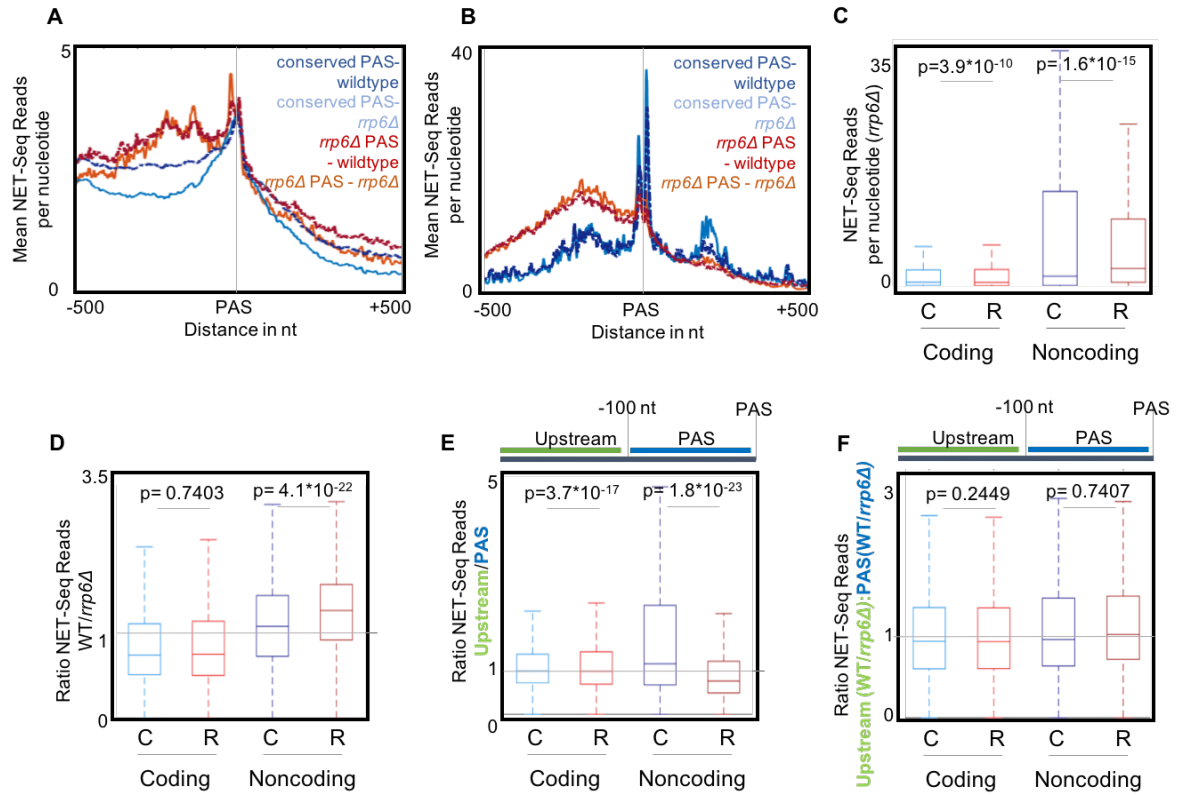


Figure 4.9. Nascent RNA transcription at *rrp6Δ*-specific and conserved PAS.

NET-Seq read densities at (A) coding and (B) non-coding genes. (C) Comparison of transcription rates per PAS at conserved and *rrp6Δ*-specific PAS of coding and non-coding genes. (D) Ratio of the relative transcription rates in wild-type and *rrp6Δ* strains. (E) Ratio of reads upstream of the PAS and proximal to the mapped PAS in *rrp6Δ* strain's NET-Seq data. (F) Ratio of relative transcription rates in wild-type and *rrp6Δ* strains upstream of the PAS and proximal to the mapped PAS. C- conserved PAS. R- *rrp6Δ*-specific PAS.

4.2.10 *rrp6Δ*-specific sites occur downstream of NDRs

PAS at the 3' ends of genes are associated with nucleosome depletion.

Nucleosomes are removed in the vicinity of these PAS to accommodate the polyadenylation machinery. For this reason, the DNA frequently contains motifs that help recruit nucleosome remodellers to the DNA (Fan et al. 2010).

The majority of both conserved coding and non-coding PAS occur within the centre of NDRs, where nucleosome depletion is most extensive (Figure 4.10 (A-D)).

The same is not the case for *rrp6Δ*-specific PAS. In fact, with the exception of PAS that are terminated at the 3' end of coding genes, *rrp6Δ*-specific PAS occur at the 5' end of NDRs (Figure 4.10 (E-G)). Hence, compared to canonical PAS, the termination of *rrp6Δ*-specific PAS occurs in comparatively nucleosome-rich regions, potentially because slower RNAP transcription speeds facilitate NNS-mediated termination.

Most nucleosome positions and occupancies are not affected by the deletion of *RRP6*. In total, only 649 nucleosomes (1.1% of all nucleosomes) showed a significant change in occupancy between the wild-type and *rrp6Δ* mutant strains. These nucleosomes mapped to 883 genes, 207 of which were non-coding genes (Figure 4.10).

590 of the affected nucleosome positions were less occupied in the *rrp6Δ* mutant. 30.5% of these mapped to gene terminators and 39.8% to promoters. In addition, 112 of the 590 nucleosomes were located within 100 nucleotides of conserved PAS. A further 127 nucleosomes were located within 100 nucleotides of *rrp6Δ*-specific PAS.

Hence, while there is some correlation with increased nucleosome depletion at polyadenylation sites, the majority of *rrp6Δ*-specific PAS is not associated with an increase in nucleosome depletion.

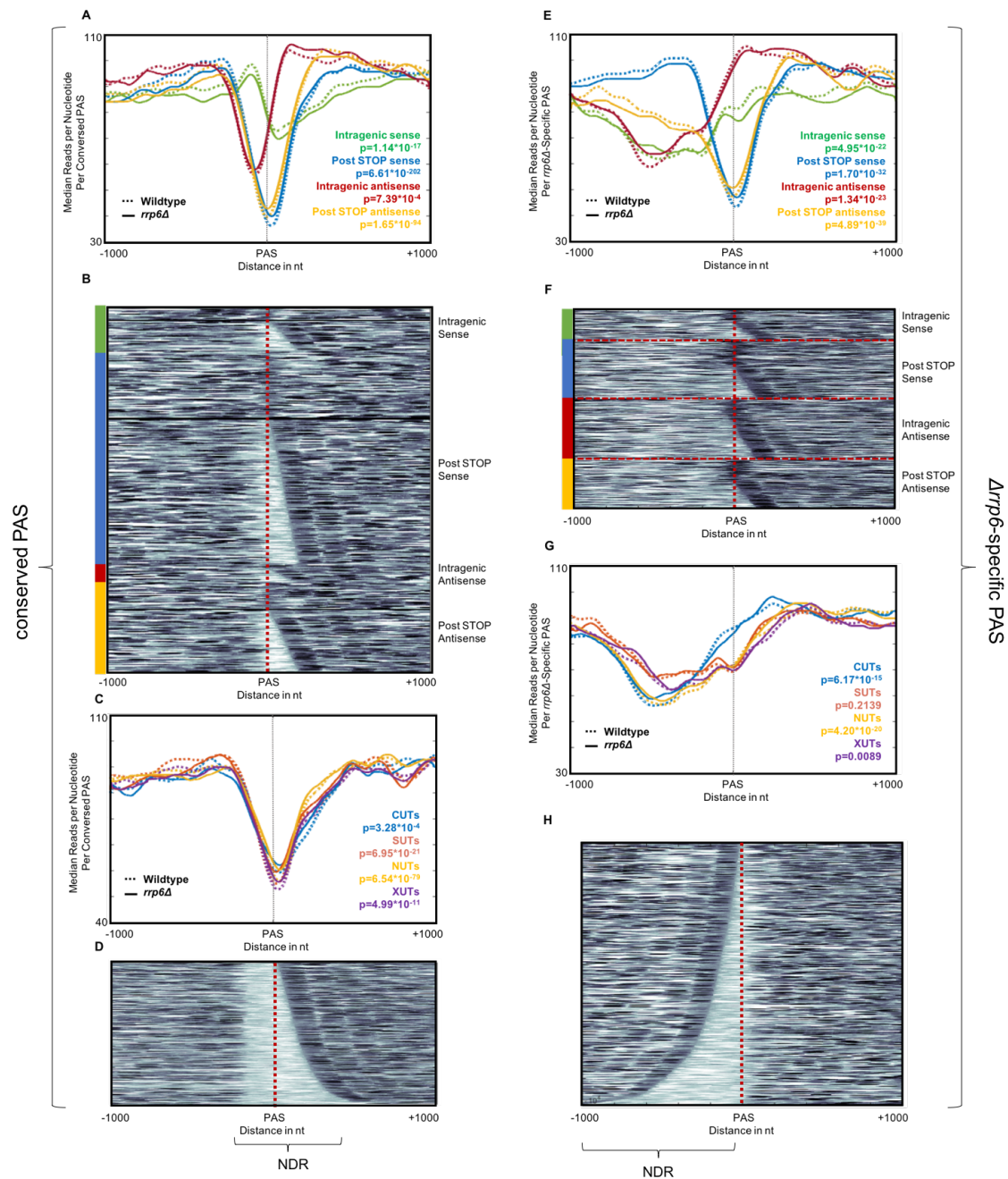


Figure 4.10. Nucleosome depletion at canonical and *rrp6Δ*-specific PAS.

(A,C,E,G) Median number of MNase reads surrounding canonical PAS at coding (A) and (C) non-coding genes and *rrp6Δ*-specific PAS at coding (E) and (G) non-coding genes. p-Values were calculated using Student t-test and represent the difference between MNase-Seq read-density in a 100 nucleotide window surround the PAS in wildtype and *rrp6Δ* strains. For the definition of the individual classes of PAS refer to Figure 2.2 (C). (B,D,F,H) Heatmaps of MNase reads surrounding canonical PAS at coding (B) and (D) non-coding genes and *rrp6Δ*-specific PAS at coding (F) and (H) non-coding genes. Dark colours indicate high nucleosome occupancy, light colours low nucleosome occupancy. PAS- mapped polyadlylation site.

4.2.11 *rrp6Δ*-specific sites occur at defined sequences

Both conserved and *rrp6Δ*-specific PAS were associated with nucleotide frequencies that are usually observed in the vicinity of canonical cleavage sites. In particular, U-rich nucleotide compositions were enriched upstream and A-rich nucleotide compositions downstream of the PAS (Figure 4.11 (A-D)). As far as specific positions were concerned, adenosines were stringently positioned two nucleotides downstream of the cleavage site. In addition, the fraction of uridines peaked at the PAS, while the fraction of cytosine and guanine frequencies peaked one nucleotide downstream of the PAS. This pattern matches previous reports of nucleotide compositions surrounding canonical cleavage sites (Ozsolak et al. 2010; Moqtaderi et al. 2013).

The same overall patterns were observed for both coding and non-coding as well as canonical and *rrp6Δ*-specific PAS (Figure 4.11 (A-D)). However, for *rrp6Δ*-specific PAS this overall pattern was less pronounced. In addition, conserved PAS were frequently positioned at the end of predicted secondary structure elements (ViennaRNA RNAFold structure prediction software; Gruber et al. 2008) , but the same was not evident for *rrp6Δ*-specific PAS (Figure 4.11 (E,F)).

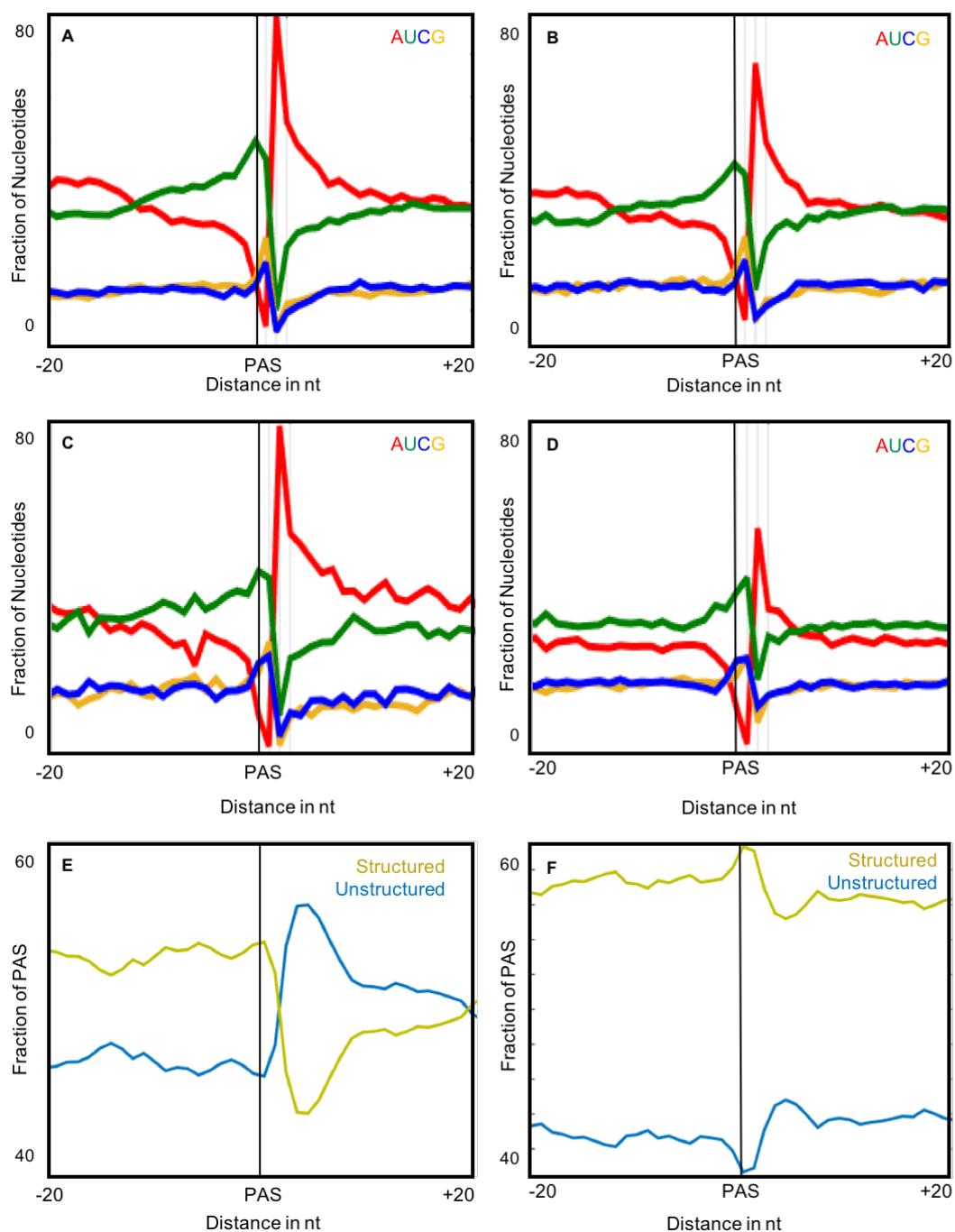


Figure 4.11. Figure: Nucleotide distributions upstream and downstream of mapped PAS.

(A-D) Profiles reflect nucleotide distributions at conserved PAS at (A) coding and (C) non-coding genes and *rrp6Δ*-specific PAS at (B) coding and (D) non-coding genes. (E-F) Amount of predicted secondary structures per PAS at (E) conserved and (F) *rrp6Δ*-specific PAS. Structures were predicted using the ViennaRNA Suite (Gruber et al. 2008).

Conserved		<i>rrp6Δ</i> -specific PAS	
Motif	p-value	Motif	p-value
UAYRUA	1.4×10^{-1349}	UCUUSB	2.2×10^{-267}
AARAARAA	6.2×10^{-446}	GUASW	2.1×10^{-169}
UUUYYUYU	6.8×10^{-387}	YAUAUAY	8.5×10^{-158}
UGUAYA	5×10^{-211}	AGGAAV	1.7×10^{-109}
AAGARR	4.5×10^{-200}	UUUHYUUU	5.8×10^{-92}
YUYYUC	5.1×10^{-172}	GVUGGYA	2.7×10^{-82}
GVUGVUG	4.2×10^{-138}	DGAARAAA	1.4×10^{-78}
GABGAMG	1.4×10^{-85}	GCBGCC	1.1×10^{-83}
CYKCC	8.8×10^{-83}	CUGAADCC	4×10^{-77}
AAGGAMR	4×10^{-74}	CNUGUAC	1.3×10^{-62}
RAWAAUA	4.2×10^{-56}	AAGUGACS	3.3×10^{-68}
SAAGV	3.1×10^{-55}	AGGCAGSW	1.6×10^{-65}
GCGYSU	1.4×10^{-40}	YSACCA	1.4×10^{-71}
AUAYAAUA	3×10^{-40}	GUNACGUA	2.7×10^{-73}
UAUUWYUA	7.6×10^{-40}	UUACCCGG	1×10^{-57}
URURUGUG	4.3×10^{-33}	CASCAG	1.2×10^{-67}
CUUYUKU	5.2×10^{-31}	CUASUCUA	1.4×10^{-54}
UCRUCR	9.9×10^{-29}	UUCWUC	3.1×10^{-99}
CUYUCC	1.4×10^{-25}	AUGABGA	1.5×10^{-60}
ACCANC	5.1×10^{-24}	CKUCCUU	9.7×10^{-76}

Table 4.2. Motifs identified within 100 nucleotides of conserved and *rrp6Δ*-specific PAS.

Motif search was conducted using the MEME-Suite (Timothy et al. 2009). A –Adenine, U- Uridine, C –Cytosine, G- Guanine, N -any base, V –ACG, D – AGU, B – CGU, H- ACU, R – AG, W- AU, S - CG, Y – CU, K – GU

4.2.12 The CFPC binding sites are found at both *rrp6Δ*-specific PAS and conserved PAS

When sequences surrounding PAS are scanned for sequence motifs, remarkably similar motifs are identified in the vicinity of both conserved and *rrp6Δ*-specific PAS (Table 4.3). The identified motifs primarily correspond to the efficiency elements, positioning elements, U-rich motifs as well as the polyadenylation-associated cleavage site itself (reviewed in Mischo & Proudfoot 2013). These motifs are, as discussed in previous Chapters 1 and 3, associated with cleavage-factor dependent termination and polyadenylation.

Looking at cleavage-factor associated binding motifs in more detail, positioning elements could be identified at the vast majority of conserved and *rrp6Δ*-specific PAS at both coding and non-coding genes, just downstream of the cleavage site. An average of 1.65 and 2.23 PEs cleavage sites were detected per conserved coding and non-coding PAS respectively, compared to 1.35 and 0.95 cleavage sites per *rrp6Δ*-specific coding and non-coding PAS. Similarly, efficiency elements were also detected at a frequency of 1.07 and 0.71 per conserved coding and non-coding PAS and 0.92 and 0.70 per *rrp6Δ*-specific coding and non-coding PAS respectively. Moreover, an average of one U-rich element could be identified per PAS. Hence, the vast majority of PAS, irrespective of whether they constitute conserved and *rrp6Δ*-specific PAS, has elements that are able to interact with the CPF machinery, although CF-CPF motifs are slightly less common at *rrp6Δ*-specific PAS and much less stringently positioned with respect to the cleavage site.

However, there are substantial differences in the positioning of these motifs relative to different types of non-coding PAS.

Firstly, for coding PAS all elements (EEs, PEs, cleavage sites and U-rich elements) were enriched at their annotated positions. Namely, EEs mapped to approximately 40-60 base pairs upstream of the PAS, while PEs were found approximately 20 base pairs upstream of the PAS. Similarly, the cleavage motif was generally found just downstream of the PAS and U-rich motifs in the regions flanking the PAS. The main difference between conserved and *rrp6Δ*-specific PAS was that the latter had, on average, fewer motifs.

By contrast, for non-coding RNAs, conserved and *rrp6Δ*-specific PAS were distinctly different. In particular, stable non-coding RNAs, *i.e.* those with conserved PAS, had stringently positioned cleavage sites just downstream of the mapped PAS. In fact, on average, non-coding conserved PAS had significantly more optimal cleavage sites proximal to the mapped site than conserved coding gene PAS. However, conserved non-coding PAS were not associated with stringent positioning of either EEs, PEs or U-rich motifs.

By comparison, not only were *rrp6Δ*-specific non-coding gene PAS associated with the lowest number of CPF binding motifs, but when the motifs were present, they seemed to be randomly distributed around the mapped PAS (Figure 4.12 (D)). Hence, while sequences surrounding all types of PAS are enriched in CF-CPF motifs, conserved and 3' end PAS are likely more able to support CF-CPF binding and efficient cleavage.

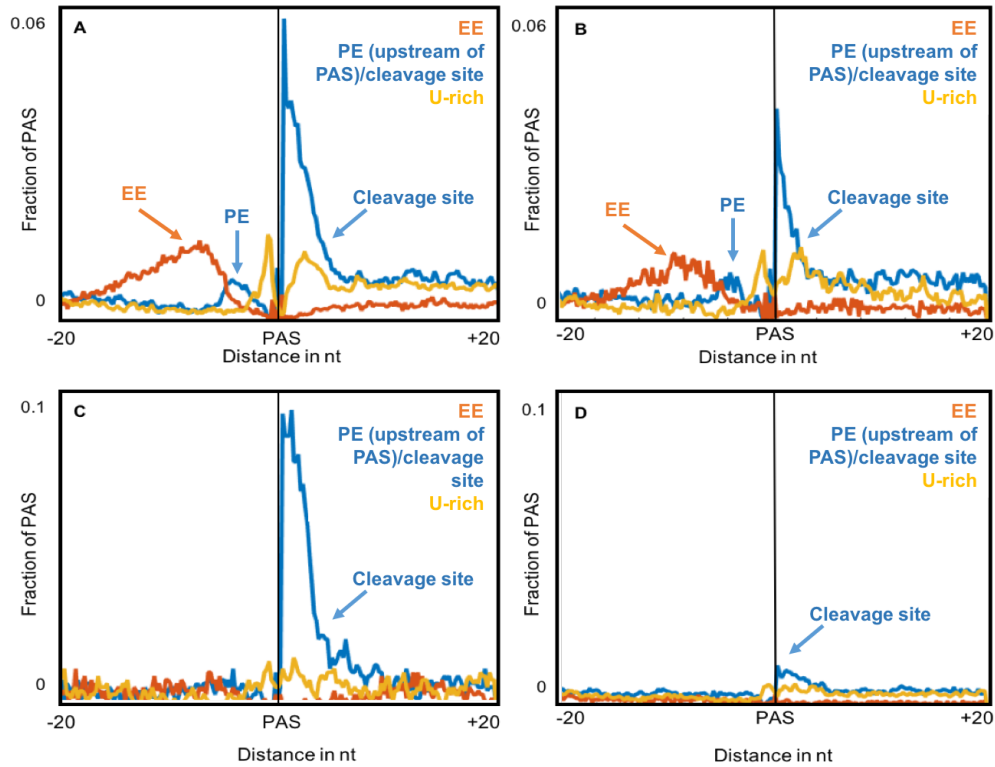


Figure 4.12. Positioning of CPF binding motifs at PAS.

Plot shown represent (A) conserved PAS and (B) *rrp6Δ*-specific PAS at coding genes and represent (C) conserved PAS and (D) *rrp6Δ*-specific PAS at non-coding genes. EE-efficiency element. PE- positioning element. U-rich – U-rich motif.

4.2.13 Nab3p and Nrd1p binding sites are enriched proximal to PAS

The NNS-termination pathway is dependent on the action of Nrd1p, Nab3p and Sen1p, which requires the NNS complex to interact with the RNAP II C-terminal domain (Vasiljeva et al. 2008; Carroll et al. 2007; Heo et al. 2013). Moreover, Nrd1p and Nab3p can bind to the RNA directly via their RRM domains, recognizing UGUA and GUAG motifs and UCUU and CTTG motifs respectively (Creamer et al. 2011; Jamonnak et al. 2011).

Nrd1 binding sites are very short and therefore very common. For this reason, most genes have Nrd1p and Nab3p binding motifs (Figure 4.13). However, Nrd1p and Nab3p binding sites are not distributed randomly across the genome (Webb et al. 2014; Schulz et al. 2013).

In fact, at coding genes Nrd1p binding sites are depleted in the coding region, but enriched in 3' UTRs in the sense and antisense direction. The STOP codon acts as a boundary and can itself form part of an Nrd1p binding site. While not all STOP codons bind to Nrd1p, 510 genes have Nrd1p binding sites overlapping with their STOP codons (Figure 4.13 (A-B)).

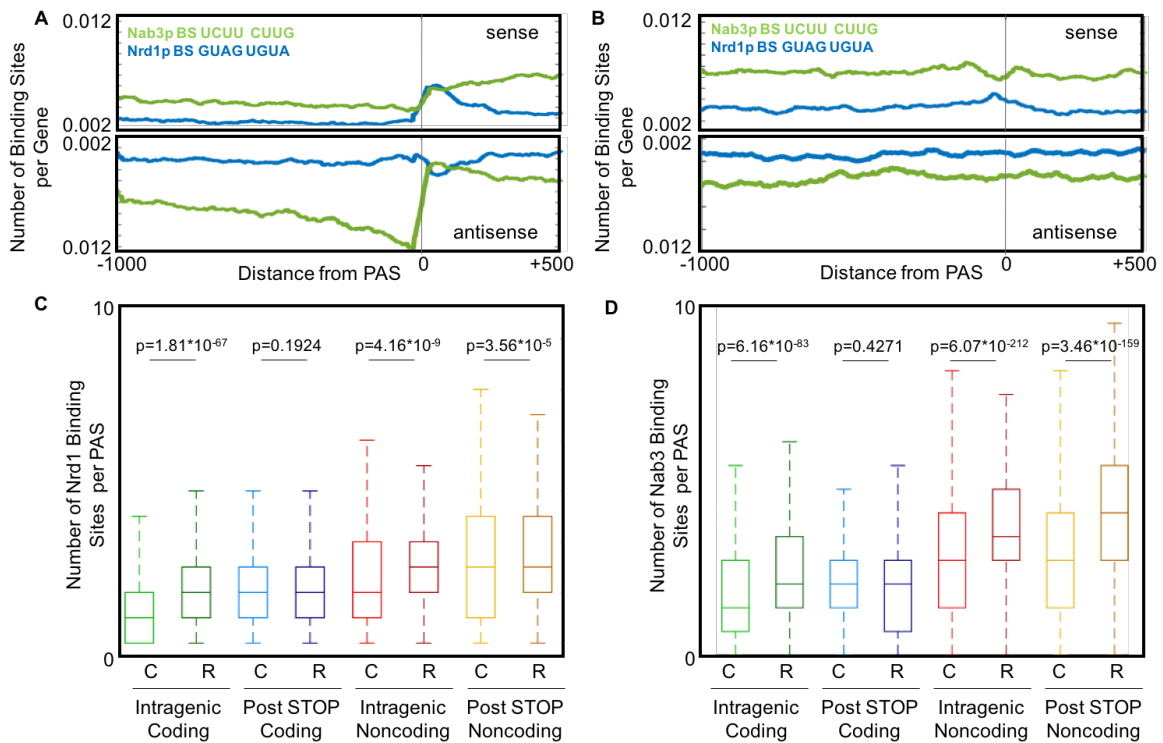


Figure 4.13. Distribution of Nab3p and Nrd1p binding sites.

(A-B) Mean distribution of Nrd1p and Nab3p binding sites per nucleotide upstream and downstream of (A) STOP codons at coding genes and (B) annotated PAS at non-coding genes. (C-D) Number of (C) Nrd1p and (D) Nab3p binding sites in a 100 basepair window upstream of mapped PAS. C- conserved. R- *rrp6Δ*-specific PAS.

By contrast, Nab3p binding sites are primarily enriched in the 3'UTR and over the gene body on the antisense strand, suggesting that Nrd1p and Nab3p have distinct distributions across coding genes. However, at non-coding genes both Nrd1p and Nab3p binding sites are enriched over the gene body as well as the 3' end.

Nrd1p and Nab3p binding sites are enriched in sequences proximal to *rrp6Δ* -specific PAS (Figure 4.13 (C, D)). More concretely, most conserved (63.5%) and *rrp6Δ* -specific (60.7%) PAS have Nrd1p or Nab3p binding sites within 50 nucleotides downstream of the mapped site (Figure 4.13 (C, D)). 10.4% of conserved and 12.7% of *rrp6Δ* -specific PAS are terminated at the exact position of a Nrd1p or Nab3p binding site (Figure 4.13, Figure 4.14 (C, D)).

Density of both Nrd1p and Nab3p binding sites increases prior to polyadenylation sites and, similarly, to CPF-CF binding sites the positioning of Nab3p and Nrd1p binding sites appears to be non-random.

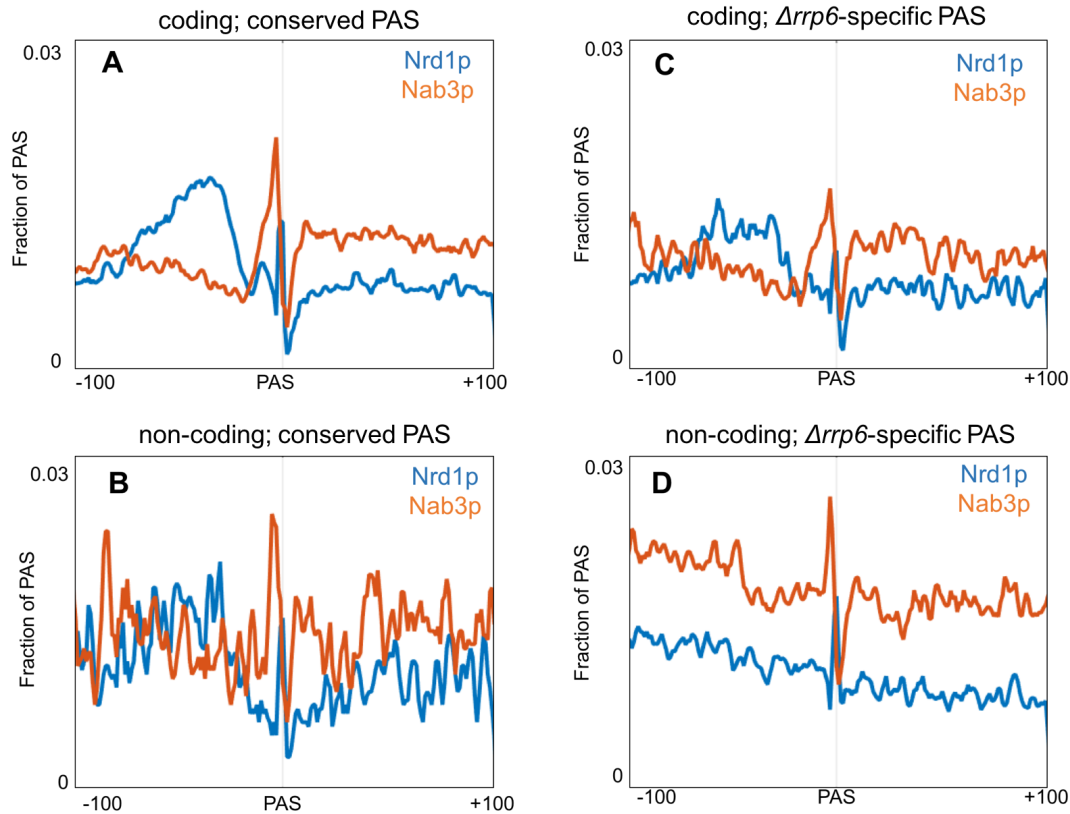


Figure 4.14. Nrd1p and Nab3p binding sites at different PAS.

Nrd1p and Nab3p binding sites at coding (A) conserved and (B) *rrp6Δ*-specific PAS and non-coding (C) conserved and (D) *rrp6Δ*-specific PAS.

In particular, for transcripts that are terminated at the 3' end of coding genes as well as conserved non-coding PAS, Nrd1p binding motifs peak approximately 40-60 nucleotides upstream of the PAS, in a region in which efficiency motifs are frequently enriched. Moreover, both Nab3p and Nrd1p binding motifs are enriched just a couple of nucleotides upstream of the mapped PAS.

By contrast, at *rrp6Δ*-specific PAS, Nab3p and Nrd1p binding motifs only spike adjacent to the PAS, suggesting that either transcription is terminated directly at the Nrd1p and Nab3p binding sites, or that additional processing occurs after termination has occurred.

Moreover, the fact that Nab3p and Nrd1p binding motifs appear to be positioned in close proximity to conserved PAS supports the notion that transcripts can be terminated at the same site by either termination pathway and is consistent with the idea of a fail-safe termination model. The precise location of motifs might play a role in this context. In particular, upstream positioning of Nrd1p motifs appears to correlate with increased transcript stabilities.

4.2.14 Pcf11p and Nrd1p are bound to terminators of both most PAS

Both Pcf11p, a subunit of the CFPC, and Nrd1p interact directly with the RNAP II complex by binding to phosphorylated Ser5 (Nrd1p) and Ser2 (Pcf11p) within the CTD of the Rpb1p subunit of the RNAP II. Owing to their direct interactions the RNAP II, both tagged Nrd1p and Pcf11p can be captured in complex with actively transcribing RNAP II and nascent RNA extracted from the RNAP II binding site. The procedure used to do this is called transcription elongation factor sequencing (TEF-Seq, data produced by Harry Fischl) and follows the same protocol as NET-Seq with the exception that termination factors rather than a subunit of RNAP II were tagged and immunoprecipitated. Assuming that the pulled-down RNA primarily stems from RNAP II –termination factor complexes, the number of sequenced reads captures the relative amount of termination factor binding to RNAP II rather than all RNAP II active transcription.

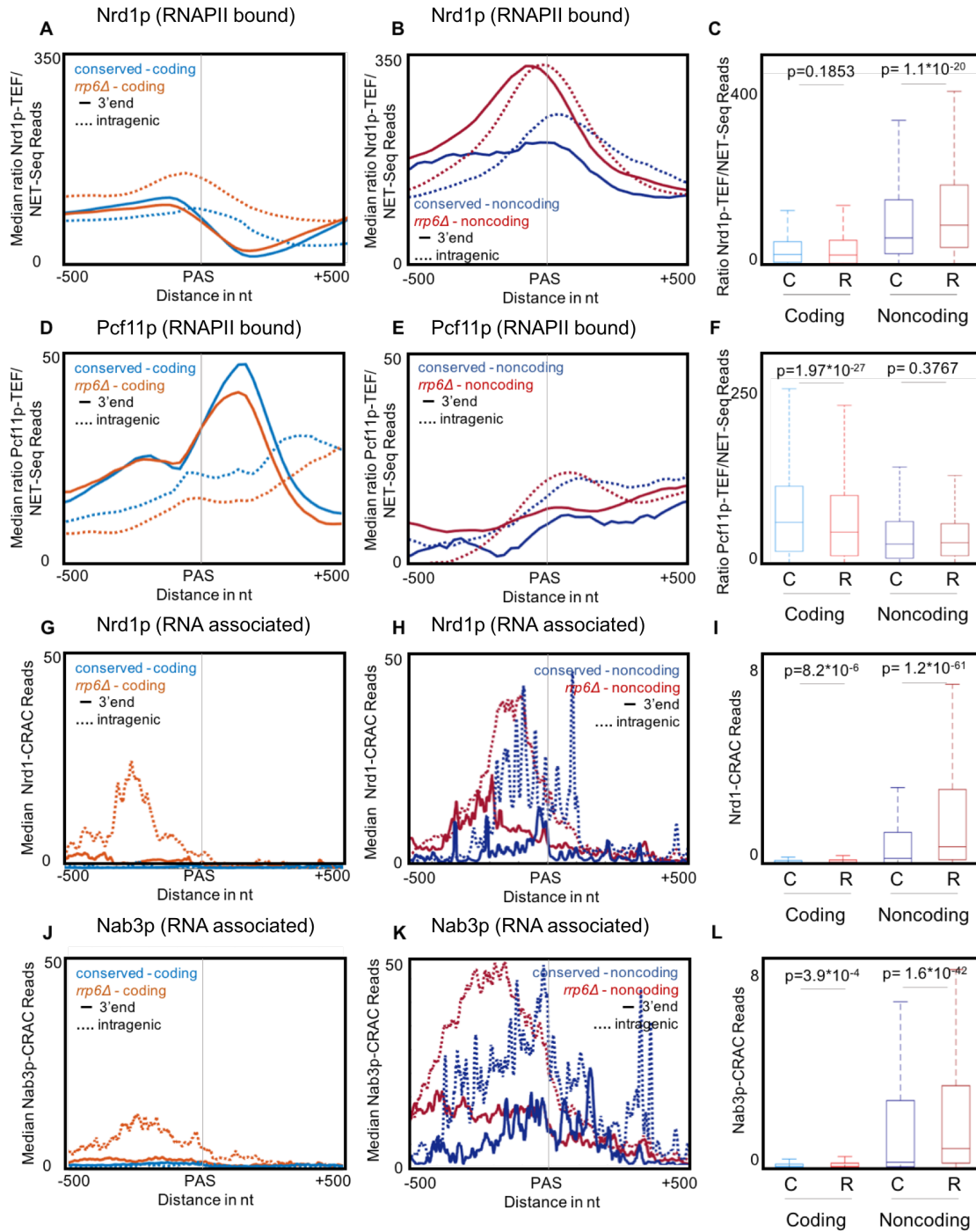


Figure 4.15. Binding Patterns of Pcf11p, Nrd1p and Nab3p.

(A,D,G,J) Binding patterns of (A) RNAPII-bound Nrd1p (normalised to transcription rates), (D) RNAP II-bound Pcf11p (normalised to transcription rates), (G) RNA-bound Nrd1p and (J) RNA-bound Nab3p at coding gene PAS. (B,E,I,K) Binding patterns of (B) RNAPII-bound Nrd1p, (E) RNAP II-bound Pcf11p, (I) RNA-bound Nrd1p and (K) RNA-bound Nab3p at non-coding gene PAS. (C,F,I,L) Enrichment of (C) RNAPII-bound Nrd1p (F) RNAP II-bound Pcf11p, (I) RNA-bound Nrd1p and (L) RNA-bound Nab3p within a 100bp window of the mapped PAS. C- conserved PAS. R- *rrp6Δ*-specific PAS. p-Values were calculated using Wilcoxon ranksum test.

Overall, RNAP II primarily binds Pcf11p at coding genes (Figure 4.15 (E)) and Nrd1p at non-coding genes upstream of both conserved and *rrp6Δ*-specific PAS (Figure 4.15 (A)), although some binding of both Nrd1p and Pcf11p occurs at most mapped PAS (Figure 4.15 (A-F)). However, conserved PAS at coding genes bind significantly more Pcf11p downstream of the mapped site (Figure 4.15 (F)), while *rrp6Δ*-specific PAS at non-coding genes bind significantly more Nrd1p (Figure 4.15 (C)). Moreover, when looking at the ratio between Nrd1p binding compared to Pcf11p binding to the RNAP II, more relative binding of Nrd1p could be detected upstream of *rrp6Δ*-specific PAS (Figure 4.15 (C, F)).

Conversely, when looking at Nrd1p- (Figure 4.15 (G-I)) and Nab3p-CRAC data sets (Figure 4.15 (J-L)), which capture the amount of RNA binding to these factors, both Nrd1p and Nab3p are significantly enriched upstream of non-coding genes (Figure 4.15 (I, L)). Although *rrp6Δ* -specific are comparatively more enriched than conserved PAS, this suggests that all non-coding PAS are primarily targeted by NNS.

4.2.15 CPF subunits are depleted from *rrp6Δ* -specific PAS non-coding PAS

With the exception of Hrp1p of Nab2p (Figure 4.16 (A, F)) , which are associated with nuclear surveillance and degradation of non-coding transcripts (Tuck & Tollervey 2013), conserved PAS were consistently bound by more CF and CPF components (including Rna15p, Mpe1, Ysh1 and Pab1p; Figure 4.16 (B, C, D, E)) than *rrp6Δ*-specific PAS at both coding and non-coding genes.

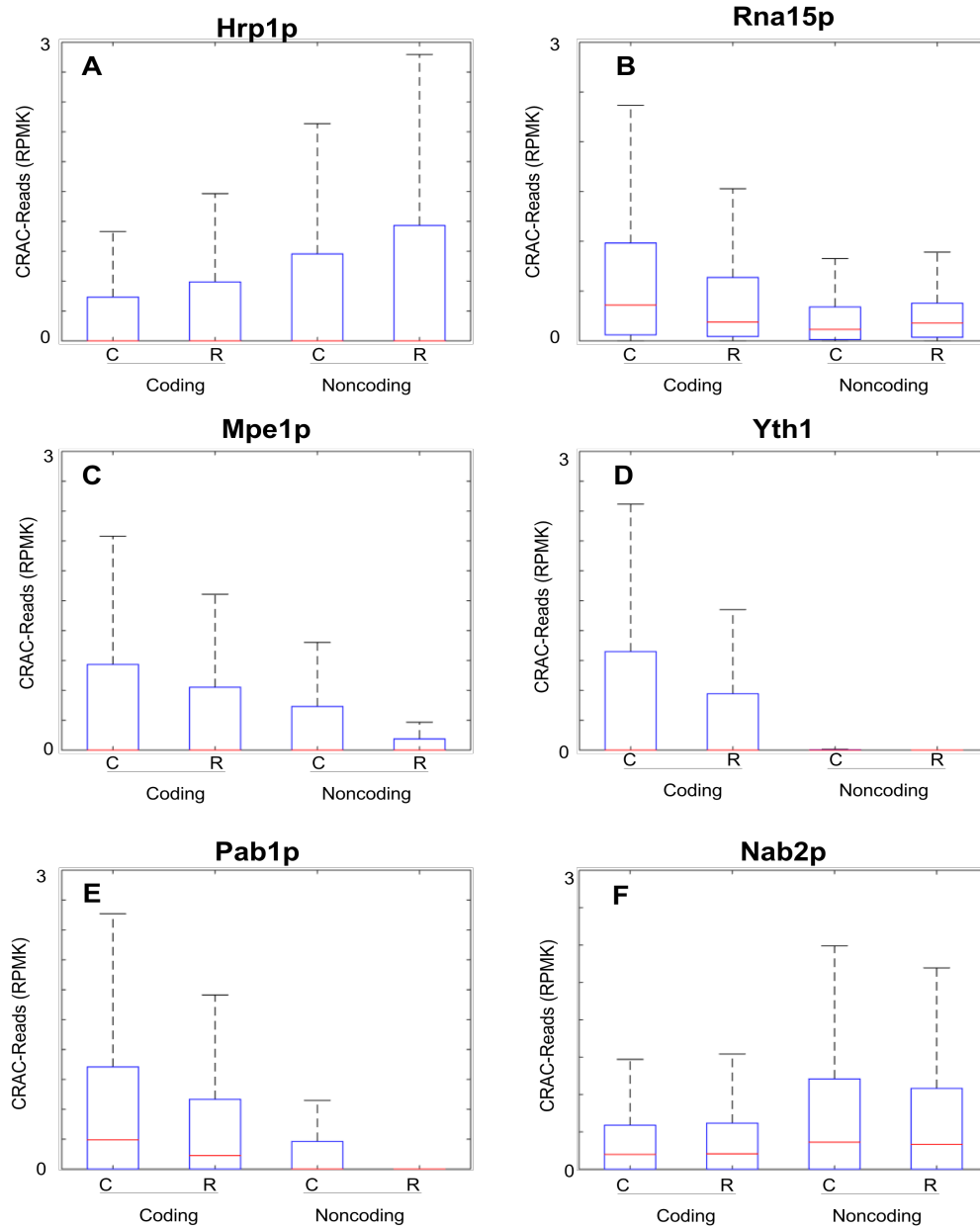


Figure 4.16. Binding of different CF-CPF subunits to sequences flanking the mapped PAS.

Amount of (A) Hrp1p, (B) Rna15p, (C) Mpe1p, (D) Yth1p, (E) Pab1p and (F) Nab2p to conserved (C) and *rrp6Δ*-specific (R) PAS as measured by CRAC. The data was taken from Baejen et al. 2014.

Consistent with the data presented earlier in this Chapter, this confirms that different PAS are to different degrees targeted by the CF-CPF and stabilised though the binding of Pab1p.

4.2.16 *rrp6Δ* is more susceptible to environmental stresses

rrp6Δ has previously been described to have longer doubling times (O'Duibhir et al. 2014). The same effect of the *RRP6* deletion was not observed under optimal growth conditions in this study. More specifically, neither the percentage of cells in G2 (Figure 4.17 (A)) nor the growth rate on agar plates was significantly different in *rrp6Δ* mutants compared to wildtypes. However, when *rrp6Δ* was put under stress, the mutation did have a negative effect on growth rates under all examined conditions (Figure 4.17 (B)). Hence, the deletion of *RRP6* affects the yeast primarily under conditions when resources are scarce or rapid adaptation to a new environment is required.

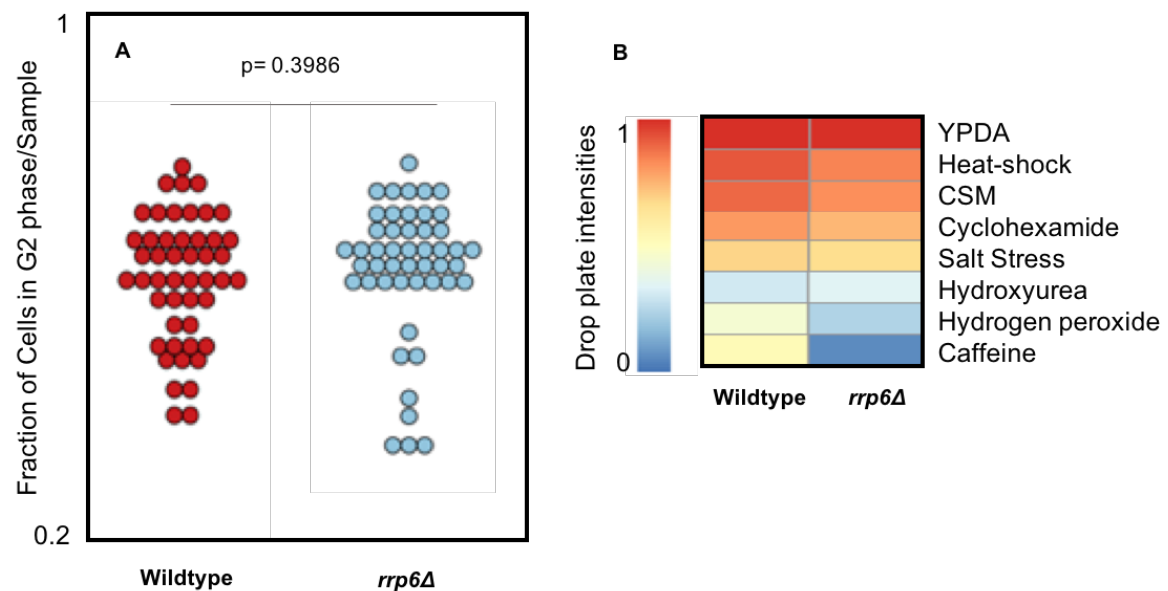


Figure 4.17. *rrp6Δ* growth rates.

(A) Fraction of cells in the G2 phase of the cell cycle. Dots represent individual samples grown to exponential growth phase and analysed by measuring the DNA content using Fluorescence Activated Cell Sorting (FACS). 10000 cells were sampled for each replicate. (B) Relative growth rates of wild-type and *rrp6Δ* strains grown on drop plates under optimal growth conditions and after exposure to different stresses (n=3). The intensities shown were normalised to the growth rates measured in YPDA.

4.3 Discussion

The aim of this Chapter was to further characterise the 3' ends of transcripts that are normally degraded by Rrp6p. The data was gathered under the assumption that 3' ends, which were only detected in the *rrp6Δ* strain, correspond to NNS-terminated transcripts. The reasoning behind this was that Rrp6p specifically interacts with the NNS machinery to remove unstable RNAs (Grzechnik & Kufel 2008; Carneiro et al. 2007; Heo et al. 2013). Hence, comparing *rrp6Δ*-specific 3' ends to those that occurred in the wildtype should be approximately equivalent to comparing NNS-terminated to CF-CPF-terminated transcripts.

Consistent with this hypothesis, sequences upstream of *rrp6Δ*-specific polyadenylation sites had an increased number of Nrd1p and Nab3p binding sites and relatively low number of CF-CPF binding sites compared to conserved PAS. Moreover, the presence of binding sites correlated with high amounts Nab3p and Nrd1p binding in the vicinity of the *rrp6Δ*-specific PAS. Conversely, flanking sequences of PAS mapping to the 3' ends of protein coding genes were bound by high amounts of CF-CPF proteins.

This confirms that *rrp6Δ*-specific PAS are targeted by NNS and conserved PAS by CF-CPF and shows that polyadenylation site mapping in *rrp6Δ* strains and wildtype yeast can be used to distinguish NNS and CF-CPF targets.

However, a recent study reported that transcriptional termination is impaired in *rrp6Δ* (Lemay et al. 2014; Fox et al. 2015). Therefore, it is possible that the NNS termination sites observed in *rrp6Δ* strain do not correspond to wildtype NNS

termination sites. For this reason, it was important to test whether the mapped polyadenylation sites are the product of transcriptional read-through. The hypothesis was tested using NET-Seq data, but, as far as could be determined, transcriptional read-through was not a wide-spread phenomenon in the *rrp6Δ* strain. This suggests that the observed *rrp6Δ*-specific sites correspond to “true” NNS termination targets.

Furthermore, the data presented in this Chapter suggests that NNS-targeted transcripts are not always terminated at random positions.

In particular, unlike CF-CPF terminated transcripts, NNS-terminated PAS are generally located at the end of nucleosome depleted regions. Hence, NNS-mediated termination is not *per se* associated with maximal nucleosome depletion. Instead, the data suggest that NNS termination is most favourable when the polymerase is about the exit an NDR. This means that NNS termination occurs after the polymerase starts to encounter barriers in the form of nucleosomes that are thought to result in a reduction in polymerase speed (Churchman & Weissman, 2011). Therefore, the reduction in RNAP II speed might play a role in NNS termination by facilitating the recruitment of NNS subunits to the transcribing RNAP II and is consistent with an earlier model, which suggested that termination by NNS primarily affects slow polymerases (Hazelbaker et al. 2013).

Moreover, the 3' ends of NNS-terminated transcripts, just as with CF-CPF processed transcripts, tend to occur at reproducible positions across different samples. That said, the sequence requirements for NNS termination are less stringent than for CF-CPF terminated transcripts.

Efficient CF-CPF termination requires CF-CPF binding motifs as well as secondary hairpin structures to be present at defined positions relative to the cleavage site (Mischo & Proudfoot 2013; Ozsolak et al. 2010; Moqtaderi et al. 2013; Dichtl & Keller 2001). By contrast, NNS targets are primarily associated with Nab3p and Nrd1p binding motifs that are positioned in very close proximity to the polyadenylation site, although CF-CPF sites could also be detected in the vicinity of the PAS of NNS-targets.

Why Nab3 and Nrd1p binding sites would be enriched this close to PAS is somewhat unclear at this point. In fact, as nascent RNAs of approximately 20 nucleotides can be accommodated in the active site of the RNAP II, one would expect NNS motifs to be enriched slightly further upstream of the polyadenylation site to allow NNS binding to RNA after it has emerged from the active site of the RNAP II. As this pattern was not observed, it is possible that NNS can support transcript processing in the absence of Rrp6p, for instance through the association with nuclear endonucleases, although no mechanism is evident at this point in time. Alternatively, the absence of Rrp6p could lead to the stabilisation of transcripts and result in termination and RNAPII independent RNA cleavage of nuclear retained transcripts after NNS-dependent termination has occurred. However, the exact mechanism will have to be investigated.

Although not fully conclusive, some of the data presented in this Chapter is consistent with the potential recruitment of CF-CPF to some NNS targets. In particular, *rrp6Δ*-specific PAS at coding genes appear to be targeted by both the NNS and CF-CPF termination complexes. In addition, conserved non-coding PAS appear to be bound by high amounts of NNS-subunits, but to also

be processed and polyadenylated by the CPF complex. It is noteworthy that Tuck and Tollervey (2013) reported a similar observation.

Two potential mechanisms for this can be envisioned.

In the first one, Nrd1p and Nab3p would recruit the CPF to the RNA after termination, resulting in the processing of the transcript. Consistent with this, part of the CPF, called the APT complex (Nedea et al. 2003), binds to Nab3 and mediates processing of the RNAs. However, the APT complex lacks Ysh1p as well Pap1p, which catalyse RNA cleavage and polyadenylation respectively. Hence, it is unclear which enzymes would catalyse the processing of RNAP II transcribed RNAs.

In the second mechanism, NNS and the CF-CPF machinery would compete for the transcript. Since both NNS and CF-CPF motifs occur in close proximity to most mapped PAS, these transcripts might, in principle, be terminated via either the NNS or the CF-CPF pathway. Such a “fail-safe” mechanism has previously been shown to contribute to transcriptional regulation (Webb et al. 2014, Rondón et al. 2009, Lemay & Bachand 2015).

In agreement with this, subunits of both the CF-CPF and the NNS bind to sequences flanking *rrp6Δ*-specific and conserved PAS. This suggests that the choice of termination pathway is probabilistic and influenced by the relative amounts of CF-CPF and the NNS binding to the terminator. Since the relative amount of CF-CPF and NNS subunit binding is variable and NNS-targeting correlates with transcript instability, this explains why some non-coding RNAs are more stable than others. More concretely, while the majority of SUTs are likely to be degraded during the early stages of transcription, some SUTs

escape NNS termination and subsequent degradation and are targeted by CPFC at a later stage.

To conclude, polyadenylation site mapping can be used to characterise the 3' ends of NNS-targets, which are, similarly to CF-CPF targets associated with particular sequence features and might support post-transcriptional processing in *rrp6Δ* mutants. Moreover, most transcripts can probably be terminated via either the CPFC or the NNS termination pathway. Which pathway is chosen will depend on a variety of factors including gene length, RNAP II phosphorylation and the occurrence of binding motifs, which determine whether Paf1p/CPF or Nrd1p are recruited to the transcribing polymerase to trigger termination.

5 Xrn1p is a major indirect repressor of untimely transcription

5.1 Introduction

XRN genes encode a family of 5'→3' exoribonucleases, which have been conserved from yeast to human (Jones et al. 2012). With the exception of plants, which carry four *XRN* genes, all other studied eukaryotic model organisms have two orthologous *XRNs* coding for the cytoplasmic *XRN1* and the nuclear *RAT1* or their respective orthologues (Nagarajan et al. 2013). The N-terminus, which is highly conserved across all *XRN* genes, has two domains that associate to form the active site (Jinek et al. 2011; Xiang et al. 2009; Chang et al. 2011). In addition, Xrn1p also has an extended, less well-conserved and unstructured C-terminus. This domain is absent in Rat1p, but required for Xrn1p catalytic activity *in vivo* (Chang et al. 2011).

As discussed in the general introduction, Xrn1p acts as the main cytoplasmic RNA decay enzyme in eukaryotes (van Dijk et al. 2011; Haimovich et al. 2013; Anderson & Parker 1998; Collier & Parker 2004). Hence, the absence of Xrn1p has severe consequences for both unicellular and multicellular organisms. In yeast, the deletion of *XRN1* results in the stabilisation of non-coding RNAs (XUTs) as well as mRNAs (van Dijk et al. 2011; Malabat et al. 2015).

Phenotypically, *xrn1Δ* strains are slow-growing and susceptible to stress, have increased cell volumes and abnormal bud morphology and are characterised by defective filamentous growth, reduced sporulation, microtubule destabilisation, fusion defects in the nuclear envelope and decreased fitness (Larimer &

Stevens 1990; Jorgensen et al. 2002; Kim 2002; Kim et al. 1990; Sinturel et al. 2012; Tishkoff et al. 1995; Tishkoff et al. 1991; Interthal et al. 1995).

In invertebrates, *XRN1* has important roles in development and both *Caenorhabditis elegans* and *Drosophila melanogaster* embryos are inviable in the absence of their respective Xrn1p orthologue. More specifically, in *C. elegans* *xrn-1* is required for ventral closure (Newbury & Wollard 2004), while in *D. melanogaster* the *XRN1* orthologue *pacman* is required for male and female fertility, imaginal disc formation as well as dorsal closure and epithelial sheet sealing. As PACMAN (dsXrn1) regulates the expression of transcripts coding for apoptotic proteins (Zabolotskaya et al. 2008; Lin et al. 2008; Jones et al. 2013; Jones et al. 2016), the *D. melanogaster* lethal phenotype is a consequence of the dysregulation of these RNAs' concentrations.

Similarly, in plants, *XRN4* deletion results in serrated leaves, larger rosettes, and late flowering due to an insensitivity to ethylene, a plant hormone that regulates plant development and stress responses (Olmedo et al. 2006; Potuschak et al. 2006; Gregory et al. 2008; Merret et al. 2013).

In higher eukaryotes no studies investigating *XRN1*'s role in development have been conducted to date. However, *XRN1* appears to have a physiological role in translational repression and signal transmission at synapses of neurons (Luchelli et al. 2015). In addition, *XRN1* has been linked to disease development in humans. Although *XRN1* has also been shown to act as a tumour suppressor gene of osteosarcomas and prostate cancer (Ding et al. 2015; Zhang et al. 2002), *XRN1* activity has primarily been studied during several types of viral infections. *XRN1* has been shown to restrict the replication

of the hepatitis C virus (Thibault et al. 2015; Li et al. 2015; Ding et al. 2015), suggesting XRN1 plays a role in the suppression of viral infections. In agreement with this, flaviviruses, yellow fever and pathogenic dengue viruses have developed specific mechanisms that allow their RNAs to escape XRN1-dependent degradation (Chapman et al. 2014; Silva et al. 2010; Moon et al. 2012). By contrast, other viruses are able to hijack the host degradation machinery for their own purposes. For instance, both vaccinia and influenza A viruses stimulate XRN1-dependent degradation of host RNAs to facilitate their own multiplication (Burgess & Mohr 2015; Khapersky et al. 2016). Hence, depending on the virus, XRN1 can either suppress or exacerbate viral infections.

Up until recently, the *xrn1Δ* phenotype was thought to be a direct by-product of altered cellular RNA concentrations. However, several recent studies have shown that Xrn1p is more directly involved in the regulation of gene expression in *S. cerevisiae*. In fact, four different main models for Xrn1p's role in transcription have emerged and are outlined below.

The first model, referred to as the transcriptional activator model, proposes that Xrn1p increases transcription rates through direct interactions with the transcription apparatus (Haimovich et al. 2013; Medina et al. 2014). According to the model, Xrn1p enhances transcription rates after nuclear import by binding to the PIC upstream of the TSS. Moreover, subsequent interactions with the RNAP II complex prevent RNAP II backtracking during transcription, thereby increasing elongation rates (Figure 5.1. (A)).

Even though RNA degradation in the cytoplasm is a prerequisite for Xrn1p's nuclear import, Xrn1p's interactions with the RNAP II and the PIC itself are

thought to occur independently of its catalytic domain. Instead, Xrn1p probably binds to the RNAP II and/or PIC via its unstructured C-terminus. Although not directly involved in RNA hydrolysis itself, the C-terminus, whose overexpression is toxic to yeast, has a role in cell cycle regulation, and contains several putative DNA binding and protein-interaction motifs (Pathak et al. 2005; Vad et al. 1998; Jones et al. 2012; Chang et al. 2011). According to the model, Xrn1p acts as a sensor of RNA concentrations, which can stimulate RNA transcription after the degradation of its targets. In fact, the model proposes that Xrn1p somehow uses the RNAs in its active site to identify genes in the nucleus. Xrn1p then acts as a transcription factor that binds AAAAAA motifs at the promoter of its targets, most of which are components of the cytoplasmic translation machinery (Medina et al. 2014; Haimovich et al. 2013).

While no further studies mechanistically supporting the transcriptional activator model have been published to date, Xrn1p-dependent gene activation after glucose-depletion has been reported (Braun et al. 2014). Xrn1p is phosphorylated by Snf1p, a kinase that is required for the relief of glucose repression after glucose depletion (phosphorylation dependence model). Phosphorylation of Xrn1p promotes RNA degradation of Snf1p-targets when glucose concentrations are high (Young et al. 2012). By contrast, when glucose is depleted, Snf1p is imported into the nucleus and phosphorylation of Xrn1p is reversed, resulting in changes in substrate preference (Figure 5.1. (B)).

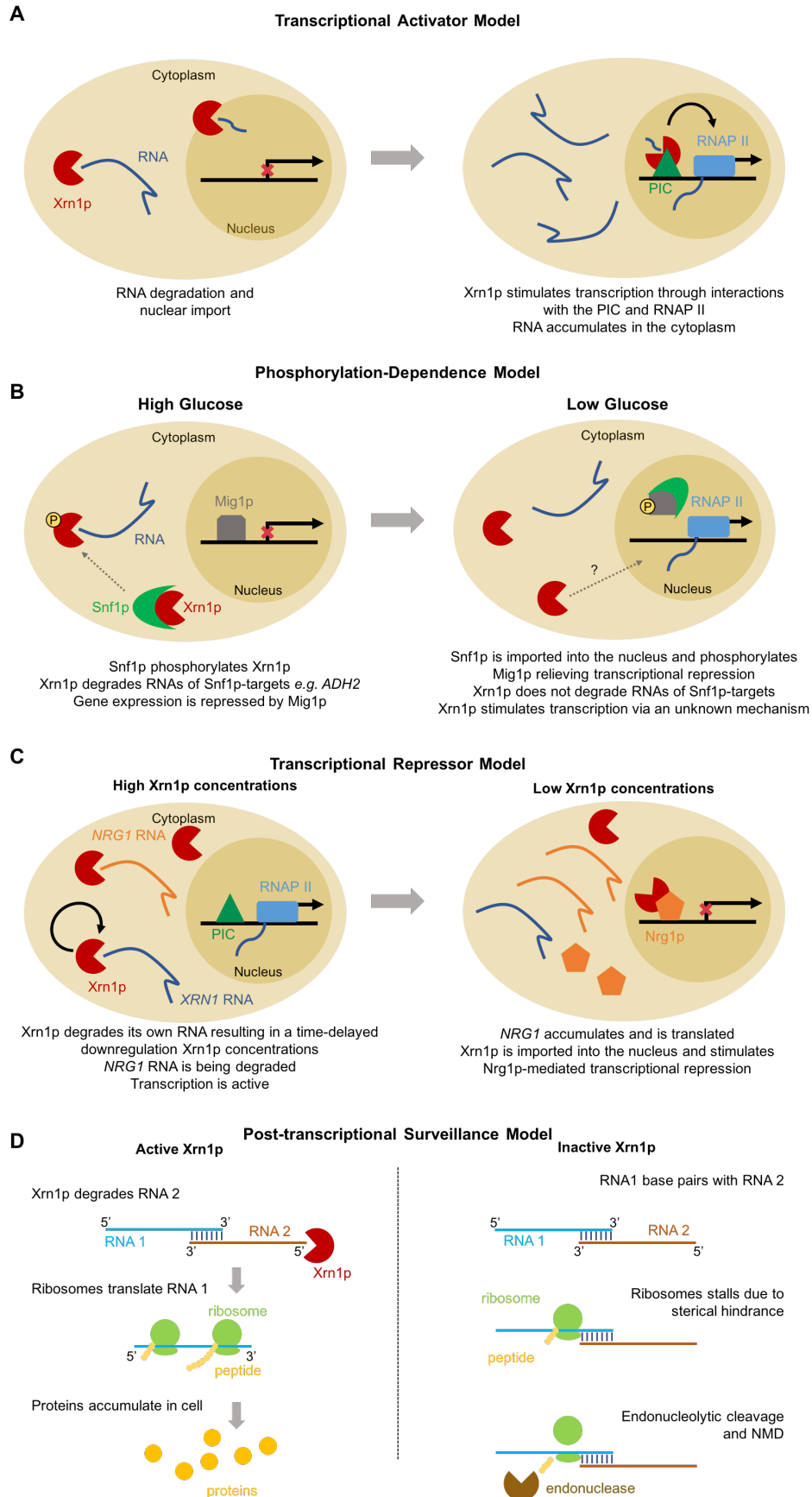


Figure 5.1. Models of Xrn1p-mediated regulation of gene expression.

Unexpectedly, the deletion of *XRN1* results in reduced concentrations of Snf1p-dependent transcripts when glucose is depleted, suggesting that Xrn1p plays a role in the accumulation RNAs of Snf1p-targeted genes during glucose depletion. Although the precise mechanism remains to be investigated, the authors propose that the phosphorylation of Xrn1p changes its affinity for some of its binding partners, in particular decapping factors, resulting in an altered substrate specificity.

The third model, called the transcriptional repressor model, proposes that Xrn1p's primary function is to buffer RNA concentrations, whereby the protein is able to regulate both degradation and transcription rates (Sun et al. 2013). According to the model, Xrn1p regulates its own concentration by degrading *XRN1* RNA in a negative feedback loop. As Xrn1p degrades most RNAs, this mechanism buffers cytoplasmic degradation rates. In addition, low Xrn1p activity leads to the translation of the transcriptional repressor Nrg1p and the downregulation of transcriptional rates (Figure 5.1. (C)). Although low Xrn1p activity in the cytoplasm is a prerequisite for the accumulation of Nrg1p, this mechanism also requires the nuclear import of Xrn1p, as Nrg1p repression appears to be less effective when Xrn1p is depleted from the nucleus (Sun et al. 2013), although changes in Nrg1p concentrations might be explained by the tethering of Xrn1p to the ribosome due to the experimental set-up of the conducted anchor-away experiments, resulting in altered *NRG1* RNA degradation rates.

In the fourth model, dubbed the post-transcriptional surveillance model, Xrn1p controls expression levels by forestalling the accumulation of specific RNAs post-transcriptionally. In this model, Xrn1p specifically degrades one RNA out of

a complementary RNA pair, while leaving the second RNA intact (Figure 5.1 (D)). Since the concentration of one RNA is being kept at very low levels, this prevents base pairing between the complementary RNAs and indirectly stimulates translation of the RNA, which is unaffected by Xrn1p activity (Sinturel et al. 2015). Conversely, in the absence of Xrn1p, overlaps in 3' UTRs of the complementary RNAs cause base pairing between the two transcripts. As a result, translating ribosomes are stalled on the RNA, leading to the recruitment of the NGD machinery, endonucleolytic cleavage and degradation of the transcripts.

Although regulation of complementary transcripts has only been confirmed at the *OCA2/POR1* locus, at which Xrn1p-mediated decay of *OCA2* is required for the induction of *POR1* during glycerol-induced stress (Sinturel et al. 2012), Xrn1p-mediated fine-tuning of RNA concentrations could affect several hundred candidate genes and result in condition-specific expression patterns of these RNAs. In addition to canonical cytoplasmic decay, targeting of RNAs to Xrn1p can also occur via the Nonsense-Mediated-Decay pathway. In fact, 70% of XUTs are degraded in the cytoplasm after the translation of uORFs (Wery et al. 2016). Since most XUTs are inherently complementary to mRNAs, degradation via the NMD might constitute a mechanism for the swift recognition of complementary RNAs.

In summary, although Xrn1p is thought to regulate RNA expression, the exact mechanism remains controversial. The aim of this chapter is to further investigate Xrn1p's function in the regulation of transcription.

5.2 Results

5.2.1 The absence of Xrn1p results in elevated RNA concentrations in most RNAP II-transcribed RNAs

Previous studies have shown that the absence of Xrn1p causes a global stabilisation of RNAP II-transcribed, polyadenylated RNAs (van Dijk et al. 2011; Malabat et al. 2015; Sun et al. 2013).

Consistent with previous studies (correlations are shown in Figure 5.2 (A)), the analysis of RNA-Seq data, collected in *xrn1Δ* mutants and wild-type yeast, shows that total RNA concentrations are elevated 2.23 fold in the *xrn1Δ* yeast.

While annotated Xrn1p- unstable transcripts (XUTs) showed the steepest increase in RNA concentrations, all types of polyadenylated RNAs are stabilised in *xrn1Δ* (Figure 5.2 (B, C)). In spite of this, the relative increase in RNA concentration was transcript-specific and not correlated with the RNA concentrations measured in wild-type yeast ($R = -0.0581$).

Overall, concentrations of 2133 (22.68%) RNAs were increased more than two fold and, based on data from three replicates, statistically different in the *xrn1Δ* compared to the wild-type (Table 5.1). These RNAs will be referred to as *xrn1Δ*-stabilised RNAs in the remainder of the text. While *xrn1Δ*-stabilised RNAs contain some annotated XUTs, they majority of these RNAs are coding RNAs. By contrast, RNAs that change less than 1.2-fold will serve as the control group (N= 1606).

	Genotype	Total (N=9005)	Coding (N=5668)	CUT (N=368)	SUT (N=460)	XUT (N=1098)	NUT (N=1411)
Fold-change	<i>xrn1Δ</i>	2.23	2.67	1.72	2.11	2.94	1.93
	<i>rrp6Δ</i>	1.12	0.95	3.61	1.78	1.58	4.81
	<i>sua7-1</i>	0.48	0.48	0.43	0.46	0.59	0.43
Number of up- regulated genes	<i>xrn1Δ</i>	2133 (22.68%)	1666 (29.39%)	29 (7.88%)	77 (16.732%)	233 (21.22%)	128 (9.07%)
	<i>rrp6Δ</i>	505 (5.61%)	24 (0.42%)	73 (19.83%)	17 (3.69%)	29 (2.64%)	362 (25.65%)
	<i>sua7-1</i>	0	0	0	0	0	0
Number of down- regulated genes	<i>xrn1Δ</i>	6 (0.07%)	2 (0.03%)	0	1 (0.22%)	0	3 (0.21%)
	<i>rrp6Δ</i>	2 (0.02%)	2 (0.03%)	0	0	0	0
	<i>sua7-1</i>	120 (1.23%)	70 (1.23%)	12 (3.26%)	6 (1.30%)	9 (0.82%)	23 (0.02%)

Table 5.1 Fold-changes in RNA concentrations in *xrn1Δ*, *rrp6Δ* and *sua7-1* mutants compared to wild-type yeast.

Fold-change estimates were based on a minimum of three replicates and calculated using DESeq2. Genes were defined as up-regulated or down-regulated if their RNAs increased or decreased by more than two-fold and had a Benjamini adjusted p-value of less than 0.05. Genes were only incorporated into the analysis if they could be detected in at least one data set. *rrp6Δ* mutants were included in this table as a control for exosome activity and *sua7-1* mutants as a control for slow growth (discussed in section 5.2.4).

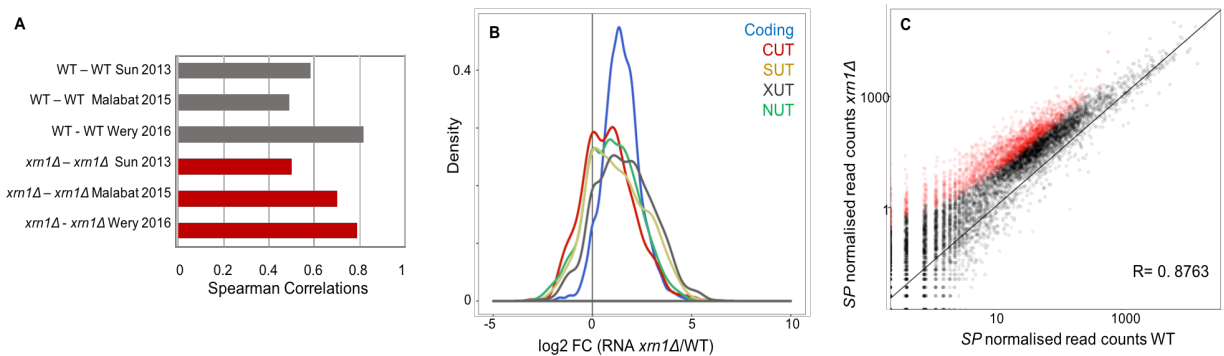


Figure 5.2. Changes in poly(A) + RNA abundance in *xrn1Δ* compared to wild-type under optimal growth conditions.

A) Spearman correlations with published RNA-Seq data sets. (B) DESeq2-estimated fold changes in RNA abundance in *xrn1Δ* compared to wild-type yeast. (C) Scatter plot of *S. pombe* -normalised read counts in wild-types compared to *xrn1Δ* mutants. RNA concentrations that change significantly are shown in red. The R-value represents the Spearman correlations between the two data sets.

5.2.2 Increases in RNA concentrations correlate with an elevated number of RNA isoforms

The observed fold-changes in RNA concentrations correlated with an increase in the number of observed 3' ends ($R = 0.6365$). The distribution of RNA 3' ends in the vicinity of the STOP codon is shown in Figure 5.3. (A). Overall, the total number of detectable 3' ends rose 2.8- fold in *xrn1Δ* mutants. This increase in the total number of 3' ends did not alter which isoform was detected most frequently for a particular gene in the majority of cases (74.55%). In fact, on average, *xrn1Δ*-specific isoforms accumulated a lower fraction of the total number of reads ($p = 1.3 \times 10^{-15}$) than the 3' ends detected in wild-type yeast. This suggests that Xrn1p has a role in removing minor isoforms.

No unifying feature could be identified among the isoforms that were only detected in *xrn1Δ* mutants. Specifically, there was no significant difference in either nucleotide frequencies (not shown) or the observed 3' end secondary RNA structures between *xrn1Δ*-specific isoforms and those, which were present in both the wild-type and *xrn1Δ* yeast (Figure 5.3.(B)).

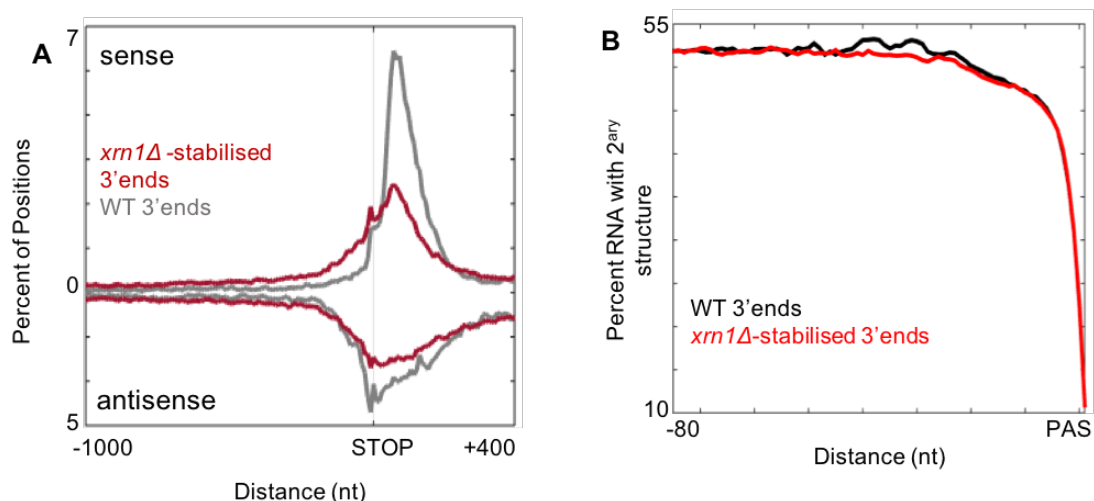


Figure 5.3. Distribution of 3' end isoforms and associated secondary structures.

(A) Distribution of detectable conserved (WT) and *xrn1Δ*-stabilised 3' ends in *xrn1Δ* mutants. (B) Comparison of predicted secondary structure of conserved (WT) and *xrn1Δ*-stabilised 3' ends. Structure prediction was carried out using the Vienna Structure Prediction Tool (Gruber et al. 2008).

5.2.3 Increases in RNA concentrations do not correlate with

Xrn1p binding

To test whether stabilised RNAs are more frequently targeted by Xrn1p, the amount of Xrn1p binding to *xrn1Δ*-stabilised and control RNAs was assessed using a published Xrn1p-CRAC data set (taken from Tuck & Tollervey 2013). As only RNAs crosslinked to the protein of interest are sequenced in CRAC data sets, the number of detected Xrn1p reads is indicative of Xrn1p binding to the RNA (Bohnsack et al. 2012).

However, although 60.1% of RNAs had detectable Xrn1p-CRAC signal at their 3' ends where Xrn1p is loaded onto the RNA (Parker 2012), no statistically significant difference between control and *xrn1Δ*-stabilised RNAs could be detected. Hence, Xrn1p binding to RNAs is not predictive of the observed changes in RNA concentration in the absence of Xrn1p (Figure 5.4 (A)).

In addition, the transcriptional activator model predicts that Xrn1p primarily regulates transcription rates of genes, whose RNAs it degrades in the cytoplasm (Haimovich et al. 2013; Medina et al. 2014). Therefore, genes whose RNAs are stabilised in *xrn1Δ* mutants should have an increased number of Xrn1p proteins binding to their promoters. In order to test this hypothesis, the number of Xrn1p ChIP-exo reads per promoter was compared in the control and *xrn1Δ*-stabilised gene groups (Figure 5.3 (B), ChIP-exo data set taken from Haimovich et al. 2013). Just as for the Xrn1p RNA binding data, there was no discernible difference between the control and *xrn1Δ*-stabilised gene groups (Figure 5.4 (B)). Hence, this finding contradicts the transcriptional activator model, suggesting that there is no direct link between RNA degradation in the cytoplasm and Xrn1p binding to the promoter in the nucleus.

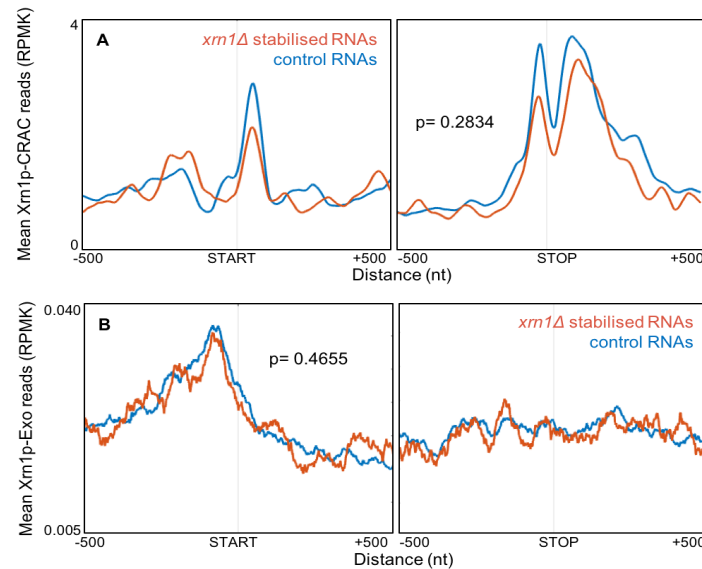


Figure 5.4. Xrn1p binding to RNA and DNA.

(A) Xrn1p binding to *xrn1Δ*-stabilised and control RNAs as measured by Xrn1p-CRAC (Tuck & Tollervey 2013). (B) Xrn1p binding to the promoter of *xrn1Δ*-stabilised and other RNAs as measured by Xrn1p ChIP-exo (Haimovich et al. 2013). p-values were calculated using Student t-test.

5.2.4 The absence of Xrn1p affects RNA concentrations of specific groups of genes

Although 85.1% of RNAs are stabilised in *xrn1Δ* mutants, it remains unclear whether all groups of genes are equally affected, independently of their molecular function. In particular, it is important to test whether (a) RNAs, whose concentrations are significantly increased in *xrn1Δ* yeast, originate from a random set of genes, (b) the observed changes in RNA concentration are equivalent to those observed in mutants of the exosome and (c) whether the observed changes in RNA concentration are primarily indicative of a generic slow-growth signature.

To confirm that the changes observed in *xrn1Δ* mutants are distinct from the ones in exosome mutants, the *xrn1Δ* strain was compared to a *rrp6Δ* strain. This comparison shows that the two mutants affect two distinct sets of genes (Figure 5.5). In fact, only the concentrations of 71 RNAs (0.0079% of RNAs with increased concentrations in *xrn1Δ*) are significantly increased relative to the wild-type in both the *xrn1Δ* and *rrp6Δ* strains. Hence, in contradiction to a previous study, which reported that many CUTs and SUTs are targeted by both Xrn1p and Rrp6p (Marquardt et al. 2011), these results suggest that Xrn1p and Rrp6p target an almost entirely distinct set of genes.

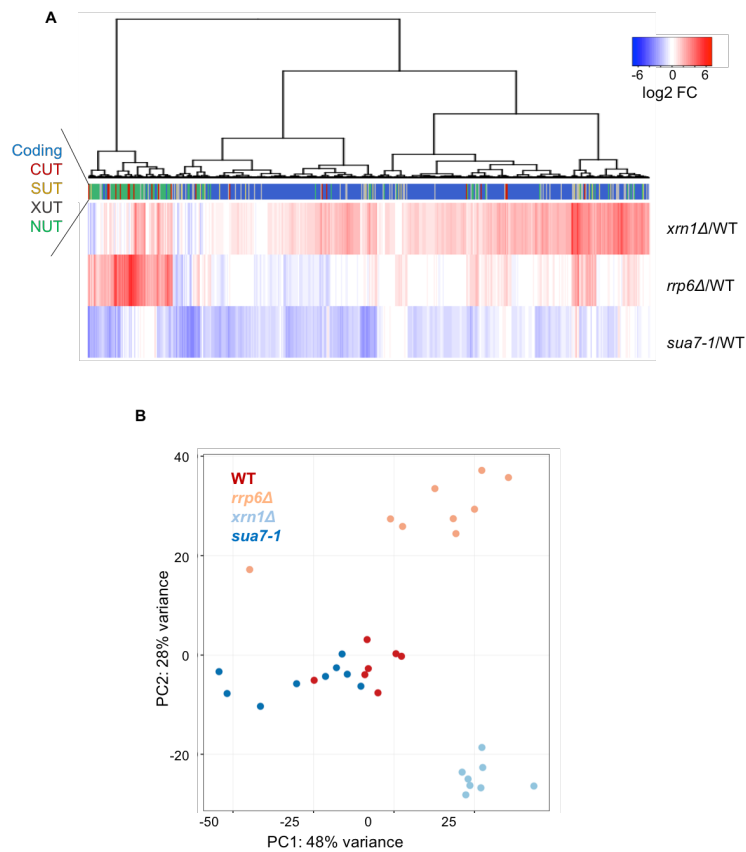


Figure 5.5. Comparison of RNA concentrations in *xrn1Δ* with *rrp6Δ* and *sua7-1* mutants.

(A) Heatmap of log2 fold-changes (FC) in RNA abundance in *xrn1Δ*, *rrp6Δ* and *sua7-1* mutants relative to the wild-type. The bar below the dendrogram indicates the gene type. (H) Principal component analysis of RNA-Seq data. Each point represents one replicate.

xrn1Δ was also compared to a slow-growth mutant. Slow growth mutants are characterised by distinct RNA expression patterns, which are reminiscent of the ones observed during the environmental stress response (O'Duibhir et al. 2014; Gasch et al. 2000; Brauer et al. 2008; Lu et al. 2009). Since the deletion of *XRN1* leads to increased doubling times (Larimer & Stevens 1990), it is important to test whether the changes in RNA concentration are characteristic of a slow growth signature or, alternatively, are indicative of a more substantial dysregulation of RNA metabolism.

To test this, RNA-Seq data from a slow growing *sua7-1* strain was compared to the *xrn1Δ* mutant. The *sua7-1* strain contains a mutation that affects the yeast orthologue of TFIIB (Chew & Lehming 2007; Pinto et al. 1994) and, due to Sua7p's importance in the transcriptional process, results in a 2.08-fold reduction in total RNA concentrations (Table 5.1, Figure 5.5 (A)). In spite of this, *sua7-1* samples cluster with the wild-type in a principal component analysis of RNA-Seq data (Figure 5.5 (B)), suggesting that RNA concentrations in *sua7-1* mutants are more similar to the wild-type than *xrn1Δ* yeast. Furthermore, fold-changes in RNA concentrations in *xrn1Δ* and *sua7-1* mutants are only poorly correlated ($R = 0.1498$). What is more, no significant enrichment of ESR genes could be detected in the *xrn1Δ*-stabilised group. This means that the *xrn1Δ* transcriptome cannot simply be explained by a stress signature or a proportional stabilisation of RNAs, suggesting that the observed changes in the transcriptome are mostly unrelated to the strains slow-growth phenotype.

	Annotation	GO ID	Type	p-Value
Genes Enriched in the List of <i>xrn1Δ</i> -stabilised Transcripts	Response to stimulus	GO:0050896	BP	5.98E-12
	Chromosome	GO:0005694	CC	5.39E-08
	Cell cycle - yeast	KEGG:04111	keg	0.000003 17
	DNA binding	GO:0003677	MF	0.000056 9
	DNA-dependent ATPase activity	GO:0008094	MF	0.000516
	Cell periphery	GO:0071944	CC	0.000942
	Starch and sucrose metabolism	KEGG:00500	keg	0.00153
	Catabolic process	GO:0009056	BP	0.00256
	Ubiquitin ligase complex	GO:0000151	CC	0.00386
	Aryl-alcohol dehydrogenase (NAD ⁺) activity	GO:0018456	MF	0.00525
	Protein modification by small protein conjugation or removal	GO:0070647	BP	0.0125
	Homologous recombination	KEGG:03440	keg	0.0131
	RNA polymerase II transcription factor activity, sequence-specific DNA binding	GO:0000981	MF	0.0175
	Piccolo NuA4 histone acetyltransferase complex	GO:0032777	CC	0.0193
	DNA repair complex	GO:1990391	CC	0.0238
	DNA secondary structure binding	GO:0000217	MF	0.0302
	Spliceosomal complex	GO:0005681	CC	0.0378
Genes Enriched in the List of Control-Group RNA	Ribosome biogenesis	GO:0042254	BP	6.67E-19
	Cytosolic ribosome	GO:0022626	CC	5.96E-18
	Structural constituent of ribosome	GO:0003735	MF	1.53E-14
	RNA binding	GO:0003723	MF	3.89E-14
	Ribosome	KEGG:03010	keg	6.83E-14
	Transposition, RNA-mediated	GO:0032197	BP	0.00005 28
	Factor: GCR1; motif: AWANTGGAAGCYSRAA; match class: 0	TF:M01729_0	tf	0.00005 52
	RNA-directed DNA polymerase activity	GO:0003964	MF	0.00016 6
	Nucleobase-containing small molecule metabolic process	GO:0055086	BP	0.00183
	Ribonuclease activity	GO:0004540	MF	0.0143
	Establishment of protein localization to mitochondrion	GO:0072655	BP	0.0174
	Factor: Rap1p; motif: RCACCCANNCMNC; match class: 1	TF:M07442_1	tf	0.0206
	Aspartate family amino acid metabolic process	GO:0009066	BP	0.0387

Table 5.2. Gene Ontology terms associated with coding factions of *xrn1Δ*-stabilised transcripts.

Genes were selected based on a DESeq2 analysis (Fold-change >2, adjusted p-value <0.05) The gene ontology search was carried out using DAVID (Huang et al. 2008). p-values are Benjamini-Hochman adjusted to account for multiple testing.

The deletion of *XRN1* does not affect all RNAs equally. This is revealed by the fact that not all RNAs show equivalent fold-changes in RNA concentration upon the deletion of *XRN1*. Stabilised coding RNA are enriched in genes that are required for cell cycle regulation, affect the cell's responsiveness to stimuli, the activity of transcription factors and metabolism. At the same time, RNAs, whose concentrations are unaffected by the deletion of *XRN1*, are enriched in ribosomal proteins genes (Table 5.2.). Specifically, RNA concentrations of 49% of all cell cycle regulated genes, 30.3% of transcription factors and 35.6% of chromatin remodellers increase more than two-fold in *xrn1Δ*. By comparison, only 1.5% of all cytoplasmic ribosome subunits and 7.3% of all genes with annotated functions in translation are affected by the deletion of *XRN1* (Figure 5.6). Hence, Xrn1p either targets particular RNA or, alternatively, some genes are more prone to being regulated by feedback loops, which result in an increase in exosome-mediated degradation rates or a reduction in transcription rates, in the absence of Xrn1p.

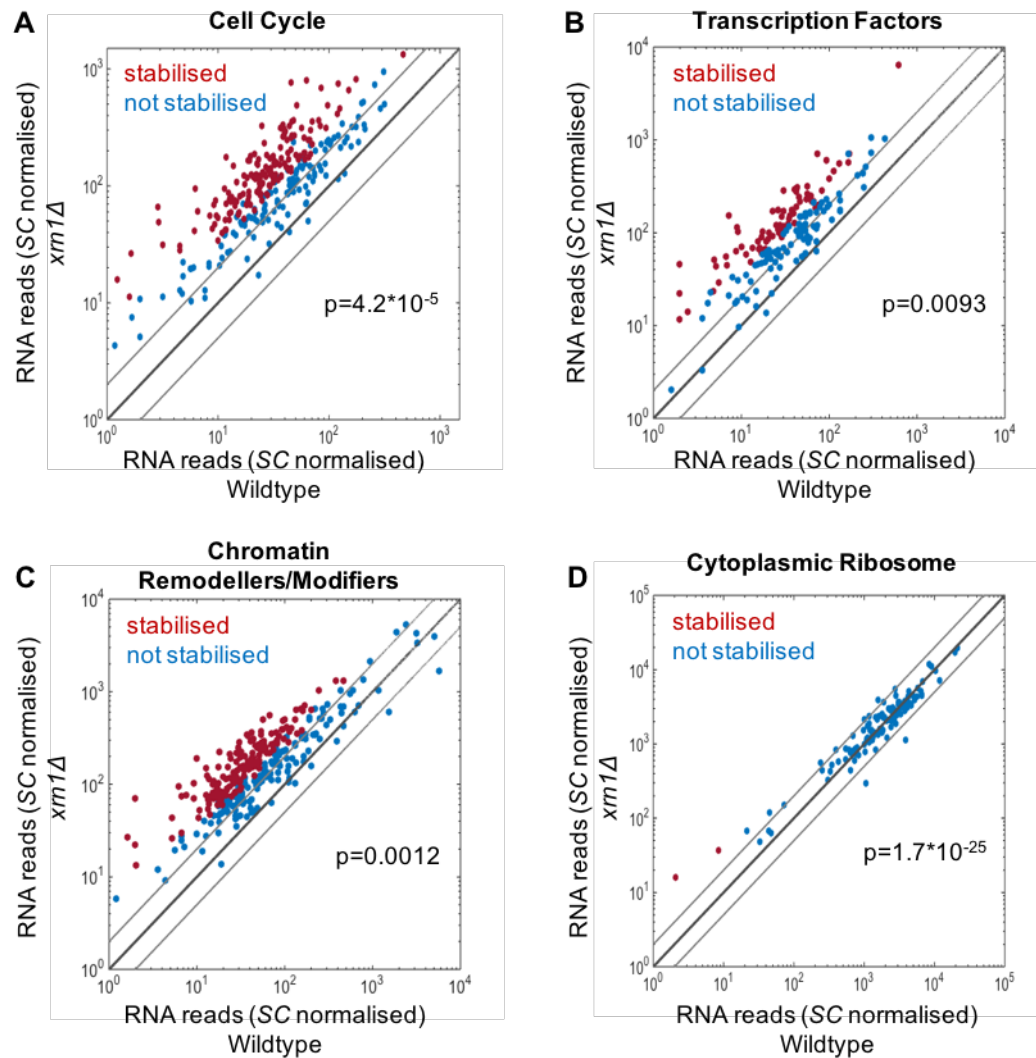


Figure 5.6. Changes in RNA concentration in selective groups of genes.

Fold changes in RNA concentrations in (A) cell cycle genes, (B) transcription factors, (C) chromatin remodellers and (D) components of the cytoplasmic ribosome. RNAs within this group, which are significantly stabilised in *xrm1Δ* mutants are shown in red. RNAs, which are not significantly stabilised, are shown in blue. p-values were calculated using Student t-tests and indicate whether the observed fold-changes in the plotted subset are statistically different from the fold-changes observed in remainder of RNAs.

5.2.5 The deletion of *XRN1* leads to the accumulation of additional transcripts

In addition to RNAs whose concentrations increase in *xrn1Δ* mutants, an additional 222 RNAs originating from coding genes and 1026 RNAs originating from non-coding genes were detected in the *xrn1Δ*, but not in wild-type yeast.

Most RNAs that were only detected in *xrn1Δ* mutants are non-coding RNAs with unidentified functions and have previously been annotated as Xrn1p-unstable transcripts (XUTs; van Dijk et al. 2011). Moreover, 37 polyadenylated tRNA as well as more than 200 additional coding RNAs could be detected in *xrn1Δ* but not in wild-type yeast. As Xrn1p is involved in the removal of hypoacetylated tRNAs, it is likely that the observed tRNAs are defective pre-cursors, which were exported to the cytoplasm by accident (Kramer & Hopper 2013; Chernyakov et al. 2008).

Coding RNAs that were only detected in *xrn1Δ* mutants were enriched in genes involved in carbohydrate metabolism, sporulation, meiosis, DNA binding and transmembrane transport (Table 5.3). The majority of these genes are not normally expressed under optimal growth conditions.

For instance, 13 out of the 23 members of the seripauperin multigene family are expressed in *xrn1Δ* mutants. As this family of genes is linked to anaerobic growth and their expression is repressed by oxygen (Rachidi et al. 2000), it is likely that the expression of seripauperin genes is induced by mistake in aerated *xrn1Δ* mutant cultures (Haber 2012; Luo & Van Vuuren 2008). Similarly, RNAs coding for carbon transporters *HXT3*, *HXT8*, *HXT10*, *HXT11* and *HXT17*,

whose expression is normally repressed in the presence of glucose(Weinhandl et al. 2014), are detectable in *xrn1Δ* yeast. Hence, *xrn1Δ* appears to be unable to sense the presence of glucose and repress gene expression accordingly. Consistent with this, these genes are in large part regulated by Snf1p which, as discussed in the introduction, requires Xrn1p for the silencing of its targets when glucose concentrations are high (Braun & Young 2014).

Moreover, the expression of genes required for sporulation and meiosis is induced in *xrn1Δ* mutants. Under normal circumstances, sporulation is a process that is only activated in response to carbon or nitrogen starvation and involves the controlled transition of a diploid cell to four haploid spores. As the laboratory strains used in this study are haploid, they are unable to undergo meiosis and sporulate. In addition, even if the cells were able to form spores, the glucose and nitrogen present in the growth medium should lead to the transcriptional silencing of meiotic and sporulation genes by Ume6p, Sum1p and Ssn6p/Tup1p (Strich et al. 1994; Xie et al. 1999; Mizuno et al. 1998).

In spite of this, *xrn1Δ* accumulates sporulation and meiosis specific transcripts including *SPO11*, *SPO13*, *SPO19*, *SPR2*, *OSW1*, *SPS4*, *ECM11*, *ZIP2* and *SAE3*. It is likely that the expression of the meiosis specific genes is driven by the presence of Ime1p and Ime2p, which are transcriptionally repressed under optimal growth conditions (Shimizu et al. 1997), but are required for the initiation of meiosis (Mandel et al. 1994). Consistent with this, RNA concentrations of *IME1* and *IME2* are increased 10.7- and 8.6-fold in *xrn1Δ* mutants compared to the wild-type. Although to a lesser extent, the same is also true for the transcription factor Ndt80p, another regulator of meiotic gene expression

(Kassir et al. 2003), which is showed a 2.3-fold increase in RNA concentration in *xrn1Δ*. Hence, it is possible that the accumulation of RNAs coding for meiotic and sporulation factors is driven by the presence of regulatory proteins that induce their transcription.

Annotation	P-value
Starch and sucrose metabolism	3.15×10^{-10}
Sporulation resulting in formation of a cellular spore	7.80×10^{-6}
Glucose import	2.36×10^{-5}
GPI-anchor	4.08×10^{-5}
Glycogen biosynthesis	1.13×10^{-4}
Glucose transmembrane transporter activity	1.22×10^{-4}
Substrate-specific transmembrane transporter activity	1.36×10^{-4}
Transmembrane transport	1.40×10^{-4}

Table 5.3. Gene Ontology terms associated with transcripts that were only detected in *xrn1Δ* yeast.

The gene ontology search was carried out with DAVID (Huang et al. 2008). p-values are Benjamini-Hochman adjusted to account for multiple testing.

5.2.6 RNAs with increased concentrations have increased transcription rates

To test whether transcription rates were affected by the deletion of *XRN1*, a NET-Seq data set was collected in *xrn1Δ* yeast (Figure 5.7., all NET-Seq data sets in this chapter were collected by Fischl, H.).

Although transcription rates in wild-type and *xrn1Δ* yeast are highly correlated ($R=0.9842$), some distinct differences were observed.

Firstly, NET-Seq reads are shifted from the 5' end to the 3' end of the gene in *xrn1Δ* mutants. In particular, fewer reads map to the first 300 nucleotides after the START codon in *xrn1Δ* compared to wild-type yeast. Since the number of reads at the 5' end of the gene is an indicator of transcription initiation

frequency, this could mean that transcription initiation rates are lower in the *xrn1Δ* mutant.

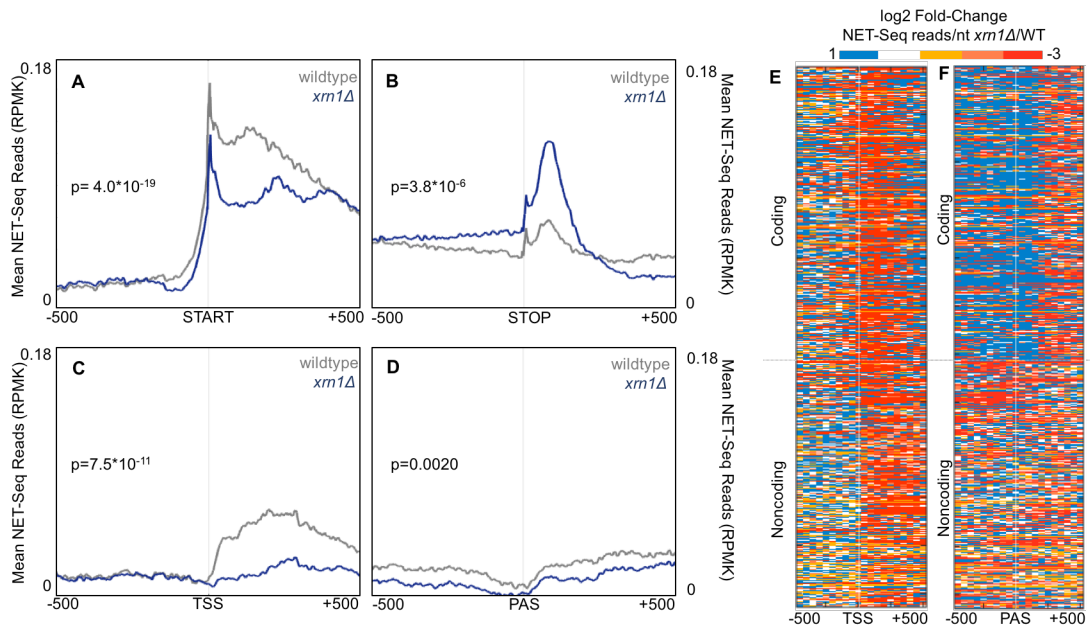


Figure 5.7. NET-Seq profiles in wild-type and *xrn1Δ* yeast.

(A-D) Metagene profile of NET-Seq reads upstream and downstream of the (A) START codon and (B) STOP codon of coding genes and (C) the TSS and (D) TES of non-coding genes in *xrn1Δ* (blue) and wild-types (grey). (E-F) Heatmaps of read distribution at individual genes at their (E) 5' and (C) 3' ends in wild-type and *xrn1Δ* yeast. Red indicates a higher number of reads in wild-types, blue a higher number in *xrn1Δ* mutants.

Secondly, NET-Seq reads peak approximately 200 nucleotides downstream of the START codon in wild-types, a point at which capping factors are exchanged for elongation factors in the RNAP II complex (Mayer et al. 2010). By contrast, NET-Seq reads in the *xrn1Δ* mutant peak 100 nucleotides further downstream, at approximately 300 nucleotides after the START codon. Assuming that the peak in NET-Seq reads is indeed indicative of a remodelling of the RNAP II elongation complex as has been discussed in Chapter 3, this suggests that capping factors are exchanged for elongation factors at a later point during the transcription cycle in *xrn1Δ* mutants (at approximately 300nt after the START

codon) compared to wildtypes (where the exchange occurs approximately 200nt after the START codons).

In addition, at least in coding genes, more reads map to the polyadenylation-associated termination zone at the 3' end of the gene (Figure 5.7 (B)). It is possible that this reflects an increased residence time for RNAP II and a reduced termination efficiency in *xrn1Δ*.

Furthermore, fewer NET-Seq reads map to non-coding genes in *xrn1Δ*. More specifically, 64% of non-coding genes were characterised by decreased NET-Seq read numbers in *xrn1Δ*. Since cytoplasmic RNA degradation rates are affected in *xrn1Δ* mutants, this could imply that transcription rates of most non-coding genes are downregulated to compensate for the reduced RNA degradation rates (Figure 5.8.).

However, this apparent feedback mechanism is not active at all genes. In fact, transcription rates of genes, whose RNAs concentrations increase in the absence of *XRN1*, are elevated in a majority of cases (Figure 5.8). By contrast, transcription rates of genes, whose RNA concentrations remain unaltered in *xrn1Δ*, are significantly downregulated.

Therefore, *xrn1Δ* may be able to adjust the transcription rates of a subgroup of genes in response to its altered RNA degradation rates. This reduction in transcription facilitates the maintenance of wild-type RNA concentrations of these RNAs.

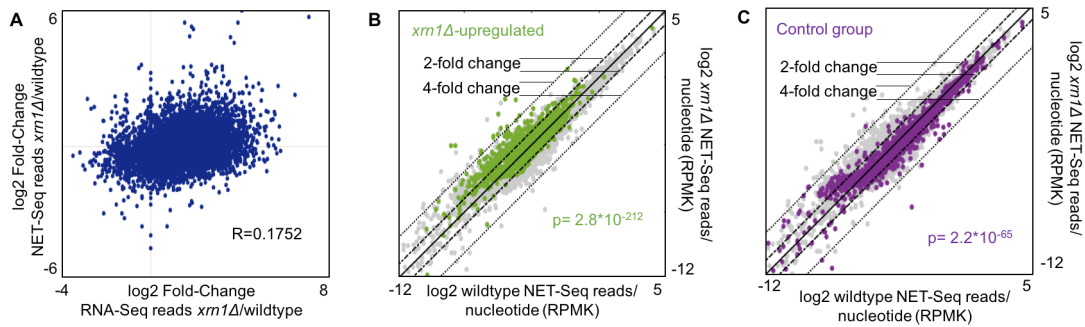


Figure 5.8. Changes in RNA transcription rates.

(A) Correlation of fold-changes in RNA transcription and degradation rates. (B-C) Transcription rates of (B) *xrn1Δ*-stabilised and (C) unaffected control RNAs.

By contrast, for genes, whose RNAs concentrations increase in *xrn1Δ*, the measured upregulation in transcription rates likely contributes the accumulation of their transcripts in the cytoplasm.

As fold-changes in RNA concentration and RNA transcription rates are only weakly correlated, differences in the relative increases in RNA concentrations can only partially be explained by altered transcription rates. Hence, in agreement with the transcription repression model, this suggests that RNA-specific changes in both degradation and transcription rates contribute to the accumulation of RNAs in *xrn1Δ*.

5.2.7 *xrn1Δ* are characterised by decreases in nucleosome occupancy at transcription factor binding sites

Although most nucleosomes are unaffected by the deletion of *XRN1*, changes in nucleosome occupancy were observed at promoters and terminators (Figure 5.9). In particular, 673 promoter-associated nucleosomes are depleted from the DNA in *xrn1Δ* mutants (Figure 5.9 G) and, consistent with this, NDRs of coding

genes span, on average, a larger region in *xrn1Δ* yeast than in wild-types (Figure 5.9 (A,C,E)).

By contrast, the average size of terminator-associated NDRs was considerably smaller in *xrn1Δ* yeast (Figure 5.9 (B,D,F)). Hence, the deletion of *XRN1* affects nucleosome remodelling at promoters and terminators in opposite ways.

Similar to transcription rates, promoters of RNAs whose concentrations are increased in *xrn1Δ* mutants, were significantly more nucleosome depleted in *xrn1Δ* compared to wild-type yeast than promoters of transcripts, whose concentrations were not affected by the deletion of *XRN1* ($p=5.1 \times 10^{-16}$).

Moreover, and consistent with the observed downregulation in their transcription rates, promoters and terminators of non-coding genes are less nucleosome depleted in *xrn1Δ* than in wild-types ($p= 0.0005$).

Therefore, changes in RNA transcription rates correlate with chromatin remodelling, especially at the promoter.

Transcription is, to a large part, driven by the recruitment of specific transcription factors to the promoter of their target genes. DNA-bound transcription factors recruit chromatin remodellers, resulting in the eviction of nucleosomes from the DNA and the establishment of an NDR (Floer et al. 2010; Walter et al. 1995; Elfving et al. 2014a; Yarragudi et al. 2004; Sherriff et al. 2007).

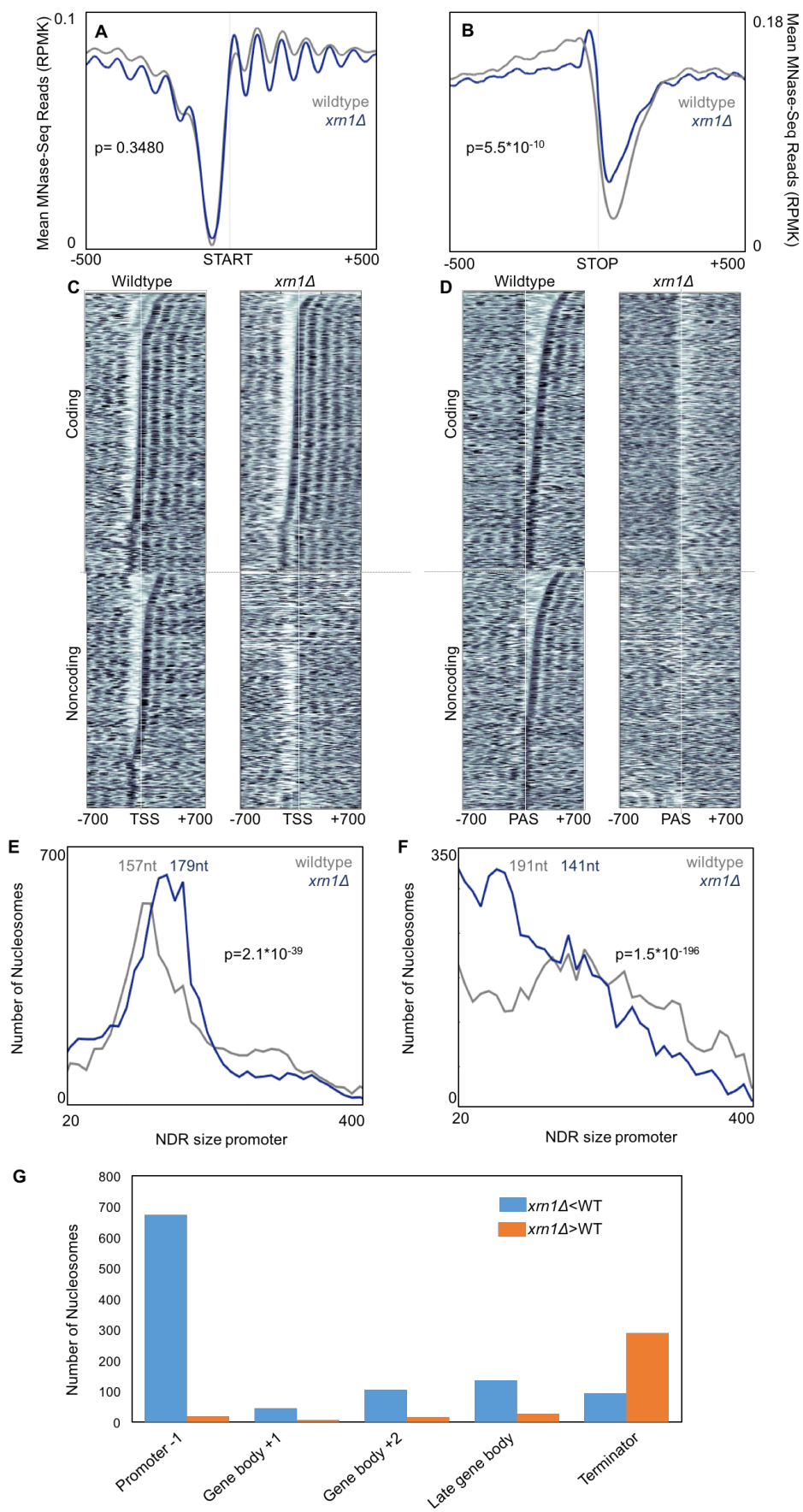


Figure 5.9. Nucleosome occupancy in *xrn1Δ*.

(A-B) Nucleosome depletion at the (A) promoter and (B) terminator in *xrn1Δ* (blue) and wild-type yeast (grey) at coding genes. (C-D) Heatmaps of nucleosome occupancy at the (C) promoter and (D) terminator. Dark colours indicate high nucleosome occupancy, light colours low nucleosome occupancy. The plot was normalised for each gene individually relative to the maximum and minimum read counts measured at each gene. (E-F) Distribution of NDR sizes at the (E) promoter and (F) terminator in *xrn1Δ* (blue) and wild-type yeast (grey) of coding genes. (G) Nucleosome occupancy at different positions within the gene and at the promoter and terminator in *xrn1Δ* compared to wild-type yeast. Increases in nucleosome occupancy in *xrn1Δ* mutants are indicated in red, decreases are indicated in blue. Nucleosomes were counted if their occupancy changed more than two-fold and had a p-value of less than 0.05. Differences in nucleosome occupancy were assessed with the danpos2 programme (Chen et al. 2013).

In *xrn1Δ* mutants, concentrations of more than 30% of RNAs coding for transcription factors are increased more than two-fold. Assuming that some of these RNAs are translated, it is possible that an increased number of transcription factors are present in *xrn1Δ* cells and that these factors contribute to the nucleosome depletion observed at promoters.

To test this, nucleosome occupancy at Gal4p, Msn2p, Fkh2p, Abf1p, Reb1p and Rap1p transcription factor binding sites was compared in *xrn1Δ* and wild-types.

Nucleosomes were significantly more depleted at binding sites of Gal4p, Msn2p, Fkh2p and Abf1p in *xrn1Δ* mutants (Figure 5.10). By contrast, changes in nucleosome occupancy at binding sites of Reb1p and Rap1p were not significant, although *xrn1Δ* was characterised by increases in nucleosome occupancy just upstream of the Reb1p and Rap1p binding sites.

Gal4p, Msn2p, Fkh2p and Abf1p all target genes which are not constitutively expressed, but whose expression is induced in response to changes in cell metabolism, stress or are required for cell cycle progression. Specifically, the transcription factors have roles in the induction of expression of the *GAL* regulon (Gal4p, Traven et al. 2006), stress-response genes (Msn2p, Elfving et

al. 2014; Stewart-Ornstein et al. 2016), cell-cycle genes (Fkh2p, Pic-Taylor et al. 2004; Ostrow et al. 2014) and nitrogen metabolism, sporulation, meiosis and ribosomal protein genes (Abf1p, Packham et al. 1996; Schroeder & Weil 1998; de Winde & Grivell 1992; Silve et al. 1992).

By comparison, Rap1p is primarily involved in telomere maintenance, as well as the transcriptional regulation of ribosomal protein genes, glycolytic genes and genes coding for components of the transcription machinery (Planta et al. 1995; Lieb et al. 2001) and Reb1p is a weak driver of transcription at a multitude of genes. Hence, while Gal4p, Msn2p, Fkh2p and, to some extent, Abf1p primarily regulate genes whose RNA concentrations are increased in the absence of *XRN1*, Rap1, and to some extent Reb1p, are associated with genes whose transcript concentrations are mostly unaffected by the deletion of *XRN1*.

While this does not prove that transcription factors bind to and recruit RNAP II to promoters of genes whose transcription rates are upregulated in *xrn1Δ* mutants, the data show that the observed increases in nucleosome depletion and transcription rates in *xrn1Δ* yeast could be driven by the binding of transcription factors to their target promoters. Moreover, the data indicate that this mechanism could be active even in the absence of a concrete stimulus.

For instance, Gal4p promoters appear depleted of nucleosomes in *xrn1Δ* mutants even when glucose is the only available carbon source. However, in order to limit Gal4p-stimulated transcription from *GAL* regulon promoters, the transcription of the *GAL4* gene is repressed in the presence of glucose in wild-type yeast (Griggs & Johnston 1991). Hence, in wild-types grown in glucose very little Gal4p is thought to be present in the cell. However, as *GAL4*

transcript concentrations are elevated in *xrn1* Δ even in glucose, one can hypothesise that low concentrations of Gal4p also accumulate in *xrn1* Δ . The Gal4p would then binds to UAS_{GAL} sequences leading to the depletion of nucleosomes of *GAL* regulon genes. An example of differences in nucleosome occupancy at the *GAL1/GAL10* locus is shown in Figure 5.11.

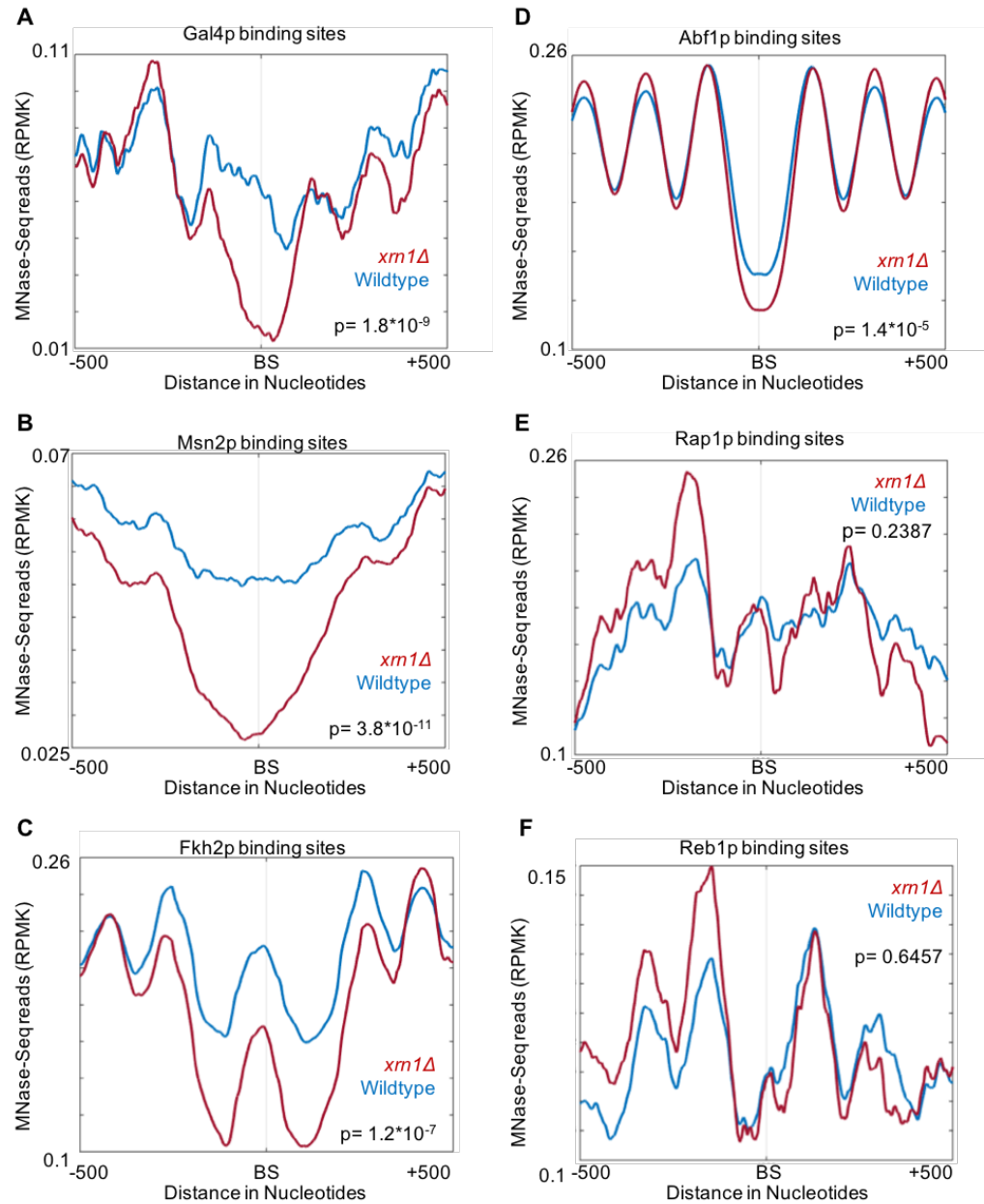


Figure 5.10. Nucleosome occupancy at transcription factor binding sites in *xrn1Δ* and wild-type strains.

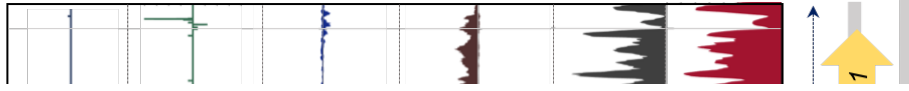
Nucleosome occupancies at (A) Gal4p, (B) Msn2p, (C) Fkh2p, (D) Abf1p, (E) Rap1p and (F) Reb1p binding sites (BS) in *xrn1Δ* (red) and wild-type (blue) yeast. p-values, calculated using Student ttests, indicate the statistical significance between differences in read numbers in the two strains in a window of -50 nucleotides upstream to +50 nucleotides downstream of the annotated binding site (BS).

5.2.8 Glucose to galactose shifts result in a global decrease in RNA concentrations

To test whether the observed remodelling of *GAL* promoters correlated with an increased responsiveness of galactose genes in *xrn1Δ* mutants, the yeast were subjected to a glucose to galactose shift. RNA-Seq, NET-Seq and MNase-Seq samples were collected early on during the yeast's adaptation to galactose, namely at 15 minutes and 60 minutes after the transfer from glucose to galactose. An example of the data is shown in Figure 5.11 (*N.B.* MNase-Seq data is not discussed in detail). Strikingly, the absence of Xrn1p led to an early induction of *GAL* gene transcription, as well as the accumulation of stable *GAL* RNA.

Intracellular RNA concentrations decreased by approximately 30% within 60 minutes of the glucose-to-galactose shift (Figure 5.12. (A)). Although RNA concentrations remained at elevated levels in *xrn1Δ* relative to wild-type yeast, this reduction in RNA concentrations was evident in both the *xrn1Δ* and the wild-type yeast, suggesting that RNA concentrations are actively regulated in the *xrn1Δ* mutants in response to acute stress.

Moreover, as NET-Seq read counts of 28% of genes in wild-types and 18% of genes in *xrn1Δ* mutants were either increased- or decreased 15 minutes after the glucose to galactose shift, the glucose to galactose shift causes major changes in transcription profiles of yeast.



281000

Figure 5.11. Differences and similarities in nucleosome occupancy, RNA transcription rates and stable RNA concentrations in *xrn1Δ* and wild-type yeast at the *GAL1/GAL10* locus.

Green arrows highlight differences between *xrn1Δ* and wild-types in polyadenylation and transcription rates in glucose. Blue dashed arrows highlight differences between *xrn1Δ* and wild-types in polyadenylation and transcription rates after 60 minutes in galactose. Red arrows highlight differences between *xrn1Δ* and wild-type yeast in NET-Seq reads after 60 minutes in galactose. All NET-Seq and MNase-Seq data was drawn on the same scale. For plots showing polyadenylation sites (PAS), the glucose data is depicted on a ten-fold lower scale than the galactose data. The NET-Seq data was collected by Fischl, H.

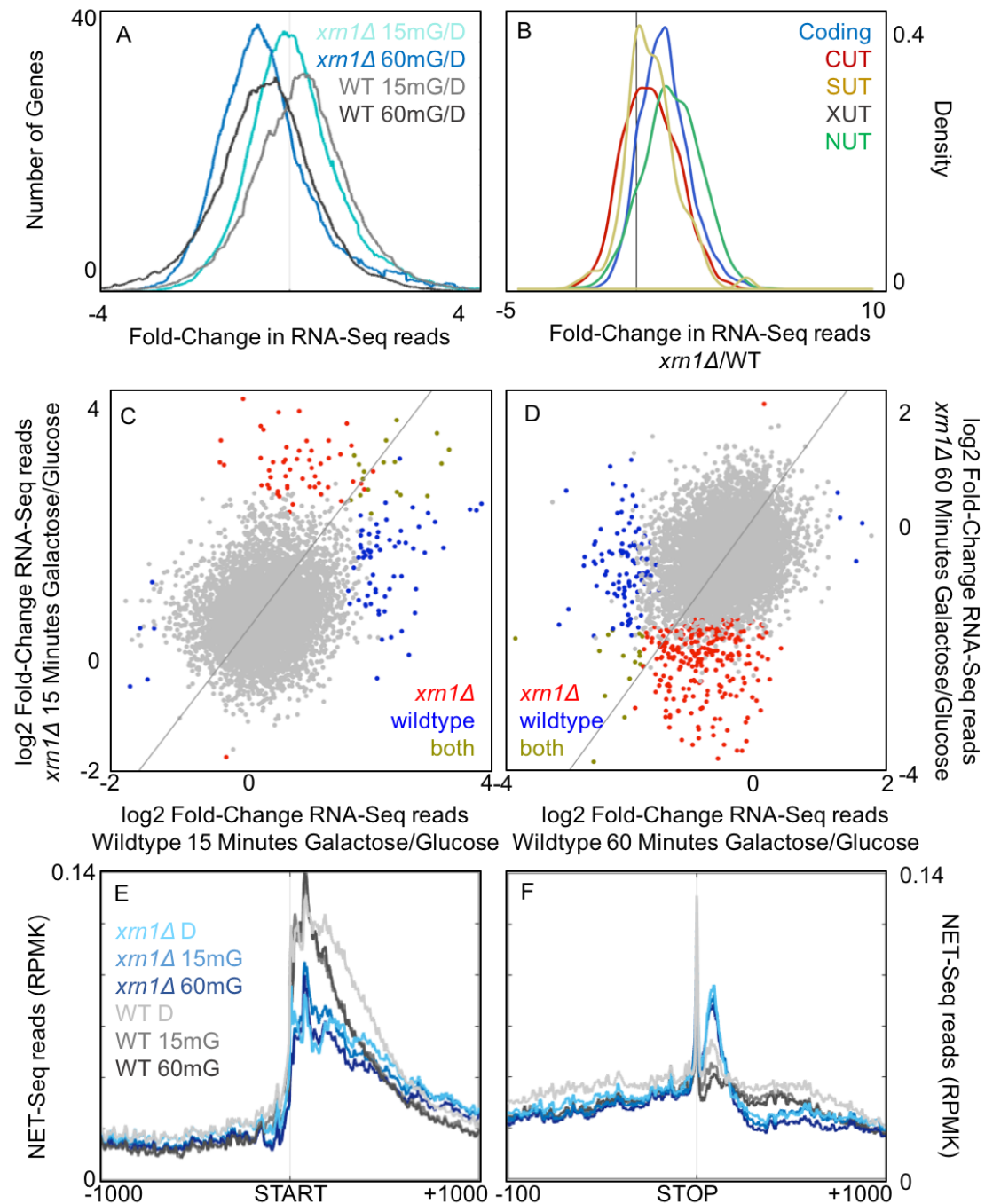


Figure 5.12. Changes in RNA concentrations and transcription rates after a glucose-to-galactose-shift.

(A) Distribution of fold-changes in RNA concentration in *xrn1Δ* and wild-type yeast 15 and 60 minutes after a glucose to galactose shift relative to RNA concentrations measured in glucose. (B) Fold changes in different types of RNAs in *xrn1Δ* relative to wild-type yeast 60 minute after a glucose-to –galactose shift. (C-D) Scatter plot of fold-changes in RNA concentration in after (C) 15 and (D) 60 minutes in galactose in *xrn1Δ* compared to wild-type yeast. Fold changes were calculated relative to the concentrations measured in glucose in *xrn1Δ* and wild-type yeast respectively. (E-F) Metagene plot of the mean number of NET-Seq reads surrounding the (E) START codon and (E) STOP codon of coding genes measured at different time points during the glucose-to-galactose shift. *xrn1Δ* mutants are shown in blue and wild-type yeast in grey. 15mG- 15 minutes in galactose, 60mG- 60 minutes in galactose, D-glucose

5.2.9 Changes in RNA concentrations and transcription rates in during a glucose-to-galactose shift are indicative of a stress response

In total, RNA concentrations of 230 coding genes in *xrn1Δ* and 286 coding genes in wild-type changed more than two-fold after a glucose to galactose shift. RNAs, whose concentrations increased, primarily code for proteins involved in glycolysis, carbohydrate metabolism and aerobic metabolism. By contrast, RNAs, whose concentrations were reduced, are associated with cytoplasmic translation (Table 5.5.).

Consistent with previous studies (Gasch et al. 2000; Nguyen et al. 2014), this suggests that the yeast undergo acute starvation upon the withdrawal of glucose. The reason for this is that yeast are unable to metabolise galactose immediately after the glucose-to-galactose shift, as they lack Gal proteins. Hence, to compensate for the loss of their primary carbon source, the yeast reduce their energy consumption through a downregulation of translation rates. The reduction in translation rates is primarily mediated by a reduction in the number of ribosomes, which coincides with the degradation of ribosomal protein transcripts and a reduction in the transcription rates of genes coding for components of the translation machinery (Miller et al. 2011). At the same time, yeast maximise their energy output by switching to from aerobic glycolysis to respiration, which is disfavoured during exponential growth (Slavov et al. 2014).

Although the starvation response occurs in both the wild-type and *xrn1Δ* yeast, fold-changes in individual RNA concentrations are only poorly correlated ($R=0.2003$ after 15 minutes in galactose). In fact, only RNA concentrations of 17

transcripts are upregulated in both the wild-type and *xrn1Δ* after 15 minutes in galactose (Figure 5.12 (C-D)). This corresponds to only 9.1% of the total number of affected transcripts in *xrn1Δ*.

The reason for this appears to be a difference in the timing of the response to galactose. While *xrn1Δ* mutants accumulate transcripts required for oxidative metabolism after 15 minutes in galactose, the same group of genes is enriched in wild-types 60 minutes after the glucose to galactose shift. The changes observed in the concentration of stable RNAs were correlated with changes in transcription rates. 15 minutes after the switch in carbon source both strains had upregulated transcription rates of genes involved in aerobic respiration and the catabolism of alternative carbon sources (Table 5.5, Table 5.6). At the same time, transcription of ribosomal protein genes and genes involved in rRNA and tRNA processing were considerably reduced in both *xrn1Δ* and wild-type yeast.

However, while the transcriptional response was mostly correlated 15 minutes after the glucose to galactose shift, at 60 minutes the response in *xrn1Δ* and wild-type yeast had become asynchronous. In particular, while wild-types were in the process of reversing the transcriptional repression of RPGs, the same was not observed in *xrn1Δ* mutants. Instead, *xrn1Δ* mutants had upregulated the transcription of the *GAL* regulon, which remained repressed in wild-types (Table 5.7.). No association with downregulated genes was found after 60 minutes in galactose.

		Term	Bonferroni
15 Minutes Galactose	<i>xrn1Δ</i> (N=185)	Oxidation-reduction process	0.001705
		Biosynthesis of secondary metabolites	0.002661
		Oxidoreductase activity	0.035317
		Glycolysis / Gluconeogenesis	0.015234
	WT (N=73)	Starch and sucrose metabolism	0.025046
60 Minutes Galactose	<i>xrn1Δ</i> (N=145)	Ribosome	0.000125
		Galactose metabolism	0.002635
		Translation	0.023011
	WT (N=113)	Starch and sucrose metabolism	0.000000
		Oxidation-reduction process	0.000021
		Metabolic pathways	0.000011
		Trehalose biosynthetic process	0.000966
		Mitochondrion	0.000321
		Oxidoreductase	0.003185
		Biosynthesis of secondary metabolites	0.003135
		Carbohydrate metabolic process	0.017836
		Mitochondrial electron transport, ubiquinol to cytochrome c	0.017977
		Glycolysis / Gluconeogenesis	0.004127
		Glycogen biosynthetic process	0.024730
		Catalytic activity	0.030615
		Respiratory chain	0.012923
		Electron transport	0.037346
		Galactose metabolism	0.014089
		Oxidative phosphorylation	0.019357

Table 5.4. Gene Ontology terms associated with up- and downregulated transcripts after a glucose to galactose shift.

The gene ontology search was carried out with DAVID (Huang et al. 2008). p-values are Benjamini-Hochman adjusted to account for multiple testing. N- number of genes with significant changes in RNA concentrations.

	Term	Benjamini
Wild-type Upregulated Genes 15 Minutes Galactose/Glucose	Oxidation-reduction process	7.52E-19
	Oxidative phosphorylation	3.96E-17
	Mitochondrion	3.70E-16
	Metabolic pathways	1.33E-14
	Starch and sucrose metabolism	1.59E-12
	Tricarboxylic acid cycle	8.18E-08
	Aerobic respiration	1.47E-06
	Biosynthesis of antibiotics	8.35E-06
	Peroxisome	8.83E-06
	Electron transport	2.52E-05
	Glucose import	2.72E-05
	Cellular response to oxidative stress	2.92E-05
	Galactose metabolism	4.27E-05
	Glycogen biosynthetic process	5.30E-05
	Glucose transmembrane transporter activity	1.99E-04
Wild-type Downregulated Genes 15 Minutes Galactose/Glucose	Ribosome biogenesis	1.0234E-118
	rRNA processing	9.40495E-90
	Cytoplasmic translation	3.64194E-78
	Ribonucleoprotein complex	1.83119E-64
	Cytosolic large ribosomal subunit	2.19698E-44
	Poly(A) RNA binding	3.89196E-36
	RNA binding	4.08534E-23
	tRNA processing	8.50267E-20
	S-adenosyl-L-methionine	3.74145E-14
	Methyltransferase	1.12355E-13
	Methylation	3.37064E-13
	snoRNA binding	1.86184E-12
	DNA-directed RNA polymerase	1.97522E-11
	Purine metabolism	9.17332E-10
	RNA secondary structure unwinding	3.28536E-07

Table 5.5. Gene Ontology terms associated with changes in transcription rates in wild-types 15 minutes after a glucose to galactose shift.

	Term	p-Value
<i>xrn1Δ</i> Upregulated Genes 15 Minutes Galactose/Glucose	Oxidation-reduction process	8.28E-23
	Carbon metabolism	2.68E-12
	Metabolic pathways	1.12E-09
	Biosynthesis of secondary metabolites	8.04E-09
	Starch and sucrose metabolism	1.26E-08
	Tricarboxylic acid cycle	1.14E-06
	Biosynthesis of antibiotics	1.65E-07
	NAD	1.87E-06
	Glycogen biosynthetic process	1.70E-05
	Cellular response to oxidative stress	2.65E-05
	Peroxisome	3.47E-06
	Electron transport	2.61E-05
	Peroxisome	2.71E-05
	Aerobic respiration	1.33E-04
	Stress response	5.94E-05
	Glycolysis / Gluconeogenesis	2.13E-04
	Fructose and mannose metabolism	2.20E-04
	Pyruvate metabolism	3.02E-04
<i>xrn1Δ</i> Downregulated Genes 15 Minutes Galactose/Glucose	Ribosome biogenesis	5.88E-101
	rRNA processing	8.47E-79
	90S preribosome	2.07E-48
	Small-subunit processome	1.70E-28
	Ribonucleoprotein complex	9.49E-27
	Ribosomal large subunit biogenesis	2.66E-22
	RNA-binding	8.38E-11
	Methyltransferase	9.72E-10
	ATP-dependent RNA helicase activity	1.11E-07
	tRNA processing	6.05E-06

Table 5.6. Gene Ontology terms associated with changes in transcription rates in *xrn1Δ* 15 minutes after a glucose to galactose shift.

	Term	p-Value
Wild-type Upregulated Genes 60 Minutes Galactose/Glucose	Nucleolus	2.03884E-59
	Ribosome biogenesis	3.83523E-56
	rRNA processing	1.38015E-44
	90S preribosome	8.31182E-29
	Preribosome, large subunit precursor	4.44089E-15
	snoRNA binding	3.64173E-07
	tRNA processing	7.11913E-06
	Ribonucleoprotein complex	5.17888E-06
	Fungal-type cell wall	8.01922E-05
	Extracellular region	0.000156632
<i>xrn1Δ</i> Upregulated Genes 60 Minutes Galactose/Glucose	Galactose metabolic process	0.000655258
	Galactose catabolic process via UDP-galactose	0.006752686

Table 5.7. Gene Ontology terms associated with changes in transcription rates in *xrn1Δ* and wild-types 60 minutes after a glucose to galactose shift.

5.2.10 The transcription of ribosomal protein genes is less dynamic in *xrn1Δ*

As indicated by the gene ontology search, RNAs coding for components of ribosome are regulated differently during a stress response in *xrn1Δ* and wild-type yeast. More concretely, *xrn1Δ* mutants are unable to suppress transcription of ribosomal protein genes (RPGs) as efficiently as wild-type yeast during the early stages of the stress response. In addition, the *xrn1Δ* mutant is unable to reverse the transcriptional repression of RPGs on the same time-scale as the wild-type.

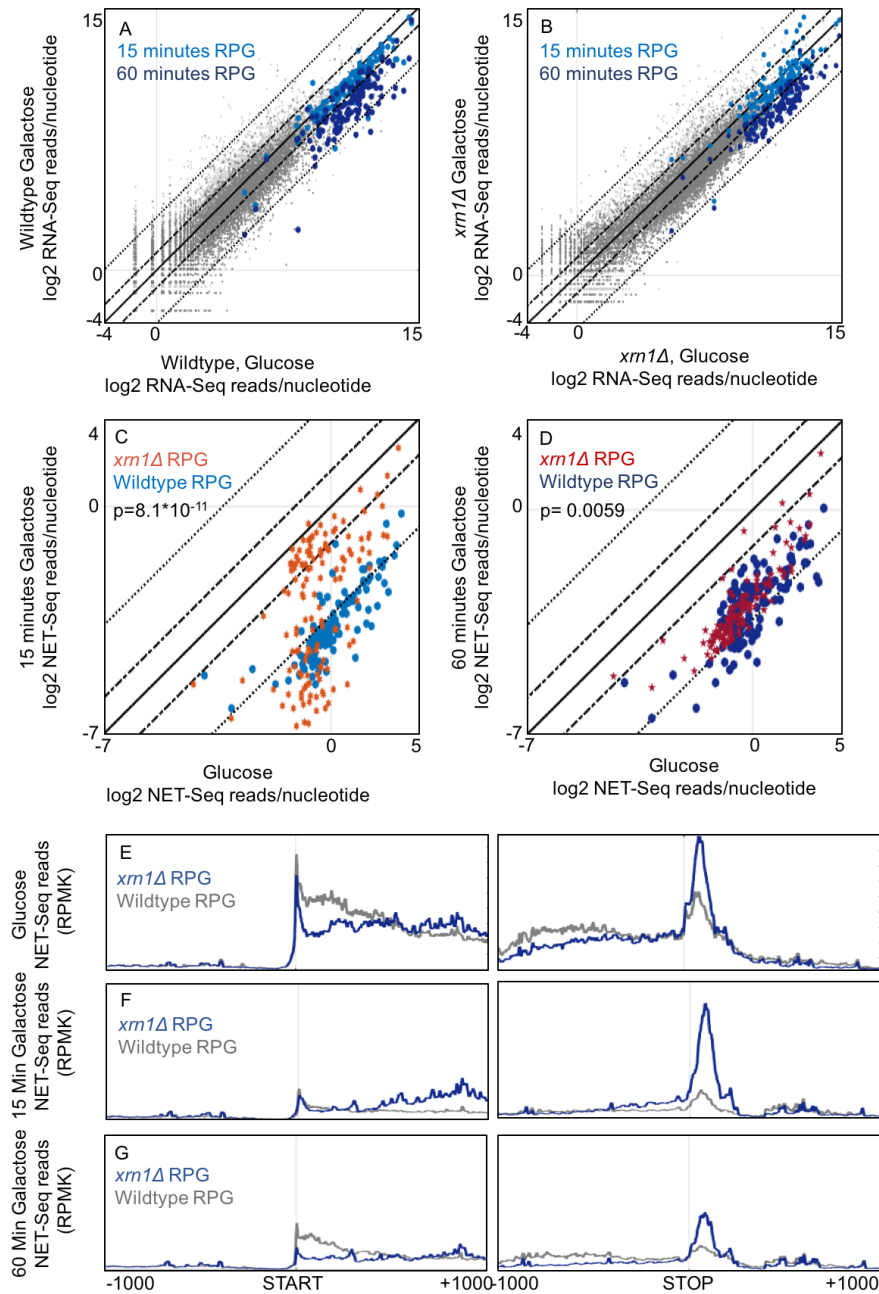


Figure 5.13. Regulation of ribosomal protein genes during a glucose-to-galactose shift.

(A-B) RNA concentrations of RPGs in (A) wild-type yeast and (B) *xrm1Δ* mutants 15 minutes (light blue) and 60 minutes (dark blue) after a glucose-to-galactose shift compared to concentrations in glucose. The grey dots in the background indicate the distribution of all other RNA concentrations in glucose compared to the 60 minutes-time point. Dashed lines represent a two-fold and dotted lines a four-fold change in RNA concentrations. (C-D) Scatter plots of the number of NET-Seq reads per nucleotide at RPGs in *xrm1Δ* (red) and wild-types (blue) (C) 15 minutes and (D) 60 minutes after a glucose-to-galactose shift. Dashed lines represent a two-fold and dotted lines a four-fold change in RNA concentrations. p-values were calculated using a Wilcoxon ranksum test and indicate the statistical significance in the difference between the fold-changes in transcription rates observed at RPGs in *xrm1Δ* and wild-type yeast. (E-G) Metagene profiles of mean NET-Seq reads at RPGs in (E) glucose and after (F) 15 minutes and (G) 60 minutes in galactose. *xrm1Δ* data is shown in blue, wild-type data in grey. NET-Seq data was collected by Fischl, H.

Evidence of this can be seen in Figure 5.13. (C). The plot shows that in *xrn1Δ* mutants 15 minutes after the switch in carbon source, a subgroup of RPGs is being transcribed at very similar rates to the ones observed in glucose. By contrast, at this time, the wild-type has downregulated the transcription of almost all RPGs. These comparatively high transcription rates of RPGs in *xrn1Δ* mutants are mirrored in the RNA concentrations of transcripts coding for RPGs measured at the 15 minutes time point. While the vast majority of transcripts coding for RPGs is downregulated in wild-types after 15 minutes in galactose compared to glucose, only 62% of RNAs coding for RPGs have lower RNA concentrations in *xrn1Δ* mutants.

Similarly, as transcription of some RPGs is induced after 60 minutes in galactose in wild-types but not in *xrn1Δ* mutants, concentrations of these transcripts relative to glucose levels are slightly increased in wild-type compared to *xrn1Δ* yeast. Hence, it appears that the regulatory response that controls the transcription of ribosomal proteins genes is delayed in the *xrn1Δ* strain.

5.2.11 Transcription of galactose genes is induced more quickly in *xrn1Δ* mutants

xrn1Δ mutants are characterised by the early transcription of the *GAL* regulon compared to wild-type yeast. This early transcription of *GAL* genes leads to the subsequent rapid accumulation of *GAL* RNAs, as evidenced by the RNA-Seq data (Figure 5.14). In particular, both RNA transcription rates and

concentrations of *GAL10*, *GAL1*, *GAL7* and *GAL4* were increased several fold in *xrn1Δ* compared to wild-type yeast.

Interestingly, both *GAL1* and *GAL4* RNAs are present at low concentrations in *xrn1Δ* in glucose and both of which probably contribute to the early gene induction observed in *xrn1Δ* yeast.

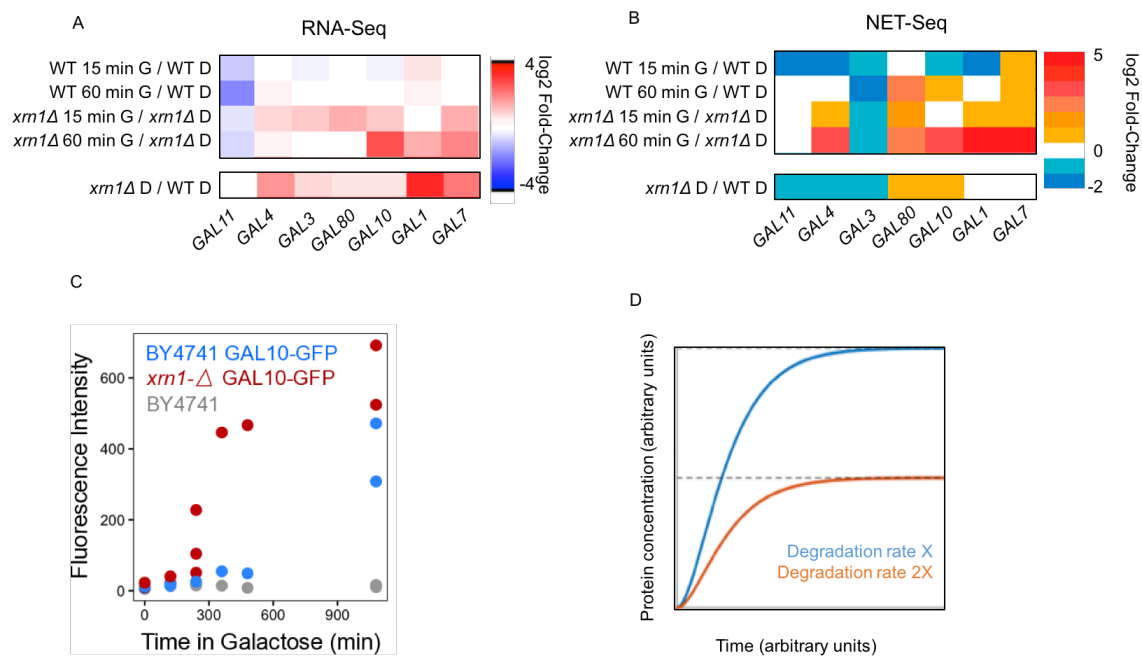


Figure 5.14. Regulation of *GAL* genes during a glucose-to-galactose shift.

(A-B) Fold-changes in (A) RNA concentrations of *GAL* transcripts and (B) NET-Seq reads per nucleotide of *GAL* genes in galactose (G) compared to glucose (D) wild-type and *xrn1Δ* yeast relative to glucose. (C) Average Fluorescence Activated Cell Sorting (FACS)-measured GFP intensity per cell (arbitrary units) in *GAL10-GFP* tagged wild-type (blue), *GAL10-GFP* tagged *xrn1Δ* (red) and untagged wild-type (grey) yeast. Each dot represents the average of two samples of 10000 cells each. The FACS data was collected by Phillip Lorenz. (D) Theoretical protein accumulation in cells with two hypothetical RNA degradation rates. In case of the red line the degradation rate was twice as high as in the case of the blue line. All other parameters (protein degradation rates, RNA transcription rates and translation rates) were kept constant.

Cells with even extremely low concentrations of Gal1p are able to induce transcription of the *GAL* regulon much more rapidly than cells, which have never been exposed to galactose and do not express *GAL1* (Zacharioudakis et al. 2007). Hence, assuming that some of the *GAL1* and *GAL4* RNA is translated in *xrn1Δ* cells, it is likely that the early expression of *GAL* genes in *xrn1Δ* mutants is driven by the presence of Gal1p as well as Gal4p.

To test whether the increased *GAL* RNA concentrations also leads the early accumulation of proteins in the cell, protein concentrations of GFP-tagged Gal10p were measured using Fluorescence-Activated-Cell-Sorting (FACS). The data show that in the absence of Xrn1p, the accumulation of *GAL10* RNA correlates with an increase in Gal10p-GFP concentrations. More precisely, Gal10p-GFP protein accumulated in *xrn1Δ* from approximately three hours after the glucose-to-galactose-shift. By contrast, Gal10p-GFP protein concentrations remained low for several hours after the three-hour point in wild-types. Overall, proteins accumulated 3.91-fold more rapidly in the *xrn1Δ* than in wild-type yeast and saturated at approximately 1.4-fold higher protein concentrations than in wild-types. Assuming that other *GAL* RNAs are similarly affected, this suggests that the *xrn1Δ* strain is able to metabolise galactose much earlier than the wild-type.

5.2.12 Lower degradation rates can lead to increases in protein concentrations

To test whether changes in transcript degradation rates can lead, at least theoretically, to increases in protein concentrations, a very basic mathematical model was formulated.

The model assumes that RNA transcription and RNA degradation, as well as protein translation and degradation, occur at a constant rate and was formulated the following way

$$\frac{dRNA}{dt} = -[RNA] * dr + tr$$

$$\frac{dProtein}{dt} = -[Protein] * dp + [RNA] * tp$$

whereby *dr* is the RNA degradation rate, *tr* the RNA transcription rate, *dp* the protein degradation rate and *tp* the protein translation rate.

The model shows that, in the absence of additional regulation (*i.e.* if *dp*, *tp* and *tr* are kept constant), lower degradation rates lead to an accumulation of RNAs and a subsequent increase in total protein concentrations (Figure 5.14.(D)). In fact, unless RNA concentrations are limited through transcriptional feedback loops or protein levels are controlled through changes in translation or protein degradation rates, a two-fold decrease in RNA degradation rates ultimately causes a two-fold increase in protein concentrations. Hence, the absence of Xrn1p could lead to significant changes in the proteome.

However, in living cells, the relation between changes in degradation rates is unlikely to be linear, as deadenylation rather than RNA degradation is

frequently the rate limiting step and poly(A) tail length affects translation dynamics of intact coding RNAs.

5.2.13 *xrn1Δ* are more resistant to some forms of stress

To test whether the early adaptation to galactose is unique or whether the *xrn1Δ* strain is also able to adapt to other stresses more quickly than the wild-type, *xrn1Δ* mutants were exposed to a variety of stresses and their growth rate compared to wild-types. When grown on drop plates, the mutant is more resistant to acute hydrogen-peroxide-induced as well as hydroxyurea-induced stress. Hence, it is possible that *xrn1Δ*, in addition to its faster response to galactose, is also able to tolerate the exposure certain stresses better than the wild-type.

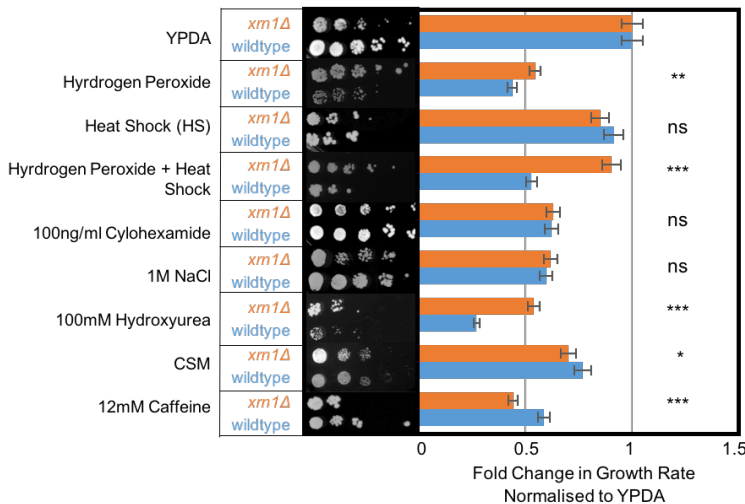


Figure 5.15. *xrn1Δ* phenotype.

Growth of *xrn1Δ* and wild-type yeast on different drop plates. Growth rates displayed in the bar chart were normalised to the measured growth rates on YPDA plates (N=3). For details on the measurements and normalisation see Methods. Images of the drops shown were taken three days after plating. p-values were calculated using Student t-tests. ns- not significant, * p<0.05, ** p<0.01, *** p<0.001

5.3 Discussion

The absence of Xrn1p leads to a global dysregulation in RNA metabolism, which ultimately results in a more than two-fold increase in total RNA concentrations and the accumulation of a majority of RNAs through both direct and secondary effects.

Previous models have suggested that Xrn1p regulates RNA concentrations both through RNA degradation and through mechanisms that affect RNA transcription rates. While the data presented here also show that Xrn1p is a regulator of transcription rates, none of the published models are consistent with the data presented in this Chapter.

Firstly, although more than 1500 transcripts with complementary 3' UTRs were detected in the RNA Seq-data, only two transcripts were characterised by decreased RNA concentrations for overlapping TUs, suggesting that the absence of Xrn1p does not lead to wide-spread reductions in the concentrations of complementary RNAs, as has been proposed by the post-transcriptional regulator model. Therefore, it is unlikely that concentrations of most RNAs are a result of inappropriate degradation via the NMD and NGD pathways in *xrn1Δ* mutants.

Secondly, in the case of the transcriptional activator model, catalytically inactive *XRN1* mutants are unable to induce stress-response genes at the same rate as wild-types. However, the strain used in the current study is characterised by an accelerated accumulation of *GAL* transcripts as well as transcripts coding for respiratory genes after a glucose-to-galactose shift. This suggests that Xrn1p represses these genes in wild-type yeast.

At the same time, the absence of Xrn1p led to the prolonged repression of ribosomal protein genes after a glucose-to-galactose shift. Hence, in contradiction with the transcriptional repressor model, Xrn1p is required for the re-activation of transcription at RPGs.

This suggests that Xrn1p directly or indirectly regulates transcriptional induction and repression. In fact, the data is consistent with a role for Xrn1p in the general surveillance of the transcriptome.

In this alternative model, outlined in Figure 5.16, Xrn1p prevents the accumulation of specific transcripts in a condition-dependent manner. Hence, rather than interacting with the transcription machinery itself, Xrn1p regulates transcription indirectly through the specific cytoplasmic degradation of RNA transcripts coding for proteins involved in transcriptional regulation.

There are several pieces of evidence that support this hypothesis.

Firstly, Xrn1p is known to limit the RNA concentrations of particular transcripts in a condition-dependent manner. For instance, during glucose repression, Snf1p targets are degraded by Xrn1p (Braun et al. 2014). Similarly, Xrn1p restricts OCA2 concentrations during respiratory metabolism but not aerobic glycolysis (Sinturel et al. 2012; Sinturel et al. 2015). Condition-specific affinities for RNAs in this context are determined by the phosphorylation of Xrn1p or the binding of regulatory proteins (Sinturel et al. 2012; Braun et al. 2014).

Consistent with this role in transcriptome surveillance, several hundred transcripts, which are below detectable levels in wild-types, accumulate in *xrn1Δ* mutants. Most of these transcripts originate from loci that are repressed under

optimal growth conditions, for instance from silenced meiotic and sporulation genes. In addition, many RNAs, whose concentrations increase in *xrn1Δ*, are not constitutively expressed. One example of this are cell-cycle genes, which are expressed periodically, but whose RNA concentrations are elevated several-fold in *xrn1Δ* mutants.

Secondly, increases in the concentration of specific RNAs at inappropriate times can have functional consequences in *xrn1Δ* mutants. For example, the absence of Xrn1p results in a cell-cycle defect during the exit from mitosis due to a failure to degrade *CLB2* (Pathak et al. 2005). In addition, in *D. melanogaster pacman* mutants, the failure to degrade specific RNAs coding for apoptotic proteins causes a lethal phenotype during development (Zabolotskaya et al. 2008; Jones et al. 2016). Therefore, Xrn1p is required to fine-tune the cell's transcriptome by restricting which RNAs are present in the cell. A failure to do so results in altered cellular signalling and has adverse consequences for the organism.

Thirdly, it is likely that an increase in RNA concentration leads to a subsequent increase in protein concentration. In general, it is assumed that the RNAs that accumulate in the *xrn1Δ* mutant are decapped and therefore non-functional (Hsu & Stevens 1993). However, while the fraction of decapped RNAs increases, there is little evidence that all the additional RNAs present in the cell are non-functional. In fact, when looking at sucrose gradients in *xrn1Δ* mutants, the majority of RNAs are bound to ribosomes (Hu et al. 2009), suggesting that they are being translated. Moreover, Xrn1p has been shown to directly stimulate decapping through interactions with the decapping complex (Chang et al. 2014;

Braun et al. 2012) and several studies have suggested that Xrn1p is required to maintain adequate decapping rates (Braun et al. 2014; Horigome et al. 2011; Braun et al. 2012). Therefore, it seems likely that both decapped and functional RNA concentrations are elevated in *xrn1Δ* yeast.

Assuming the concentration of functional RNAs is indeed increased, it is probable that increases in RNA concentrations are mimicked by increases in protein concentrations. This notion is supported by the Gal10p-GFP FACS data, which show that Gal10p-GFP not only accumulates more quickly due to the early induction of *GAL10* expression, but that Gal10p protein concentrations saturate at 1.4-fold higher levels than in the wild-type. As increases in RNA concentrations are sufficient to cause changes in protein concentrations, this strongly suggests that protein concentrations are increased in absence of Xrn1p.

Lastly, one of the groups of genes, whose transcripts increase most in the absence of Xrn1p, are chromatin remodellers and transcription factors. Accumulation of these RNAs in glucose correlates with increased nucleosome depletion from the promoter.

Changes in the abundance of transcriptional activators, in particular, could result in transcription factor binding to the promoter of their target genes and, ultimately, de-repression of the locus. Consistent with this, the accumulation of *GAL4* RNA prior to exposure to galactose correlates with, and likely is the main driver of, the early induction of *GAL* gene expression.

Similarly, overexpression of repressors could result in reduced transcription rates at other genes, including ribosomal protein genes.

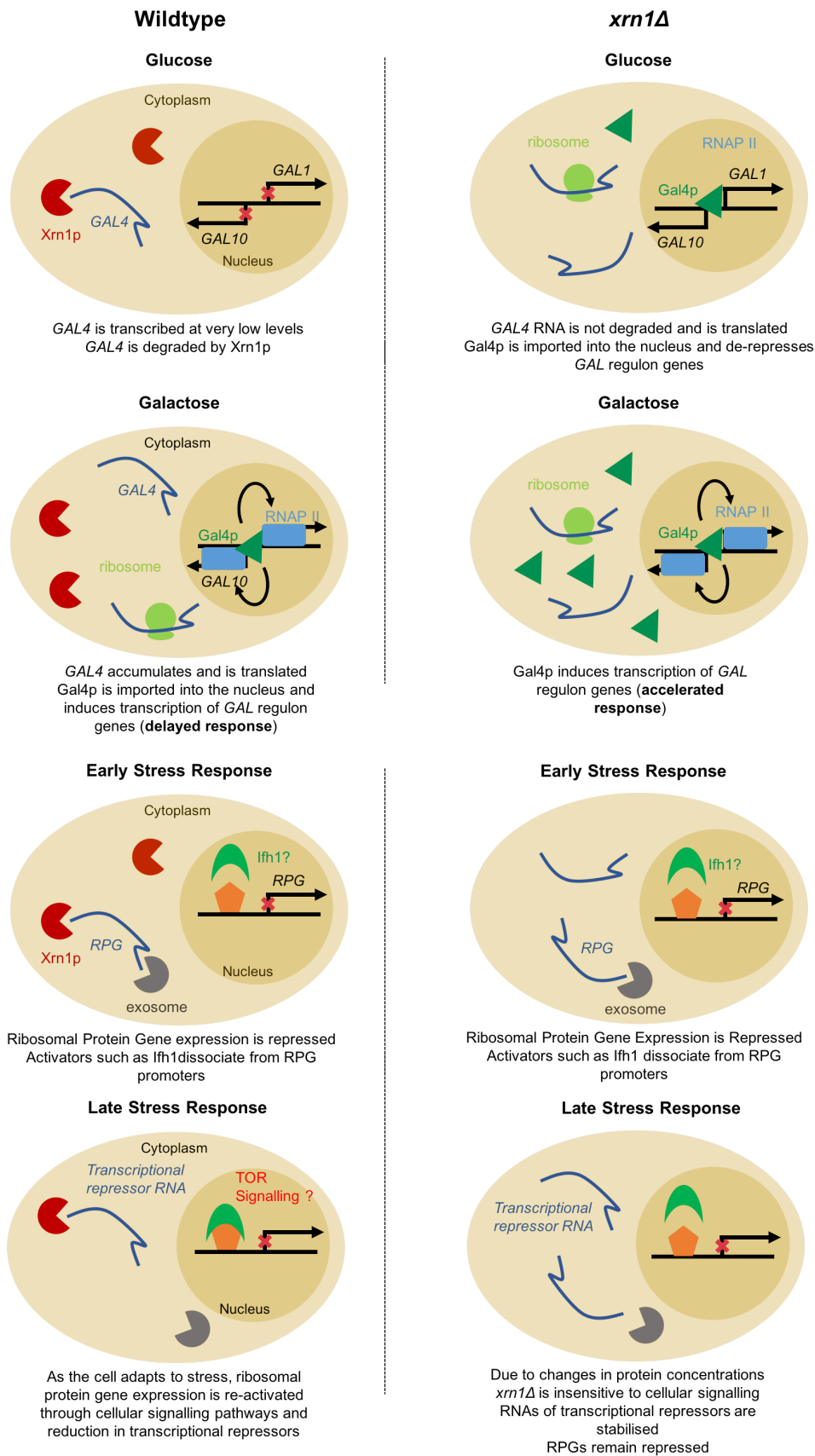


Figure 5.16. Model of Xrn1p transcriptional regulation.

Alternatively, dysregulation could also occur at the level of intercellular signalling. For instance, ribosomal protein gene expression is partially regulated by the TOR-signalling pathway. As some TOR regulators are synthetically lethal with *xrn1Δ*, it is possible that altered signalling of TOR contributes to differences in ribosomal gene expression in *xrn1Δ* (Brown et al. 2000; Loewith & Hall 2011). However, the exact mechanism of gene repression will have to be further investigated.

In summary, in the model proposed here, Xrn1p is required to maintain cellular RNA concentrations and prevent the accumulation of inappropriately expressed transcripts. This condition-specific degradation of RNAs also restricts concentrations of transcripts coding for transcription factors and chromatin remodellers, thereby helping the cell maintain transcriptional feedback loops. The absence of Xrn1p causes a dysregulation of these feedback loops leading to genome-wide differences in transcription rates. Importantly, unlike in the transcriptional activator model and the transcriptional repressor models, this model does not require Xrn1p to be imported into the nucleus.

6 Conclusion and Future Work

The aim of this thesis was to (a) integrate information from different genome-wide data sets to gain information about the transcriptional process, in particular with respect to termination and (b) to further investigate the functions of the two exoribonucleases Rrp6p and Xrn1p during the transcriptional process.

Chapter 3 showed that transcriptional termination is primarily conducted in transcription termination zones, which are frequently overlapping between genes and that the process is inherently linked to chromatin remodelling. Moreover, it showed that minor transcript isoforms are frequently terminated within the gene bodies, suggesting that intragenic transcripts constitute a previously uncharacterised group of molecules. It is likely that these molecules are, at least in part, terminated due to the occurrence of roadblock-like transcriptional terminators bound to the DNA.

Chapter 4 focussed on the function of Rrp6p and showed that polyadenylation site mapping can be used to characterise differences between stable RNAs and unstable Rrp6p-targeted isoforms. In addition, the data showed that Rrp6p targets a variety of non-coding transcripts. These frequently do not seem to have defined 3' ends, but are terminated across their annotated gene bodies in a NNS-dependent manner. Interestingly, unlike cleavage-factor-dependent termination, NNS does not occur within the central region of nucleosome depleted regions, suggesting that NNS termination does not underlie the same steric constraints as cleavage-factor dependent processing. In spite of the use of the NNS termination pathway, the data is consistent with the cleavage and

processing of non-coding transcripts in the absence of Rrp6p. This might be triggered by the prolonged residence time of Rrp6p-targets in the nucleus.

Chapter 5 investigated the function of Xrn1p during a stress-response.

Consistent with previous reports, the absence of Xrn1p resulted in increased RNA concentrations as well as altered chromatin packaging and transcription profiles. In particular, the absence of Xrn1p led to the early transcriptional induction of the *GAL* regulon, as well as of the late reactivation of *RPG* expression, suggesting that Xrn1p functions in both the repression and activation of transcription.

More broadly speaking, the data are consistent with a model in which Xrn1p's primary role is in the surveillance of the transcriptome. By removing RNAs coding for transcriptional activators and repressors, Xrn1p is able to modulate RNA transcription in the nucleus and fine-tune cellular signalling. The absence of Xrn1p leads to a dysregulation of these processes, probably due to the inappropriate accumulation of regulatory proteins.

Hence, transcription can be regulated via a multitude of mechanisms and feedback loops that ultimately shape the cell's transcriptome.

Future work will focus on confirming the results presented in this thesis. For Chapter 3 it will be important to show that intragenic transcripts can be captured by methods other than next-generation sequencing, for instance, through northern blotting. Moreover, to confirm that roadblock mechanisms play a role in the termination of specific transcripts candidate transcription factor binding sites will be mutated to see whether this leads to transcriptional read-through.

For Chapter 4, future work will try to elucidate a mechanism by which non-coding RNAs could be processed. In particular, it will be investigated whether Ysh1 plays a role in transcript processing.

With respect to Xrn1p's role in transcriptional regulation, future work will focus on gathering more concrete evidence that changes in transcription dynamics are driven by altered cellular concentrations of transcription factors. At the time of the writing of this thesis a proteome was in preparation, which will address the question whether protein concentrations are affected globally in *xrn1Δ* mutants. However, at this point in time the results are still pending. Moreover, ChIP-Seq experiments will be performed in *xrn1Δ* and wild-type yeast to confirm that transcription factor binding is indeed altered in the absence of Xrn1p. Combined, this will test whether transcriptional regulation in *xrn1Δ* mutants is indeed triggered by differences in the abundance of transcriptional regulators.

7 Bibliography

- Adelman, K. & Lis, J.T., 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet*, 13(10), pp.720–731.
- Alen, C. et al., 2002. A role for chromatin remodeling in transcriptional termination by RNA polymerase II. *Mol Cell*, 10, pp.1441–1452.
- Amrani, N. et al., 2004. A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature*, 432(7013), pp.112–118.
- Amrani, N. et al., 1997. PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I. *Molecular and cellular biology*, 17(3), pp.1102–9.
- Amrani, N. et al., 1997. Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro. *Molecular and cellular biology*, 17(7), pp.3694–3701.
- Anderson, J.S. & Parker, R.P., 1998. The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J*, 17(5), pp.1497–1506.
- Baejen, C. et al., 2014. Transcriptome Maps of mRNP Biogenesis Factors Define Pre-mRNA Recognition. *Molecular Cell*, 55(5), pp.745–757.
- Balagopal, V. & Parker, R., 2009. Polysomes, P-bodies and stress granules: states and fates of eukaryotic mRNAs. *Curr Opin Cell Biol*, 21(3), pp.403–408.
- Barrass, J.D. et al., 2015. Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. *Genome Biology*, 16(1), pp.1–17.
- Belotserkovskaya, R. et al., 2003. FACT Facilitates Transcription-Dependent Nucleosome Alteration. *Science*, 301(5636), pp.1090–1093.
- Bernecky, C. & Taatjes, D.J., 2012. Activator–Mediator Binding Stabilizes RNA Polymerase II Orientation within the Human Mediator–RNA Polymerase II–TFIIF Assembly. *Journal of Molecular Biology*, 417(5), pp.387–394.
- Boeck, R. et al., 1996. The Yeast Pan2 Protein Is Required for Poly(A) -binding Protein-stimulated Poly(A) -nuclease Activity *. , 271(1), pp.432–438.
- de Boer, C.G. et al., 2014. A unified model for yeast transcript definition. *Genome research*, 24(1), pp.154–66.
- Bohnsack, M.T., Tollervey, D. & Granneman, S., 2012. Chapter Thirteen - Identification of RNA Helicase Target Sites by UV Cross-Linking and Analysis of cDNA. *RNA helicases*.

- Academic Press, pp. 275–288.
- Bonneau, F. et al., 2016. The Yeast Exosome Functions as a Macromolecular Cage to Channel RNA Substrates for Degradation. *Cell*, 139(3), pp.547–559.
- Borja, M.S. et al., 2011. Dcp1 links coactivators of mRNA decapping to Dcp2 by proline recognition. *RNA*, 17(2), pp.278–290.
- Bouveret, E. et al., 2000. A Sm-like protein complex that participates in mRNA degradation. *The EMBO Journal*, 19(7), pp.1661–1671.
- Brauer, M.J. et al., 2008. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell*, 19(1), pp.352–367.
- Braun, J.E. et al., 2012. A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. *Nat Struct Mol Biol*, 19(12), pp.1324–1331.
- Braun, K.A. et al., 2014. Phosphoproteomic analysis identifies proteins involved in transcription-coupled mRNA decay as targets of Snf1 signaling. *Science Signaling*, 7(333), p.ra64.
- Braun, K.A. & Young, E.T., 2014. Coupling mRNA synthesis and decay. *Molecular and cellular biology*, 34 VN-r(22), pp.4078–4087.
- Bregman, A. et al., 2011. Promoter elements regulate cytoplasmic mRNA decay. *Cell*, 147(7), pp.1473–83.
- Brickner, D.G. et al., 2007. H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol*, 5(4), p.e81.
- Brown, C.E. & Sachs, A.B., 1998. Poly(A) Tail Length Control in *Saccharomyces cerevisiae* Occurs by Message-Specific Deadenylation. *Mol Biol Cell*, 18(11), pp.6548–6559.
- Brown, J.T., Yang, X. & Johnson, A.W., 2000. Inhibition of mRNA turnover in yeast by an *xrn1* mutation enhances the requirement for eIF4E binding to eIF4G and for proper capping of transcripts by Ceg1p. *Genetics*, 155(1), pp.31–42.
- Brune, C. et al., 2005. Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export. *RNA*, 11(4), pp.517–531.
- Buchan, J.R. & Parker, R., 2009. Eukaryotic stress granules: the ins and outs of translation. *Mol Cell*, 36(6), pp.932–941.
- Buratowski, S., 2003. The CTD code. *Nat Struct Biol*, 10(9), pp.679–680.
- Burgess, H.M. & Mohr, I., 2015. Cellular 5'-3' mRNA Exonuclease Xrn1 Controls Double-

- Stranded RNA Accumulation and Anti-Viral Responses. *Cell Host & Microbe*, 17(3), pp.332–344.
- Burnetti, A.J., Aydin, M. & Buchler, N.E., 2016. Cell cycle Start is coupled to entry into the yeast metabolic cycle across diverse strains and growth rates *Molecular Biology of the Cell*, 27(1), pp.64–74.
- Cacace, F. et al., 2012. Stochastic Modeling of Expression Kinetics Identifies Messenger Half-Lives and Reveals Sequential Waves of Co-ordinated Transcription and Decay. *PLoS Comput Biol*, 8(11), p.e1002772.
- Cao, D. & Parker, R., 2003. Computational modeling and experimental analysis of nonsense-mediated decay in yeast. *Cell*, 113(4), pp.533–545.
- Carneiro, T. et al., 2007. Depletion of the yeast nuclear exosome subunit Rrp6 results in accumulation of polyadenylated RNAs in a discrete domain within the nucleolus. *Mol Cell Biol*, 27(11), pp.4157–4165.
- Carroll, K.L. et al., 2004. Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol Cell Biol*, 24(14), pp.6241–6252.
- Carroll, K.L. et al., 2007. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA*, 13(3), pp.361–373.
- Castells-Roca, L. et al., 2011. Heat shock response in yeast involves changes in both transcription rates and mRNA stabilities. *PLoS ONE*, 6(2), e17272.
- Chang, C. Te et al., 2014. The activation of the decapping enzyme DCP2 by DCP1 occurs on the EDC4 scaffold and involves a conserved loop in DCP1. *Nucleic Acids Research*, 42(8), pp.5217–5233.
- Chang, J.H. et al., 2011. Structural and biochemical studies of the 5'→3' exoribonuclease Xrn1. *Nature structural & molecular biology*, 18(3), pp.270–6.
- Chapman, E.G. et al., 2014. RNA structures that resist degradation by Xrn1 produce a pathogenic dengue virus RNA. *eLife*, 2014(3), pp.1–25.
- Chapman, E.G. et al., 2014. The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science*, 344(6181), pp.307–310.
- Chen, K. et al., 2013. DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res*, 23(2), 341-51.
- Chernyakov, I. et al., 2008. Degradation of several hypomodified mature tRNA species in

- Saccharomyces cerevisiae* is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. *Genes & development*, 22(10), pp.1369–80.
- Chew, B.S. & Lehming, N., 2007. TFIIB/SUA7(E202G) is an allele-specific suppressor of TBP1(E186D). *Biochem J*, 406(2), pp.265–271.
- Chong, Y.T. et al., 2015. Yeast proteome dynamics from single cell imaging and automated analysis. *Cell*, 161(6), pp.1413–1424.
- Chowdhury, A., Kalurupalle, S. & Tharun, S., 2014. Pat1 contributes to the RNA binding activity of the Lsm1-7–Pat1 complex. *RNA*, 20(9), pp.1465–1475.
- Chowdhury, A., Mukhopadhyay, J. & Tharun, S., 2007. The decapping activator Lsm1p-7p–Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA*, 13(7), pp.998–1016.
- Christiano, R. et al., 2016. Global Proteome Turnover Analyses of the Yeasts. *Cell Reports*, 9(5), pp.1959–1965.
- Churchman, L.S. & Weissman, J.S., 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, 469(7330), pp.368–373.
- Cloutier, S.C. et al., 2013. Long noncoding RNAs promote transcriptional poising of inducible genes. *PLoS biology*, 11(11), p.e1001715.
- Cloutier, S.C. et al., 2016. Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation. *Molecular Cell*, 61(3), pp.393–404.
- Cole, S.E. et al., 2009. A Convergence of rRNA and mRNA Quality Control Pathways Revealed by Mechanistic Analysis of Nonfunctional rRNA Decay. *Molecular Cell*, 34(4), pp.440–450.
- Colin, J. et al., 2014. Roadblock Termination by Reb1p Restricts Cryptic and Readthrough Transcription. *Molecular Cell*, 56(5), pp.667–680.
- Coller, J. & Parker, R., 2004. Eukaryotic mRNA decapping. *Annu Rev Biochem*, 73, pp.861–890.
- Coller, J. & Parker, R., 2005. General translational repression by activators of mRNA decapping. *Cell*, 122(6), pp.875–886.
- Copela, L.A. et al., 2008. Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. *RNA*, 14(6), pp.1214–1227.
- Corden, J.L., 2016. Pol II CTD Code Light. *Molecular Cell*, 61(2), pp.183–184.

- Creamer, T.J. et al., 2011. Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet*, 7(10), p.e1002329.
- Czaplinski, K. et al., 1995. Purification and characterization of the Upf1 protein: a factor involved in translation and mRNA degradation. *RNA*, 1(6), pp.610–623.
- David, L. et al., 2006. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A*, 103(14), pp.5320–5325.
- Davis, C.A. & Ares Jr., M., 2006. Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, 103(9), pp.3262–3267.
- Decker, C.J. & Parker, R., 1993. A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev*, 7(8), pp.1632–1643.
- Derrien, T. et al., 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, 22(9), pp.1775–1789.
- Derrien, T., Guigó, R. & Johnson, R., 2011. The Long Non-Coding RNAs: A New (P)layer in the “Dark Matter.” *Frontiers in Genetics*, 2, p.107.
- Dichtl, B. & Keller, W., 2001. Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. *The EMBO Journal*, 20(12), pp.3197–3209.
- van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., et al., 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature*, 475(7354), pp.114–117.
- Dimitrova, L.N. et al., 2009. Nascent peptide-dependent translation arrest leads to Not4p-mediated protein degradation by the proteasome. *Journal of Biological Chemistry*, 284(16), pp.10343–10352.
- Ding, M. et al., 2015. A dual yet opposite growth-regulating function of miR-204 and its target XRN1 in prostate adenocarcinoma cells and neuroendocrine-like prostate cancer cells. *Oncotarget*, 6(10), pp.7686–7700.
- Dion, M.F. et al., 2007. Dynamics of replication-independent histone turnover in budding yeast. *Science*, 315(5817), pp.1405–1408.
- Djebali, S. et al., 2012. Landscape of transcription in human cells. *Nature*, 489(7414), pp.101–108.
- Doma, M.K. & Parker, R., 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in

- translation elongation. *Nature*, 440(7083), pp.561–564.
- Dori-Bachash, M. et al., 2012. Widespread promoter-mediated coordination of transcription and mRNA degradation. *Genome biology*, 13(12), p.R114.
- Draker, R. et al., 2012. A Combination of H2A.Z and H4 Acetylation Recruits Brd2 to Chromatin during Transcriptional Activation. *PLoS Genet*, 8(11), p.e1003047.
- Durand, S. & Lykke-Andersen, J., 2013. Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. *Nature structural & molecular biology*, 20(6), pp.702–9.
- Dziembowski, A. et al., 2007. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat Struct Mol Biol*, 14(1), pp.15–22.
- Eddy, S.R., 2014. Computational Analysis of Conserved RNA Secondary Structure in Transcriptomes and Genomes. *Annual Review of Biophysics*, 43(1), pp.433–456.
- Ehrensberger, A.H., Kelly, G.P. & Svejstrup, J.Q., 2013. Mechanistic Interpretation of Promoter-Proximal Peaks and RNAPII Density Maps. *Cell*, 154(4), pp.713–715.
- Elfving, N. et al., 2014. A dynamic interplay of nucleosome and Msn2 binding regulates kinetics of gene activation and repression following stress. *Nucleic Acids Research*, 42(9), pp.5468–5482.
- Escalante-Chong, R. et al., 2015. Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. *Proceedings of the National Academy of Sciences of the United States of America*, 112(5), pp.1636–1641.
- Eser, P. et al., 2014. Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. *Molecular Systems Biology*, 10, 717.
- Eulalio, A., Behm-Ansmant, I. & Izaurralde, E., 2007. P-bodies: at the crossroads of post-transcriptional pathways. *Nature Reviews Molecular Cell Biology*, 8(1), pp.9–22.
- Ezeokonkwo, C. et al., 2012. Novel interactions at the essential N-terminus of poly(A) polymerase that could regulate poly(A) addition in *Saccharomyces cerevisiae*. *FEBS letters*, 586(8), pp.1173–1178.
- Fan, X. et al., 2010. Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. *Proc Natl Acad Sci U S A*, 107(42), pp.17945–17950.
- Fasken, M.B., Laribee, R.N. & Corbett, A.H., 2015. Nab3 Facilitates the Function of the TRAMP

Complex in RNA Processing via Recruitment of Rrp6 Independent of Nrd1. *PLoS Genetics*, 11(3), p.e1005044.

- Fasken, M.B., Stewart, M. & Corbett, A.H., 2008. Functional significance of the interaction between the mRNA-binding protein, Nab2, and the nuclear pore-associated protein, Mlp1, in mRNA export. *J Biol Chem*, 283(40), pp.27130–27143.
- Feng, L. & Niu, D.-K., 2007. Relationship Between mRNA Stability and Length: An Old Question with a New Twist. *Biochemical Genetics*, 45(1), pp.131–137.
- Floer, M. et al., 2010. A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. *Cell*, 141(3), pp.407–18.
- Floor, S.N., Jones, B.N. & Gross, J.D., 2008. Control of mRNA Decapping by Dcp2: an open and shut case? *RNA biology*, 5(4), pp.189–192.
- Formosa, T. et al., 2001. Spt16-Pob3 and the HMG protein Nhp6 combine to form the nucleosome-binding factor SPN. *EMBO J*, 20(13), pp.3506–3517.
- Fox, M.J. et al., 2015. The Exosome Component Rrp6 Is Required for RNA Polymerase II Termination at Specific Targets of the Nrd1-Nab3 Pathway J. Corden, ed. *PLoS Genetics*, 11(2), p.e1004999.
- Frischmeyer, P.A. et al., 2002. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science*, 295(5563), pp.2258–2261.
- Fromm, S.A. et al., 2012. The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. *The EMBO Journal*, 31(2), pp.279–290.
- García-Martínez, J. et al., 2016. The cellular growth rate controls overall mRNA turnover, and modulates either transcription or degradation rates of particular gene regulons. *Nucleic Acids Research*, 44(8), pp.3643–3658.
- Gardner, L.B., 2008. Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response. *Molecular and Cellular Biology*, 28(11), pp.3729–3741.
- Gasch, A.P. et al., 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell*, 11, pp.4241–4257.
- Geisberg, J. V et al., 2014. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell*, 156(4), pp.812–24.
- Geisler, S. et al., 2012. Decapping of long noncoding RNAs regulates inducible genes. *Molecular cell*, 45(3), pp.279–91.

- Ghaemmaghami, S. et al., 2003. Global analysis of protein expression in yeast. *Nature*, 425(6959), pp.737–741.
- Ghazy, M.A. et al., 2012. The interaction of Pcf11 and Clp1 is needed for mRNA 3'-end formation and is modulated by amino acids in the ATP-binding site. *Nucleic Acids Res*, 40(3), pp.1214–1225.
- Goldstrohm, A.C. et al., 2006. PUF proteins bind Pop2p to regulate messenger RNAs. *Nat Struct Mol Biol*, 13(6), pp.533–539.
- Granovskaia, M. V et al., 2010. High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol*, 11(3), p.R24.
- Gregory, B.D. et al., 2008. A link between RNA metabolism and silencing affecting *Arabidopsis* development. *Dev Cell*, 14(6), pp.854–866.
- 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*, 33(23), pp.4718–4731.
- Griggs, D.W. & Johnston, M., 1991. Regulated expression of the *GAL4* activator gene in yeast provides a sensitive genetic switch for glucose repression. *Proc Natl Acad Sci U S A*, 88(19), pp.8597–8601.
- Grimaldi, Y., Ferrari, P. & Strubin, M., 2014. Independent RNA polymerase II preinitiation complex dynamics and nucleosome turnover at promoter sites *in vivo*. *Genome Research*, 24(1), pp.117–124.
- Grohmann, D. et al., 2011. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell*, 43(2), pp.263–274.
- Gruber, A.R. et al., 2008. The Vienna RNA Websuite. *Nucleic Acids Research*, 36(2), pp.W70–W74.
- Grzechnik, P., Gdula, M.R. & Proudfoot, N.J., 2015. Pcf11 orchestrates transcription termination pathways in yeast. *Genes & Development*, 29(8), pp.849–861.
- Grzechnik, P. & Kufel, J., 2008. Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol Cell*, 32(2), pp.247–258.
- Grzechnik, P., Tan-Wong, S.M. & Proudfoot, N.J., 2014. Terminate and make a loop: regulation of transcriptional directionality. *Trends in Biochemical Sciences*, 39(7), pp.319–327.
- Gudipati, R.K. et al., 2012. Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells. *Molecular Cell*, 48(3), pp.409–421.

- Guo, Z. & Sherman, F., 1996. 3'-end-forming signals of yeast mRNA. *Trends in Biochemical Sciences*, 21(12), pp.477–481.
- Gupta, I. et al., 2014. Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA–protein interactions. *Molecular Systems Biology*, 10(2), p.719.
- Guttman, M. et al., 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 458(7235), pp.223–227.
- Haber, J.E., 2012. Mating-Type Genes and MAT Switching in *Saccharomyces cerevisiae*. *Genetics*, 191(1), pp.33–64.
- Haimovich, G. et al., 2013. Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell*, 153(5), pp.1000–11.
- Haimovich, G., Choder, M., et al., 2013. The fate of the messenger is pre-determined: a new model for regulation of gene expression. *Biochim Biophys Acta*, 1829(6–7), pp.643–653.
- Hainer, S.J. & Martens, J.A., 2011. Identification of histone mutants that are defective for transcription-coupled nucleosome occupancy. *Mol Cell Biol*, 31(17), pp.3557–3568.
- Harigaya, Y. & Parker, R., 2010. No-go decay: a quality control mechanism for RNA in translation. *Wiley Interdiscip Rev RNA*, 1(1), pp.132–141.
- Harlen, K.M. et al., 2016. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Reports*, 15(10), pp.2147–2158.
- Hartley, P.D. & Madhani, H.D., 2009. Mechanisms that specify promoter nucleosome location and identity. *Cell*, 137(3), pp.445–458.
- Hazelbaker, D.Z. et al., 2013. Kinetic Competition between RNA Polymerase II and Sen1-dependent Transcription Termination. *Molecular Cell*, 49(1), pp.55–66.
- He, F. & Jacobson, A., 2015. Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annual Review of Genetics*, 49(1), pp. 339–366.
- Heo, D.-H. et al., 2013. The RNA Polymerase II C-terminal Domain-Interacting Domain of Yeast Nrd1 Contributes to the Choice of Termination Pathway and Couples to RNA Processing by the Nuclear Exosome. *The Journal of biological chemistry*, 288(51), pp.36676–36679
- Hilleren, P. et al., 2001. Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature*, 413(6855), pp.538–542.
- Hilleren, P. & Parker, R., 2001. Defects in the mRNA export factors Rat7p, Gle1p, Mex67p, and Rat8p cause hyperadenylation during 3'-end formation of nascent transcripts. *RNA*, 7(5), pp.753–764.

- Hilleren, P.J. & Parker, R., 2003. Cytoplasmic degradation of splice-defective pre-mRNAs and intermediates. *Molecular Cell*, 12(6), pp.1453–1465.
- Hirtreiter, A. et al., 2010. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res*, 38(12), pp.4040–4051.
- Hobor, F. et al., 2011. Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J Biol Chem*, 286(5), pp.3645–3657.
- Hobson, D.J. et al., 2012. RNA Polymerase II Collision Interrupts Convergent Transcription. *Molecular Cell*, 48(3), pp.365–374.
- Hogan, D.J. et al., 2008. Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System. *PLoS Biology*, 6(10), p.e255.
- Holland, M.J., 2002. Transcript abundance in yeast varies over six orders of magnitude. *Journal of Biological Chemistry*, 277(17), pp.14363–14366.
- Hongay, C.F. et al., 2006. Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell*, 127(4), pp.735–745.
- van Hoof, A. et al., 2002. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science*, 295(5563), pp.2262–2264.
- Horigome, C. et al., 2011. Ribosome biogenesis factors bind a nuclear envelope SUN domain protein to cluster yeast telomeres. *The EMBO journal*, 30(18), pp.3799–811.
- Houseley, J. et al., 2008. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Mol Cell*, 32(5), pp.685–695.
- Houseley, J. & Tollervey, D., 2009. The many pathways of RNA degradation. *Cell*, 136(4), pp.763–76.
- Hsu, C.L. & Stevens, A., 1993. Yeast cells lacking 5'→3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Mol Cell Biol*, 13(8), pp.4826–4835.
- Hu, W. et al., 2009. Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature*, 461(7261), pp.225–229.
- Huang, D.W., Sherman, B.T. & Lempicki, R.A., 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.*, 4(1), pp.44-57.
- Huang, H. et al., 2013. The nucleosome regulates the usage of polyadenylation sites in the human genome. *BMC Genomics*, 14(1), pp.1–10.

- Huang, H., Liu, H. & Sun, X., 2013. Nucleosome distribution near the 3' ends of genes in the human genome. *Biosci Biotech Bioch*, 77(10), pp.2051-2055.
- Huebert, D.J. et al., 2012. Dynamic changes in nucleosome occupancy are not predictive of gene expression dynamics but are linked to transcription and chromatin regulators. *Molecular and cellular biology*, 32(9), pp.1645–53.
- Huisinga, K.L. & Pugh, B.F., 2004. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol Cell*, 13(4), pp.573–585.
- Imbalzano, A.N. et al., 1994. Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature*, 370(6489), pp.481–485.
- Interthal, H. et al., 1995. A role of Sep1 (= Kem1, Xrn1) as a microtubule-associated protein in *Saccharomyces cerevisiae*. *EMBO J*, 14(6), pp.1057–1066.
- Jackson, J.S. et al., 2004. Recruitment of the Puf3 protein to its mRNA target for regulation of mRNA decay in yeast. *RNA*, 10(10), pp.1625–1636.
- Jamonnak, N. et al., 2011. Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA*, 17(11), pp.2011–2025.
- Jiao, X. et al., 2010. Identification of a quality control mechanism for mRNA 5'-end capping. *Nature*, 467(7315), pp.608–611.
- Jinek, M., Coyle, S.M. & Doudna, J. a, 2011. Coupled 5' nucleotide recognition and processivity in Xrn1-mediated mRNA decay. *Molecular cell*, 41(5), pp.600–8.
- Jones, C.I. et al., 2016. RNA-seq reveals post-transcriptional regulation of *Drosophila* insulin-like peptide dilp8 and the neuropeptide-like precursor Nplp2 by the exoribonuclease Pacman/XRN1. *Nucleic Acids Research*, 44(1), pp.267–280.
- Jones, C.I. et al., 2013. The 5'-3' exoribonuclease Pacman (Xrn1) regulates expression of the heat shock protein Hsp67Bc and the microRNA miR-277-3p in *Drosophila* wing imaginal discs. *RNA Biol*, 10(8).
- Jones, C.I., Zabolotskaya, M.V. & Newbury, S.F., 2012. The 5' → 3' exoribonuclease XRN1/Pacman and its functions in cellular processes and development. *Wiley interdisciplinary reviews. RNA*, 3(4), pp.455–68.
- Jorgensen, P. et al., 2002. Systematic identification of pathways that couple cell growth and division in yeast. *Science*, 297(5580), pp.395–400.
- Kadaba, S., Wang, X. & Anderson, J.T., 2006. Nuclear RNA surveillance in *Saccharomyces*

- cerevisiae*: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA*, 12(3), pp.508–521.
- Kaplan, N. et al., 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236), pp.362–6.
- Kasinathan, S. et al., 2013. High-resolution mapping of transcription factor binding sites on native chromatin. *Nature methods*, 11(2), pp.203–209
- Kassir, Y. et al., 2003. Transcriptional regulation of meiosis in budding yeast. *Int Rev Cytol*, 224, pp.111–171.
- Kayikci, O. & Nielsen, J., 2015. Glucose repression in *Saccharomyces cerevisiae*. *Fems Yeast Research*, 15(6), p.8.
- Khapersky, D.A. et al., 2016. Selective Degradation of Host RNA Polymerase II Transcripts by Influenza A Virus PA-X Host Shutoff Protein P. Palese, ed. *PLoS Pathogens*, 12(2), p.e1005427.
- Kharerin, H. et al., 2016. Role of transcription factor-mediated nucleosome disassembly in PHO5 gene expression. *Scientific Reports*, 6, p.20319.
- Kim, H. et al., 2010. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol*, 17(10), pp.1279–1286.
- Kim, J., 2002. KEM1 is involved in filamentous growth of *Saccharomyces cerevisiae*. *FEMS Microbiol Lett*, 216(1), pp.33–38.
- Kim, J.-H. et al., 2013. The glucose signaling network in yeast. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1830(11), pp.5204–5210.
- Kim, J., Ljungdahl, P.O. & Fink, G.R., 1990. kem mutations affect nuclear fusion in *Saccharomyces cerevisiae*. *Genetics*, 126(4), pp.799–812.
- Kim, M. et al., 2006. Distinct pathways for snoRNA and mRNA termination. *Mol Cell*, 24(5), pp.723–734.
- Kim, T., 2000. Mechanism of ATP-Dependent Promoter Melting by Transcription Factor IIH. *Science*, 288(5470), pp.1418–1421.
- Kim, T. et al., 2012. Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. *Cell*, 150(6), pp.1158–69.
- Kim, M. et al., 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, 432(7016), pp.517–522.

- Klauer, A.A. & van Hoof, A., 2012. Degradation of mRNAs that lack a stop codon: A decade of nonstop progress. *Wiley interdisciplinary reviews. RNA*, 3(5), pp.649–660.
- Kopcewicz, K. A., O'Rourke, T. W., & Reines, D., 2007. Metabolic Regulation of IMD2 Transcription and an Unusual DNA Element That Generates Short Transcripts *Molecular and Cellular Biology*, 27(8), 2821–2829.
- Kramer, E.B. & Hopper, A.K., 2013. Retrograde transfer RNA nuclear import provides a new level of tRNA quality control in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, 110(52), pp.21042–21047.
- Kruk, J.A. et al., 2011. The multifunctional Ccr4 – Not complex directly promotes transcription elongation. *Genes Dev*, 25(6), pp.581–593.
- Kuang, Z. et al., 2014. High temporal-resolution view of transcription and chromatin states across distinct metabolic states in budding yeast. *Nature structural & molecular biology*, 21(10), pp.854–863.
- Kundu, S., Horn, P.J. & Peterson, C.L., 2007. SWI/SNF is required for transcriptional memory at the yeast *GAL* gene cluster. *Genes Dev*, 21(8), pp.997–1004.
- Kuras, L. et al., 2000. TAF-Containing and TAF-independent forms of transcriptionally active TBP *in vivo*. *Science*, 288(5469), pp.1244–1248.
- Langmead, B., 2011. Aligning short sequencing reads with Bowtie. *Current protocols in bioinformatics*, pp.1–24.
- Larimer, F.W. & Stevens, A., 1990. Disruption of the gene *XRN1*, coding for a 5'----3' exoribonuclease, restricts yeast cell growth. *Gene*, 95(1), pp.85–90.
- Lavy, T. et al., 2012. The Gal3p transducer of the *GAL* regulon interacts with the Gal80p repressor in its ligand-induced closed conformation. *Genes and Development*, 26(3), pp.294–303.
- Lemay, J.F. & Bachand, F., 2015. Fail-safe transcription termination: Because one is never enough. *RNA Biol.*, 12(9), pp.927-932.
- Lemay, J.F. et al., 2014. The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nat Struct Mol Biol*, 21(10), pp.919-926.
- Lenstra, T.L. et al., 2016. Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. *Molecular Cell*, 60(4), pp.597–610.
- Li, Y., Yamane, D. & Lemon, S.M., 2015. Dissecting the roles of the 5' exoribonucleases Xrn1 and Xrn2 in restricting hepatitis C virus replication. *Journal of virology*, 89(9), pp.4857–65.

- Lieb, J.D. et al., 2001. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet*, 28(4), pp.327–334.
- Lin, M.D. et al., 2008. *Drosophila* processing bodies in oogenesis. *Dev Biol*, 322(2), pp.276–288.
- Loewith, R. & Hall, M.N., 2011. Target of rapamycin (TOR) in nutrient signaling and growth control. *Genetics*, 189(4), pp.1177–1201.
- Lohr, D. & Lopez, J., 1995. GAL4/GAL80-dependent nucleosome disruption/deposition on the upstream regions of the yeast GAL1-10 and GAL80 genes. *J Biol Chem*, 270(46), pp.27671–27678.
- Longtine, M.S. et al., 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast*, 14, pp.953–961.
- Lu, C., Brauer, M.J. & Botstein, D., 2009. Slow growth induces heat-shock resistance in normal and respiratory-deficient yeast. *Mol Biol Cell*, 20(3), pp.891–903.
- Luchelli, L., Thomas, M.G. & Boccaccio, G.L., 2015. Synaptic control of mRNA translation by reversible assembly of XRN1 bodies. *Journal of cell science*, pp.1542–1554.
- Lunde, B.M. et al., 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol*, 17(10), pp.1195–1201.
- Luo, W., Johnson, A.W. & Bentley, D.L., 2006. The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev*, 20(8), pp.954–965.
- Luo, Z. & Van Vuuren, H.J.J., 2008. Stress-induced production, processing and stability of a seripauperin protein, Pau5p, in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, 8(3), pp.374–385.
- Lykke-Andersen, J. & Bennett, E.J., 2014. Protecting the proteome: Eukaryotic cotranslational quality control pathways. *Journal of Cell Biology*, 204(4), pp.467–476.
- Lykke-Andersen, S. & Heick Jensen, T., 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nature Publishing Group*, 16.
- Machanick, P. & Bailey, T.L., 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics (Oxford, England)*, 27(12), pp.1696–7.
- Makino, D.L. et al., 2015. RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature*, 524(7563), pp.54–58.

- Makino, D.L., Baumgartner, M. & Conti, E., 2013. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature*, 495(7439), pp.70–75.
- Malabat, C. et al., 2015. Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife*, 2015(4), pp.1–59.
- Mandal, S.S. et al., 2004. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *PNAS*, 101(20), pp.7572–7577.
- Mandel, C.R. et al., 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end processing endonuclease. *Nature*, 444(7121), pp.953–6.
- Mandel, S., Robzyk, K. & Kassir, Y., 1994. *IME1* gene encodes a transcription factor which is required to induce meiosis in *Saccharomyces cerevisiae*. *Dev Genet*, 15(2), pp.139–147.
- Marquardt, S., Hazelbaker, D.Z. & Buratowski, S., 2011. Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. *Transcription*, 2(3), pp.145–154.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), p.10.
- Mata, J., 2013. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biology*, 10(8), pp.1407–1414.
- Mavrich, T.N. et al., 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res*, 18(7), pp.1073–1083.
- Mayer, A. et al., 2012. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science*, 336(6089), pp.1723–5.
- Mayer, A. et al., 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol*, 17(10), pp.1272–1278.
- Mayer, A. et al., 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nature structural & molecular biology*, 17(10), pp.1272–8.
- Medina, D. a. et al., 2014. Cytoplasmic 5'-3' exonuclease Xrn1p is also a genome-wide transcription factor in yeast. *Frontiers in Genetics*, 5, pp.1.
- Meinhart, A. & Cramer, P., 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, 430(6996), pp.223–226.
- Merret, R. et al., 2013. XRN4 and LARP1 are required for a heat-triggered mRNA decay pathway involved in plant acclimation and survival during thermal stress. *Cell Reports*, 5(5), pp.1279–1293.

- Miller, C. et al., 2011. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Molecular systems biology*, 7(458), p.458.
- Mischo, H.E. & Proudfoot, N.J., 2013. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta*, 1829(1), pp.174–185.
- Mitchell, S.F. et al., 2013. Global analysis of yeast mRNPs. *Nature structural & molecular biology*, 20(1), pp.127–33.
- Mizuno, T. et al., 1998. The Tup1-Ssn6 general repressor is involved in repression of IME1 encoding a transcriptional activator of meiosis in *Saccharomyces cerevisiae*. *Curr Genet*, 33(4), pp.239–247.
- Moon, S.L. et al., 2012. A noncoding RNA produced by arthropod-borne flaviviruses inhibits the cellular exoribonuclease XRN1 and alters host mRNA stability. *RNA*, 18(11), pp.2029–2040.
- Moqtaderi, Z. et al., 2013. Species-specific factors mediate extensive heterogeneity of mRNA 3' ends in yeasts. *Proceedings of the National Academy of Sciences of the United States of America*, 110(27), pp.11073–11078.
- Moqtaderi, Z. et al., 1996. Yeast homologues of higher eukaryotic TFIID subunits. *Proc Natl Acad Sci U S A*, 93(25), pp.14654–14658.
- Morris, K. V & Vogt, P.K., 2010. Long antisense non-coding RNAs and their role in transcription and oncogenesis. *Cell Cycle*, 9(13), pp.2544–2547.
- Muhlrad, D. & Parker, R., 1992. Mutations affecting stability and deadenylation of the yeast MFA2 transcript. *Genes Dev*, 6(11), pp.2100–2111.
- Munchel, S.E. et al., 2011. Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. *Molecular biology of the cell*, 22(15), pp.2787–95.
- Murray, S.C. et al., 2012. A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic Acids Res*, 40(6), pp.2432–2444.
- Murray, S.C. et al., 2015. Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Research*.
- Nagarajan, V.K. et al., 2013. XRN 5'→3' exoribonucleases: structure, mechanisms and functions. *Biochimica et biophysica acta*, 1829(6–7), pp.590–603.
- Nedea, E. et al., 2003. Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3' ends. *J Biol Chem*, 278(35), pp.33000–33010.

- Nehlin, J.O., Carlberg, M. & Ronne, H., 1991. Control of yeast *GAL* genes by *MIG1* repressor: a transcriptional cascade in the glucose response. *EMBO J*, 10(11), pp.3373–3377.
- Neil, H. et al., 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, 457(7232), pp.1038–1042.
- Newbury, S.F. & Wollard, A., 2004. The 5'–3' exoribonuclease *xrn-1* is essential for ventral epithelial enclosure during *C. elegans* embryogenesis. *RNA*, 10(1), pp.59–65.
- Nguyen, T. et al., 2014. Transcription mediated insulation and interference direct gene cluster expression switches. *eLife*, pp.1–21.
- Nissan, T. et al., 2010. Decapping Activators in *Saccharomyces cerevisiae* Act by Multiple Mechanisms. *Molecular Cell*, 39(5), pp.773–783.
- O'Duibhir, E. et al., 2014. Cell cycle population effects in perturbation studies. *Molecular Systems Biology*, 10(6), p.732.
- O'Leary, S.E. et al., 2013. Dynamic Recognition of the mRNA Cap by *Saccharomyces cerevisiae* eIF4E. *Structure (London, England : 1993)*, 21(12), p.10.1016/j.str.2013.09.016.
- Olmedo, G. et al., 2006. ETHYLENE-INSENSITIVE5 encodes a 5'→3' exoribonuclease required for regulation of the EIN3-targeting F-box proteins EBF1/2. *Proc Natl Acad Sci U S A*, 103(36), pp.13286–13293.
- Ostling, J. & Ronne, H., 1998. Negative control of the Mig1p repressor by Snf1p-dependent phosphorylation in the absence of glucose. *Eur J Biochem*, 252(1), pp.162–168.
- Ostrow, A.Z. et al., 2014. Fkh1 and Fkh2 Bind Multiple Chromosomal Elements in the *S. cerevisiae* Genome with Distinct Specificities and Cell Cycle Dynamics. *PLoS ONE*, 9(2), p.e87647.
- Ozsolak, F. et al., 2010. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell*.
- Packham, E.A., Graham, I.R. & Chambers, A., 1996. The multifunctional transcription factors Abf1p, Rap1p and Reb1p are required for full transcriptional activation of the chromosomal PGK gene in *Saccharomyces cerevisiae*. *Mol Gen Genet*, 250(3), pp.348–356.
- Parker, R., 2012. RNA degradation in *Saccharomyces cerevisiae*. *Genetics*, 191(3), pp.671–702.
- Parker, R. & Sheth, U., 2007. P Bodies and the Control of mRNA Translation and Degradation. *Molecular Cell*, 25(5), pp.635–646.

- Pathak, R. et al., 2005. A role for *KEM1* at the START of the cell cycle in *Saccharomyces cerevisiae*. *Current Genetics*, 48(5), pp.300–309.
- Pavri, R. et al., 2006. Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell*, 125(4), pp.703–717.
- Pearson, E.L. & Moore, C.L., 2013. Dismantling Promoter-driven RNA Polymerase II Transcription Complexes *in Vitro* by the Termination Factor Rat1. *The Journal of Biological Chemistry*, 288(27), pp.19750–19759.
- Pelechano, V. et al., 2012. Genome-Wide Polyadenylation Site Mapping. *Methods Enzymol.*, 513, pp.271-296.
- Pelechano, V., Chávez, S. & Pérez-Ortín, J., 2010. A complete set of nascent transcription rates for yeast genes. *PLoS One*, 5(11), pp. e15442.
- Pelechano, V., Wei, W. & Steinmetz, L.M., 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*, 497(7447), pp.127–131.
- Pelechano, V., Wei, W. & Steinmetz, L.M., 2015. Widespread co-translational RNA decay reveals ribosome dynamics. *Cell*, 161(6), pp.1400–1412.
- Pic-Taylor, A. et al., 2004. Regulation of Cell Cycle-Specific Gene Expression through Cyclin-Dependent Kinase-Mediated Phosphorylation of the Forkhead Transcription Factor Fkh2p. *Mol. Cell. Biol.*, 24(22), pp.10036–10046.
- Pinto, I. et al., 1994. Characterization of *sua7* mutations defines a domain of TFIIB involved in transcription start site selection in yeast. *J Biol Chem*, 269(48), pp.30569–30573.
- Pinto, I., Ware, D.E. & Hampsey, M., 1992. The yeast *SUA7* gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection in vivo. *Cell*, 68(5), pp.977–988.
- Planta, R.J., Goncalves, P.M. & Mager, W.H., 1995. Global regulators of ribosome biosynthesis in yeast. *Biochem Cell Biol*, 73(11–12), pp.825–834.
- Porrua, O., Libri, D., 2013. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat Struct Mol Biol*, 20(7), 884-891.
- Porrua, O. et al., 2012. In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J*, 31(19), pp.3935–3948.
- Porrua, O. & Libri, D., 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol*, 16(3), pp.190–202.

- Potuschak, T. et al., 2006. The exoribonuclease XRN4 is a component of the ethylene response pathway in Arabidopsis. *Plant Cell*, 18(11), pp.3047–3057.
- Prescott, EM, Proudfoot, NJ. 2002. Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A*, 99(13), pp.8796-801.
- Presnyak, V. et al., 2015. Codon optimality is a major determinant of mRNA stability. *Cell*, 160(6), pp.1111–1124.
- Rachidi, N. et al., 2000. *Saccharomyces cerevisiae* PAU genes are induced by anaerobiosis. *Molecular Microbiology*, 35(6), pp.1421–1430.
- Raghavan, R., Sloan, D.B. & Ochman, H., 2012. Antisense Transcription Is Pervasive but Rarely Conserved in Enteric Bacteria. *mBio*, 3(4), pp.e00156-12.
- Rani, P.G., Ranish, J.A. & Hahn, S., 2004. RNA polymerase II (Pol II)-TFIIF and Pol II-mediator complexes: the major stable Pol II complexes and their activity in transcription initiation and reinitiation. *Molecular and cellular biology*, 24(4), pp.1709–1720.
- Ranjan, A. et al., 2013. Nucleosome-free Region Dominates Histone Acetylation in Targeting SWR1 to Promoters for H2A.Z Replacement. *Cell*, 154(6), pp.1232–1245.
- Ransom, M. et al., 2009. FACT and the proteasome promote promoter chromatin disassembly and transcriptional initiation. *Journal of Biological Chemistry*, 284(35), pp.23461–23471.
- Reeder, R.H., Guevara, P. & Roan, J.G., 1999. *Saccharomyces cerevisiae* RNA Polymerase I Terminates Transcription at the Reb1 Terminator In Vivo. *Molecular and Cellular Biology*, 19(11), pp.7369–7376.
- Reimand, J. et al., 2007. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research*, 35(Web Server issue), pp.W193–W200.
- Rhee, H.S. & Pugh, B.F., 2012a. ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Current protocols in molecular biology*, 21, pp. 21-24.
- Rhee, H.S. & Pugh, B.F., 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, 483(7389), pp.295–301.
- Richard, P. & Manley, J.L., 2009. Transcription termination by nuclear RNA polymerases. *Genes & development*, 23(11), pp.1247–69.
- Rondón, AG. et al., 2009. Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Mol Cell*, 9;36(1), pp.88-98.

- Safaei, N. et al., 2016. Interdomain Allostery Promotes Assembly of the Poly(A) mRNA Complex with PABP and eIF4G. *Molecular Cell*, 48(3), pp.375–386.
- Sainsbury, S., Bernecky, C. & Cramer, P., 2015. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol*, 16(3), pp.129–143.
- Schaeffer, D. et al., 2009. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nature structural & molecular biology*, 16(1), pp.56–62.
- Schaeffer, D. & van Hoof, A., 2011. Different nuclease requirements for exosome-mediated degradation of normal and nonstop mRNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6), pp.2366–2371.
- Schlackow, M. et al., 2017. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol Cell*, 65(1), pp.25–38.
- Schmid, M. et al., 2012. Rps6p controls mRNA poly(A) tail length and its decoration with poly(A) binding proteins. *Mol Cell*, 47(2), pp.267–280.
- Schmid, M. et al., 2012. Rps6p Controls mRNA Poly(A) Tail Length and Its Decoration with Poly(A) Binding Proteins. *Molecular Cell*, 47(2), pp.267–280.
- Schneider, C. et al., 2012. Transcriptome-wide Analysis of Exosome Targets. *Molecular Cell*, 48(3–3), pp.422–433.
- Schroeder, S.C. & Weil, P.A., 1998. Genetic tests of the role of Abf1p in driving transcription of the yeast TATA box binding protein-encoding gene, SPT15. *J Biol Chem*, 273(31), pp.19884–19891.
- Schuller, H.J., 2003. Transcriptional control of nonfermentative metabolism in the yeast *Saccharomyces cerevisiae*. *Curr Genet*, 43(3), pp.139–160.
- Schulz, D. et al., 2013. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell*, 155(5), pp.1075–87.
- Segal, E. & Widom, J., 2009. Poly(dA:dT) Tracts: Major Determinants of Nucleosome Organization. *Current opinion in structural biology*, 19(1), pp.65–71.
- Shalem, O. et al., 2011. Transcriptome kinetics is governed by a genome-wide coupling of mRNA production and degradation: a role for RNA Pol II. *PLoS genetics*, 7(9), p.e1002273.
- Sharif, H. & Conti, E., 2013. Architecture of the Lsm1-7-Pat1 Complex: A Conserved Assembly in Eukaryotic mRNA Turnover. *Cell Reports*, 5(2), pp.283–291.

- Sherriff, J.A., Kent, N.A. & Mellor, J., 2007. The Isw2 chromatin-remodeling ATPase cooperates with the Fkh2 transcription factor to repress transcription of the B-type cyclin gene CLB2. *Mol Cell Biol*, 27(8), pp.2848–2860.
- Shimizu, M. et al., 1997. Transcriptional repression at a distance through exclusion of activator binding in vivo. *Proc Natl Acad Sci U S A*, 94(3), pp.790–795.
- Shoemaker, C.J. & Green, R., 2012. Translation drives mRNA quality control. *Nature structural & molecular biology*, 19(6), pp.594–601.
- Silva, P.A. et al., 2010. An RNA pseudoknot is required for production of yellow fever virus subgenomic RNA by the host nuclease XRN1. *J Virol*, 84(21), pp.11395–11406.
- Silve, S. et al., 1992. ABF1 is a phosphoprotein and plays a role in carbon source control of COX6 transcription in *Saccharomyces cerevisiae*. *Mol Cell Biol*, 12(9), pp.4197–4208.
- Simms, C.L. et al., 2014. An active role for the ribosome in determining the fate of oxidized mRNA. *Cell reports*, 9(4), pp.1256–1264.
- Singh, G., Rebbapragada, I. & Lykke-Andersen, J., 2008. A Competition between Stimulators and Antagonists of Upf Complex Recruitment Governs Human Nonsense-Mediated mRNA Decay M. Wickens, ed. *PLoS Biology*, 6(4), p.e111.
- Sinturel, F. et al., 2012. Activation of 5'-3' exoribonuclease Xrn1 by cofactor Dcs1 is essential for mitochondrial function in yeast. *Proc Natl Acad Sci U S A*, 109(21), pp.8264–8269.
- Sinturel, F. et al., 2012. Activation of 5'-3' exoribonuclease Xrn1 by cofactor Dcs1 is essential for mitochondrial function in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 109(21), pp.8264–9.
- Sinturel, F. et al., 2015. Cytoplasmic Control of Sense-Antisense mRNA Pairs. *Cell Reports*, 12(11), pp.1853–1864.
- Skourti-Stathaki, K., Proudfoot, N.J. & Gromak, N., 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell*, 42(6), pp.794–805.
- Skowronek, E. et al., 2014. tRNA 3' processing in yeast involves tRNase Z, Rex1, and Rrp6. *RNA*, 20(1), pp.115–130.
- Slavov, N. et al., 2012. A conserved cell growth cycle can account for the environmental stress responses of divergent eukaryotes. *Mol Biol Cell*, 23(10), pp.1986–1997.
- Slavov, N. et al., 2014. Constant Growth Rate Can Be Supported by Decreasing Energy Flux and Increasing Aerobic Glycolysis. *Cell reports*, 7(3), pp.705–714.

- Slavov, N. et al., 2011. Metabolic cycling without cell division cycling in respiring yeast. *Proc Natl Acad Sci U S A*, 108(47), pp.19090–19095.
- Slavov, N. & Botstein, D., 2011. Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast. *Mol Biol Cell*, 22(12), pp.1997–2009.
- Smith, J.E. et al., 2014. Translation of Small Open Reading Frames within Unannotated RNA Transcripts in *Saccharomyces cerevisiae*. *Cell Reports*, 7(6), pp.1858–1866.
- Stewart-Ornstein, J. et al., 2016. Msn2 Coordinates a Stoichiometric Gene Expression Program. *Current Biology*, 23(23), pp.2336–2345.
- Strich, R. et al., 1994. *UME6* is a key regulator of nitrogen repression and meiotic development. *Genes Dev.*, 8, pp.796–810.
- Struhl, K., 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol*, 14(2), pp.103–105.
- Struhl, K. & Segal, E., 2013. Determinants of nucleosome positioning. *Nature structural & molecular biology*, 20(3), pp.267–73.
- Sun, M. et al., 2013. Global analysis of eukaryotic mRNA degradation reveals *xrn1*-dependent buffering of transcript levels. *Molecular cell*, 52(1), pp.52–62.
- Sun, Q. et al., 2013. R-loop stabilization represses antisense transcription at the *Arabidopsis* FLC locus. *Science*, 340(6132), pp.619–621.
- Tan-Wong, S.M. et al., 2012. Gene loops enhance transcriptional directionality. *Science*, 338(6107), pp.671–675.
- Tanny, J.C., 2015. Chromatin modification by the RNA Polymerase II elongation complex. *Transcription*, 5(5), p.e988093.
- Tavazoie, S. et al., 1999. Systematic determination of genetic network architecture. *Nat Genet*, 22(3), pp.281–285.
- Teif, V.B. et al., 2012. Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol*, 19(11), pp.1185–1192.
- Tharun, S. & Parker, R., 2001. Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol Cell*, 8(5), pp.1075–1083.
- Thibault, P.A. et al., 2015. Regulation of Hepatitis C Virus Genome Replication by Xrn1 and *MicroRNA-122* Binding to Individual Sites in the 5' Untranslated Region. *Journal of virology*, 89(12), pp.6294–311.

- Thiebaut, M. et al., 2006. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Molecular cell*, 23(6), pp.853–64.
- Thompson, D.M. & Parker, R., 2007. Cytoplasmic decay of intergenic transcripts in *Saccharomyces cerevisiae*. *Molecular and cellular biology*, 27(1), pp.92–101.
- Tishkoff, D.X. et al., 1995. The Sep1 Mutant of *Saccharomyces Cerevisiae* Arrests in Pachytene and Is Deficient in Meiotic Recombination. *Genetics*, 139(2), pp.495–509.
- Tishkoff, D.X., Johnson, A.W. & Kolodner, R.D., 1991. Molecular and genetic analysis of the gene encoding the *Saccharomyces cerevisiae* strand exchange protein Sep1. *Molecular and Cellular Biology*, 11(5), pp.2593–2608.
- Traven, A., Jelacic, B. & Sopta, M., 2006. Yeast Gal4: a transcriptional paradigm revisited. *EMBO Rep*, 7(5), pp.496–499.
- Trcek, T. et al., 2011. Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. *Cell*, 147(7), pp.1484–97.
- Treutlein, B. et al., 2012. Dynamic Architecture of a Minimal RNA Polymerase II Open Promoter Complex. *Molecular Cell*, 46(2), pp.136–146.
- Trumbly, R.J., 1992. Glucose repression in the yeast *Saccharomyces cerevisiae*. *Mol Microbiol*, 6(1), pp.15–21.
- Tsuboi, T. et al., 2012. Dom34: Hbs1 Plays a General Role in Quality-Control Systems by Dissociation of a Stalled Ribosome at the 3' End of Aberrant mRNA. *Molecular Cell*, 46(4), pp.518–529.
- Tu, B.P. et al., 2005. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science*, 310(5751), pp.1152–1158.
- Tuck, A.C. & Tollervey, D., 2013. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell*, 154(5), pp.996–1009.
- Tucker, M. et al., 2001. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell*, 104(3), pp.377–386.
- Tudek, A. et al., 2014. Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Molecular Cell*, 55(3), pp.467–481.
- Ulbricht, R.J. & Olivas, W.M., 2008. Puf1p acts in combination with other yeast Puf proteins to control mRNA stability. *RNA*, 14(2), pp.246–262.

- Vad, R. et al., 1998. High-level production of human parathyroid hormone (hPTH) by induced expression in *Saccharomyces cerevisiae*. *Protein Expr Purif*, 13(3), pp.396–402.
- Vanáčová, S. et al., 2005. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS biology*, 3(6), p.e189.
- Vanacova, S. & Stefl, R., 2007. The exosome and RNA quality control in the nucleus. *EMBO reports*, 8(7), pp.651–7.
- Vasiljeva, L., Kim, M., Mutschler, H., et al., 2008. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol*, 15(8), pp.795–804.
- Vasiljeva, L., Kim, M., Terzi, N., et al., 2008. Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. *Mol Cell*, 29(3), pp.313–323.
- Vasudevan, S. & Peltz, S.W., 2016. Regulated ARE-Mediated mRNA Decay in *Saccharomyces cerevisiae*. *Molecular Cell*, 7(6), pp.1191–1200.
- Venkatesh, S. & Workman, J.L., 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews. Molecular Cell Biology*, 16(3), pp.178–189.
- Venters, B.J. & Pugh, B.F., 2009. A canonical promoter organization of the transcription machinery and its regulators in the *Saccharomyces genome*. *Genome research*, 19(3), pp.360–71.
- De Vit, M.J., Waddle, J.A. & Johnston, M., 1997. Regulated nuclear translocation of the Mig1 glucose repressor. *Mol Biol Cell*, 8(8), pp.1603–1618.
- Vodala, S., Abruzzi, K.C. & Rosbash, M., 2008. The nuclear exosome and adenylation regulate posttranscriptional tethering of yeast *GAL* genes to the nuclear periphery. *Mol Cell*, 31(1), pp.104–113.
- Voss, T.C. et al., 2011. Dynamic Exchange at Regulatory Elements during Chromatin Remodeling Underlies Assisted Loading Mechanism. *Cell*, 146(4), pp.544–554.
- Walter, P.P. et al., 1995. Stimulation of transcription factor binding and histone displacement by nucleosome assembly protein 1 and nucleoplasmin requires disruption of the histone octamer. *Mol Cell Biol*, 15(11), pp.6178–6187.
- Wang, W. et al., 2001. The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. *The EMBO Journal*, 20(4), pp.880–890.

- Webb, S. et al., 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.*, 15.
- Weinhandl, K. et al., 2014. Carbon source dependent promoters in yeasts. *Microbial cell factories*, 13, p.5.
- Wery, M. et al., 2016. Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. *Molecular Cell*, 61(3), pp.379–392.
- West S, Gromak N, Proudfoot NJ. 2004. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016), pp.522-525.
- Whipple, J.M. et al., 2011. The yeast rapid tRNA decay pathway primarily monitors the structural integrity of the acceptor and T-stems of mature tRNA. *Genes Dev*, 25(11), pp.1173–1184.
- Williams, S.K., Truong, D. & Tyler, J.K., 2008. Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc Natl Acad Sci U S A*, 105(26), pp.9000–9005.
- de Winde, J.H. & Grivell, L.A., 1992. Global regulation of mitochondrial biogenesis in *Saccharomyces cerevisiae*: ABF1 and CPF1 play opposite roles in regulating expression of the QCR8 gene, which encodes subunit VIII of the mitochondrial ubiquinol-cytochrome c oxidoreductase. *Mol Cell Biol*, 12(6), pp.2872–2883.
- Wu, D. et al., 2014. Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. *Cell research*, 24(2), pp.233–46.
- Wurm, J.A.N.P., Overbeck, J.A.N. & Sprangers, R., 2016. The *S. pombe* mRNA decapping complex recruits cofactors and an Edc1-like activator through a single dynamic surface. *Rna*, 22(9), 1360-1372.
- Wyers, F. et al., 2005. Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. *Cell*, 121(5), pp.725–737.
- Xiang, S. et al., 2009. Structure and function of the 5'-3' exoribonuclease Rat1 and its activating partner Rai1. *Nature*, 458(7239), pp.784–788.
- Xie, J. et al., 1999. Sum1 and Hst1 repress middle sporulation-specific gene expression during mitosis in *Saccharomyces cerevisiae*. *The EMBO Journal*, 18(22), pp.6448–6454.
- Xu, Z. et al., 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 457(7232), pp.1033–1037.

- Yarragudi, A. et al., 2004. Comparison of ABF1 and RAP1 in Chromatin Opening and Transactivator Potentiation in the Budding Yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, 24(20), pp.9152–9164.
- Yen, K., Vinayachandran, V. & Pugh, B.F., 2013. SWR-C and INO80 Chromatin Remodelers Recognize Nucleosome-free Regions Near +1 Nucleosomes. *Cell*, 154(6), pp.1246–1256.
- Young, E.T. et al., 2012. The AMP-activated protein kinase Snf1 regulates transcription factor binding, RNA polymerase II activity, and mRNA stability of glucose-repressed genes in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 287(34), pp.29021–29034.
- Zabolotskaya, M. V et al., 2008. The 5'-3' exoribonuclease Pacman is required for normal male fertility and is dynamically localized in cytoplasmic particles in *Drosophila* testis cells. *Biochem J*, 416(3), pp.327–335.
- Zacharioudakis, I., Gligoris, T. & Tzamarias, D., 2007. A yeast catabolic enzyme controls transcriptional memory. *Curr Biol*, 17(23), pp.2041–2046.
- Zhang, K. et al., 2002. The human homolog of yeast SEP1 is a novel candidate tumor suppressor gene in osteogenic sarcoma. *Gene*, 298(2), pp.121–127.
- Zhou, Q., Li, T. & Price, D.H., 2012. RNA polymerase II elongation control. *Annual review of biochemistry*, 81, pp.119–43.