

Dissertation zur Erlangung des Doktorgrades  
der Fakultät für Chemie und Pharmazie  
der Ludwig-Maximilians-Universität München

Transcriptome surveillance in *S.  
cerevisiae* by mRNA synthesis and  
degradation coupling and selective  
termination of non-coding RNAs



Daniel Schulz  
aus  
Simmern, Deutschland

2013



## **Erklärung**

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Patrick Cramer betreut.

## **Eidesstattliche Versicherung**

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 11.09.2013

Daniel Schulz

Dissertation eingereicht am 13.09.2013

1. Gutachter: Prof. Dr. Patrick Cramer
2. Gutachter: PD Dr. Dietmar Martin

Mündliche Prüfung am 09.10.2013

## Acknowledgments

In the first place, I want to express my gratefulness to my dear parents Peter and Eva. Not only have they supported me throughout my whole life, but they have raised me to be the curious person I am. Questioning, not taking predefined opinions and statements for granted, being spontaneous and open minded and last but not least, being optimistic. I want to thank you for letting me make my own decisions early on and making my own experiences in life. I cannot be thankful enough for what you have given to me in life. This is why this thesis is also dedicated to you.

I am also very thankful to my supervisor, Patrick Cramer. He is undoubtedly an outstanding person and a great leader. It felt good working with him from the first day on and to see him fight through his massive daily workload, still smiling and having an open ear when dropping into his room. He is truly inspiring. This positive attitude is transferred to the Cramer lab and is reflected in good vibes and team spirit. I am thankful, that I was given the opportunity to do my thesis at this institute in the lab of Patrick Cramer.

I also want to thank the other members of my thesis committee: Dr. Dietmar Martin, Prof. Dr. Klaus Förstermann, Prof. Dr. Achim Tresch, Dr. Katja Strässer and Prof. Dr. Karl-Peter Hopfner for their support and time.

Also, I want to specifically thank some people from the gene center, with whom I have worked with closely over the past years. I want to thank Andreas Mayer, who introduced me into his world of ChIP-Chip. I want to thank Kerstin Maier for being helpful in the lab and being a positive, happy person. I want to thank Johannes Soeding and Julien Gagneur for working together with me and allowing my "quiet nature" around their offices. I want to thank Achim Tresch for being an unusual professor, whom I like to work with a lot. I also want to thank Michael Lidschreiber who has always been helpful with bioinformatics as well as Stefan Krebs for great sequencing advice and help and Alexander Graf for help with deep sequencing data handling. Finally I want to thank Mai Sun for working together with me and sharing the great cDTA project with me and Björn Schwalb. I also want to express my gratefulness to Björn Schwalb aka Dr. Poony with whom I have not only enjoyed working very much. Last but not least, I want to thank my great colleagues Christoph and Carlo for discussions, nights at the computers and "Safran Sosse Ole".

I do also want to thank all my friends in Munich, who have made these four years unforgettable. Especially to all my past and present flat mates, with Boris being the constant friend whom I share a lot with in life.

## Abstract

Eukaryotic gene transcription is highly complex and regulation occurs at multiple stages. RNA Polymerase II (Pol II) is recruited to promoter regions of the DNA to initiate transcription. Shortly after initiation, Pol II exchanges initiation factors for elongation factors. After Pol II passes termination signals, the RNA is cleaved and Pol II eventually released from the DNA template. pre-mRNAs are polyadenylated and exported to the cytosol for translation and ultimately degradation. Mechanisms regulating transcription have been studied extensively, but mechanisms of mRNA degradation are less well understood. To monitor mRNA synthesis and degradation, we developed the comparative dynamic transcriptome analysis (cDTA). cDTA provides absolute rates of mRNA synthesis and decay in *Saccharomyces cerevisiae* *Sc* cells with the use of *Schizosaccharomyces pombe* *Sp* cells as internal standard. We show that *Sc* mutants can buffer mRNA levels and that impaired transcription causes decreased mRNA synthesis rates compensated by decreased decay rates. Conversely, impairing mRNA degradation causes decreased decay rates, but also decreased synthesis rates. Thus, although separated by the nuclear membrane, transcription and mRNA degradation are coupled.

In addition to regulated mRNA synthesis, pervasive transcription can be found throughout the genome, governed by an intrinsic affinity of Pol II for DNA. These divergent non-coding RNAs (ncRNAs) stem to a large extent from bidirectional promoters. However, global mechanisms for the termination of ncRNA synthesis that could act as a transcriptome surveillance mechanism are not known. It is also unclear if such a surveillance system protects the transcriptome from deregulation. Here we show that ncRNA transcription in *Sc* is globally restricted by early termination which relies on the essential RNA-binding factor Nrd1. Depletion from the nucleus results in Nrd1-terminated transcripts (NUTs) that originate from nucleosome-depleted regions (NDRs) throughout the genome and can deregulate mRNA synthesis by antisense repression and transcription interference. Transcriptome-wide Nrd1-binding maps reveal divergent NUTs at essentially all promoters and antisense NUTs in most 3'-regions of genes. Nrd1 preferentially binds RNA motifs which are enriched in ncRNAs and depleted in mRNAs except in some mRNAs whose synthesis is controlled by transcription attenuation. These results describe a mechanism for transcriptome surveillance that selectively terminates ncRNA synthesis to provide promoter directionality and prevent transcriptome deregulation.

## Publications

Part of this thesis has been published or is in the process of being published:

**2013 Transcriptome surveillance by selective termination of non-coding RNA synthesis**

**D. Schulz\***, B. Schwalb\*, A. Kieser, C. Baejen, P. Torkler, J. Gagneur, J. Soeding, P. Cramer

(\* joint first authorship)

Under revision

D.S. and P.C. conceived and designed the study. D.S. performed ChIP and 4-tU-Seq experiments. C.B. performed PAR-CLIP experiments. B.S., D.S., J.S. and J.G. designed data analysis. B.S., A.K., P.T., D.S. and C.B. carried out data analysis. D.S. and P.C. wrote manuscript with input from all authors. P.C. supervised the project

**2012 Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation**

M. Sun\*, B. Schwalb\*, **D. Schulz\***, N. Pirkl, S. Etzold, L. Larivière, K.C. Maier, M. Seizl, A. Tresch, P. Cramer

(\* joint first authorship)

*Genome Research*, 2012 22(7):1350-9

M. Sun, B.S., L.L., and P.C. conceived and designed the study. M. Sun and D.S. implemented and evaluated experimental methods. S.E. und M. Seizl conducted the rpb1-1 experiments. B.S. and A.T. designed and carried out computational analyses. M.S., D.S., N.P., K.M., S.E., and M. Seizl designed and performed experiments. P.C. and A.T wrote the manuscript, with input from all authors.

**2012 Measurement of genome-wide RNA synthesis and decay rates with Dynamic Transcriptome Analysis (DTA)**

B. Schwalb, **D. Schulz**, M. Sun, B. Zacher, S. Duemcke, D.E. Martin, P. Cramer, A. Tresch

*Bioinformatics*, 2012 28(6):884-5

**2010 Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast**

C. Miller\*, B. Schwalb\*, K.C. Maier\*, **D. Schulz**, S. Duemcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Doelken, D.E. Martin, A. Tresch, P. Cramer

(\* joint first authorship)

*Molecular Systems Biology* 7, 458

# Contents

Erklärung . . . . .	iii
Eidesstattliche Versicherung . . . . .	iii
Acknowledgements . . . . .	iv
Abstract . . . . .	v
Publications . . . . .	vi
<b>1 Introduction</b>	<b>1</b>
1.1 General Introduction . . . . .	1
1.1.1 Gene Transcription . . . . .	1
1.2 The RNA polymerase II (Pol II) transcription cycle . . . . .	2
1.2.1 Transcription initiation . . . . .	4
1.2.2 Transcription elongation . . . . .	4
1.2.3 Transcription termination . . . . .	5
1.3 Specific Introduction . . . . .	6
1.3.1 Systems Biology . . . . .	6
1.3.2 Transcriptomics . . . . .	10
1.3.3 Pervasive transcription in Yeast . . . . .	11
1.3.4 Nrd1-dependent transcription termination in yeast . . . . .	12
1.3.5 Aims and scope of this thesis . . . . .	13
<b>2 Material and Methods</b>	<b>16</b>
2.1 Materials . . . . .	16
2.1.1 Yeast and bacterial strains . . . . .	16
2.1.2 List of Plasmids . . . . .	17
2.1.3 List of Primers . . . . .	17
2.1.4 List of Antibodies . . . . .	18
2.1.5 Growth media . . . . .	18
2.1.6 Buffers and Solutions . . . . .	18
2.2 Experimental methods . . . . .	20
2.2.1 Molecular cloning <i>E. coli</i> . . . . .	20

## CONTENTS

---

2.2.2	Molecular cloning of <i>Sc</i> . . . . .	21
2.2.3	Experimental methods for cDTA analysis . . . . .	24
2.2.4	Chromatin immunoprecipitation (ChIP) . . . . .	27
2.2.5	Deep sequencing data analysis . . . . .	29
2.2.6	PAR-CLIP . . . . .	32
<b>3</b>	<b>Results</b>	<b>33</b>
3.1	cDTA reveals a mutual feedback loop between mRNA transcription and degradation . . . . .	33
3.1.1	Establishment of cDTA . . . . .	33
3.1.2	Rate extraction from cDTA data . . . . .	34
3.1.3	cDTA supersedes conventional methods . . . . .	38
3.1.4	Comparison of mRNA metabolism in distant yeast species . . . . .	40
3.1.5	Impaired mRNA synthesis is compensated by decreased degradation	43
3.1.6	Impaired degradation is compensated by decreased synthesis . . . . .	45
3.1.7	A transcription inhibitor and degradation enhancer may buffer mRNA levels . . . . .	46
3.2	Transcriptome surveillance by selective termination of non-coding RNA synthesis . . . . .	48
3.2.1	Nrd1 nuclear localization is essential . . . . .	48
3.2.2	Nrd1 generally terminates ncRNA transcription . . . . .	48
3.2.3	NUTs are extended ncRNA transcripts . . . . .	51
3.2.4	NUTs originate from distinct PICs in NDRs . . . . .	51
3.2.5	Many NUTs are divergent and antisense transcripts . . . . .	53
3.2.6	Nrd1 and Nab3 preferentially bind divergent and antisense ncRNAs	53
3.2.7	Nrd1-binding RNA motifs are depleted in mRNA . . . . .	54
3.2.8	Yeast promoters are generally bidirectional . . . . .	57
3.2.9	Nrd1 is required for promoter directionality . . . . .	59
3.2.10	Antisense ncRNA synthesis can down-regulate transcription . . . . .	60
3.2.11	Upstream ncRNA synthesis can up-regulate transcription . . . . .	62
3.2.12	Termination of ncRNA synthesis prevents transcription interference	62
3.2.13	Transcription attenuation is rare . . . . .	68
<b>4</b>	<b>Discussion</b>	<b>71</b>
4.1	cDTA reveals a mutual feedback loop between mRNA transcription and degradation . . . . .	71
4.2	Transcriptome surveillance by selective termination of non-coding RNA synthesis . . . . .	72

## CONTENTS

---

4.3 Summary and future perspectives . . . . .	75
<b>References</b>	<b>100</b>
<b>Abbreviations</b>	<b>101</b>

# Chapter 1

## Introduction

### 1.1 General Introduction

#### 1.1.1 Gene Transcription

Gene transcription is a fundamental process in all living beings which enables the genetic information to flow from DNA to RNA to protein. This concept is known as the Central Dogma [CRICK, 1958]. Although, in specialized cases the information flow can be reversed [Temin, 1985, Koonin, 2012], the dogma holds. Its first step - gene transcription - is still one of the most studied processes in science. Transcription is the process of RNA production from a DNA template by RNA polymerases and has first been described in 1959 [Weiss and Gladstone, 1959]. After the discovery of bacterial RNA polymerases in 1960 [Hurwitz, 2005], the first eukaryotic RNA polymerases to be discovered were the multi-subunit enzymes called RNA polymerase I, II and III [Roeder and Rutter, 1969]. Shortly after their discovery their distinct functions were also described: RNA polymerase I transcribes the large ribosomal RNA (rRNA) precursor, RNA polymerase II (Pol II) transcribes all mRNA and RNA polymerase III transcribes transfer-RNAs (t-RNAs) and several small rRNAs [Zylber and Penman, 1971, Weinmann and Roeder, 1974]. All three polymerases share a structurally conserved ten subunit core with additional subunits located on the periphery [Cramer et al., 2008]. RNA polymerase II is a 514 kDa enzyme consisting of 12 subunits [Cramer et al., 2001]. Since its first description many additional classes of RNAs that are transcribed by Pol II have been described. Among them are small nuclear RNAs (snRNAs) [Cramer et al., 2008], small nucleolar RNAs (snoRNAs) [Maxwell and Fournier, 1995] as well as a plethora of functional and non-functional cryptic RNAs [Xu et al., 2009, Neil et al., 2009, van Dijk et al., 2011, Guil and Esteller,

2012]. Additionally, most eukaryotes have evolved a pathway of small silencing RNAs of which a number depend on Pol II transcription [Ghildiyal and Zamore, 2009].

In 1986 it was discovered that a small part of the mitochondrial proteome is transcribed in the mitochondria themselves by a distinct polymerase [Greenleaf et al., 1986, Kelly and Lehman, 1986, Pikaard et al., 2008]. This mitochondrial RNA polymerase (mitoPol) is a single subunit enzyme that is distantly related to the T7 bacteriophage RNA polymerase [Ringel et al., 2011]. In contrast to bacteriophage T7 polymerase, the mitochondrial transcription system requires two additional transcription factors to assist promoter dependent transcription [Asin-Cayuela and Gustafsson, 2007]. Transcription in all eukaryotes is carried out by the three RNA polymerases I, II, III and the mitochondrial RNA polymerase. In plants however, two additional, non-essential RNA polymerases have been identified in recent years [Pontier et al., 2005, Herr et al., 2005, Onodera et al., 2005]. These multi-subunit enzymes are specialized in small RNA mediated gene silencing pathways and evidence suggests that they have evolved as specialized forms of Pol II [Haag and Pikaard, 2011].

## 1.2 The RNA polymerase II (Pol II) transcription cycle

Transcription of all messenger RNAs (mRNAs) in eukaryotes is carried out by Pol II. The act of mRNA transcription can be divided into three phases which constitute the transcription cycle: Initiation, elongation, termination (Figure 1.1). Although Pol I, II and III share five of the ten core-subunits [Cramer et al., 2008] and generally show a very similar subunit configuration [Engel et al. 2013 accepted], a major difference is the large C-terminal domain (CTD) on Rpb1 of Pol II. The CTD primarily acts as a large landing platform for factors involved in the regulation of the transcription cycle [Egloff and Murphy, 2008, Buratowski, 2009, Nechaev and Adelman, 2011, Venters and Pugh, 2009, Perales and Bentley, 2009, Zhang et al., 2012b]. In yeast this domain consists of a variable number of highly conserved hepta-repeats:  $Y_1S_2P_3T_4S_5P_6S_7$  [Corden et al., 1985]. The number of hepta-repeats varies between organisms with 26 repeats in *Saccharomyces cerevisiae* (*Sc*) and 52 in humans. The unique property of the CTD to interact with a wide range of transcription associated factors stems from its structure and the possibilities of post-translational modifications. Serine 2, 5 and 7 as well as threonine 4 and tyrosine 1 can be phosphorylated, threonine and serines can be glycosylated and proline can be isomerized [Egloff and Murphy, 2008, Fuchs et al., 2009, Zeidan and Hart, 2010, Mayer et al., 2012]. All these different modifications dictate the coordinated binding of different proteins at different time points during transcription [Zhang et al., 2012b, Mischo and Proudfoot, 2013]. It is

the dynamic interplay of factors with Pol II that couple and coordinate the different steps during transcription and help Pol II progress through the transcription cycle.

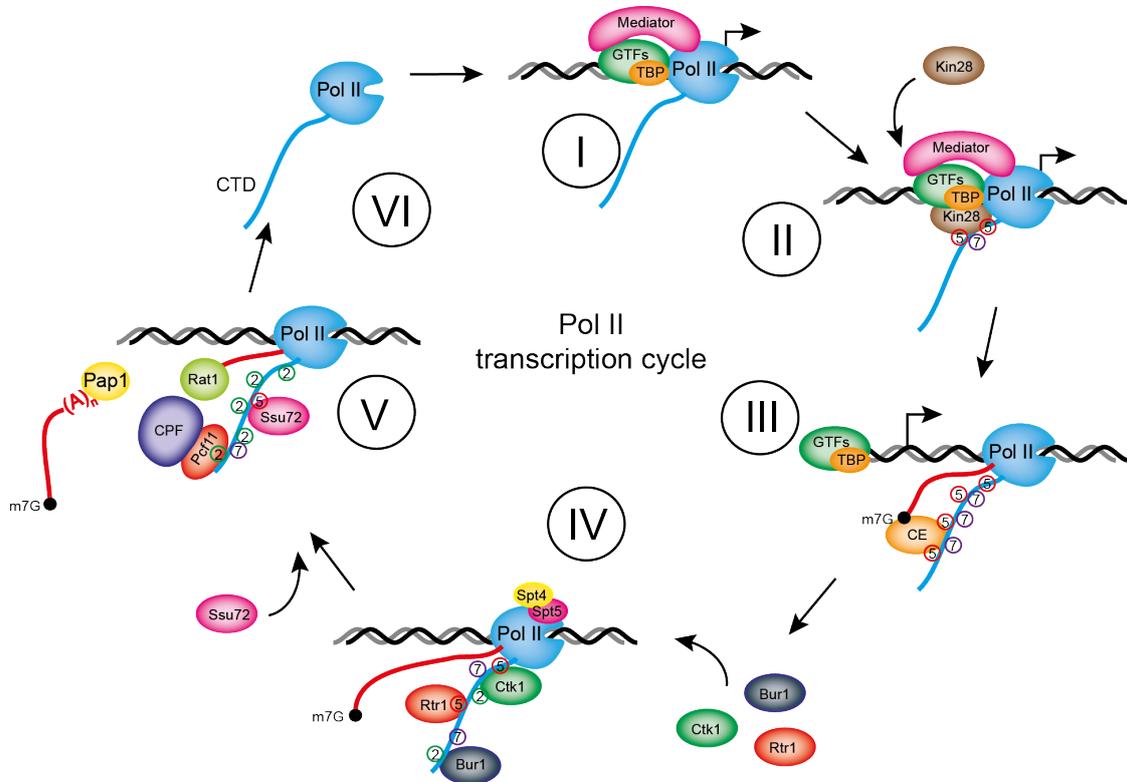


Figure 1.1: Pol II transcription cycle. (I) Pol II is recruited to the DNA template either through sequential assembly on the promoter or through binding of the transcription competent holoenzyme complex. Both pathways lead to the formation of a PIC. (II) As one of the last subunits Kin28 of the general transcription factor TFIIF is recruited and phosphorylates the CTD at Ser-5 and Ser-7 residues. This early mark of transcription allows for promoter clearance and association of the capping enzyme complex while GTFs (III) and TBP may stay associated with the DNA for initiation of additional rounds of transcription. (IV) Ser-5 phosphorylation attracts the kinases Bur1 and Ctk1 (which phosphorylate Ser-2 of the CTD), as well as Rtr1 which dephosphorylates Ser-5. Elongation factors Spt4/5 bind Pol II and render the Pol II elongation complex stable and processive. (V) Towards the 3' end of genes Ssu72 binds to the CTD and dephosphorylates Ser-5 residues leaving mostly Ser-2 phosphorylation marks on the CTD behind. Finally, the cleavage and polyadenylation machinery is recruited to the CTD (in part by Pcf11), the RNA is cleaved and polyadenylated at its 3' end, followed by release of Pol II from the template. After Pol II release from the template the free polymerase can initiate another round of transcription (VI).

### 1.2.1 Transcription initiation

Transcription initiation starts with the recruitment of Pol II to the gene promoter. Two separate models exist for this earliest step in transcription: The sequential assembly pathway and the Pol II holoenzyme pathway [Thomas and Chiang, 2006]. Both pathways do eventually lead to the assembly of a preinitiation complex (PIC) [Buratowski, 2009] and start with the recognition of upstream sequences by transcription factors. In the sequential model, general transcription factors (GTFs) assemble stepwise: TFIID first recognizes the TATA box via TBP which is followed by TFIIA, TFIIB and Mediator binding [Buratowski et al., 1989, Thomas and Chiang, 2006, Svejstrup et al., 1997]. Mediator is a large multi-subunit complex consisting of up to 31 subunits in humans [Bourbon et al., 2004]. The Mediator complex functions as a coactivator and connects transcription factors bound at regulatory elements with Pol II and the PIC [Kornberg, 2005, Malik and Roeder, 2005]. Subsequently, Mediator delivers Pol II to the emerging PIC, followed by TFIIF, then TFIIE and finally by TFIIH [Maxon et al., 1994, Thomas and Chiang, 2006]. The holoenzyme model was proposed after it was discovered that Pol II tightly interacts with subunits of the Mediator complex, specifically the Mediator head subunits and other GTFs in the absence of DNA [Koleske and Young, 1994]. The pre-assembled, transcription competent holoenzyme complex is then recruited to the DNA via TBP to activate transcription [Zhang et al., 2012b]. PIC assembly is also assisted and accompanied by recruitment of nucleosome remodelers like ISWI and SWI/SNF, chromatin modifying enzymes like SAGA, NuA4 and protein complexes like INO80/SWR1 that deposit histone variants [Venters and Pugh, 2009]. Subsequently TFIIH unwinds the DNA and thus helps in the formation of an open complex [Gruenberg et al., 2012, Hahn, 2004]. The CTD kinase subunit of TFIIH phosphorylates S5 of the CTD, an early mark of transcription which decreases as Pol II advances along the gene [Schroeder et al., 2000, Komarnitsky et al., 2000].

### 1.2.2 Transcription elongation

After the formation of the open complex Pol II starts with the initial production of RNA. In eukaryotes Pol II and TFIIB dictate the distance between the transcription start site to the TATA element [Li et al., 1994] and in *Sc* it was suggested that Pol II scans the DNA for the initiator element which is located 40-120 bp downstream of the TATA element [Kuehner and Brow, 2006, Giardina and Lis, 1993, Kostrewa et al., 2009]. An initial transcribing complex (ITC) is formed and a short RNA up to position +8 is synthesized, but at this stage transcription often is abortive [Holstege et al., 1997, Hahn, 2004]. After 12-13 bp

of transcription the nascent RNA clashes with TFIIB, which is then released, enabling Pol II to form a stable elongation complex (EC) [Sainsbury et al., 2013]. Meanwhile, S5 phosphorylation leads to Set1 recruitment which trimethylates histone H3 at lysine 4 (H3K4) and SAGA leaves marks of active transcription through acetylation of N-terminal tails of histones H3 and H4 [Kuo et al., 1996, Venters and Pugh, 2009, Ng et al., 2003]. Importantly, S5 phosphorylation also recruits the Ceg1 and Abd1 subunits of the capping enzyme complex [Schroeder et al., 2000, Cho et al., 1997] as well as Nrd1 of the Nrd1-Nab3-Sen1 (NNS) complex [Vasiljeva et al., 2008, Kubicek et al., 2012]. Capping of the nascent RNA has also been hypothesized to be an early checkpoint for transcription to ensure synthesis of mRNAs that have received a cap [Mandal et al., 2004, Kim et al., 2004a]. Hence, these studies suggest that capping enzyme sets up an early check-point for transcription elongation. If not passed, it leads to termination of Pol II transcription and degradation of the nascent RNA by the nuclear exosome [Buratowski, 2009, Wei et al., 2011, Seila et al., 2009]. In higher eukaryotes preassembled PICs can be found at inactive genes and Pol II often pauses shortly after transcription initiation until further signals release it into productive elongation [Margaritis and Holstege, 2008]. However, once the initial check-points have been overcome, initiation factors are exchanged with elongation factors like Spt4/5/6, as well as CTD kinases and phosphatases Bur1, Ctk1, Rtr1 and Ssu72 [Mayer et al., 2010, Zhang et al., 2012a, Martinez-Rucobo et al., 2011]. The transcription elongation complex helps Pol II to transcribe through the chromatin template and aids in processivity [Wada et al., 1998]. It is very stable and is proposed to be of identical composition at all genes [Mayer et al., 2010].

### 1.2.3 Transcription termination

While transcription initiation and elongation have been well studied in the past, termination by Pol II has only moved into the scientific focus more recently. Transcription termination is crucial and principally consists of two steps: (1) RNA cleavage and polyadenylation followed by (2) release of Pol II from the DNA template. Cleavage and polyadenylation occurs at the poly(A) signal (pAS), a highly conserved hexameric motif in humans (AAUAAA) [Proudfoot, 2011] which is less conserved in *Sc* [Ozsolak et al., 2010, Mischo and Proudfoot, 2013]. The pAS is recognized and bound by two highly conserved multiprotein complexes, the cleavage and polyadenylation factor (CPF) and cleavage factor 1A and B (CFIA/CFIB) [Mandel et al., 2008]. CFIA contains the proteins Rna14, Rna15, Clp1 and Pcf11 which are all essential for transcription termination to occur [Amrani et al., 1997, Minvielle-Sebastia et al., 1998, Minvielle-Sebastia and Keller, 1999]. Rna14 and Rna15 form heterodimers associated with Clp1 and Pcf11, which in turn contact the

CPF complex [Mischo and Proudfoot, 2013]. Endonucleolytic cleavage of the nascent RNA is promoted by Rna15 followed by polyadenylation by Pap1 which is part of the CPF complex [Birse et al., 1998]. Among many proteins of the termination machinery that interact with the CTD of Pol II [Zhang et al., 2012b], Pcf11 is most prominent and preferentially binds the Ser2-phosphorylated CTD [Zhang et al., 2005, Sadowski et al., 2003, Meinhart and Cramer, 2004]. It has recently been shown, that Y1-phosphorylation of the CTD impairs the recruitment of Pcf11 to the CTD during elongation [Mayer et al., 2012]. However, Y1-phosphorylation decreases just up-stream of the pAS and thus allows for Pcf11 binding and transcription termination.

Release of Pol II from the DNA template after cleavage has been proposed to follow two mechanisms: the "torpedo model" or the "allosteric model" [Connelly and Manley, 1988, Logan et al., 1987]. The 5'-3' exonuclease Rat1 (Xrn2 in human) has been shown to carry out functions similar to the Rho factor in bacteria [Kim et al., 2004b, West et al., 2004]. Rat1 preferentially binds the unphosphorylated RNA 5' ends and degrades the nascent RNA until it "catches" Pol II and promotes release. However, Rat1 alone is not sufficient to terminate Pol II [Kim et al., 2004b] and has been proposed to interact with the RNA-helicase Sen1 to efficiently promote termination [Kawauchi et al., 2008]. The allosteric model proposes that transcription through the pAS induces an exchange of elongation factors for termination factors [Logan et al., 1987]. Termination of Pol II could also be aided by decreased transcription rates after the pAS [Yonaha and Proudfoot, 1999].

Additionally, yeast uses another termination pathway required for termination of non-coding RNAs like sn/snoRNAs, stable unannotated transcripts (SUTs), cryptic unstable transcripts (CUTs). This pathway will be discussed in detail in section 1.3.3

## 1.3 Specific Introduction

### 1.3.1 Systems Biology

Modern biological research has acquired a new discipline along with classical studies which focus on one individual gene or protein at a time - Systems Biology [Snyder and Gallagher, 2009]. Instead of looking at single events in an organism, this discipline tries to explain phenomena globally, systematically and simultaneously. To this end, different types of datasets can be combined and correlated to obtain system-wide information.

Most of today's large scale approaches are based on the knowledge of the genetic code of the respective organism that is being studied. Deciphering any genetic code was in principle made possible by two technologies developed in the 1980s that allowed for the de-

termination of DNA sequence which are Sanger and Maxam-Gilbert Sequencing [Sanger et al., 1977, Maxam and Gilbert, 1977]. With *Sc* being one of the first organisms for which the whole genome sequence was available [Goffeau et al., 1996] it soon became one of the best studied model organisms. The first method developed to study the expression of many genes in parallel were microarrays [Schena et al., 1995]. The underlying principle is simple and straightforward: Single stranded DNA oligonucleotides of known sequence (probes) are immobilized on a surface in spots containing many oligonucleotides of the same type. A single stranded, fragmented, fluorescently labeled DNA or cDNA sample is then hybridized to the oligonucleotides on the surface. Excitation of the hybridized probes allows for quantification of the fluorescent signal which is, in a certain range, proportional to the DNA/cDNA amount from the original sample. However, as apparent from the setup, microarray studies require pre-existing knowledge about the DNA sequence of the organism that is being studied, in order to design the respective probes. Hence, more advanced techniques were developed.

### 1.3.1.1 Second generation sequencing (NGS)

A revolutionary breakthrough for the high-throughput era was achieved by the development of next-generation DNA sequencing platforms. Next generation sequencing led to an increase in throughput by approximately five orders of magnitude compared to Sanger sequencing accompanied by an almost equal decrease in sequencing costs per base [Liu et al., 2012]. Unlike microarray technologies, sequencing platforms do not require any knowledge about the DNA sequence of the organism that is studied which allows scientists to study new organisms with reasonable efforts. After years of development 454 Life Sciences introduced the 454 GenomeSequencer instrument (later GS FLX, Roche Applied Bioscience). This platform combines pyrophosphate sequencing [Ronaghi et al., 1996] with the possibility to amplify single DNA molecules in microliter reactors in an emulsion PCR [Dressman et al., 2003]. Thereby it became possible to sequence 250 bp (700 bp today) of 400,000 DNA fragments (approximately 1 million today) in parallel, resulting in a total of 100 megabases (Mb) of DNA sequenced (700 Mb today). The relatively long read length is the unique selling point of this platform although competitive methods applying single molecule sequencing are gaining ground (Pacific Biosciences and Oxford Nanopores).

The to date by far most applied next-generation sequencing platform is the Illumina Genome Analyzer and now the HiSeq system. The principle is based on the sequencing-by-synthesis technology where each nucleotide is labeled with a different fluorescent dye. Adapter-ligated fragments are immobilized on a surface coated with oligonucleotides

complementary to the adapter sequences. Single strand DNA molecules form a "bridge" by hybridizing with both ends. A "bridge-PCR" is performed in order to amplify the randomly distributed fragments on the surface until clusters of about 1000 molecules of clonal DNA fragments have been synthesized [Adessi et al., 2000, Fedurco et al., 2006]. During the sequencing reaction high resolution CCD cameras are used to record the fluorescent intensity of each cluster upon incorporation of labeled nucleotides. From the X-Y coordinates of each cluster the sequence of the DNA cluster can be reconstituted from the pictures that were taken throughout. Lately, this technology is able to parallel sequence up to 6 billion DNA fragments with a length of 100 bp from each side of the fragment. One sequencing run can therefore produce a total of 600 gigabases (Gb) of output which is 200 times the size of the human genome. With these numbers in mind, it becomes clear why biology has entered the "high-throughput" era and why computational biologists as well as computer scientists establishing and maintaining data structures are required to handle the ever growing, massive amounts of data being produced. Other technologies comprise the Ion Proton<sup>TM</sup> sequencer from Life Technologies. The sequencer makes use of the release of a proton upon nucleotide incorporation during the sequencing reaction which leads to a change in pH detected through a semiconductor chip. The sequencing quality is generally very stable and does not decrease for long reads, as it does in fluorescently based approaches. However, this technology is limited in throughput due to the size of the reaction wells in the semiconductor chip, which determine the amount of parallel sequencing reactions.

### 1.3.1.2 Third generation sequencing

Any of the described sequencing reactions require prior PCR amplification of the DNA fragments in order to create a specific signal strong enough for detection. Despite all the positive aspects of PCR it can also introduce biases through unequal amplification of templates in complex samples [Polz and Cavanaugh, 1998]. To overcome these problems intensive efforts have been made over past decades to develop single molecule sequencers. The platforms using these technologies are referred to as "third generation sequencers" because they differ with respect to PCR amplification and real-time detection of the sequencing signal. Currently only the Single-molecule real-time (SMRT) technology is commercially used by Pacific Biosciences and promising results have been obtained by Oxford Nanopores. SMRT uses specialized chips containing extremely small reaction containers called "Zero mode waveguides" (ZMWs). Each of these containers holds a single DNA polymerase immobilized at the bottom and illuminated with laser-light from beneath, creating a detection zone for fluorescent excitation surrounding the polymerase.

The ZMVs contain nucleotides, each labeled with a different fluorophore. Upon incorporation of a nucleotide by the DNA polymerase the fluorophore, is kept in the excitation range slightly longer than average, creating a specific detection signal which can be read out. During incorporation, the phosphate holding the fluorophore is cleaved, thus releasing the fluorophore and thereby decreasing the background noise. Pacific Biosciences can currently create reads of an average length greater than 3000 bp and approximately 30,000 reads per run. This technology is therefore ideally suited for *de novo* genome assemblies, targeted re-sequencing or even specialized applications like base modification detections. Finally, a technology that has been under development since the mid 90s is nanopore sequencing. A channel protein (nanopore) is incorporated into a membrane composed of synthetic polymers which give rise to a high electronic resistance. A potential can be applied across the membrane creating an ionic current running through the nanopore. Molecules passing the nanopore cause characteristic disruptions in the current that can be used to identify the molecule. For DNA sequencing single stranded DNA is passed through the nanopore creating a characteristic current disruption for every base that passes the pore. Oxford Nanopores has recently introduced the GridION<sup>TM</sup> system for commercial nanopore sequencing.

### 1.3.1.3 Systems biology applications

With the possibility to detect large quantities of DNA fragments simultaneously, a plethora of methods to analyze genome-wide characteristics have either been adapted or developed over the last years. In principle it is possible to analyze DNA of any kind and to study RNA, it can simply be reverse transcribed into cDNA. The most prominent genome-wide methods are RNA-Seq and ChIP-Seq which are used to study gene expression and protein-DNA interactions, respectively. In RNA-Seq, total RNA from a cell is extracted, reverse transcribed, fragmented and sequenced to obtain genome wide information on RNA abundances, splicing, transcript isoforms or non-coding RNAs. ChIP-Seq is used to sequence DNA fragments that have been bound by a protein to obtain information where this specific protein bound throughout the genome. More recently protein-RNA interactions have moved into the scientific focus and methods like Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) [Hafner et al., 2010] or individual-nucleotide resolution CLIP (iCLIP) [Konig et al., 2011b] which allow for detection of transcriptome wide protein-RNA interactions at nucleotide resolution. To gain insights into the 3D structure of chromatin in the nucleus methods have been developed to study genome-wide chromatin interactions [Lieberman-Aiden et al., 2009]. Additionally, ChIP-Exo has been developed to increase the resolution of ChIP-Seq experiments [Rhee

and Pugh, 2011] and methods to study transcription rates instead of steady-state levels of gene expression have also been developed [Doelken et al., 2008, Miller et al., 2011, Core et al., 2008, Churchman and Weissman, 2011, Rabani et al., 2011].

### 1.3.2 Transcriptomics

#### 1.3.2.1 mRNA metabolism

As mentioned under section 1.3.1.3, gene expression measurements usually reflect the steady-state level of a certain mRNA. These steady-state levels are determined by specific mRNA synthesis and degradation rates. While mRNA synthesis rates are governed by Pol II transcription in the nucleus [Fuda et al., 2009], bulk mRNA degradation occurs in the cytoplasm [Eulalio et al., 2007, Parker and Sheth, 2007, Wiederhold and Passmore, 2010]. Despite the spatial separation, evidence exists that these two processes are coordinated [Lotan et al., 2005, Lotan et al., 2007, Harel-Sharvit et al., 2010]. mRNA synthesis rates can be measured by nuclear-run on experiments [Garcia-Martinez et al., 2004] and degradation rates can be measured after blocking transcription with inhibitors [Shalem et al., 2008, Grigull et al., 2004, Lam et al., 2001] or using a temperature-sensitive yeast strain [Holstege et al., 1998]. These methods are however cell invasive and unperturbed measurements of synthesis and decay rates can be obtained via metabolic labeling of RNA and kinetic modeling [Doelken et al., 2008, Miller et al., 2011, Kenzelmann et al., 2007, Cleary et al., 2005]. This involves the incorporation of the nucleotide analog 4-thiouridinetriphosphate (4sUTP) into RNA during Pol II transcription [Melvin et al., 1978]. In yeast, 4-thiouridine (4sU) can be taken up through the expression of a human nucleoside transporter [Miller et al., 2011] and be converted into 4sUTP by the cell. Thiol-labeled, newly synthesized RNAs can be isolated via biotinylation and purification [Cleary et al., 2005, Doelken et al., 2008]. After separation of labeled RNAs from pre-existing RNAs, each fraction can be analyzed with microarrays and kinetic modeling can be used to calculate synthesis and decay rates for all mRNAs [Doelken et al., 2008, Miller et al., 2011]. This method was named Dynamic Transcriptome Analysis (DTA) in yeast and can be used to analyze dynamic changes in mRNA metabolism to study gene-regulatory networks [Miller et al., 2011].

#### 1.3.2.2 comparative Dynamic Transcriptome Analysis (cDTA)

In standard transcriptomics the assumption is made, that the global amount of RNA in a cell does not change [Loven et al., 2012]. In practice, RNA is extracted from similar

amounts of different cellular sources. Similar amounts of RNA are then processed further and analyzed via micro-arrays or RNA-Seq. The crucial information on how much RNA each cell actually contained is lost due to unknown cell lysis and RNA extraction efficiencies. Furthermore, signal intensities in microarray experiments are often normalized by centering the median of the expression profile to a common value [Bolstad et al., 2003]. This experimental setup does not allow for the detection of global RNA amounts in a cell because a proper external standard that could be used to calibrate intensities is lacking. External standards have previously been used to account for the normalization problem [Holstege et al., 1998, van de Peppel et al., 2003, Wang et al., 2002]. However, cell-lysis and RNA extraction efficiencies were not taken into account in these studies. To enable normalization between different DTA measurements of different samples in yeast we extended DTA to comparative DTA (cDTA). In cDTA, a defined number of labeled fission yeast *Schizosaccharomyces pombe* (*Sp*) cells is added to budding yeast *Sc* wild-type or perturbed samples before cell lysis and RNA preparation as an internal standard. Thereby, cDTA allows the absolute quantification and accurate comparison of mRNA synthesis and decay rates between samples.

### 1.3.3 Pervasive transcription in Yeast

Sequencing and microarray technologies revealed that the genomes of eukaryotic cells are pervasively transcribed. In human cells, about 74% of the genome gives rise to RNA transcripts, although only about 2% correspond to protein-coding mRNA genes [Djebali et al., 2012]. In the yeast *Sc*, 85% of the genome is transcribed [David et al., 2006], and hundreds of non-coding RNAs (ncRNAs) were discovered in addition to the classical 4 rRNAs, 42 tRNAs, 6 snRNAs, and 77 snoRNAs (*Saccharomyces* genome database) [Hani and Feldmann, 1998]. Pervasive transcription stems to a large extent from the poor directionality of RNA polymerase (Pol) II initiation [Core et al., 2008, Neil et al., 2009, Seila et al., 2008, Xu et al., 2009]. The existence of bi-directional transcription is supported by the observation of two adjacent pre-initiation complexes (PICs) in nucleosome-depleted regions (NDRs) of yeast [Murray et al., 2012, Rhee and Pugh, 2012].

Two mechanisms have been identified that restrict the extent of pervasive transcription in eukaryotes: Firstly, transcription initiation can be biased towards the mRNA direction by gene looping and preferred formation of PICs for mRNA transcription, thereby limiting initiation of divergent ncRNA transcription [Rhee and Pugh, 2012, Tan-Wong et al., 2012]. Secondly, ncRNAs are rapidly removed by RNA degradation [Xu et al., 2009, Neil et al., 2009]. In yeast, 925 ncRNAs called cryptic unstable transcripts (CUTs) are degraded from their 3' end by the exosome, and deletion of the nuclear exosome subunit

Rrp6 stabilizes these ncRNAs [Wyers et al., 2005, Xu et al., 2009]. Other studies even detected 1496 CUTs that we refer to as CUT\*s [Neil et al., 2009] and full inactivation of the exosome resulted in additional 1600 CUTs [Gudipati et al., 2012]. Degradation of ncRNAs also occurs from the 5' end, since deletion of the 5'-exonuclease Xrn1 stabilizes 1658 Xrn1-dependent unstable transcripts (XUTs) [van Dijk et al., 2011]. Thus, pervasive transcription can be controlled at the level of transcription initiation and by RNA degradation. It is unlikely though that these two mechanisms account for the elimination of all aberrant ncRNAs, because gene looping is not a global phenomenon and because the RNA degradation factors Rrp6 and Xrn1 are non-essential.

### 1.3.4 Nrd1-dependent transcription termination in yeast

Global ncRNA synthesis may, however, be restricted by selective termination of ncRNA [Ntini et al., 2013, Almada et al., 2013]. Termination of Pol II transcription in yeast occurs via two distinct pathways [Hsin and Manley, 2012, Kim et al., 2006, Mischo and Proudfoot, 2013]. Termination of mRNA genes requires the cleavage and polyadenylation factor which binds a polyadenylation signal (pA) in the nascent RNA (Section 1.2.3). In contrast, termination of snRNAs and snoRNAs depends on Nrd1, an essential protein that contains an RNA recognition motif (RRM) and interacts with Pol II via its CTD interaction domain (CID), preferentially with the serine-5 phosphorylated form of the CTD [Steinmetz and Brow, 1996, Vasiljeva et al., 2008] (Figure 1.2). Nrd1 binds a tetramer motif in the RNA transcript [Carroll et al., 2004, Creamer et al., 2011, Porrua et al., 2012, Wlotzka et al., 2011], and interacts with Nab3 and Sen1 to promote termination [Steinmetz et al., 2001]. The Nrd1-Nab3-Sen1 complex also interacts with cap binding proteins Cbc20 and Cbc80, as well as Trf4 of the TRAMP complex [Vasiljeva and Buratowski, 2006]. Transcription termination of several CUTs [Arigo et al., 2006b, Thiebaut et al., 2006] and a few SUTs depends on the Nrd1 pathway [Marquardt et al., 2011]. Nrd1 is also required for the removal of aberrant *Sc* transcripts that result from heterologous expression of a prokaryotic factor [Honorine et al., 2011].

Based on these results it was proposed that Nrd1-dependent termination can restrict transcription from bidirectional promoters to the sense direction by terminating divergent transcription and subjecting the divergent transcript to rapid degradation [Buratowski, 2009, Jacquier, 2009, Seila et al., 2009, Wei et al., 2011, Porrua and Libri, 2013b]. The existence of such a nuclear RNA surveillance mechanism is supported by *in vivo* RNA cross-linking of Nrd1 and Nab3 to CUTs [Wlotzka et al., 2011] and to RNA produced antisense of weakly expressed genes [Creamer et al., 2011]. Nrd1, however, also cross-

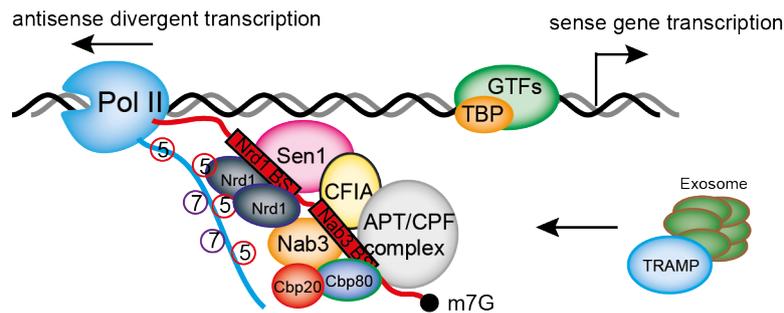


Figure 1.2: Model of Nrd1 dependent transcription termination of non-coding RNAs. Divergent transcription of Pol II is observed at many promoters in *Sc*. Nrd1 binds to Ser-5 phosphorylated forms of the CTD via its CID. Additionally, binding sites for Nrd1 and Nab3 are displayed on the nascent RNA, which are bound by Nrd1 and Nab3. The helicase Sen1 tightly interacts with the Nrd1-Nab3 complex. Additional proteins (the APT (Associated with Pta1)-subcomplex of the CPF complex, CFIA and cap binding proteins Cbc20 and Cbc80) are recruited [Mischo and Proudfoot, 2013]. After the APT-subcomplex and CFIA dependent cleavage of the RNA, Sen1 has been proposed to release Pol II from the template in a manner similar to that of Rho in bacteria [Porrua and Libri, 2013a]. Cleaved RNAs are polyadenylated by the non-canonical poly(A) polymerase Trf4 of the TRAMP complex and subsequently degraded via the nuclear exosome.

links to many mRNAs [Creamer et al., 2011, Wlotzka et al., 2011], and is recruited to mRNA genes according to chromatin immunoprecipitation (ChIP) [Mayer et al., 2012]. This raised the question whether Nrd1-dependent termination functions widely in the attenuation of mRNA transcription, as observed for mRNA genes *NRD1*, *HRP1*, and *IMD2* [Arigo et al., 2006a, Steinmetz et al., 2006b], *URA2*, *URA8*, and *ADE12* [Kuehner and Brow, 2008, Thiebaut et al., 2008], and *FKS2* ([Kim and Levin, 2011]).

### 1.3.5 Aims and scope of this thesis

This thesis was designed to answer two unrelated, generally unsolved questions:

I) What is the importance of Pol II elongation rate for mRNA synthesis and decay rates and are mRNA synthesis and mRNA decay coupled processes? It has been shown that the elongation rate can influence splicing activity in yeast [Howe et al., 2003] and that it also affects processivity of Pol II *in vivo* [Mason and Struhl, 2005]. Recently it has also been shown, that the elongation rate of Pol II influences transcription termination [Hazelbaker et al., 2013]. However, the importance of the elongation rate for the synthesis rate of mRNA has not been investigated. A genetic screen in the Kashlev Lab revealed a Pol II mutant that showed decreased elongation rates in *in vitro* transcription assays [Malagon et al., 2006]. This mutant carries a single point mutation near the active center and is therefore ideally suited to investigate the importance between elongation rate and mRNA

synthesis rate. Furthermore, although mRNA synthesis and degradation in eukaryotes occur in the nucleus and cytoplasm, respectively, evidence exists, that these processes can be coupled (Section 1.3.2.1).

Therefore, a protocol had to be established that allows for measurements of absolute synthesis and decay rates and detects global changes between different samples (Section 1.3.2.2). Synthesis and decay rates in *Sc* can be measured by DTA through metabolic RNA labeling and kinetic modeling [Miller et al., 2011]. To enable direct comparison of different DTA measurements and obtain information about absolute changes in mRNA metabolism, we extended DTA to cDTA. Absolute quantification and accurate comparison of mRNA synthesis and decay rates between samples is achieved in cDTA through addition of an internal standard. This was accomplished through addition of a defined number of *Sp* cells to the *Sc* samples before cell lysis and RNA preparation. We applied cDTA to *Sc* cells that are impaired in either mRNA synthesis (slow Pol II) or degradation. This revealed that Pol II elongation speed is critical for mRNA synthesis *in vivo* and furthermore, that the globally decreased synthesis rates were compensated by a global decrease in mRNA decay rates. Compensatory changes in mRNA synthesis rates were also observed for cells impaired in mRNA degradation which indicates that a eukaryote can buffer mRNA levels to render gene expression robust.

II) The discovery of pervasive genome transcription and ubiquitous ncRNA synthesis raised four questions related to the possibility that transcriptome fidelity is achieved by selective early termination of ncRNA synthesis and subsequent rapid RNA degradation. First, what is the origin of ncRNA transcription? Second, what is the global mechanism for ncRNA transcription termination? Third, does a failure to terminate ncRNA synthesis lead to transcriptome deregulation? Forth, how does the termination mechanism distinguish ncRNA synthesis from mRNA transcription? Answers to these questions are required to establish the concept of transcriptome surveillance.

This thesis elucidates the above raised questions through investigation of the global function of the Nrd1 dependent termination pathway in *Sc*. Nrd1 has been studied for almost two decades. Nrd1 localizes to the nucleus and the *NRD1* gene is essential and can therefore not be deleted. It is involved in transcription termination of sn/snoRNAs and some other ncRNAs, gene transcription regulation via attenuation, and it was proposed that Nrd1-dependend termination can suppress divergent ncRNA synthesis (Section 1.3.4). To investigate global Nrd1 functions we conditionally depleted Nrd1 from the nucleus using the anchor-away technique [Haruki et al., 2008] and monitored changes in RNA synthesis and Pol II occupancy. RNA synthesis was monitored through metabolic labeling of RNAs (Section 1.3.2.1, 1.3.2.2) followed by NGS (Section 1.3.1.1) and we referred to this pro-

tolocol as 4tU-Seq. Pol II occupancies were determined via ChIP-Seq using an antibody against Rpb3. Additionally, high quality Nrd1 and Nab3 RNA binding sites throughout the transcriptome were determined using PAR-CLIP.

We found that Nrd1 rarely attenuates mRNA transcription, but that it is responsible for the selective, global termination of ncRNA synthesis, including transcription antisense to known genes and divergent transcription from bidirectional promoters. Divergent ncRNAs were found at most promoters by a comprehensive mapping of Nrd1 and Nab3 onto the transcriptome. Nuclear depletion of Nrd1 resulted in aberrant transcription that deregulated the genome. Together with an analysis of RNA-binding motif occurrence, these results show that selective termination of ncRNA synthesis by the Nrd1 pathway acts as a global mechanism for transcriptome surveillance, providing transcription directionality and preventing transcriptome deregulation. Interestingly, similar findings have just recently been made in mammals [Ntini et al., 2013, Almada et al., 2013]. These studies found increased densities of pA motifs (AWTAAA) upstream of TSSs of mRNAs which serve to selectively terminate bidirectional transcription.

# Chapter 2

## Material and Methods

### 2.1 Materials

#### 2.1.1 Yeast and bacterial strains

Table 2.1: Yeast strains

Name	Genotype	Source
GRY3020	MATa, his3 $\Delta$ 1, leu2 $\Delta$ 0, lys2 $\Delta$ 0, met15 $\Delta$ 0, trp1 $\Delta$ ::hisG, URA3::CMV-tTA RPO21	Kashlev Lab
GRY3027	MATa, his3 $\Delta$ 1, leu2 $\Delta$ 0, lys2 $\Delta$ 0, met15 $\Delta$ 0, trp1 $\Delta$ ::hisG, URA3::CMV-tTA rpb1-N488D	Kashlev Lab
BY4741	MATa, his2 $\Delta$ 1, leu2 $\Delta$ 0, met15 $\Delta$ 0, ura3 $\Delta$ 0	Euroscarf
Y40343	W303 MAT $\alpha$ tor1-1 fpr1::NAT RPL13A-FKBP12 :: TRP1	Euroscarf
Nrd1AA	W303 MAT $\alpha$ tor1-1 fpr1::NAT RPL13A-FKBP12 :: TRP1 YNL251C::YNL251C-FRB-KanMX4	Generated in this study
Nrd1AA-GFP	W303 MAT $\alpha$ tor1-1 fpr1::NAT RPL13A-FKBP12 :: TRP1 YNL251C::YNL251C-FRB-GFP-KanMX4	Generated in this study
Nrd1-Tap	MATa, his2 $\Delta$ 1, leu2 $\Delta$ 0, met15 $\Delta$ 0, ura3 $\Delta$ 0 YNL251C::YNL251C::TAP::HIS3MX6	Euroscarf

Table 2.2: *E. coli* strains

XL-1 Blue	Rec1A; endA1; gyrA96; thi-1; hsdR17; supE44; relA1; lac(F' proAB lacIqZΔM15Tn10[Tet <sup>r</sup> ])
-----------	--

### 2.1.2 List of Plasmids

Table 2.3: Plasmids

Name	Insert	Selection Marker	Vector	Source
pYMS17	FRB-GFP	KanMX	pFA6a	Euroscarf
pYMS19	FRB	KanMX	pFA6a	Euroscarf

### 2.1.3 List of Primers

Table 2.4: Primers

ID	Name	Sequence	Length
207	Nrd1aa fw	ATTCTTTGATGAATATGCTTAACCAACAG CAGCAGCAACAACAACAAAGCCGGATC CCCGGGTTAATTAA	70
208	Nrd1aa rev	GGTAGATTAGTTTTATGTACTATGAGCAA ATAAAGGGTGGAGTAAAGATCGAATTC GAGCTCGTTTAAAC	70
209	Nrd1 ORF rev	GATGCCTACTGATTCTGGC	19
-	FRB ctrl rv2	GATGTTTCCTTCAGAGTCTGG	21
-	YER fw	TGCGTACAAAAAGTGTCAAGAGATT	25
-	YER rev	ATGCGCAAGAAGGTGCCTAT	20
-	ADH1 5' fw	TTTCCTTCCTTCATTCACGCACA	24
-	ADH1 5' rev	TCAAGTAACTGGAAGGAAGGCCGTA	25
-	ADH1 ORF fw	AGCCGCTCACATTCCTCAAG	20
-	ADH1 ORF rev	ACGGTGATAACCAGCACACAAGA	22
-	ADH1 3' fw	CCTGTAGGTCAGGTTGCTTT	20
-	ADH1 3' rev	CGGTAGAGGTGTGGTCAA	18
-	HSP12 ORF fw	CAAGGTCGCTGGTAAGGTTC	20
-	HSP12 ORF rev	AGAGTCGTGGACACCTTGGA	20

### 2.1.4 List of Antibodies

Table 2.5: Antibodies

Target	Host	Source	Type
Rpb3 (Pol II)	Mouse	NeoClone	monoclonal

### 2.1.5 Growth media

Table 2.6: Growth media

Medium	Description	Species
YPD	1% (w/v) yeast extract; 2% (w/v) peptone; 2% (w/v) glucose (additional 2% (w/v) agar for plates)	<i>Sc</i>
YES	0.5% (w/v) yeast extract; 3% (w/v) glucose (additional 2% (w/v) agar for plates)	<i>Sp</i>
LB	10g Bacto <sup>TM</sup> Tryptone; 5g Bacto <sup>TM</sup> Yeas Extract; 10g NaCl; add ddH <sub>2</sub> O to 1L	<i>E. coli</i>

### 2.1.6 Buffers and Solutions

Table 2.7: Buffers and Solutions

Name	Composition	Application
1 x PBS	2mM KH <sub>2</sub> PO <sub>4</sub> ; 4mM Na <sub>2</sub> HPO <sub>4</sub> ; 140 mM NaCl; 3 mM KCl; pH 7.4 (25°C)	
1 x TBS	20 mM TrisHCl, pH 7.5 at 4°C; 150 mM NaCl	ChIP
1 x TAE	4.84 g Tris; 1.14 ml glacial acetic acid; 0.37 g EDTA; add ddH <sub>2</sub> O to 1 L	
1 M HEPES	238.5 g HEPES dissolved in 800 ml ddH <sub>2</sub> O; adjust pH to 7.5 with KOH; add ddH <sub>2</sub> to 1 L	
6 x Loading Dye (Fermentas)	1.5 g/L Bromphenol blue; 1.5 g/L Xylene cyanol; 50% (v/v) Glycerol	

Continued on next page

Table 2.7: Buffers and Solutions

Name	Composition	Application
TFB-1 Buffer	30 mM KOAc; 50 mM MnCl <sub>2</sub> ; 100 mM RbCl; 10 mM CaCl <sub>2</sub> ; 15% (v/v) Glycerol; pH 5.8 at 25°C	<i>E. coli</i> competent cells
TFB-2 Buffer	10 mM MOPS, pH 7.0 at 25°C; 10 mM RbCl; 75 mM CaCl <sub>2</sub> ; 15% (v/v) Glycerol	<i>E. coli</i> competent cells
LiAc	5.1 g LiOAc; pH 7.5 at 25°C; add ddH <sub>2</sub> O to 1 L	<i>Sc</i> competent cells
PEG3350	25 g PEG 3350; add sterile ddH <sub>2</sub> to 1 L	<i>Sc</i> competent cells
WB transfer Buffer	25 mM Tris; 192 mM Glycine; 20% (v/v) Ethanol	Western blotting
WB blocking Buffer	2% (w/v) milk powder in 1 x PBS	Western blotting
Protease inhibitor (PI)	1 mM Leupeptin, 2 mM Pepstatin A, 100 mM Phenylmethylsulfonyluoride, 280 mM Benzamidine	ChIP
TE Buffer	10 mM TrisHCl, pH 7.4 at 4°C; 1 mM EDTA	ChIP
FA lysis Buffer	50 mM HEPESKOH, pH 7.5 at 4°C; 150 mM NaCl; 1 mM EDTA; 1% (v/v) Triton X100; 0.1% (v/v) Na deoxycholate; 0.1% (v/v) SDS; PI	ChIP
FA lysis high salt Buffer	FA lysis Buffer with 500 mM NaCl instead of 150 mM NaCl	ChIP
ChIP wash Buffer	10 mM TrisHCl, pH 8.0 at 4°C; 0.25 M LiCl; 1 mM EDTA; 0.5% (v/v); NP40; 0.5% (v/v) Na deoxycholate	ChIP
ChIP elution Buffer	50 mM TrisHCl, pH 7.5 at 25°C; 10 mM EDTA; 1% (v/v) SDS	ChIP
RNase free 10 x biotinylation Buffer	100 mM Tris pH 7.4; 10 mM EDTA	RNA biotinylation
Washing Buffer	100mM Tris pH7.5; 10mM EDTA; 1M NaCl; 0.1% Tween20	RNA biotinylation

Continued on next page

Table 2.7: Buffers and Solutions

<b>Name</b>	<b>Composition</b>	<b>Application</b>
Elution Buffer (DTT)	100 mM Dithiothreitol in H <sub>2</sub> O	RNA biotinylation
Biotin-HPDP (Thermo Scientific)	1mg/ml in dimethylformamide (DMF)	RNA biotinylation
Poly-L lysine (Sigma)	0.1% in H <sub>2</sub> O	

## 2.2 Experimental methods

### 2.2.1 Molecular cloning *E. coli*

#### 2.2.1.1 Chemically competent *E. coli*

Chemically competent *E. coli* cells for transformations were prepared as follows: Single colonies of *E. coli* were inoculated in LB medium and grown o.N. at 37°C. In the morning 200 ml of LB medium were inoculated from the o.N. culture and grown to an OD<sub>600</sub> of 0.4 to 0.5 and incubated on ice for 10 min. Cells were harvested by centrifugation at 4°C, washed with 50 ml TFB-1 Buffer, harvested by centrifugation again and resuspended in 4 ml TFB-2 Buffer. 50 µl Aliquots of competent cells were flash frozen in liquid nitrogen and stored at -80°C.

#### 2.2.1.2 *E. coli* transformation

Aliquots of chemically competent *E. coli* cells (Section 2.2.1.1) were used. 1 to 2 µg of DNA were added to cell aliquots on ice and incubated for 20 min. Cells were heated to 42°C for 40 s and transferred back on ice for 2 min. 250 µl of LB medium were added and cells incubated at 37°C for 1 h (not required for Ampicillin resistance). Cells were centrifuged and the supernatant was removed partially. Cells were resuspended in the remaining medium, spread on culture dishes containing the respective antibiotics and incubated at 37°C.

### 2.2.2 Molecular cloning of *Sc*

#### 2.2.2.1 Chemically competent *Sc*

Chemically competent *Sc* were made from single colonies of respective *Sc* cells. O.N. cultures were used to inoculate 50 ml of medium and the cultures were grown for at least two doublings. Cells were harvested by centrifugation at 30°C and resuspended in 25 ml of sterile H<sub>2</sub>O. After a second centrifugation cells were resuspended in 1 ml of fresh 100 mM LiAc and transferred to a 1.5 ml tube. Cells were sedimented for 15 s at top speed and the pellet was resuspended in 400 µl of 100 mM LiAc. Aliquots of 100 µl were made and in 1.5 ml tube. Aliquots were directly used for transformations or could be stored at -80°C.

#### 2.2.2.2 *Sc* transformation

For transformations with linear or plasmid DNA, 100 µl of chemically competent *Sc* cells (Section 2.2.2.1) were used. 50 µl of previously boiled and rapidly cooled salmon sperm DNA, approximately 1 µg of DNA, 240 µl of PEG3350 and 36 µl of 1 M LiAc were added. Cells were vortexed vigorously for 1 min, incubated at 30°C for 30 min, heat shocked at 42°C for 15 min, sedimented at 7000 rpm for 15 s and resuspended in 200 µl TE Buffer. For antibiotic resistance markers cells were recovered in YPD for 2 h before spreading on selection plates.

#### 2.2.2.3 Cryo-stock generation

Single colonies from a *Sc* strain were streaked on YPD plates and incubated at 30°C for two days. Cells were collected from the plate and dissolved in 1 ml of 30% glycerol and stored at -80°C.

#### 2.2.2.4 Polymerase Chain Reactions

**Polymerase Chain Reactions (PCRs)** were used to create and amplify desired DNA fragments. According to the different polymerase enzyme, different buffers and supplements were used for PCR reactions. Taq-PCR reactions generally contained 10-150 ng of DNA template, 0.2 mM dNTPs, 0.04 mM MgCl<sub>2</sub>, 0.5 µl, 2.5 µl 10x Taq buffer, 0.75 µl Taq polymerase and 16.15 µl H<sub>2</sub>O. PCR was performed with a T3000 Thermocycler (Biometra) using the following protocol: 95°C for 3 min, (95°C for 30 sec, 52°C for 30

sec, 72°C for 1 min) x 29, 72°C for 10 min.

Phusion polymerase (Finnzymes) was generally used for templates which required higher polymerase fidelity. Reactions typically contained 1-200 ng of DNA template, 0.5  $\mu$ l of each primer, 1.5  $\mu$ l DMSO, 25  $\mu$ l Phusion 2 x mastermix and 17.5  $\mu$ l H<sub>2</sub>O. PCR program was as follows: 98°C for 30 sec, (98°C for 10 sec, 50-65°C for 30 sec, 72°C for 15-60 sec) x 20-30, 72°C for 10 min. Depending on the PCR product, parameters like annealing temperature, MgCl<sub>2</sub> concentration DMSO concentration were varied.

Colony PCRs were performed to control for correct insertion of DNA fragments into the *Sc* genome. Therefore a single colony was picked and solved in 100  $\mu$ l 20 mM NaOH and 50  $\mu$ l of glass beads added. The reaction was incubated at 95° for exactly 5 min shaking vigorously. The sample was sedimented at top speed for 15 s and stored at 4°. 5  $\mu$ l of the supernatant were used for Taq-PCRs.

For overlap PCRs, in order to fuse fragments, DNA templates were mixed equimolar and primers were designed to have annealing temperatures above 60°C.

### 2.2.2.5 Reverse transcription

First strand DNA synthesis was carried out with the ProtoScript II Reverse Transcriptase (New England Biolabs, Cat No. M0368). 20 ng of RNA were used in combination with random primers following the manufacturers protocol. First strand cDNA was stored at -20°C for further use.

### 2.2.2.6 DNA purification

If not specified otherwise, DNA was always purified using Qiagen purification systems following the manufacturers instructions: QIAquick PCR purification Kit; QIAquick Gel extraction Kit or depending on the elution volume QIAquick Minelute PCR purification. For plasmid isolation QIAquick Miniprep Kit was used. Genomic DNA isolation was carried out using QIAgen DNeasy Blood and Tissue Kit with slight adaptations:

### 2.2.2.7 *In vivo* protein tagging

Generation of *Sc* strains containing epitope tagged proteins was carried out by homologous recombination [Longtine et al., 1998]. Phusion PCRs (Section 2.2.2.4) were carried out to amplify an epitope sequence followed by a selection marker. The primers contained approximately 50 bp long overhangs homologous to the region just upstream and downstream of the ORF, replacing the stop codon of the ORF. The PCR fragment was purified

(Section 2.2.2.6) and transformed into competent *Sc* cells (Section 2.2.2.2), resulting in the insertion of the PCR fragment in frame just behind the last codon of the ORF. Correct insertion into the genome was verified via colony PCR (Section 2.2.2.4) and sequencing.

### **2.2.2.8 Conditional depletion of proteins from the nucleus**

Conditional depletion of proteins from the nucleus was done using the parental Y40343 strain (Table 2.1) for insertion of the FKBP12-rapamycin-binding (FRB)-tag (Section 2.2.2). Correct insertion of the FRB-tag was tested via colony PCR (Section 2.2.2.4) and growth was tested on YPD agar plates containing 1  $\mu\text{g/ml}$  rapamycin. Over night cultures of positive clones were grown in YPD at 30°C in replicates. Cultures were diluted to an OD<sub>600</sub> of 0.1 in the next morning and grown until an OD<sub>600</sub> of 0.6. Rapamycin was added to the cell culture for 60 min with a final concentration of 1  $\mu\text{g/ml}$ . Either cross-linking for chromatin immunoprecipitation (Section 2.2.4) or metabolic RNA labeling (Section 2.2.3 2.2.3.2) were performed after 60 min of rapamycin treatment.

### **2.2.2.9 Yeast microscopy**

Cells were grown until OD<sub>600</sub> 0.6, splitted in half and one half supplemented with 1  $\mu\text{g/ml}$  rapamycin. After 60 min 6. 1.5 ml of each culture were transferred to a 2 ml Eppendorf tube and treated with 500  $\mu\text{l}$  fresh paraformaldehyde solution (10 % paraformaldehyde, 13 mM NaOH, 150 mM Phosphate buffered saline (PBS)) for 10 min at room temperature. Cells were pelleted, washed with PBS once and resuspended in 100  $\mu\text{l}$  PBS. Glass slides were coated with Poly-L-lysine (Sigma, No. P8920) for 10 min. Poly-L-lysine was aspirated off and 20  $\mu\text{l}$  of fixed cells were applied to the glass slide for 10 min. Cell suspension was aspirated off and 20  $\mu\text{l}$  PBS with 1 ng/ $\mu\text{l}$  Dapi were pipetted on the fixed cells and incubated for 2 min. Slides were washed with PBS twice, covered with cover glass and analyzed under a microscope.

### **2.2.2.10 Electrophoretic separation of DNA**

Separation of DNA mixtures was performed through agarose gel electrophoreses. DNA samples were mixed with 6 x Loading Dye (Fermentas) and loaded onto 1 x TAE agarose gels containing 0.8 % to 2% agarose (depending on DNA size) and SYBR Safe<sup>®</sup> (0.01  $\mu\text{g/ml}$ ; Invitrogen). Additionally 5  $\mu\text{l}$  of 100 bp or 1 kb GeneRuler<sup>TM</sup> ladders (Fermentas) were loaded onto the gel. DNA was separated at 110 V for varying times and visualized under UV light.

### 2.2.3 Experimental methods for cDTA analysis

#### 2.2.3.1 Metabolic labeling of *Sc* and *Sp*

Metabolic labeling for comparative dynamic transcriptome analysis (cDTA) was done as described [Sun et al., 2012]. Generally, cells were streaked from cryo-stocks onto YPD plates and incubated for two days at 30°C. A single colony was picked, transferred to 20 ml YPD and grown o.N. at 30°C. On the next morning 50 ml of YPD were inoculated from the o.N. culture with a starting OD<sub>600</sub> of 0.1. At an OD<sub>600</sub> of 0.8 4-thiouracil (4-tU, Sigma, 2M in DMSO) was added to a final concentration of 50 mM. After 6 min of labeling, cells were rapidly transferred to a centrifuge and sedimented by centrifugation at 2465 x g 30°C for 1 min. The supernatant was discarded and the pellet resuspended in 500 µl RNAlater solution (Ambion/Applied Biosystems). 50 µl of the cell suspension were used to determine the cell concentration with a Cellometer N10 (Nexus) and the remaining cell suspension was flash frozen in liquid nitrogen and stored at -80°C for further use. Similarly, *Sp* cells were grown in YES medium instead of YPD, labeled with 4-thiouridine (4tU, Sigma, 50 mM in ddH<sub>2</sub>O) for 6 min at a final concentration of 500 µM and harvested for 3 min by centrifugation. Other steps were as for *Sc*. Also, a 4 L culture of *Sp* was labeled to generate a large stock of labeled *Sp* cells.

#### 2.2.3.2 Total RNA extraction and isolation of labeled RNA

For cDTA, a 1:3 mixture of *Sc* and *Sp* cells was made using  $2.25 \times 10^8$  *Sc* cells and  $0.75 \times 10^8$  *Sp* cells. All following steps were carried out with the cell mixture. Total RNA was extracted by RiboPure<sup>TM</sup> (LifeTechnologies<sup>TM</sup>). Instead of the manufacturers protocol, cell lysis was performed with a FastPrep24 (MP Biomedicals) and acid washed glass beads. Cells were lysed 8 x 40 s at 6.5 m/s with 1 min on ice between each lysis step. Total RNA was stored at -20°C for further use.

Isolation of labeled RNA was performed as previously described [Miller et al., 2011]. 100 µg of total RNA were chemically biotinylated in 1 x Biotinylation Buffer with a final concentration of 100 ng/µl Biotin-HPDP for 1.5 h at RT. Unbound biotin was removed via chloroform/isoamylalcohol (24:1) extraction with Phase-lock-gel tubes (Eppendorf). Total RNA was precipitated through addition of 1/10 volume of NaCl an, equal volume of isopropanol and sedimentation at max speed and 4°C in a table top centrifuge (Eppendorf). The pellet was resuspended in 100 µl RNase-free water and denatured at 65°C for 10 min. The total RNA was incubated with 100 µl of µMACS streptavidin (Miltenyi) beads for 15 min at RT to bind labeled RNA to the beads. Columns (Miltenyi) were used

to capture the labeled RNA and the total RNA in the flow-through. Columns were washed 5 times with increasing volumes of Washing Buffer (500 to 1000  $\mu$ l) before the labeled RNA was eluted with 100  $\mu$ l of Elution Buffer (DTT). After 5 min a second round of elution was performed with 100  $\mu$ l. 700  $\mu$ l of RLT Buffer were added and the labeled RNA was purified using the RNeasy MinElute Cleanup Kit, following the manufacturers protocol and eluted in 20  $\mu$ l of RNase-free water.

For the cDTA analysis of *rpb1-1* strains, overnight cultures were diluted in fresh medium to an OD<sub>600</sub> of 0.15 (125 ml cultures, 160 rpm shaking incubator, 30°C). At an OD<sub>600</sub> of 0.9 (time point: 18 min), RNA was labeled. Eighteen minutes later (time point 0 min) cultures were shifted to 37°C by adding the same volume of 42°C tempered medium. RNA was again labeled 18 min and 60 min after heat shock (time points +24 min and +66 min, respectively).

### 2.2.3.3 Microarray hybridization for cDTA

Gene expression analysis with microarrays was carried out with 100 ng of total or labeled RNA using the GeneChip 3'IVT labeling assay (Affymetrix). Samples were hybridized to GeneChip 2.0 microarrays containing probesets for *Sc* and *Sp* following the manufacturers instructions (Affymetrix).

### 2.2.3.4 cDTA data analysis outline

Data was pre-processed array-wise using *expresso* (R/Bioconductor) with the RMA background correction method. We created our own probe annotation environment (*cdf*), which excludes probes in probesets that show cross-hybridization between *Sc* and *Sp*. 8708 annotated *Sc* probes and 13,317 annotated *Sp* probes out of a total of 120,855 probes showed cross-hybridization when a conservative intensity cut-off of 4.5 (log intensity values after preprocessing) was used. Cross-hybridizing probes were excluded from further analysis. This included 16 whole probe sets (Figure 3.2a). Note that the standard GC-RMA method is not suitable for our purposes, since its bias model cannot handle bimodal intensity distributions, as caused by the simultaneous hybridization of *Sc* and *Sp* transcripts with global differences in RNA abundance (Figure 3.2b). Labeling bias estimation and correction was done as described [Miller et al., 2011]. Between-array normalization of arrays containing mixed *Sc* and *Sp* total RNA was done by proportional rescaling, such that the median *Sp* gene expression level was 1 (Figure 3.5b). Accordingly, between-array normalization of arrays containing mixed *Sc* and *Sp* labeled RNA was done by proportionally scaling the array to a median labeled *Sp* gene expression level of *c* (Figure

3.5a). The constant  $c$  scales the median half-life of all experiments. We calibrated  $c$  in a way that the resulting median  $Sc$  wild type mRNA half-life equaled that observed previously [Miller et al., 2011]. Now, all  $Sc$  RNA levels, no matter if total or labeled, no matter from which experiment, can be compared on an absolute level. Decay rates and synthesis rates were obtained as described [Miller et al., 2011]. We assume that the labeled RNA fraction is subject to degradation from the very time it is synthesized. In contrast, [Rabani et al., 2011] (Supplementary Methods therein) assume that the labeled RNA fraction is mostly nuclear and not degraded at all. We compared the synthesis rate estimates resulting from both alternatives. Given our labeling time, the differences of both approaches are negligible. The whole analysis workflow has been carried out using the open source R/Bioconductor package DTA.

### 2.2.3.5 Kinetic model

Our model has been cast as Equation (1) provided in the main text. The steady-state mRNA levels predicted by this model are  $g = \frac{\mu_g}{\lambda_g} \cdot \frac{f(s)}{h(d)}$ , from which we deduce that regulation imposed by S or D is always global, i.e., total mRNA levels are shifted by a common factor  $f(s)/h(d)$ . Since the mRNA levels in the deadenylation mutants globally increase, we conclude that the mRNA level  $s_9$  of S in the deadenylation mutants is higher than in wild-type (level  $s$ ). At the same time, we can estimate the quotient  $f(s')/f(s)$  by

$$\begin{aligned} \frac{f(s')}{f(s)} &= \text{median} \left( \frac{\text{synthesis rate of } g \text{ in the mutant}}{\text{synthesis rate of } g \text{ in the wildtype}}, g \in \text{genes} \right) \\ &= \begin{cases} 0.4 \text{ for } \Delta \text{ pop2} \\ 0.5 \text{ for } \Delta \text{ ccr4} \end{cases} \end{aligned}$$

(see Supplemental Fig. S8 for a rigorous derivation). Together,  $s' \geq s$  and  $f(s') \geq f(s)$  imply that  $f$  acts as a transcription inhibitor. Similar considerations show that  $d' \geq d$  holds in the Pol II mutant, and that

$$\begin{aligned} \frac{h(d')}{h(d)} &= \frac{\text{total mRNA mutant}}{\text{total mRNA wildtype}} \cdot \text{median} \left( \frac{\text{decay rate of } g \text{ in the mutant}}{\text{decay rate of } g \text{ in the wildtype}}, g \in \text{genes} \right) \\ &= 0.31 < 1 \end{aligned}$$

From  $d' < d$  and  $h(d') < h(d)$  we conclude that D is a degradation enhancer.

### 2.2.3.6 4-thiouracil labeling coupled to RNA-Seq (4tU-Seq)

Labeled RNA from 50 ml cultures of anchor away experiments (Section 2.2.2.8) were used for strand specific RNA-Seq library generation. Depletion of ribosomal RNA was

carried out using the Ribo-Zero rRNA removal Kit (Epicentre, Cat No. MRZH116) following the manufacturers instructions using 1  $\mu$ g of labeled and total RNA as input. Multiplexed libraries were prepared with the NuGEN Encore Complete RNA-Seq Library System (NuGEN, Part No. 0312) using 15 ng of rRNA depleted RNA as input following the manufacturers protocol. Final libraries were quantified on a Qubit 1.0 and qualified on an Agilent Bioanalyzer 2100. Libraries were pooled and 36 bp single end sequencing and 6 bp barcode sequencing were performed on an Illumina Genome Analyzer IIX.

### 2.2.4 Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitations were in principle carried out as described [Mayer et al., 2010] with minor changes. For ChIP followed by qPCR (Section 2.2.4.1) only 50 ml of *Sc* cell culture were inoculated at OD<sub>600</sub> of 0.1 from an over night culture. 40 ml of each culture were cross-linked with 1% formaldehyde for 20 min at 20 °C. The cross-linking reaction was quenched with 375 mM glycine for 10 min and cells were sedimented for 5 min at 4000 rpm. The supernatant was discarded and cells were washed with ice cold TBS. After another centrifugation cell pellets were resuspended in 2 ml of ice cold FA-lysis buffer and transferred to a 2 ml tube. Cells were sedimented at 3000 rpm for 5 min at 4 °C, supernatant removed and cell pellets flash frozen in liquid nitrogen and stored at -80°C for further use.

For immunoprecipitations cell pellets were thawed and resuspended in 1.5 ml FA-lysis buffer. Cell suspensions were transferred to screw-cap lysis tubes for the MP-Biosciences FastPrep<sup>®</sup>-24 machine. Cell lysis was performed 8 x 40 s with 6.5 m/s and 1 min on ice in between each lysis step. Cell lysat was transferred to 15 ml sonication tubes and chromatin was fragmented 35 x 30 s at 4°C with high intensity settings. Sonified chromatin was transferred to 2 ml tubes and centrifuged 10 min at 13000 rpm and 4°C. The supernatant containing the soluble fragmented chromatin was transferred to a new 2 ml tube and centrifuged for 30 min at 13000 rpm and 4°C. For tap-tagged proteins 700  $\mu$ l of chromatin in FA-lysis buffer were used for immunoprecipitation with 25  $\mu$ l sepharose IgG beads. IP was carried out on a turning wheel at 4°C for 60 min. For untagged proteins, chromatin was incubated with an experimentally determined amount of the respective antibody over night at 4°C. The IP was performed with 25  $\mu$ l of protein A and G sepharose beads in FA-lysis buffer for 90 min at 4°C on a turning wheel. 30  $\mu$ l of chromatin solution were taken as input and 100  $\mu$ l were taken for determination of chromatin fragment size before the IP. After the IP, samples were washed in filter-tubes (Millipore) with FA-lysis buffer three times, with FA-lysis buffer (500 mM NaCl) two times, with ChIP-wash buffer two times and with TE-buffer once. Sepharose beads were resuspended in 120  $\mu$ l ChIP

elution buffer and transferred to a 1.5 ml tube. Proteins were eluted from beads for 10 min at 65 °C. The supernatant, containing the eluted chromatin fragments, was treated with Proteinase K for 2 h at 37 °C followed by 65 °C up to 16 hours. Finally, the DNA fragments were purified using the PCR-purification Kit (Section 2.2.2.6) and eluted in 50  $\mu$ l of H<sub>2</sub>O. Samples were stored at -20°C for further use.

### 2.2.4.1 quantitative PCR (qPCR)

qPCR was either performed to detect the amount of bound DNA from a protein (ChIP) (Section 2.2.4) or to detect the amount of RNA (reverse transcribed) (Section 2.2.2.5) that was present in a cell. For ChIP appropriate primers spanning a region of roughly 65-80 bp and primers for the amplification of a control locus (YER) were used to set up qPCR reactions. The PCR efficiency of primer pairs was determined using a serial dilution of fragmented genomic DNA (from previous experiments). This resulted in a standard curve from which the amplification efficiency was calculated with the Bio-Rad CFX Manager software version 1.1. qPCR reactions were set up with the SensiFAST<sup>®</sup> SYBR No-Rox Kit (Bioline) in 96 well plates. Each reaction contained 1x enzyme supermix, 1  $\mu$ l of DNA and 0.8 $\mu$ l of forward and reverse primers in a total volume of 20  $\mu$ l. Plates were sealed and run with the following protocol: 95°C 2 min, 40x (95°C 5 s, 61°C 10 s, 72°C 15 s). Threshold cycle (Ct) values were calculated using the Bio-Rad CFX Manager software version 1.1 using the "Regression" settings. The *fold enrichments* (FE), representing the enrichment of a given DNA region over the control region [Fan et al., 2008], were calculated as follows:

$$FE = \frac{1.9^{CtIP_{control} - CtInput_{control}}}{1.9^{CtIP_{locus} - CtInput_{locus}}}$$

For reverse transcribed RNA, the SensiFAST<sup>®</sup> SYBR No-Rox Kit (Bioline) or the SSO-fast Evagreen Supermix (Bio-Rad) or the were used to set up 20  $\mu$ l reactions containing 10  $\mu$ l 2x Supermix, 1  $\mu$ l of each primer and 1  $\mu$ l of the reverse transcription reaction (Section 2.2.2.5).

### 2.2.4.2 Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq)

Chromatin immunoprecipitations for deep sequencing were essentially prepared as described in (Section 2.2.4) except that more cells had to be used for cross-linking reactions. 400 ml of cell culture were used for cross-linking. Cultures were splitted in equal portions after quenching to fit 250 ml tubes for centrifugation. Cell pellets were washed with 100 ml of ice cold TBS twice before resuspension in 10 ml FA-lysis buffer. Each cell pellet

was flash frozen separately in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  for cell lysis. Each pellet was thawed and resuspended in FA-lysis buffer for cell lysis and the approximate lysis efficiency was determined measuring the  $\text{OD}_{600}$  before and after cell lysis. Importantly, before sonication, each cell lysate was centrifuged in a 2 ml tube for 15 min at 13000 rpm and  $4^{\circ}\text{C}$ . The supernatant was removed and the pellets were resuspended in fresh FA-lysis buffer. Following steps were carried out as above (Section 2.2.4), treating every tube as a distinct sample. After DNA purification each sample was treated with RNase A (10 mg/ml) for 20 min at  $37^{\circ}\text{C}$ . Samples were purified again and eluted in  $15\ \mu\text{l}$  of  $\text{H}_2\text{O}$ . Samples that were splitted originally were pooled and used for library preparation with the NEB-Next ChIP-Seq Library Reagent Set for Illumina Sequencing (NEB, Cat No. E6200) in combination with Multiplex Oligos (NEB, Cat No. 7335). Libraries were prepared following the manufacturers protocol. Size selection was performed using 2% agarose gels. Purifications were performed with Qiagen Minelute columns except for the purification of the final library which was done using 1.2x Agencourt AMPure XP beads (Beckman Coulter, Cat No. A63880) following the manufacturers protocol. Libraries were quantified on a Qubit 1.0 and qualified on an Agilent Bioanalyzer 2100, pooled accordingly. 36 bp single end sequencing and 6 bp barcode sequencing was performed on an Illumina GAIIX platform.

### 2.2.5 Deep sequencing data analysis

#### 2.2.5.1 Basic data analysis

Sequencing data were obtained in the raw Fastqsanger format and all basic processing was done using Galaxy [Goecks et al., 2010, Blankenberg et al., 2010, Giardine et al., 2005]. ChIP-Seq and 4tU-Seq (RNA-Seq) files were in principle treated identical with minor differences explained later in this section. Unfiltered Fastqsanger files were demultiplexed to obtain data for each sample. Each dataset was then filtered according to read quality. Therefore reads with Phred scores lower than 30 were discarded. Trimming was done from both ends of reads and the final read length had to be at least 30 bp. Short reads were then mapped to the *Saccharomyces cerevisiae* genome build SacCer3 (April 2011) with Bowtie 1.0 [Langmead et al., 2009] using the following options: (Number of mismatches: -1; Try hard: yes; Report up to n valid alignments per read: -1, Report all valid alignments: No; Suppress all alignments if more than n alignments exist: 1; sBestoption: 1; Use strata option). The resulting SAM files were then converted to BAM files using SAMtools [Li et al., 2009]. The number of reads for every genomic position was calculated using the pileup function from SAMtools. To obtain information about expression levels, read

counts per annotated feature were calculated using the DEseq HTseq count (version 1.0) function with "intersection strict" settings [Anders and Huber, 2010]. For 4tU-Seq data mapping was carried out with the strand-specific option of Bowtie resulting in two files per sample, one for the Watson and one for the Crick strand.

### 2.2.5.2 4tU-Seq data analysis for the Nrd1 project

Further processing of the 4tU-Seq data was carried out using the R/Bioconductor environment. Piled-up read counts for every genomic position were summed up over replicates. Size factors for each condition were calculated as described in [Anders and Huber, 2010] using only reads falling into ORF-T regions [Xu et al., 2009] and used to correct for library size and sequencing depth variations. We chose to normalize to ORF-T regions after we observed strong increases in ncRNA abundances when inspecting the data manually, explaining why similar numbers of genes show increased and decreased expression within gene-boundaries. We chose to normalize to ORF-T regions after we observed strong increases in ncRNA abundances when inspecting the data manually, explaining why similar numbers of genes show increased and decreased expression within gene-boundaries. Differential profiles were calculated as the log<sub>2</sub> ratio of Nrd1 depleted and wild-type read count pileups. One pseudo count was added to each position prior to division in order to prevent singularities. The 4tU-Seq data had been created without the use of actinomycin D and therefore had an antisense bias probably created during reverse transcription [Perocchi et al., 2007]. The antisense bias was estimated to be 10% using mid to high expressed regions without antisense annotation. The real number of reads  $s$  in a region of interest was calculated according to the following formula:

$$s = \frac{S - cA}{1 - c^2}$$

where  $S$  and  $A$  are the observed number of reads in a given window on the sense and antisense strand.  $A$  is shifted +100 bp with respect to the location of  $S$ , which was the estimated offset resulting from fragment size and the number of bps sequenced.  $c$  gives the ratio of spurious reads originating from the opposite strand.

Differential expression analysis was done using the R/Bioconductor package "DESeq". Transcripts with a fold-change of at least 1.5 and multiple testing adjusted p-value lower than 0.1 were considered differentially expressed. Reads per kilobase (rpk) were calculated upon bias corrected read counts falling into the region of an annotated feature divided by transcript-length in kilobases.

### 2.2.5.3 NUTs annotation

Position based 4tU-Seq fold-changes upon nuclear depletion of Nrd1 were used to identify regions of a minimum fold of 1.25 and a minimal length of 100 bp. Consecutive regions with gaps smaller than 25 bp were merged. The automatically identified segments were manually curated (has been done previously [Xu et al., 2009] to yield high quality transcript boundaries resulting in a defined set of NUTs. The previously defined characteristics of transcript curation might have slightly been altered due to manual assessment. In order to verify the high quality of transcript annotation we defined 3 criteria: 1. Minimal length of a NUT is 100 bp (length of NUTs ranges from 137 to 8313 bp). This criterion is met by all NUTs. 2. Minimal fold-change upon nuclear depletion of Nrd1 is 1.25. This criterion is met by over 99% of NUTs (84% even significantly with adj. p-value 0.1). Note that many NUTs are not apparent before nuclear depletion of Nrd1, which can therefore lead to non significant fold-change observations. This fact suggests the last criterion. 3. NUTs should at least have a coverage of 2 after nuclear depletion of Nrd1.

### 2.2.5.4 ChIP-Seq data analysis

Replicate handling, size factor correction and calculation of differential profiles of the ChIP-Seq data was carried as for 4tU-Seq data. For each annotated transcript the predicted termination site was estimated by finding the border between two segments (transcript-TSS proximal region (PPF) and transcript body (TBF)) via fitting a piecewise constant curve to the differential profile between transcript-TSS and either the TSS + (transcript length / 2) or maximally the TSS + 1000 bp using the segment method from the R/Bioconductor package "tilingArray". Escape Indices (EIs) were subsequently calculated as the ratio of median transcript body fold-change (second segment) and median transcript-TSS proximal region fold-change (first segment). EIs were weighted to yield coverage-dependent quantities by the following factor:

$$\sqrt{\frac{26}{l_{PPF}} \left( \frac{1}{\sum r_{ND_{PPF}}} + \frac{1}{\sum r_{U_{PPF}}} \right) + \frac{26}{l_{TBF}} \left( \frac{1}{\sum r_{ND_{TBF}}} + \frac{1}{\sum r_{U_{TBF}}} \right)}$$

where  $l$  is the length of the segment in bp,  $r$  is the number of readcounts, ND is the Nrd1 depleted sample and U the untreated sample. Thresholds for EI selection were chosen as the 0.95% quantiles of weighted EIs calculated from within replicate measurements in both conditions.

### **2.2.6 PAR-CLIP**

PAR-CLIP was done by Carlo Baejen and was done similar to previous publications [Creamer et al., 2011] with minor changes. A detailed protocol can be found in Carlo Baejens publications.

# Chapter 3

## Results

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

#### 3.1.1 Establishment of cDTA

To measure changes in mRNA synthesis and decay rates between different strains of budding yeast *Sc*, we included the distantly related fission yeast *Sp* in our DTA protocol as an internal standard (Figure 3.1). We counted *Sc* sample cells and *Sp* control cells and mixed them in a defined ratio (Section 2.2.3). The resulting cell mixture was lysed, total mRNA extracted, labeled RNA purified, and microarrays were hybridized as described [Miller et al., 2011]. The RNA mixture was quantified on a microarray that contains probes for both *Sc* and *Sp* transcripts (Affymetrix GeneChip Yeast Genome 2.0 Array) [Miller et al., 2011]. We used 4-thiouracil (4tU) instead of 4sU for *Sc* RNA labeling, because it is taken up by *Sc* [Munchel et al., 2011] without expression of a nucleoside transporter [Miller et al., 2011]. 4tU labeling did not affect normal cell physiology [Sun et al., 2012] and allowed growth of yeast in YPD instead of selective medium. We quantified only labeled and total RNA, because the unlabeled fraction was not required for rate extraction. We refer to this protocol as comparative DTA (cDTA).

We first tested whether the *Sc* sample showed cross-hybridization to *Sp* array probes and vice versa. When either a *Sc* or *Sp* sample was hybridized to the array, cross-hybridization occurred for a minor fraction of the probes (Section 2.2.3.4) when a conservative intensity cut-off of 4.5 (log intensity values after preprocessing) was used (Figure 3.2a). Cross-hybridizing probes were excluded from further analysis, leading to loss of only 16

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

out of 10,799 probe sets (Section 2.2.3.4). The mixing ratio of *Sc:Sp* cells was tuned to 3:1 to maximize the overlap of the *Sc* and *Sp* expression intensity distributions (Figure 3.5a). This ensures that after calibration most *Sc* and *Sp* probe intensities are in the linear measurement range of the microarray, an important prerequisite for our calculations. We restricted our analysis to RNAs with log intensity signals above 4.5 and below 8 (Figure 3.2b).

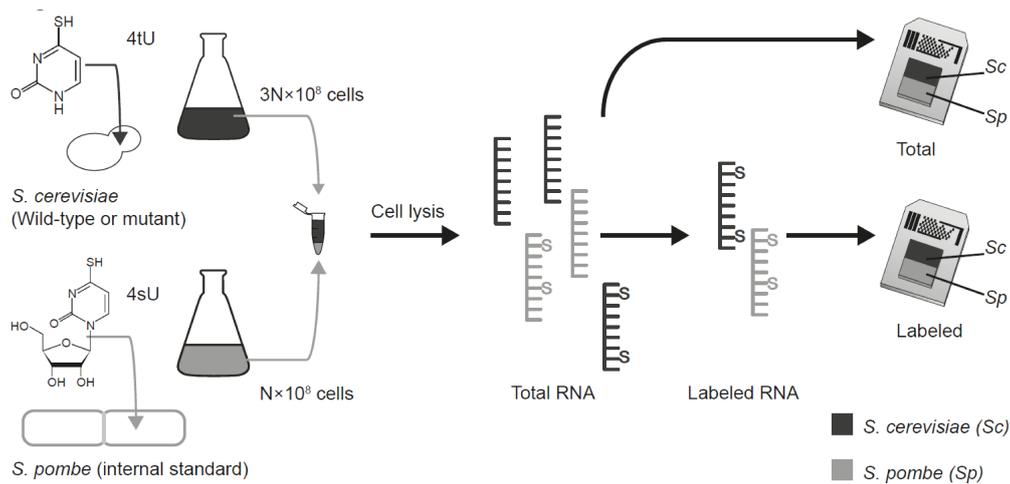


Figure 3.1: Design of a cDTA experiment. The *Sc* cells are labeled by adding 4tU into the media, whereas *Sp* cells are labeled by adding 4sU. The cells are then counted. *Sc* cells from different experiments are always mixed with the same amount of labeled *Sp* cells from a single batch. Cells are then lysed, RNA is extracted, biotinylated, and labeled RNA separated. Microarrays containing probes against both *Sc* and *Sp* transcripts are then used to quantify both total and labeled RNA.

#### 3.1.2 Rate extraction from cDTA data

To obtain absolute synthesis and decay rates for *Sc* and *Sp*, we derived the ratios of labeled to total RNA intensities  $cSc$  and  $cSp$  for *Sc* and *Sp*, respectively. These ratios set the global median level of synthesis and decay rates and rely on a robust previous estimate of the median *Sc* half-life [Miller et al., 2011] for which labeled, total, and unlabeled RNA fractions were available. Once  $cSp$  is known, the measured levels of the *Sp* standard can be used to calibrate the *Sc* data (Figure 3.6). This new normalization method allows rate estimation from labeled and total quantities alone (Methods). Our published median half-life for *Sc* mRNAs [Miller et al., 2011] enabled determination of the median *Sp* half-life relative to *Sc* (Figure 3.3). We measured growth curves and obtained a doubling time of 90 min for *Sc* in YPD medium at 30°C and 116 min for *Sp* in YES medium at 32°C [Sun et al., 2012]. These doubling times were used in kinetic modeling [Miller et al., 2011]. We

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

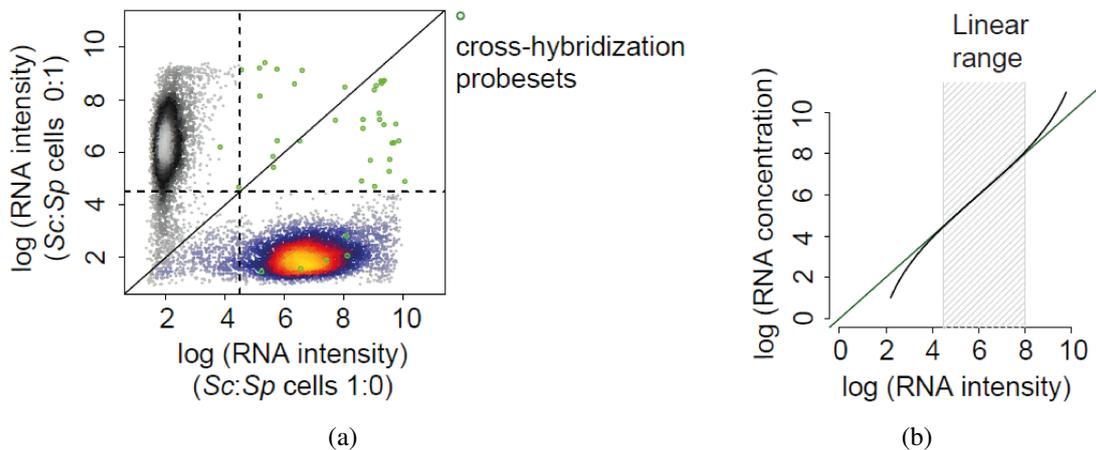


Figure 3.2: Establishing the cDTA protocol. (a) Assessment of cross-hybridization. Scatterplot of log intensities of 10,928 Affymetrix probe sets. The values on the x- resp. y-axis are obtained as the mean of two pure *Sc* resp. *Sp* replicate samples that were hybridized to the arrays. *Sc* and *Sp* probe sets (heat colored and gray scaled, respectively) can be separated almost perfectly. A total of 23 out of 5771 *Sc* probe sets show intensities above a (log) background intensity threshold of 4.5 in the *Sp* sample, whereas eight out of 5028 *Sp* probe sets were above background in the *Sc* sample. These 31 probe sets are regarded as affected by cross-hybridization (green circles). Of these, 16 probe sets were excluded from analysis because all probes were affected by cross-hybridization (Section 2.2.3). (b) Linear measurement range. Exemplary illustration showing that the relation of mRNA concentration (real amount) and mRNA intensity (fluorescent scanner readout) follows the Langmuir adsorption model (Hekstra et al. 2003; Held et al. 2003, 2006; Skvortsov et al. 2007). The green line indicates linearity. (Black line) Sigmoidal behavior, resulting from noise at low-hybridization levels and saturation effects at high hybridization levels. (Gray stripe) Linear measurement range that we defined as an intensity range of 4.5–8 (natural log basis) based on noise signals below 4.5, for example, for probes that detect transcripts of genes that were knocked out and based on observed saturation effects above 8.

confirmed that the rates obtained by cDTA are essentially the same as the ones previously obtained by DTA (Table 3.1).

RNA halfives that were recently determined by 4tU pulse-chase labeling in *Sc* are 1.5-fold longer [Munchel et al., 2011], likely because a very long labeling time was used that allowed for thionucleotide reincorporation after mRNA decay. We calculated mRNA synthesis rates as the number of complete transcripts made per cell and per 90 min (the cell cycle time for wild-type *Sc*), using a new estimate of 60,000 transcripts per yeast cell [Zenklusen et al., 2008] instead of the previously used, older, and fourfold lower estimate [Hereford and Rosbash, 1977]. For *Sp*, we estimated the number of transcripts from the observed 2.51-fold cumulative total RNA level to be 150,801. Our rate estimates are unaffected by the efficiency of 4tU labeling, which varies between strains and experiments

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

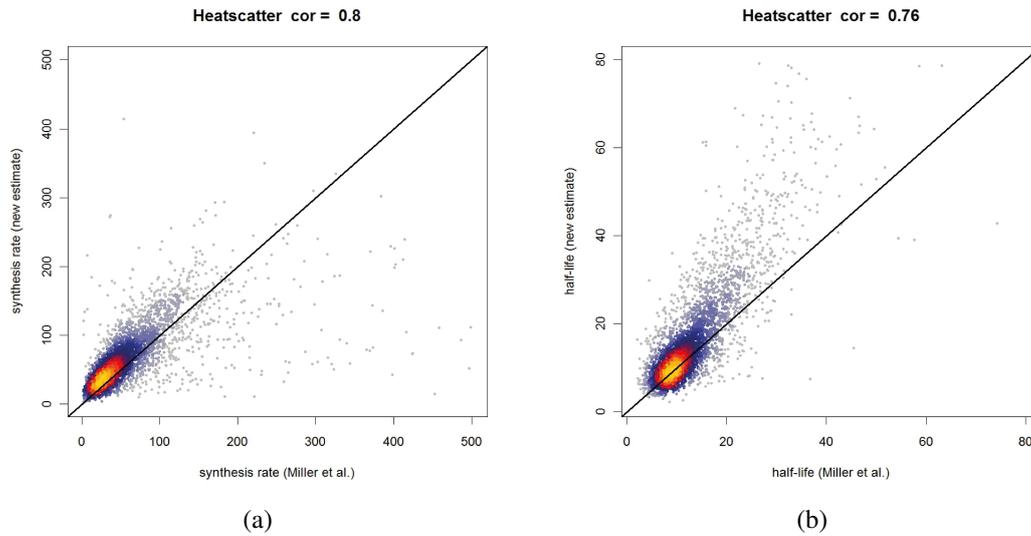


Figure 3.3: The scatterplots compare the *Sc* synthesis (a) and half-lives (b) as obtained by cDTA (y-axis) and those obtained from our data using the method from [Miller et al., 2011]. Spearman correlations are 0.8 and 0.76 respectively.

(Figure 3.4).

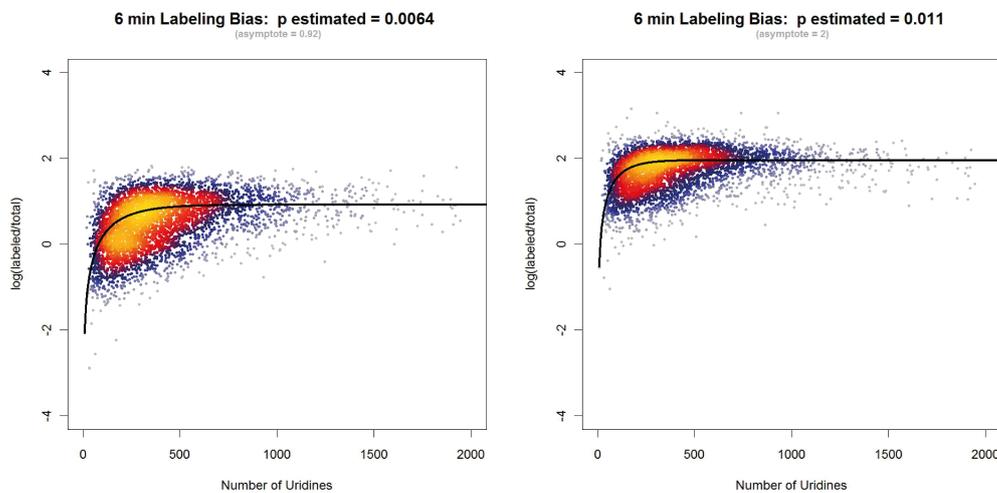


Figure 3.4: The number of uridines is plotted versus the log-ratio of L and T. The black line shows that the labeling bias curve estimated as in [Miller et al., 2011]. Left: labeling bias plot for the slow Pol II mutant.  $p_{lab} = 0.0064$  means that approximately every 156th uridine residue is replaced by 4tU and afterwards attached to a biotin molecule. Right: labeling bias plot for the wild-type.  $p_{lab} = 0.011$  means that approximately every 90th uridine residue is replaced by 4tU and afterwards attached to a biotin molecule. The shifted asymptotes indicate the observed fold of the decay rate comparing these two conditions.

For normalization between different *Sc* samples, we linearly rescaled all array intensities such that the total and labeled *Sp* fractions have a median intensity of 1 or  $cSp$  (Figure

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

3.5a). We assessed the accuracy of the cDTA procedure by estimating the intensity ratios of *Sc:Sp* cells that were mixed at 1:1, 3:1, and 10:1. The correct values were recovered with an accuracy of 5% (Figure 3.5b).

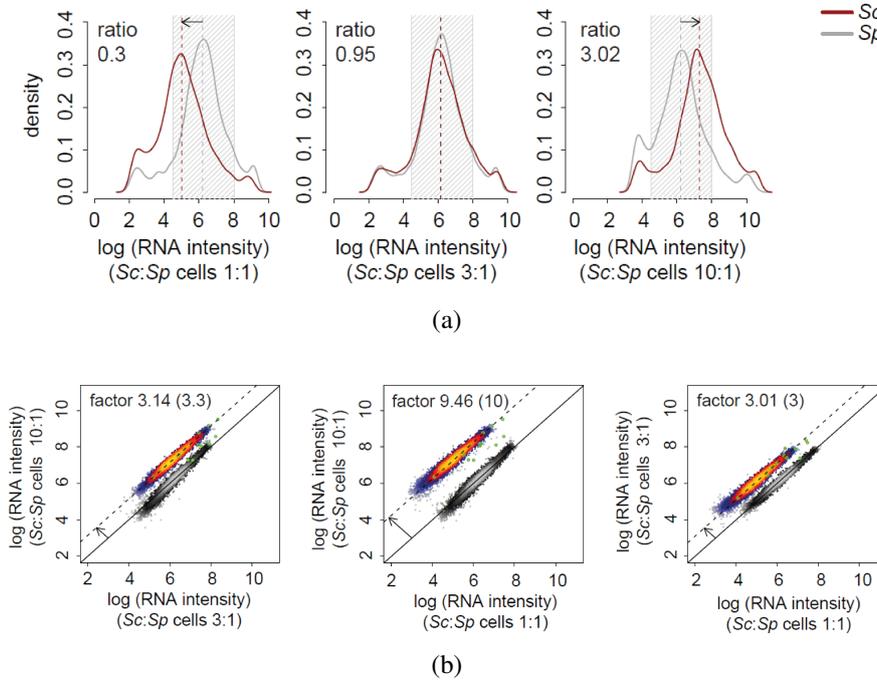


Figure 3.5: Establishing the cDTA protocol. (a) Calibration of *Sc:Sp* cell mixture ratio. The optimal cell mixture ratio has been chosen to maximize the number of probes for both *Sc* and *Sp* that fall into the linear measurement range. *Sc* and *Sp* cells were mixed in *Sc:Sp* ratios of 1:1, 3:1, and 10:1. The respective median mRNA level ratios are 0.3, 0.95, and 3.02. Log (RNA intensity) distributions of *Sc* (red) and *Sp* (gray) are shown. The median intensity level of *Sp* is approximately three times higher than that of *Sc*. As a consequence, a *Sc:Sp* cell mixture ratio of 3:1 was used. (b) Comparison of the three different cell mixtures of (a) in pairwise log–log scatter plots. All arrays are normalized to a common median of 4052 *Sp* probe sets (gray-scaled). A total of 4475 *Sc* probe sets (those in the linear measurement range) are shown in heat colors. The parallel offset of the *Sc* probe sets from the main diagonal measures their RNA level differences of *Sc* in the three cell mixtures. The differences should be 3.3, 10, and 3 when we plot *Sc:Sp* ratios of 10:1 vs. 3:1, 10:1 vs. 1:1, and 3:1 vs. 1:1, respectively. The corresponding measured offsets are 3.14, 9.46, and 3.01, and thus in very good agreement.

Selected mRNA levels of the 1:1 and 10:1 ratio mixtures were additionally quantified by RT–qPCR (Methods). The expected ratio of the four tested *Sc* transcripts was recovered within a relative error of 9% when normalized to two housekeeping *Sp* genes (data not shown). In summary, cDTA normalization removes the major sources of experimental differences between samples in RNA-labeling efficiency, cell lysis, RNA extraction, RNA biotinylation and labeled RNA purification, and array hybridization. cDTA detects global changes between *Sc* samples, in contrast to standard normalization procedures that

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

Table 3.1: Median mRNA half-lives and synthesis rates of *Sc* and *Sp* transcripts

	Species	cDTA	DTA
Median mRNA half.life (min)	<i>Sc</i>	12	11.5
	<i>Sp</i>	59	N.A.
Median mRNA synthesis rate (mRNAs per cell and cell cycle time) <sup>a</sup>	<i>Sc</i>	53 <sup>a</sup>	18 (72) <sup>a</sup>
	<i>Sp</i>	44	N.A.

The cDTA contains the estimates obtained from using the labeled:total ratio of the complementary strain and the known total and labeled *Sc:Sp* ratios to calculate the missing labeled:total ratio, i.e.,  $L_{Sc}/T_{Sc} = (L_{Sp}/T_{Sp}) \cdot (T_{Sp}/T_{Sc}) \cdot (L_{Sc}/L_{Sp})$ . The DTA column shows the *Sc* half-life estimate obtained from Miller et al. (2011). Note that the *Sc* estimates are virtually identical to ours, although we used a different labeling technique (4tU instead of 4sU) and had spiked-in *Sp* controls in the sample. <sup>a</sup>Please note that we previously used in our calculations a total number of transcripts per cell of 15,000 according to an old estimate [Hereford and Rosbash, 1977], whereas we now used a recent estimate of 60,000 [Zenklusen et al., 2008]. If the same number of transcripts is used, the median synthesis rate obtained by DTA would be 72, comparable to our new estimate obtained by cDTA, despite the difference in media and cell cycle time [Miller et al., 2011].

eliminate global changes, because they assume constant median RNA levels.

#### 3.1.3 cDTA supersedes conventional methods

Conventional methods measure mRNA half-lives by inducing transcription arrest and following changes in mRNA levels over time. Transcription arrest has been achieved by adding the transcription inhibitor 1,10-phenanthroline [Dori-Bachash et al., 2011] or by shifting the temperature-sensitive mutant strain *rpb1-1*, which carries point mutations in the largest subunit of Pol II [Nonet et al., 1987], to the restrictive temperature [Holstege et al., 1998, Wang et al., 2002, Grigull et al., 2004, Shalem et al., 2008]. To investigate whether the latter method yields reliable data or whether it perturbs mRNA metabolism, we regenerated the *rpb1-1* strain and analyzed it with cDTA using published growth parameters [Holstege et al., 1998] (Section 2.2.3.2). This revealed that mRNA synthesis rates were decreased globally by a factor of 2.7 already at the permissive temperature of 30°C (Figure 3.7a). After 24min at the restrictive temperature, mRNA synthesis rates had decreased further by a factor of 3.4, but recovered essentially to the rates measured at the permissive temperature after 66 min (Figure 3.7a). These observations indicated that the mRNA metabolism in the *rpb1-1* strain is already perturbed at the permissive temperature, and that the temporary changes in mRNA metabolism observed at the restrictive temperature are mainly due to a heat-shock response. To test this, we conducted a corresponding heat-shock experiment on wild-type cells. We analyzed the total mRNA from this experiment together with the data from the *rpb1-1* mutant by conventional decay time

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

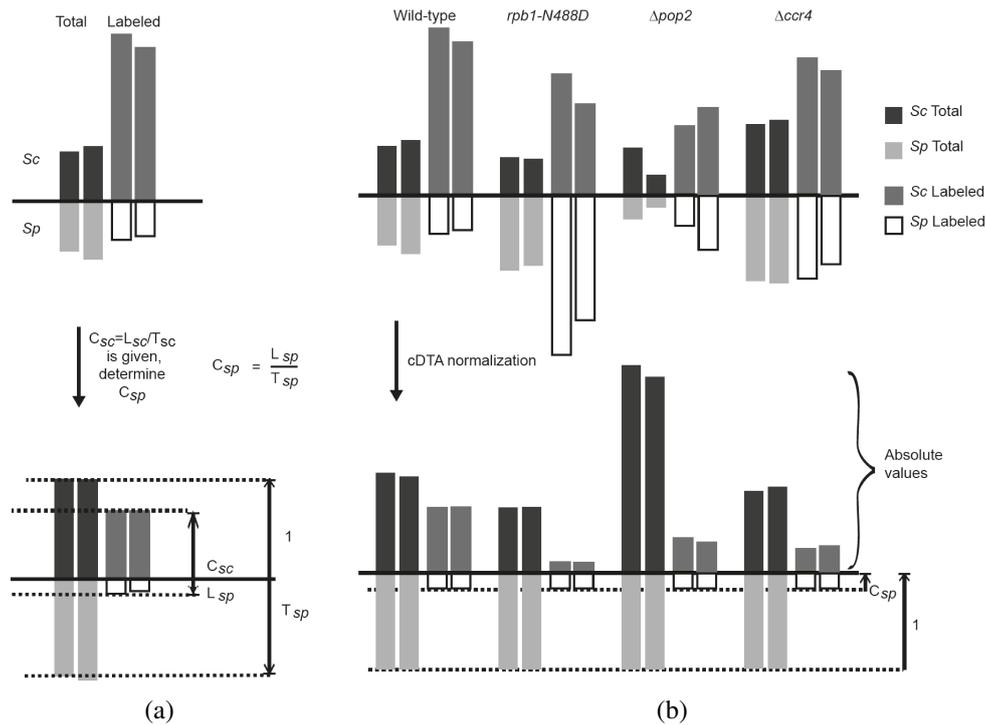


Figure 3.6: cDTA normalization reveals global changes. (a) Determination of  $cSp$ , the ratio of labeled over total *Sp* mRNA. To obtain absolute synthesis and decay rates for *Sc* and *Sp*, we derived ratios of labeled to total RNA  $cSc$  and  $cSp$  for *Sc* and *Sp*, respectively. The  $cSc$  ratio was obtained in our previous study [Miller et al., 2011]. To determine  $cSp$ ,  $L_{Sc}$  and  $T_{Sc}$  are set to  $cSc$  and 1, respectively.  $L_{Sp}$  and  $T_{Sp}$  are then linearly rescaled. The resulting  $L_{Sp}/T_{Sp}$  is defined as  $cSp$  and then used in the further experiments as the global cDTA normalization factor. (b) cDTA normalization uses *Sp* signals as an internal standard. The bars indicate the median intensities of the array probe sets. Due to our experimental design, the ratio of labeled to total *Sp* RNA ( $cSp = L_{Sp}/T_{Sp}$ ) must be the same in all experiments. To correct for differences in cell lysis, RNA extraction efficiency, and RNA purification efficiencies, the levels of *Sp* total and labeled mRNA are rescaled to the same values in all experiments. The *Sc* RNA levels are then corrected by median centering of *Sp* RNA levels. This normalization allows for a direct comparison of *Sc* data between experiments. Shown are both replicates for each of the four cDTA experiments.

series analysis [Holstege et al., 1998, Wang et al., 2002, Grigull et al., 2004, Shalem et al., 2008]. The obtained mRNA half-lives during heat shock correlated very well with data derived from the *rpb1-1* mutant strain and with published half-lives obtained with this strain (Figure 3.7b). The obtained half-lives were longer than the half-lives measured in unperturbed cells, likely because mRNA degradation was down-regulated during the stress response. There was also a good correlation with half-lives obtained after adding 1,10-phenanthroline and even with our previous data obtained during the osmotic stress response [Miller et al., 2011], if processed in the conventional way. This indicates that all of these data are dominated by perturbations in mRNA metabolism that result from a

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

general stress response. In contrast, published half-lives derived from metabolic RNA labeling [Munchel et al., 2011] and our cDTA-derived half-lives do not correlate with data obtained by perturbing conventional methods. We conclude that conventional methods for estimating mRNA half-lives using the *rpb1-1* mutant strain or transcription inhibition cannot be used to obtain reliable half-life estimations.

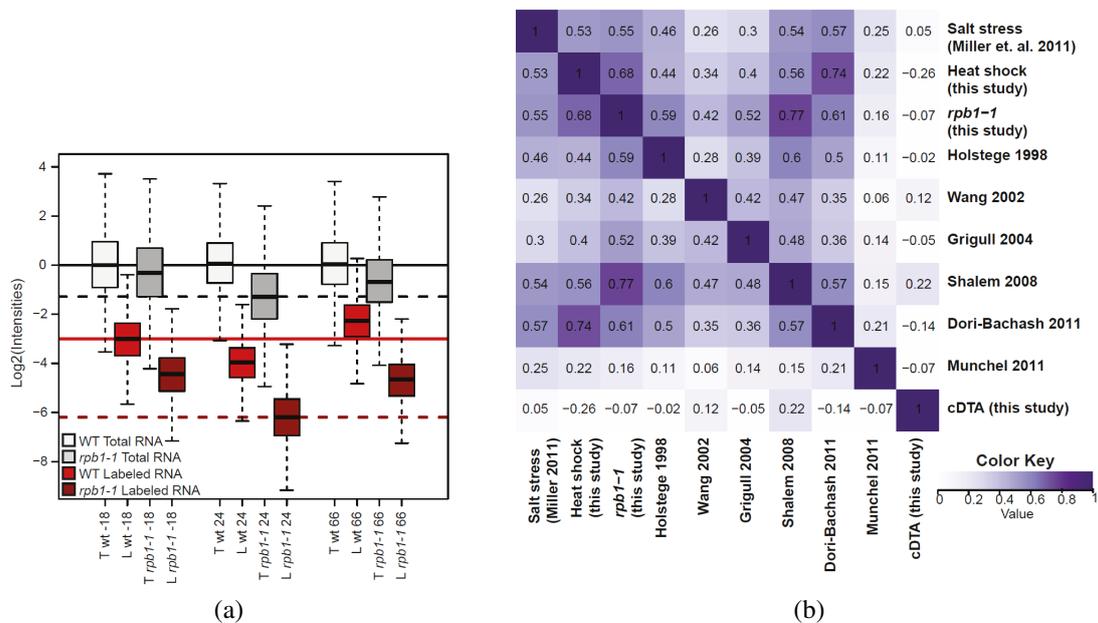


Figure 3.7: Comparison of cDTA with conventional methods. (a) Box plots of the expression distributions of the total and the labeled (newly synthesized) mRNA after cDTA normalization, obtained from the wildtype and the *rpb1-1* mutant before, and 24 and 66 min after the shift to restrictive temperature. Transcriptional activity is roughly restored in both strains after 66 min. The global shifts in labeled expression are only slightly more pronounced in the *rpb1-1* mutant, indicating a dominant role of heat shock in the profiles of *rpb1-1*. (b) Correlation analysis of mRNA half-life measurements. The heatmap shows pairwise Spearman correlation coefficients of half-life measurements (white: negative or zero correlation; purple: perfect correlation). The published half-life estimates except for [Munchel et al., 2011] were obtained by experiments using transcriptional arrest. The estimates of [Holstege et al., 1998, Wang et al., 2002, Grigull et al., 2004, Shalem et al., 2008] were obtained using a yeast strain containing the Pol II temperature sensitive mutant *rpb1-1*. Dori-Bachash et al. (2011) used the transcription inhibitor 1,10-phenanthroline.

#### 3.1.4 Comparison of mRNA metabolism in distant yeast species

As an immediate result, cDTA reveals similarities and differences in the mRNA metabolism of *Sc* and *Sp*. First, the median mRNA synthesis rates are very similar in *Sc* and *Sp* (Figure 3.8a). The median synthesis rate was 53 mRNAs per cell and 90 min for wild-type *Sc*,

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

and 44 mRNAs per cell and 90 min for *Sp*. Second, *Sp* mRNAs have about fivefold longer half-lives on average than *Sc* mRNAs, with a median of 59 min (Figure 3.8a), compared with 12 min for *Sc*.

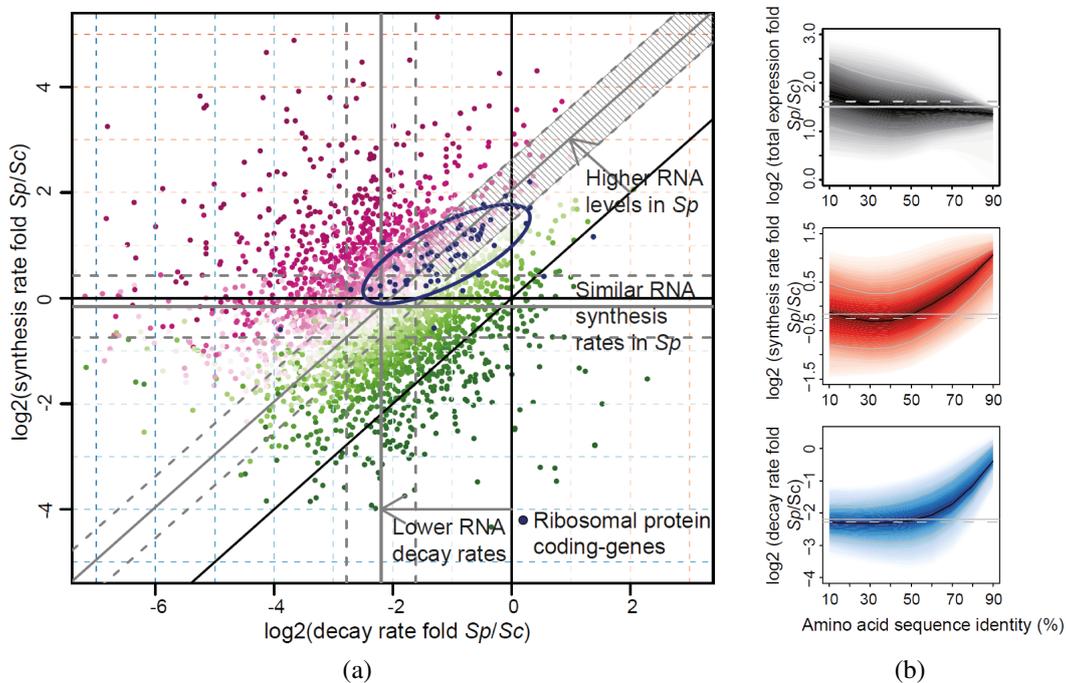


Figure 3.8: Comparison of mRNA metabolism in *Sp* and *Sc*. (a) Scatter plot comparing mRNA decay rate folds versus synthesis rate folds of *Sp* and *Sc* transcripts encoding protein orthologs (>25% amino acid sequence identity). The offset of gray lines to parallel black lines indicates *Sp*:*Sc* ratios of median decay rates, synthesis rates, or total mRNA (0.20/0.83/2.72). Dashed gray lines indicate 1.5-fold changes from the median (gray lines). Color scheme corresponds to folds in total mRNA (magenta, positive log fold; green, negative log fold). A set of genes that show higher decay and synthesis rates (1.5-fold and adjusted P-value 0.5%), but almost unchanged (<1.5-fold) total mRNA (93 transcripts, striped area) was selected and tested with a Bayesian network-based gene-set analysis (MGSA) (Bauer et al. 2010). In this gene set, the ribosomal protein genes were enriched (blue dots; ellipse shows the 75% region of highest density). (b) Plots show  $\log_2$  fold distributions of total mRNA (gray), synthesis rate (red), and decay rate (blue) of *Sp* versus *Sc* transcripts encoding orthologous proteins as a function of amino acid sequence identity (%). Transcripts encoding highly conserved proteins such as ribosomal proteins are located on the right. They show more rapid turnover (synthesis and decay) in *Sp*, resulting in similar mRNA levels. (Solid black lines) Median  $\log_2$  fold; (shaded bands) central 80% regions. (Solid/ dashed gray lines) Median  $\log_2$  fold of all orthologs/all genes.

As expected, the cDTA-derived *Sp* half-lives show a fair correlation with half-lives obtained by another nonperturbing metabolic labeling [Amorim et al., 2010]. Furthermore, reprocessing the data of [Amorim et al., 2010] with our cDTA algorithm, which takes into

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

account the labeling bias and an additional parameter to correct for cell growth, increases the correlation to our results and leads to a median half-life of 50 min, in good agreement with an estimate of 59 min in our study [Sun et al., 2012]. Third, the overall mRNA levels in *Sp* were about 3.1-fold higher than in *Sc*. Since the haploid *Sc* cells with a median volume of 42  $\mu\text{m}^3$  are approximately two- to threefold smaller than *Sp* cells with a median cell volume of 115  $\mu\text{m}^3$  [Jorgensen et al., 2002, Neumann and Nurse, 2007], the higher mRNA levels apparently lead to similar cellular mRNA concentrations. The change in mRNA levels is mainly a global multiplicative change ( $R^2 = 0.82$ , [Sun et al., 2012]). Taken together, these data suggest that *Sp* cells generally contain more stable mRNAs than *Sc* cells to reach similar mRNA concentrations at similar mRNA synthesis rates, despite their larger volume.

We investigated whether mRNA sequence conservation correlates with a conservation of total RNA levels, synthesis rates, or decay rates (Figure 3.8); [Sun et al., 2012]. This analysis revealed a conservation of the relative total levels of mRNAs that encode orthologous proteins in *Sc* and *Sp*. The levels of mRNAs that encode proteins with an amino acid sequence identity of at least 25% (2568 mRNAs) show a high Spearman correlation of 0.69. Synthesis rates correlate well between both species Spearman correlation 0.61), but the half-lives show only a fair correlation Spearman correlation 0.4). Although the data suggest that *Sp* cells have globally shifted decay rates, to reach similar cellular mRNA concentrations, there is a minor fraction of transcripts that behave exceptionally.

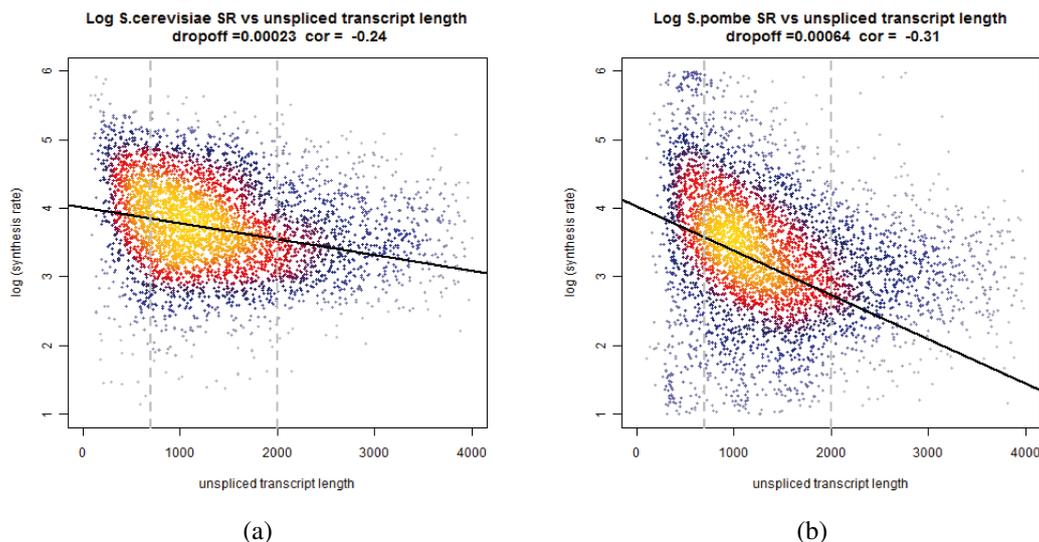


Figure 3.9: Pol II dropoff rate. Correlation of (log) synthesis rates with length for *Sc* (a) and *Sp* (b). The linear regression and the Pearson correlation were calculated for the transcripts with a length between 700 and 2000 nucleotides.

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

In particular, 93 *Sp* transcripts show almost unchanged mRNA levels (<1.5-fold), but significantly higher synthesis and decay rates (>1.5-fold), and are enriched for ribosomal protein genes (Figure 3.8a). More generally, transcripts that encode highly conserved proteins show similar levels, but a faster turnover in *Sp* (Figure 3.8b). We also assessed the correlation of synthesis rates with transcript lengths, and revealed a substantially higher Pol II drop-off rate in *Sp* (Figure 3.9).

#### 3.1.5 Impaired mRNA synthesis is compensated by decreased degradation

We applied cDTA to the question of whether the speed of Pol II is relevant for setting the cellular rates of mRNA synthesis. We used a yeast strain that carries the nondisruptive point mutation N488D in the largest Pol II subunit Rpo21 (also known as Rpb1) (*rpb1-N488D*). This mutation slows down Pol II speed in RNA elongation assays in vitro [Malagon et al., 2006] and is located near the active site [Cramer et al., 2001]. We subjected this strain and an isogenic wild-type strain to cDTA, and collected two biological replicates that showed a Spearman correlation of 0.99 for total and labeled RNA [Sun et al., 2012]. We measured cell-doubling times, and used these in the kinetic modeling to correct synthesis rates for a change in doubling time [Sun et al., 2012]. In the *rpb1-N488D* mutant strain, mRNA synthesis rates were globally decreased 3.9-fold (Figure 3.10). This is consistent with the observed two to 4.5-fold decrease in Pol II speed measured in vitro [Malagon et al., 2006]. We observed a Pol II drop-off rate similar to that described previously [Jimeno-Gonzalez et al., 2010], but quantitative modeling excludes drop-off of Pol II during elongation as the cause for the decreased synthesis rates (Figure 3.11b).

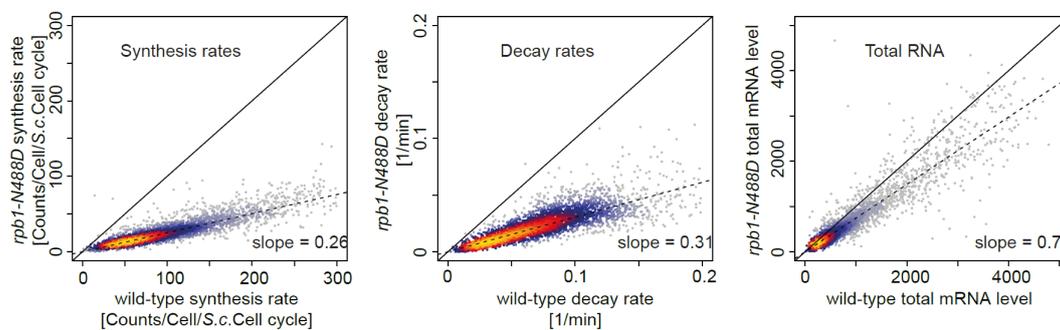


Figure 3.10: Linear scatter plots (heat-colored) of mRNA synthesis rates, decay rates, and total mRNA levels in wild-type and mutant *rpb1-N488D* yeast strains as measured by cDTA. Slopes indicate global shift ratios of median synthesis rates, decay rates, and total mRNA of the *rpb1-N488D* mutant strain compared with wild type (0.26/0.31/0.75).

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

Despite the lower synthesis rates, global mRNA levels were not changed very much in the slow Pol II mutant strain (Figure 3.10). This resulted from a strong decrease in mRNA decay rates of 3.2-fold on average. Synthesis and decay rates of all mRNAs were shifted by approximately the same factor, independent of their wild-type expression level, synthesis rate, or decay rate. The globally increased mRNA half-lives apparently compensated for the decreased mRNA synthesis rates to buffer cellular mRNA levels, which were decreased 1.3-fold only. The measured total RNA levels agreed well with total mRNA levels calculated from the changed synthesis and decay rates (Figure 3.11a). These results show that cells with a strong defect in mRNA synthesis can maintain nearly normal mRNA levels by compensatory changes in mRNA decay rates.

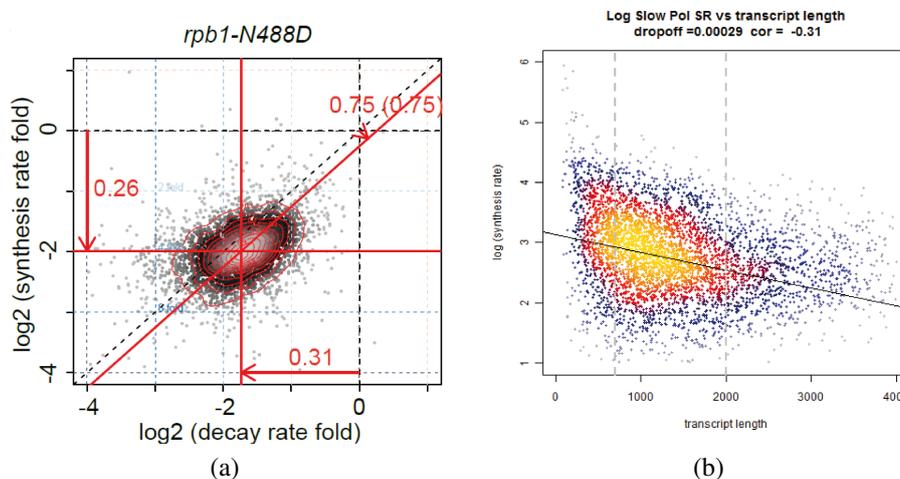


Figure 3.11: Changes in synthesis and degradation rate in a slow Pol II mutant strain  
(a) Alternative representation of the data from (Figure 3.10) *rpb1-N488D* mutant strain compared with the wild-type strain. Each point corresponds to one mRNA. The density of points is encoded by their brightness (grayscale). Contour lines define regions of equal density. The center of the distribution is located at  $(-1.8, -1.6)$ , indicating that there is a global shift in the median synthesis rate by a factor of 0.26 (shift of the horizontal red line relative to the dashed x-axis line), and a global shift in the median decay rate by a factor of 0.31 (shift of the vertical red line relative to the dashed y-axis line). The global change in total mRNA levels is predicted by the offset of the diagonal red line from the dashed main diagonal, which corresponds to a change by a factor of 0.75. The number in brackets following this number (0.75) is the global change as it has been observed in the total mRNA measurements, which agrees well with the predicted number. The changes in total RNA levels do not exactly equal the quotient of synthesis and decay rate changes, due to an additional parameter for cell growth. (b) Dependence of synthesis rates on transcript length for the slow Pol II mutant.

### 3.1.6 Impaired degradation is compensated by decreased synthesis

The observed synthesis-decay compensation implies that cells buffer total mRNA levels. If true, cells should also be able to compensate for a mutation that impairs mRNA degradation with a change in mRNA synthesis rates. To investigate this, we applied cDTA to mutant yeast strains with global defects in mRNA degradation. The choice of mutant was difficult, since RNA degradation involves multiple enzymes in the nucleus and cytoplasm [Houseley and Tollervey, 2009]. We decided to use mutant strains that lack either one of the two catalytic subunits of the Ccr4–Not complex, Ccr4 or Pop2, which show a defect in mRNA deadenylation, a ratelimiting step in mRNA degradation [Tucker et al., 2002]. As predicted, mRNA decay rates were globally decreased in the  $\Delta ccr4$  and  $\Delta pop2$  strains, and changed on average by a factor of 0.43 and 0.16, respectively. This suggests that Ccr4 and Pop2 mRNA degradation factors are used globally.

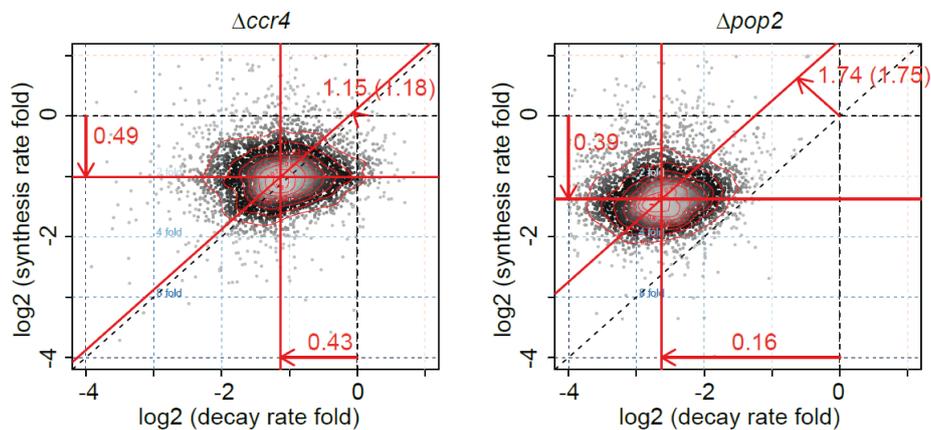


Figure 3.12: Changes in mRNA metabolism for a slow Pol II mutant strain. Scatter plots as in (Figure 3.11a) comparing synthesis rates, decay rates, and total mRNA levels of  $\Delta ccr4$  and  $\Delta pop2$  mutant strains to wild-type yeast. Ratios of median synthesis rates, decay rates, and total mRNA of the  $\Delta ccr4/\Delta pop2$  mutant strain compared with wild type are 0.49/0.39, 0.43/0.16, and 1.15/1.74, respectively.

In both degradation-deficient knock-out strains, an unexpected decrease in mRNA synthesis rates was observed (Figure 3.12). Synthesis rates were changed by a factor of 0.49 and 0.38 in the  $\Delta ccr4$  and  $\Delta pop2$  strains, respectively, limiting the increase in total mRNA levels due to highly defective degradation to a factor of only 1.18 and 1.75, respectively (Figure 3.12). This effect could be observed directly in the labeled fractions of the  $\Delta ccr4$  and  $\Delta pop2$  strains. Only 62% or 46% of the RNA was labeled within the same labeling time, indicating lower synthesis rates. Thus, the defects in RNA degradation in these strains are at least partially compensated by decreased mRNA synthesis rates in order to buffer mRNA levels. This mutual compensation cannot be explained by measurement variance. A variation analysis for the estimation of the median synthesis and decay rates

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

(Figure 3.13); [Sun et al., 2012]) shows that the 95% confidence regions of the median synthesis and decay rate estimates are clearly disjoint.

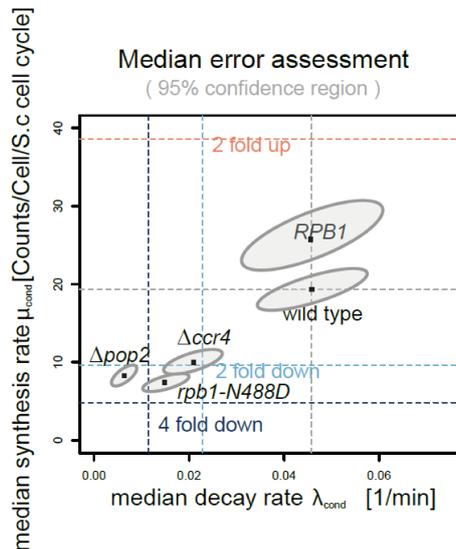


Figure 3.13: Coupling of synthesis and decay rates, on the absolute level. For each condition, the median synthesis rate (y-axis) and degradation rate (x-axis) is shown (dark dots). (Dashed lines) Fold induction/repression relative to wild type. The dots lie approximately on a line with positive slope, indicating synthesis-decay compensation. A variation analysis for the estimation of the median synthesis and decay rates with cDTA has been performed. The ellipses show the 95% bootstrap confidence regions in each condition. The main axes of the ellipses reveal that the errors in the estimation of synthesis and decay rates are not independent, yet small enough to prove that the coupling is not due to estimation variance.

#### 3.1.7 A transcription inhibitor and degradation enhancer may buffer mRNA levels

The above data show that yeast cells can compensate for impaired mRNA synthesis with decreased mRNA decay rates, and for impaired degradation by decreased mRNA synthesis rates. Yeast cells thus have mechanisms to buffer mRNA levels by mutual negative feedback between nuclear mRNA synthesis and cytoplasmic mRNA decay. To explore this further, we extended our model for mRNA turnover under steady-state conditions. The mRNA of a gene  $G$  is synthesized at a gene-specific constant rate  $\mu_g$ , and is degraded at a gene-specific rate  $g \cdot \lambda$ , with  $g$  being the mRNA amount resulting from gene  $G$ . We assume that there is a transcription modulator  $S$  and a degradation modulator  $D$  that globally

### 3.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

affect the synthesis rate (SR) and decay rate (DR) by factors  $f(s)$  and  $h(d)$ , respectively:

$$\frac{dg}{dt} = SR(g, s) - DR(g, d) = \mu_g \cdot f(s) - g\lambda_g \cdot h(d) \quad (3.1)$$

The important and plausible assumption of this model is that  $f$  and  $h$  are monotonic functions. However, we do not assume that mRNA levels translate linearly into protein levels, or that the degree of modulation is a linear function of the underlying mRNA concentrations of  $S$  and  $D$ . One might think of  $S$  and  $D$  as proteins, whose activity is a function of their mRNA concentrations  $s$  and  $d$ . From the model (Section 3.1), we inferred the regulatory logic of the observed feedback, as outlined below. A rigid derivation and an extensive discussion of the model's assumptions are given (Section 2.2.3.5); Supplemental Methods in [Sun et al., 2012]. Here, we compare synthesis and decay rates of a gene between two conditions  $C$  and  $C'$ :

$$\frac{SR'(g', s')}{SR(g, s)} = \frac{\mu'_g f(s')}{\mu_g f(s)} \quad (3.2)$$

$$\frac{DR'(g', s')}{DR(g, s)} = \frac{g' \lambda'_g h(d')}{g \lambda_g h(d)} \quad (3.3)$$

The left-hand sides of Equations 3.2 and 3.3 can be evaluated by cDTA. The left-hand side of Equation 3.2 is substantially smaller than 1 for virtually all measurements  $g$ ,  $g'$ , and for both deadenylation mutants (Fig. 6B). For these mutants, we also know that  $\mu_g = \mu'_g$ , and consequently  $f(s') < f(s)$ . We also observe that  $g' > g$  and  $s' > s$ , from which we conclude that  $f$  is monotonically decreasing. This implies that  $S$  acts as a transcription inhibitor. In the slow Pol II mutant, we observe  $\lambda_g = \lambda'_g$ . Using a similar argument as above, Equation (Section 3.3), and cDTA data of the slow Pol II mutant, we conclude that  $h$  is monotonically increasing, implying that  $D$  is a degradation enhancer. These conclusions could only be derived because cDTA enables the comparison of global synthesis and decay rates. The results would be identical if  $S$  and  $D$  were the same molecule. Thus, the most simple explanation of our observations is the existence of a factor that serves as an inhibitor of transcription and an enhancer of degradation and shuttles between the nucleus and cytoplasm.

## 3.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis

### 3.2.1 Nrd1 nuclear localization is essential

To investigate the roles of pA-independent transcription termination in genome expression, we conditionally depleted Nrd1 from the *Sc* nucleus using the anchor-away method [Haruki et al., 2008]. Nrd1 was tagged with the FKBP12 rapamycin-binding domain (FRB) and depleted from the nucleus by rapamycin treatment. Rapamycin forms a ternary complex with Nrd1-FRB and FKBP12-RPL13A fusion proteins, which is efficiently exported out of the nucleus. Strains expressing Nrd1-FRB from the endogenous *NRD1* promoter grew normally, but did not grow in the presence of 1  $\mu\text{g/ml}$  rapamycin (Figure 3.14a). Fluorescence microscopy showed that the Nrd1-FRB fusion protein was exclusively localized in the nucleus, and that rapamycin treatment led to nuclear depletion after 60 minutes (Figure 3.14b). These results demonstrate that Nrd1 is essential for nuclear function.

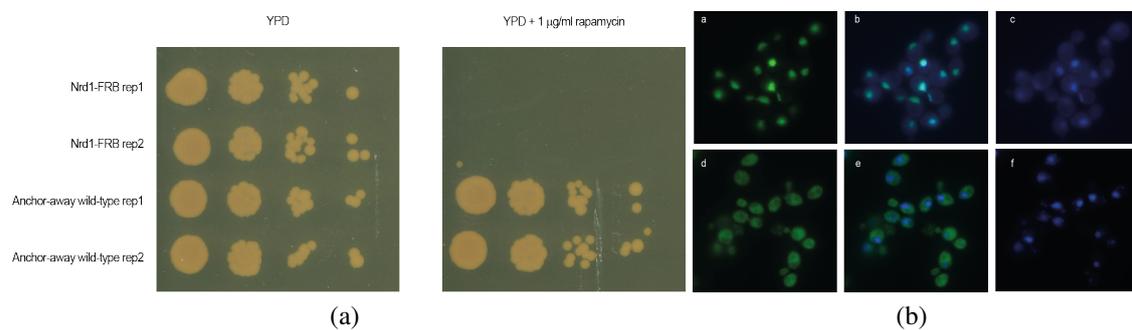


Figure 3.14: Nrd1 nuclear function is essential. (a) Growth control for strain harbouring Nrd1-FRB construct. Two replicate cultures of the anchor-away wild-type strain and the strain harbouring the Nrd1-FRB construct were used for spot-dilutions. Growth of the Nrd1-FRB strain on plates containing rapamycin at the concentration used for all experiments was not detectable at all. (b) Nrd1 localization in *Sc* in exponentially growing cells (panel a-c) and 60 min after rapamycin addition (panel d-f). GFP signal corresponds to Nrd1 tagged with FRB-GFP (green signal). Nuclear staining was carried out with Dapi (blue signal). Merged signals are shown in the middle panels.

### 3.2.2 Nrd1 generally terminates ncRNA transcription

To monitor RNA synthesis in yeast cells, we metabolically labelled newly synthesized RNA for 6 minutes with 4-thiouracil (4tU), purified labelled RNA as described [Sun et al.,

2012], and subjected purified labelled RNA to deep sequencing (Section 2.2.3 2.2.3.2 2.2.3.6). We refer to this highly sensitive method for global estimation of transcription activity as 4tU-Seq, in agreement with the previously reported 4sU-Seq method that uses 4-thiouridine (4sU) labelling in human cells [Rabani et al., 2011, Windhager et al., 2012]. High correlations between biological replicates demonstrated the high reproducibility of 4tU-Seq experiments (Figure 3.15).

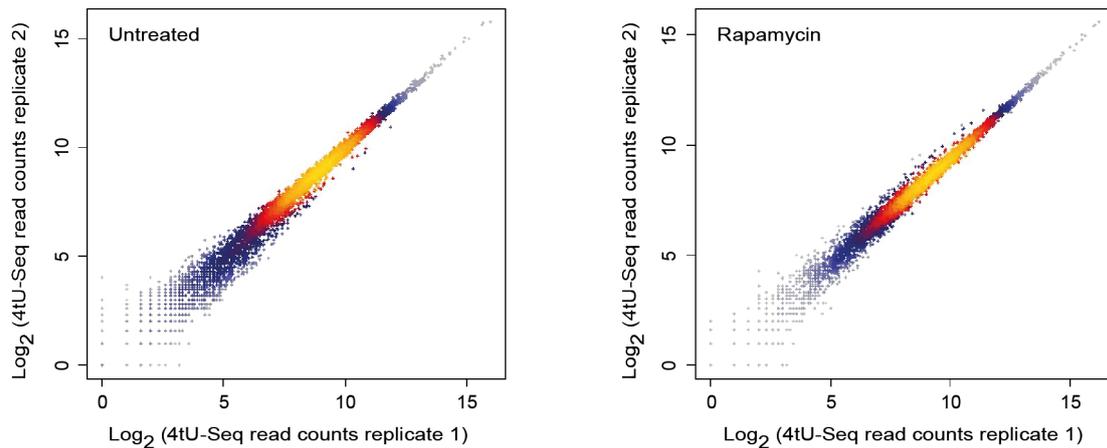


Figure 3.15: 4tU-Seq reproducibility. Comparison of replicate measurements for 4tU-Seq of the untreated samples and rapamycin treated samples. The scatterplot compares read counts of ORF-Ts, SUTs and CUTs. Spearman correlations are 0.99.

To follow changes in RNA synthesis upon nuclear depletion of Nrd1, we carried out 4tU-Seq before and after treatment of cells with 1  $\mu\text{g/ml}$  rapamycin for 60 minutes. After treatment with rapamycin at  $\text{OD}_{600}=0.6$ , cells continued to grow until  $\text{OD}_{600}$  3. Visual inspection of 4tU-Seq data revealed that sn/snoRNAs were generally not terminated and RNA signals were up-regulated in regions encoding sn/snoRNAs (Figure 3.16a) and CUTs (Figure 3.16b). The normalized read counts for all annotated genomic features [Anders and Huber, 2010] revealed an up-regulation of 80% of sn/snoRNAs and many CUTs by  $> 1.5$ -fold (FDR 0.1) but only of 4% of transcribed protein-coding regions (ORF-Ts) (Figure 3.17a).

To examine the termination defects globally, we determined the amount of read-through transcription upon nuclear depletion of Nrd1 by calculating the difference in the number of reads in a 250 base pair (bp) window downstream of each feature (Figure 3.17b). Whereas mRNAs were generally not affected, termination defects were observed for 80% of sn/snoRNAs, 68% of CUTs, and 58% of stable unannotated transcripts (SUTs) that are observed in wild-type yeast cells [Xu et al., 2009]. These results indicate that the Nrd1 pathway generally terminates ncRNA transcription.

### 3.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis

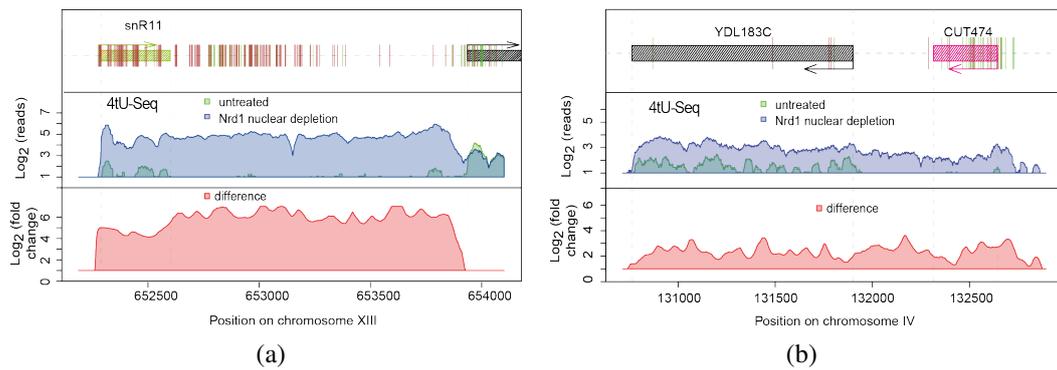


Figure 3.16: Nuclear depletion of Nrd1 leads to defective termination of snRNA11 transcription. (a) Genome browser view of log<sub>2</sub> counts of reads measured by 4tU-Seq before (green) and after (blue) nuclear depletion of Nrd1, and the fold-change between these signals (red) for every genomic position. Vertical green and brown lines depict RNA-binding sites of Nrd1 and Nab3 as determined by PAR-CLIP. (b) Genome browser view as in (a) but for CUT474.

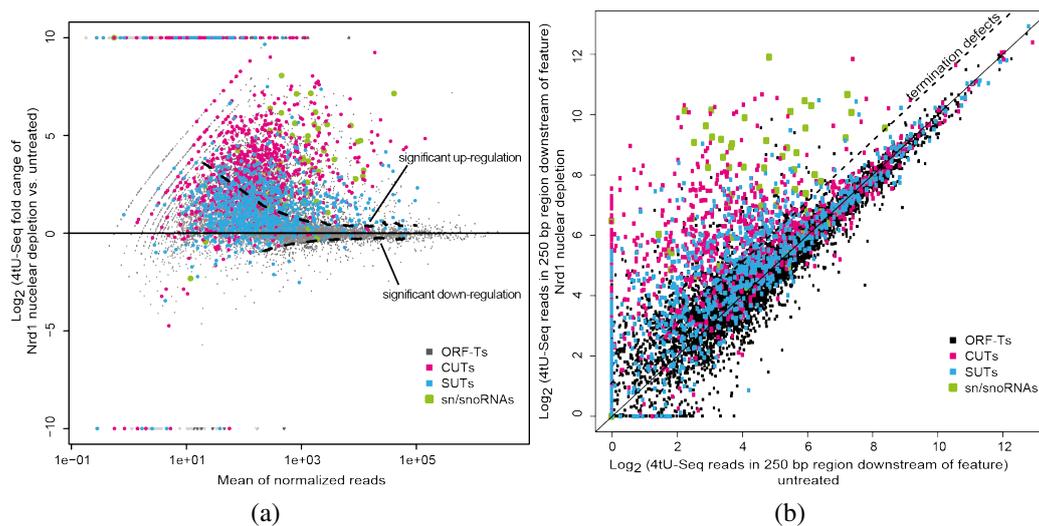


Figure 3.17: Up-regulation and termination defects of most sn/snoRNAs and CUTs upon nuclear depletion of Nrd1. (a) Points mark each transcript's log<sub>2</sub> fold-change upon nuclear depletion of Nrd1 versus the normalized mean read count across replicates and conditions [Anders and Huber, 2010]. Transcripts above or below the dashed line are significantly up- or down-regulated as calculated by DE-Seq. SUTs, CUTs, sn/snoRNAs, and mRNAs from ORF-Ts are in magenta, blue, green, and grey, respectively. (b) Log<sub>2</sub> of normalized read counts in a 250 pb region downstream of annotated genomic feature upon nuclear depletion of Nrd1 versus the same measure in untreated cells.

### 3.2.3 NUTs are extended ncRNA transcripts

To describe the dramatic changes in the transcriptome upon nuclear depletion of Nrd1, we annotated a total of 1526 new transcripts and called them Nrd1-dependent un-terminated transcripts, or NUTs (Material and Methods). Many previously annotated ncRNAs overlapped by at least 50% with NUTs, namely 625 CUTs, 314 SUTs, 620 XUTs, 45 sn/snoRNAs, and 658 CUT\*s (Figure 3.18). NUTs were on average 3.8-fold longer than the overlapping annotated ncRNAs. Only 120 NUTs (8%) overlapped with mRNAs and 213 NUTs (14%) did not overlap with known genomic features. Therefore NUTs are distinct from, although often overlapping with, previously annotated ncRNAs and generally arise from a defect in Nrd1-dependent termination of ncRNA transcription.

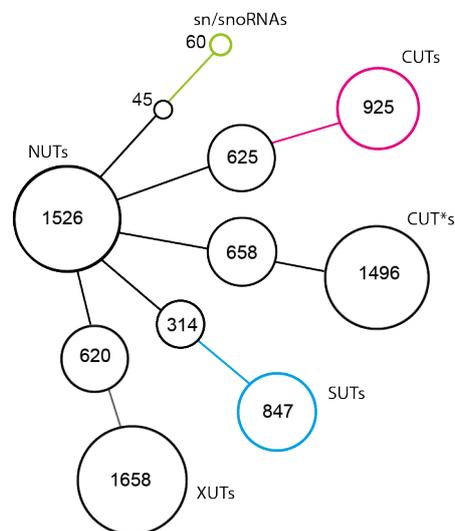


Figure 3.18: Overlap of NUTs with CUTs and SUTs from [Xu et al., 2009], XUTs, snRNAs, and CUT\*s from [Neil et al., 2009]. NUTs were counted to be overlapping when they covered at least 50% of a previously annotated transcript.

### 3.2.4 NUTs originate from distinct PICs in NDRs

The majority of NUTs (896, 59%) originated from previously defined 5' and 3' NDRs flanking known genes [Mavrigh et al., 2008], whereas 339 NUTs originated from intergenic regions, and 291 NUTs originated from within ORF-Ts (Figure 3.19a). All NUTs showed similar levels of nucleosome depletion at their origin (Figure 3.19b). On average, NDRs with NUT origins were almost as pronounced as NDRs containing the transcription start sites (TSSs) of ORF-Ts (Figure 3.19b). To determine the preference of NUTs to originate from 5' or 3' NDRs, we analyzed ORF-T pairs with a distance of at least 452 bp, the minimum distance required to distinguish 5' and 3' NDRs (Xu et al., 2009). 153 NUTs

originated from the 5'-region of ORF-Ts exclusively, whereas 106 NUTs originated from 3'-regions (Figure 3.20). Thus NUTs generally originate from NDRs and often terminate in promoter-associated NDRs.

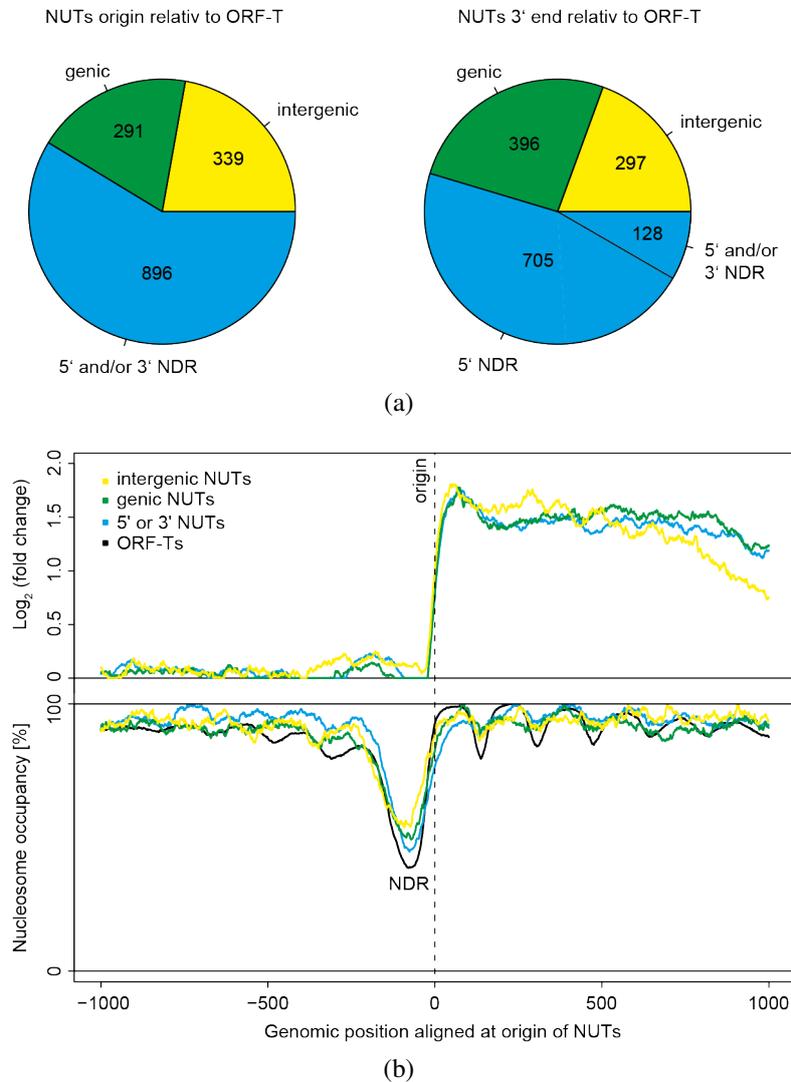


Figure 3.19: NUTs originate from NDRs (a) Fraction of NUT 5' origins and 3' ends in NDRs (blue), genic regions in ORF-Ts (green, from TSS + 50 bp to pA site - 50 bp) or intergenic regions (yellow). For 3' end positions in NDRs those in the 5' NDR of ORF-Ts in antisense or sense direction are shown. (b) Average 4tU-Seq log<sub>2</sub> fold-changes upon nuclear depletion of Nrd1 (upper chart) and averaged nucleosome occupancies (lower chart) (reference) around the NUT 5' origin for the three categories defined in (A) and for all ORF-Ts (black line, lower chart).

We could assign the origins of 690 NUTs (45%) to experimentally mapped PICs [Rhee and Pugh, 2012], of which 257 were unassigned (corresponding to 33% of all orphans), 318 were assigned to CUTs, and 147 to SUTs. NUT transcription initiation thus explained one third of all unassigned mapped PICs [Rhee and Pugh, 2012]. NUTs with mapped

PICs showed a 1.6-fold higher median RNA synthesis than NUTs lacking mapped PICs. The 3' ends of 60% of all NUTs were found in a 5'-NDR of an ORF-T, maybe due to the presence of a PIC for ORF-T transcription (Figure 3.19a).

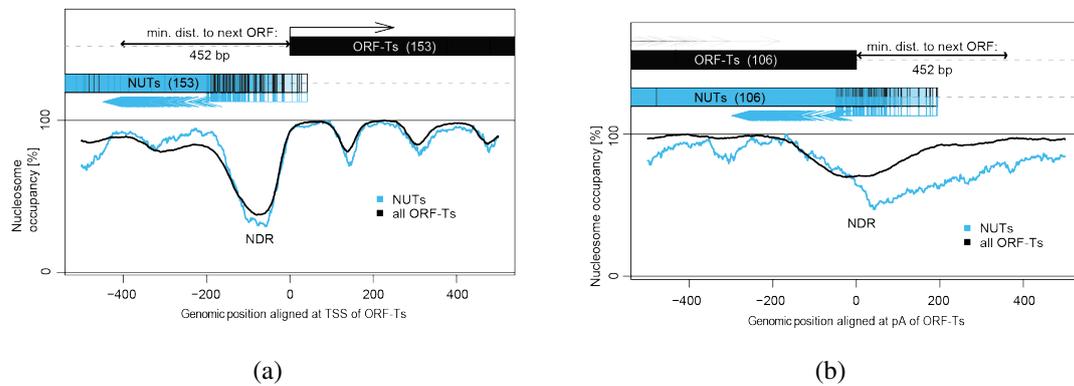


Figure 3.20: NUTs originate in distinct 5' and 3' NDRs (a) Nucleosome occupancies (as in 3.19b) for 153 selected NUTs (blue box with arrows indicating origins) that originate strictly from 5' NDRs and run antisense to ORF-Ts (black box). Plot was aligned at the TSS of those 153 ORF-Ts. (b) Nucleosome occupancies (as in 3.19b) for 106 selected NUTs that originate strictly from 3' NDRs of ORF-Ts.

#### 3.2.5 Many NUTs are divergent and antisense transcripts

A total of 845 NUTs (55%) were divergent transcripts arising from bidirectional promoters. There was no correlation between RNA transcription levels of divergent NUT and ORF-T transcripts arising from the same bidirectional promoter (not shown). This is consistent with previous findings [Murray et al., 2012, Yassour et al., 2010] and with the suggestion that transcription activity is set by independent PICs for divergent transcripts, and not by the amount of nucleosome depletion [Rhee and Pugh, 2012]. Many NUTs originated upstream and antisense of ORF-Ts either from the 5' NDR or an overlapping 3' NDR of an upstream ORF-T. The NUT origin in 5' NDRs is on average 180 bp upstream of the TSS of ORF-Ts. These results show that NUTs often run antisense to known genes and often originate from bidirectional promoters as divergent transcripts.

#### 3.2.6 Nrd1 and Nab3 preferentially bind divergent and antisense ncRNAs

To examine why Nrd1 preferentially terminates Pol II that transcribes in divergent direction, and why it generally does not terminate ORF-T transcription, we globally and

comprehensively mapped RNA interactions of Nrd1 and its partner Nab3 in growing yeast with the use of PAR-CLIP [Hafner et al., 2010]. We optimized PAR-CLIP for the yeast system (Baejen and Cramer, unpublished), and developed a computational pipeline to analyse PAR-CLIP data (Torkler and Soeding, unpublished). We defined an RNA-binding event as the occurrence of at least two overlapping reads with T-C nucleotide conversion [Hafner et al., 2010]. This identified approximately 267,000 Nrd1-binding sites and 223,000 Nab3-binding sites in the yeast transcriptome in a strand-specific manner.

To estimate relative binding affinities of Nrd1 and Nab3 over the transcriptome, we normalized the PAR-CLIP data with transcript occurrence as described [Kishore et al., 2011, Konig et al., 2011a]. Normalization is necessary because the cellular concentration of regular transcripts is much higher than that of rapidly degraded ncRNAs. Normalization was carried out with 4tU-Seq data upon nuclear depletion of Nrd1 because ncRNAs transcripts are barely detected under normal conditions (Torkler and Soeding, unpublished). Alternatively, normalization may be carried out with the use of Pol II ChIP-Seq data, but we refrained from doing this because these data are not strand-specific.

Analysis of the normalized PAR-CLIP data revealed that binding of Nrd1 and Nab3 to 5123 ORF-T transcripts was weak, whereas binding to divergent antisense ncRNA transcripts was much stronger, in particular within the first few hundred nucleotides (Figure 3.21) (Figure 3.22). We also observed strong antisense binding of Nrd1 around the pA site of ORF-Ts (Figure 3.22), consistent with antisense transcripts originating from the 3'-region of ORF-Ts. We note that the observed, about two-fold preference of ncRNA binding versus ORF-T transcript binding (Figure 3.22a) is apparently a gross underestimate, because we could only normalize with ncRNA 4tU-Seq signals measured upon Nrd1 nuclear depletion which are much higher than under normal conditions. These results show that Nrd1 and Nab3 preferentially bind divergent and antisense ncRNA.

#### **3.2.7 Nrd1-binding RNA motifs are depleted in mRNA**

We speculated that the different Nrd1/Nab3 binding densities observed between ORF-T transcripts, antisense ncRNAs, and intergenic ncRNAs may be a result of different motif compositions of these transcript classes. Analysis of the PAR-CLIP sites shows that the reported specific RNA-binding motifs UGUA/GUAG for Nrd1 and UCUU/CUUG for Nab3 [Carroll et al., 2004, Creamer et al., 2011, Porrua et al., 2012, Wlotzka et al., 2011] were overrepresented and explained 27% and 21%, respectively, of all binding events (Figure 3.23a). In addition, we found several related motifs that also showed a high frequency of Nrd1/Nab3 binding (Figure 3.23a). The best binding motif for Nab3 (UCUU)

frequently occurred in a window of 21 bp around Nrd1-binding sites (Figure 3.23a), consistent with a functional complex of Nrd1 and Nab3. This analysis revealed that Nrd1 and Nab3 have binding preferences for several RNA motifs rather than strict specificity for a single motif, generally consistent with an analysis by [Porrua et al., 2012]. Among the preferred motifs were several found recently by [Wlotzka et al., 2011].

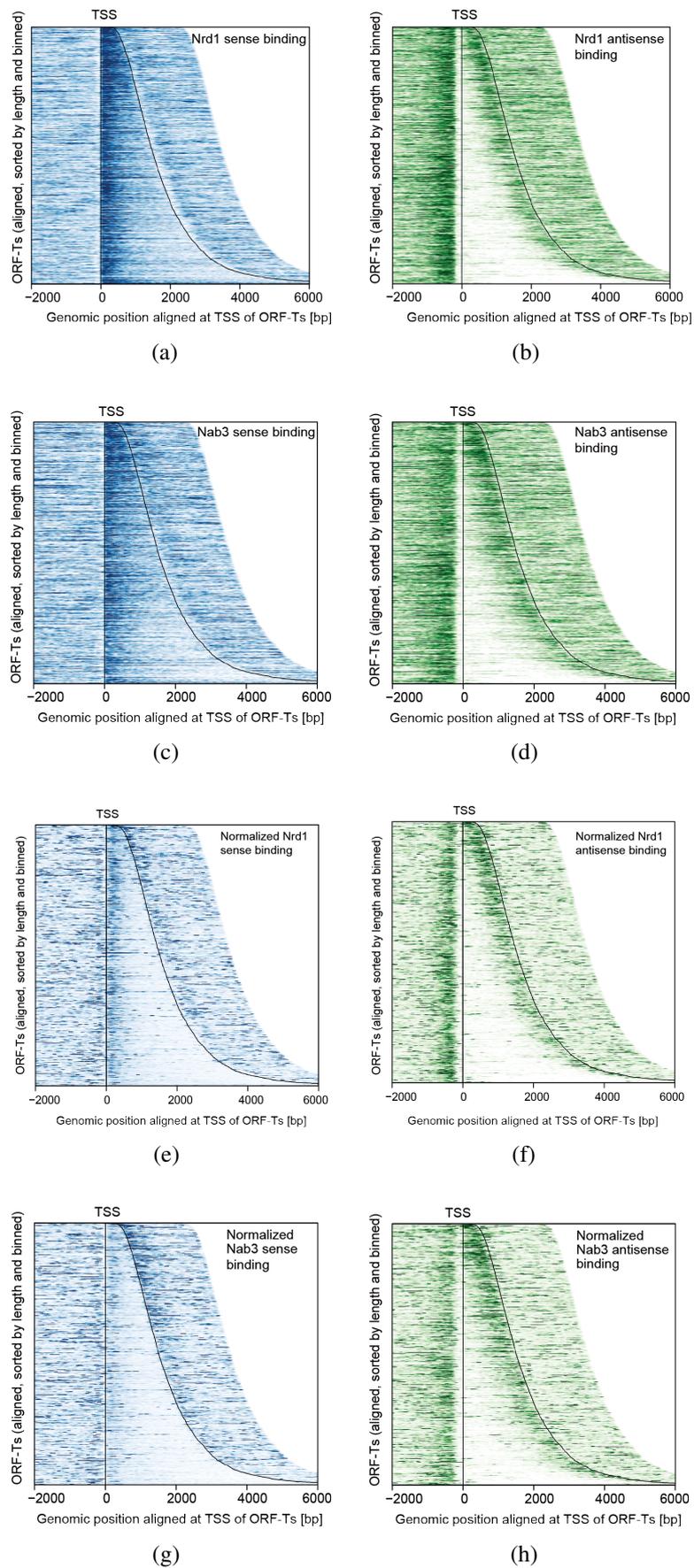
The preferred binding of Nrd1 and Nab3 to divergent and antisense ncRNAs suggested that mRNAs contain less Nrd1-binding motifs than ncRNAs. To investigate this, we calculated an apparent Nrd1-binding affinity for each of the 256 tetramer motifs from their relative frequency near the Nrd1-binding sites observed by PAR-CLIP. We then calculated apparent Nrd1 binding affinities along the yeast genome. Strikingly, mRNAs were markedly depleted in additive apparent Nrd1-binding affinity, a factor of 1.5 lower than antisense ncRNAs (Figure 3.23b). Intergenic ncRNAs were also enriched in overall Nrd1-binding affinity with respect to mRNAs, by a factor of 1.3 (Figure 3.23c). The real binding preference in vivo is likely much higher than the observed differences in apparent binding affinity because multiple copies of Nrd1 are likely recruited in a cooperative manner, and because Nrd1 forms a complex with Nab3, which binds neighboring sites in the RNA and likely contributes to cooperative effects. Consistent with this, Nab3 also showed an increased apparent binding affinity for ncRNA, with values similar to that for Nrd1.

These results indicate that Nrd1 generally binds with higher affinity to ncRNAs than to mRNAs, because the preferred Nrd1-binding motifs are depleted from mRNAs. These observations may be the result of two evolutionary processes. Nrd1 may have evolved to bind RNA motifs that do not occur in coding mRNA, and yeast genes may have evolved to preferentially use codons that do not give rise to Nrd1 motifs. The higher motif occurrence

---

Figure 3.21 (*following page*): Nrd1 preferentially binds divergent and antisense ncRNAs (a) Heat map of Nrd1 RNA-binding sites as derived by PAR-CLIP in sense direction for all ORF-Ts. ORF-Ts were sorted by length and aligned at their TSS [Xu et al., 2009]. The curved line on the right represents the pA sites. Strength of binding is coded from white (no binding) to dark blue (strong binding). (b) Heat map of Nrd1 RNA binding sites as in 3.21a but for the antisense direction. (c) Heat map of Nab3 RNA binding sites as in 3.21a for the sense direction. (d) Heat map of Nab3 RNA binding sites as in 3.21a for the antisense direction. (e) Expression-normalized heat map of Nrd1 RNA-binding sites as derived by PAR-CLIP in sense direction for all ORF-Ts. ORF-Ts were sorted by length and aligned at their TSS [Xu et al., 2009]. The curved line on the right represents the pA sites. Strength of binding is coded from white (no binding) to dark blue (strong binding). (f) Expression-normalized heat map of Nrd1 RNA-binding sites as in 3.21e but for the antisense direction. (g) Expression-normalized heat map of Nab3 RNA-binding sites as in 3.21e for the sense direction. (h) Expression-normalized heat map of Nab3 RNA-binding sites as in 3.21e for the antisense direction.

### 3.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis



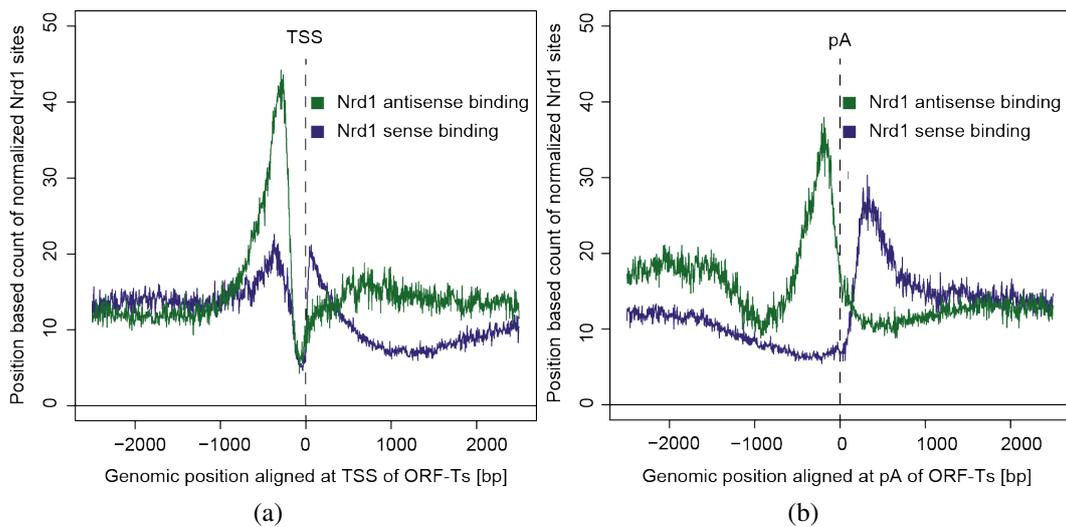


Figure 3.22: Accumulated Nrd1 binding around TSS and pA (a) Expression-normalized Nrd1 RNA-binding site distribution around the TSS of all ORF-Ts for the sense (blue) and antisense direction (green) with respect to ORF-Ts. The y-values are proportional to the occupancy of Nrd1 on the transcripts. (b) As in 3.22a but around the pA site of all ORF-Ts.

in ncRNAs explains why ncRNAs are preferred over mRNA transcripts as substrates for Nrd1-dependent termination. Higher motif occurrence and PAR-CLIP site density was also detected downstream of ORF-Ts, which can account for a known fail-safe mechanism for mRNA termination [Rondon et al., 2009]. A positive control is provided by analysis of sn/snoRNAs, which contained on average 7.7 times more PAR-CLIP sites per nucleotide compared to ORF-Ts (Figure 3.24a), and a very high density of Nrd1/Nab3-binding motifs (Figure 3.24b).

### 3.2.8 Yeast promoters are generally bidirectional

Of all 5123 ORF-Ts in the annotation file we used [Xu et al., 2009] 1712 are divergent ORF-T pairs with a maximum distance of 452 bp between them. Of the remaining 3411 ORF-Ts we detected at least two PAR-CLIP sites upstream and antisense within 452 bp for 1898 ORF-Ts (for 2272 ORF-Ts at least one PAR-CLIP site was observed). These 1898 ORF-Ts had no other ORF-T annotated upstream and divergent within 452 bp. Since PAR-CLIP signals reflect RNA cross-links, these sites show that divergent ncRNAs must have existed. Based on these data, a total of 3610/3984 (70%/78%) of *Sc* promoters are bidirectional, assuming the detection of at least two/one PAR-CLIP sites/site on the divergent ncRNA. Using the same criteria 845 of the 3411 non-divergent ORF-Ts had an antisense NUT assigned. This finding is consistent with the PAR-CLIP results but reflects

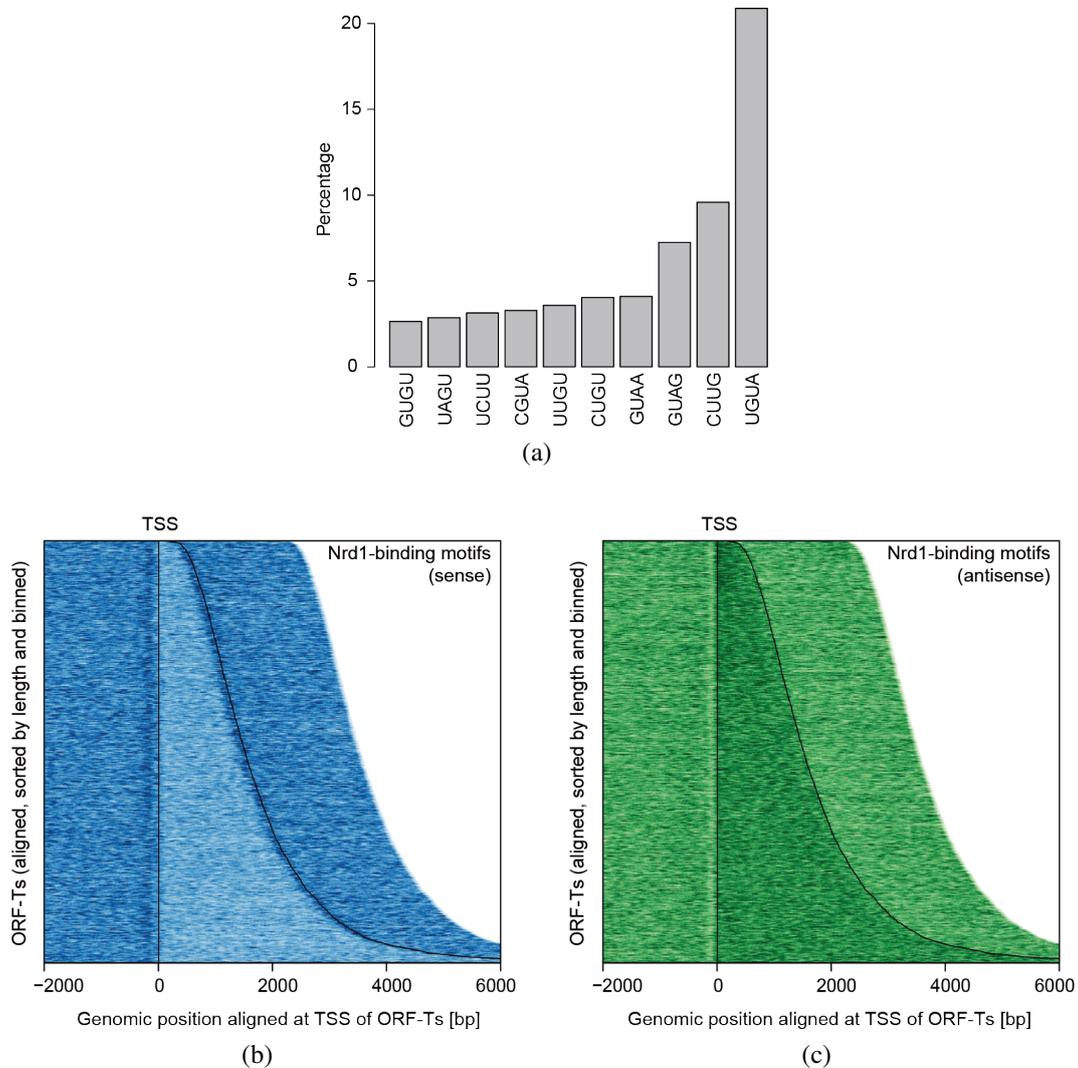


Figure 3.23: Nrd1 RNA binding motifs (a) Barplot shows the top 10 Nrd1 tetramers with the highest odd-ratios in percentage of contribution to the PAR-CLIP binding event. (b) Heat map of tetrameric motif occurrence and binding preference of Nrd1 in sense direction for all ORF-Ts. The occurrence of tetramers was weighted by the likelihood of Nrd1 binding. ORF-T alignment and coloring like in panel (Figure 3.21a). (b) As in (Figure 3.23b) but in antisense direction.

### 3.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis

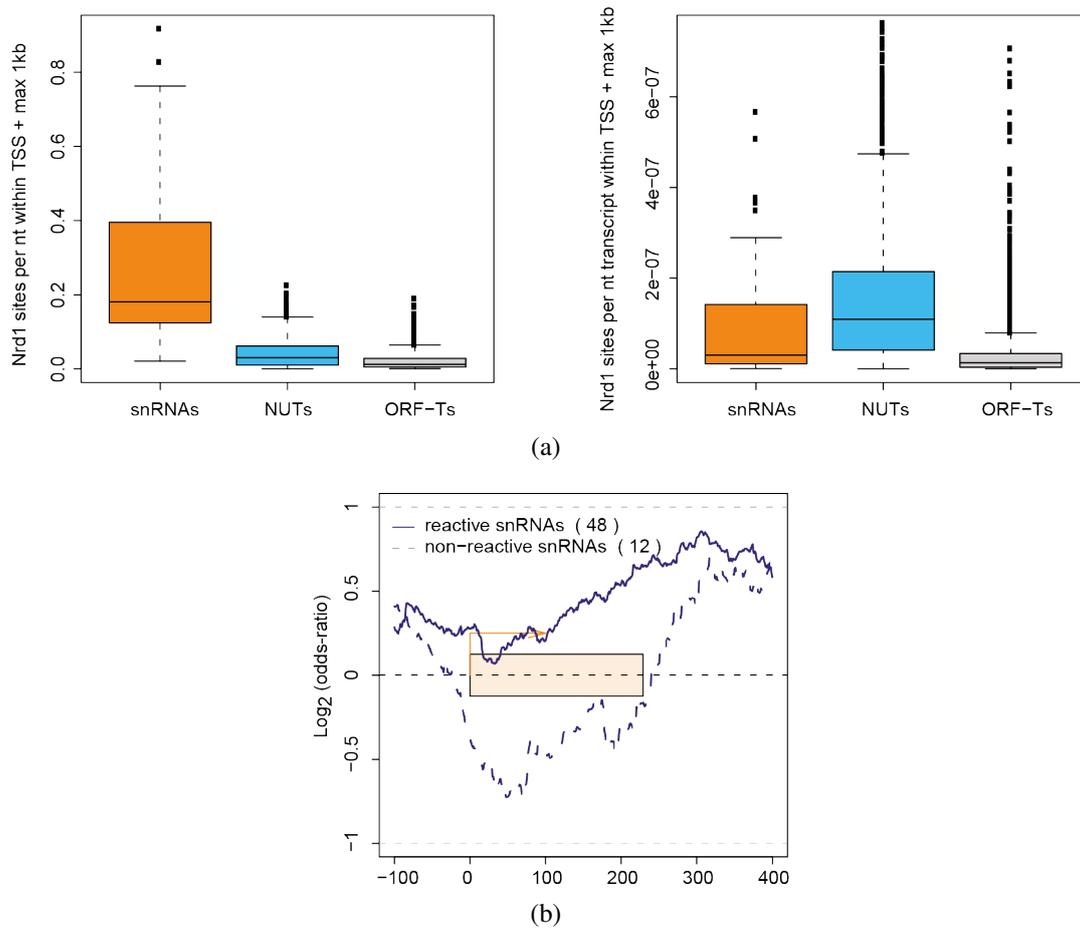


Figure 3.24: Nrd1 RNA binding sites are enriched on sn/snoRNAs and NUTs (a) Left panel: Nrd1 sites per nucleotide in the first 1000 bp of sn/snoRNAs, NUTs and ORF-Ts. Right panel: Nrd1 sites per nucleotide and transcript in the first 1000 bp of sn/snoRNAs, NUTs and ORF-Ts. We detected on average 8.2 times more Nrd1 sites per nt transcript in the first 1000 bp of NUTs compared to ORF-Ts and 2.3 times more sites per nt transcript on sn/snoRNAs compared to ORF-Ts. (b) Reactive sn/snoRNAs show an enrichment in Nrd1 preferred tetramers in their downstream region. Plot shows Nrd1 tetrameric motif preferences (odds-ratios) on a log-scale for reactive and non-reactive sn/snoRNAs scaled to a common length and aligned at the TSS.

a higher sensitivity of PAR-CLIP over 4tU-Seq for detecting short-lived ncRNAs.

#### 3.2.9 Nrd1 is required for promoter directionality

The above results provide strong evidence that yeast promoters are generally bidirectional and generate both mRNA and divergent ncRNA, and that the divergent ncRNA preferentially binds Nrd1. This is consistent with the idea that selective Nrd1-dependent termina-

tion of divergent ncRNA transcription is important for setting promoter directionality. To investigate whether Nrd1 depletion leads to a partial loss of promoter directionality, we plotted sense and antisense 4tU-Seq signals around all TSSs of ORF-Ts (Figure 3.25). This revealed that Nrd1 depletion leads to a two-fold average increase in divergent transcription, demonstrating a partial loss of promoter directionality. Nrd1 depletion also increases antisense transcription in ORF-Ts and sense transcription upstream of ORF-Ts, consistent with a global transcriptome surveillance mechanism that restricts ncRNA synthesis by Nrd1-dependent termination.

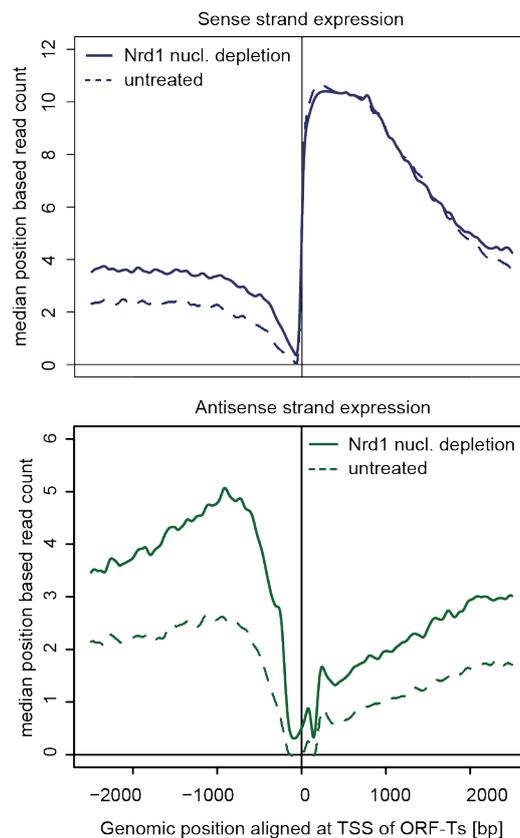


Figure 3.25: Loss of transcription directionality upon nuclear depletion of Nrd1. Top panel: Sense strand expression (median position based read count) of ORF-Ts aligned at their TSS measured by 4tU-Seq before (dashed line) and after (solid line) nuclear depletion of Nrd1. Bottom panel: as on top but with antisense strand expression.

#### 3.2.10 Antisense ncRNA synthesis can down-regulate transcription

To investigate whether defects in ncRNA termination induced by nuclear depletion of Nrd1 can deregulate genome transcription, we tested whether NUT transcription antisense to ORF-Ts influences sense transcription (Figure 3.26a). Antisense transcription was shown to regulate several yeast loci [Camblong et al., 2007, Hongay et al., 2006, Houseley

### 3.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis

et al., 2008, Xu et al., 2011]. A total of 942 NUTs were antisense to annotated ORF-Ts (antisense NUT class). We plotted changes in 4tU-Seq signals in ORF-Ts over changes in antisense signals in the promoter region of ORF-Ts upon nuclear depletion of Nrd1 (Figure 3.26b).

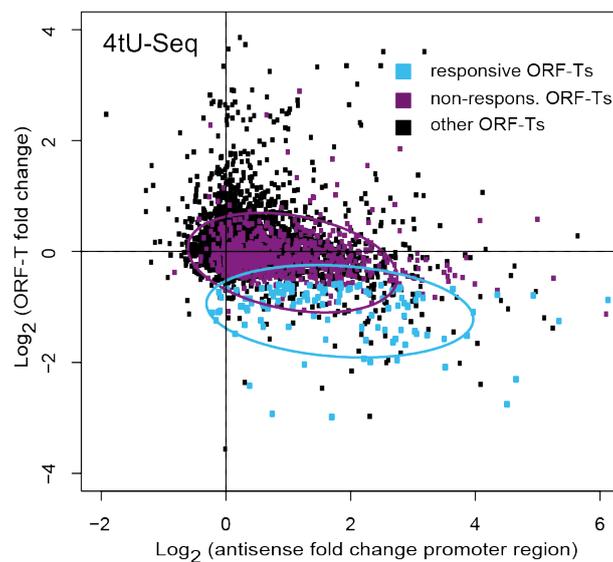
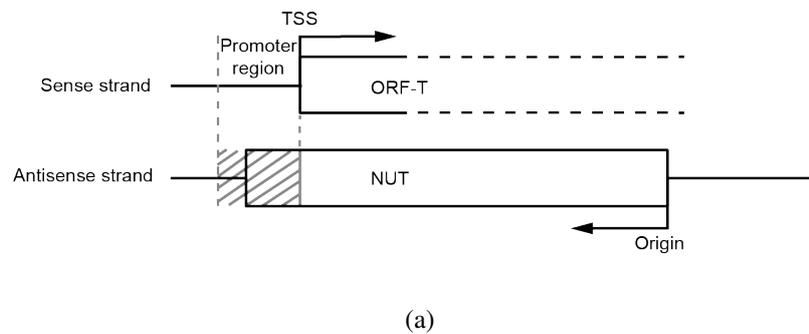


Figure 3.26: Antisense NUTs can interfere with ORF-T transcription (a) Definition of antisense NUT class. A NUT belongs to the antisense class when its origin of is downstream of the overlapping ORF-T and the overlap is at least 200 bps. Dashed vertical lines delineate the ORF-T promoter region (-200 bp to TSS). (b) Some ORF-Ts are repressed upon antisense NUT transcription. Fold-change of ORF-T 4tU-Seq signal versus fold-change in antisense signal over the ORF-T promoter region (compare 3.26a). ORF-Ts with a decrease in their 4tU-Seq signal (114 responsive ORF-Ts) are in cyan, 828 non-responsive ORF-Ts are in purple, and all others are in black.

We found that increasing levels of antisense transcription in the promoter region of ORF-Ts correlates with down-regulation of ORF-T transcription. In total 114 (56%) of the 202 significantly down-regulated genes had an annotated antisense NUT that explained down-

regulation. The 114 ORF-Ts that were down-regulated (responsive) showed stronger antisense transcription, with the most pronounced difference in the promoter region of ORF-Ts (Figure 3.27a). Antisense NUTs of responsive ORF-Ts extended further into the promoter region than those of the 828 non-responsive ORF-Ts. (Figure 3.27b). These results are consistent with previous findings [Xu et al., 2011] and show that antisense NUT transcription can down-regulate sense ORF-T transcription when it reaches a certain level, apparently by interfering with transcription initiation at the promoter, maybe due to a steric interference of converging Pol II enzymes as observed *in vitro* [Hobson et al., 2012].

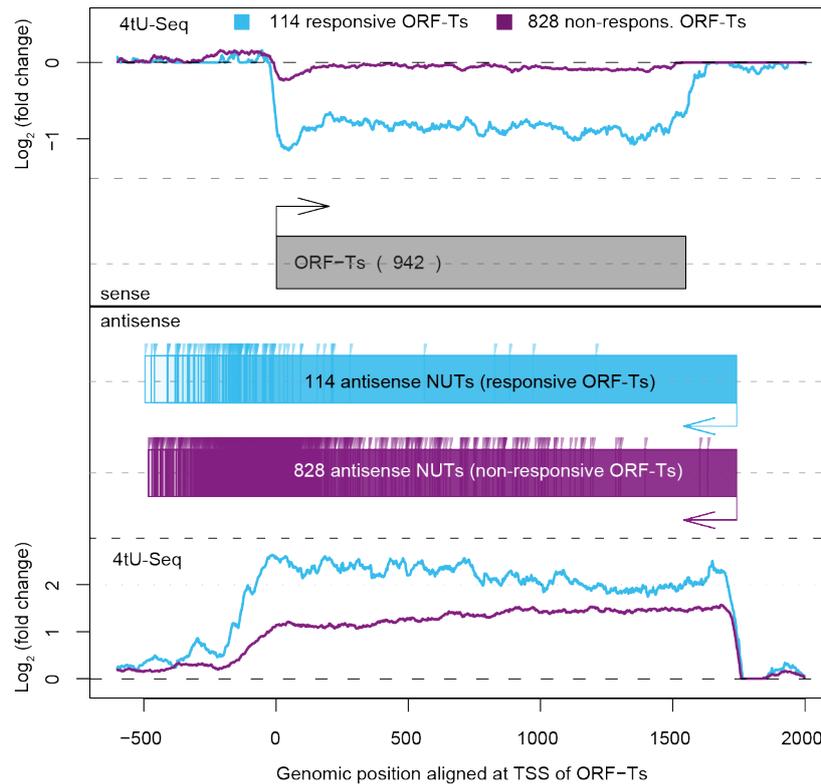
### 3.2.11 Upstream ncRNA synthesis can up-regulate transcription

Another possible mechanism for deregulation of ORF-T transcription involves upstream synthesis of ncRNAs, which was shown for selected genes to interfere with ORF-T transcription by Nrd1-dependent mechanisms [Colin et al., 2011]. Of all NUTs, one third is found upstream of ORF-Ts with a median NUT origin approximately 1000 bp upstream of ORF-Ts. We selected all NUTs upstream of ORF-Ts with a maximum distance of 100 bp between the ORF-T and the NUT (sense NUT class, 459 NUTs) (Figure 3.28a). Downstream of sense NUTs, 106 ORF-Ts were up-regulated 3-4-fold in our 4tU-Seq data, whereas the remaining 353 ORF-Ts were unchanged (Figure 3.28b). The up-regulated ORF-Ts showed lower median RNA synthesis than unchanged ORF-Ts (Figure 3.28c), and the associated upstream NUTs showed higher levels than the remaining NUTs in the sense class (Figure 3.28c). Upstream NUT synthesis was responsible for up-regulation of 37% of a total of 287 significantly up-regulated ORF-Ts. Only in 28 exceptional cases, when the ORF-T was transcribed at high levels, NUT synthesis repressed ORF-T transcription slightly (Figure 3.28d). Taken together, strong ncRNA synthesis upstream of ORF-Ts can up-regulate weak ORF-T transcription.

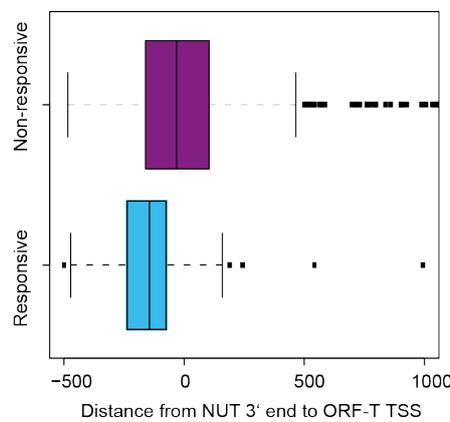
### 3.2.12 Termination of ncRNA synthesis prevents transcription interference

All above results suggested that early Nrd1-dependent termination of aberrant ncRNAs prevents genome deregulation by NUT synthesis. To further investigate this, we determined termination sites of aberrant ncRNAs by mapping Pol II over the genome before and after nuclear depletion of Nrd1. We used chromatin immunoprecipitation (ChIP) as described [Mayer et al., 2010] coupled to deep sequencing (ChIP-Seq, Material and Meth-

### 3.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis



(a)



(b)

Figure 3.27: Antisense NUTs interfere with corresponding ORF-Ts at the promoter (a) Top: Distribution of  $\log_2$  fold-changes in 4tU-Seq signal upon nuclear depletion of Nrd1 for responsive (cyan) and non-responsive ORF-Ts (purple) defined in (Figure 3.26b). Bottom: 4tU-Seq signals for antisense NUTs corresponding to responsive (cyan) and non-responsive (purple) ORF-Ts. All 942 ORF-Ts are scaled to a median length and aligned at their TSS. NUTs are aligned at their median origin (blue and purple boxes with arrows). Vertical lines indicate NUT 3' ends. (b) Extension of antisense NUTs into the promoter region of their corresponding sense ORF-T correlates with transcription repression. Color code as in (B). The distance between the antisense NUT 3' end and the TSS of its corresponding sense TSS is plotted on the horizontal axis.

ods). The obtained ChIP-Seq replicates correlated very well and had a high correlation with RNA synthesis monitored by 4tU-Seq, showing that Pol II ChIP occupancy is a good proxy for transcription activity (Figure 3.29).

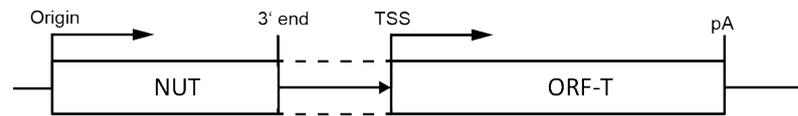
We analyzed changes in Pol II occupancy upon nuclear depletion of Nrd1 by determining changes in an Escape Index [Brannan et al., 2012]. For every NUT transcription unit, we calculated an Escape Index (EI) as the ratio of Pol II occupancy fold-change in the promoter-distal versus the promoter-proximal region, where the promoter proximal region was defined as the region between NUT origin and transcript termination site (TTS) (Figure 3.30a). An increased EI after nuclear depletion of Nrd1 indicates defective termination because more Pol II moves to the promoter-distal region, and this was also observed in averaged Pol II occupancy profiles for all NUTs (Figure 3.30b).

We estimated the TTS for a ncRNA as the point downstream of which the density of Pol II increases upon nuclear depletion of Nrd1. Technically, we determined the point downstream of the NUT origin (max. distance 1000 bp) at which the profile of log<sub>2</sub> fold change in Pol II occupancy was best approximated by a two-segment piecewise constant function (Figure 3.30) (Section 2.2.5.4). We determined the TTS for 283 ncRNAs that upon nuclear depletion of Nrd1 give rise to sense NUTs, which are upstream of ORF-Ts that contain a mapped PIC [Rhee and Pugh, 2012]. We then calculated the distance of each TTS to the PIC of the down-stream ORF-T (Figure 3.30d). This revealed that ncRNA synthesis is generally terminated before the transcribing polymerase would clash with the PIC at the downstream ORF-T, apparently to prevent transcription interference.

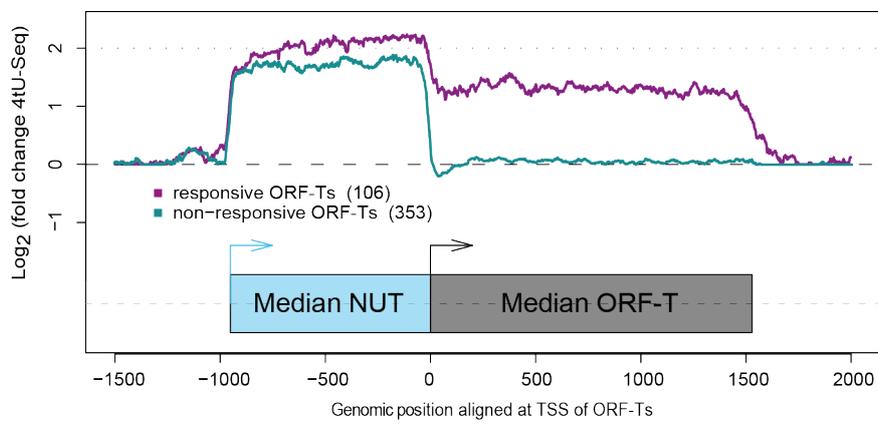
---

Figure 3.28 (*following page*): Sense NUTs can deregulate transcription of down-stream ORF-Ts. (a) Definition of the sense NUT class. A NUT belongs to the sense class when its origin is upstream of the TSS of the downstream ORF-T and the distance between the NUT and the ORF-T was not more than 100 200 bps. (b) A fraction of sense class NUTs up-regulates downstream ORF-T transcription. Median log<sub>2</sub> fold-change in 4tU-Seq signal upon nuclear depletion of Nrd1 was plotted for responsive (purple) and non-responsive (green) ORF-Ts of the sense NUT class. 3.28c ORF-Ts that are responsive to upstream sense NUT transcription are weakly transcribed, whereas the corresponding NUTs are highly transcribed. Color code as in 3.28b. (d) Median expression fold-change shown for 28 repressed ORF-Ts of the sense NUT class. The blue box below represents the median sense NUT followed by the ORF-T (black box). ORF-Ts were aligned at the TSS and the distribution of log<sub>2</sub> fold-changes in 4tU-Seq signal upon nuclear depletion of Nrd1 for the 28 down-regulated genes is shown. (red shaded area; black line indicates median, grey lines indicate first and third quartiles).

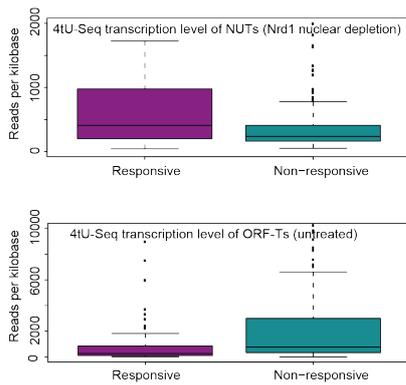
### 3.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis



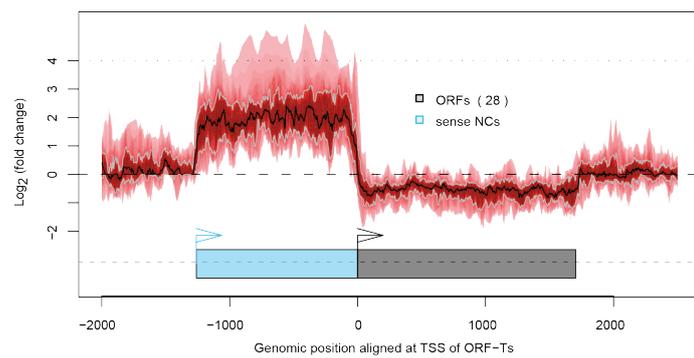
(a)



(b)



(c)



(d)

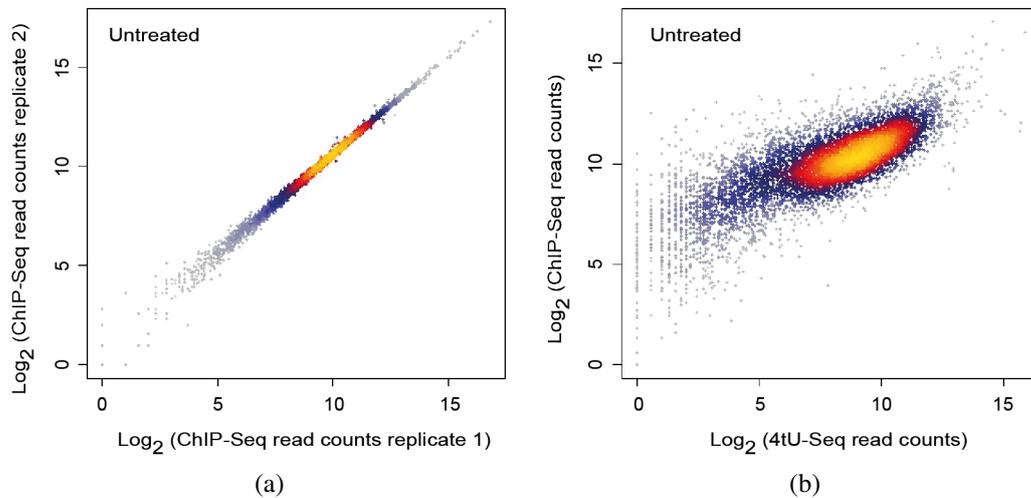
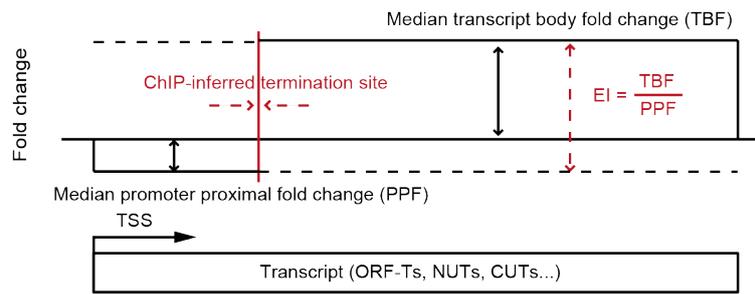


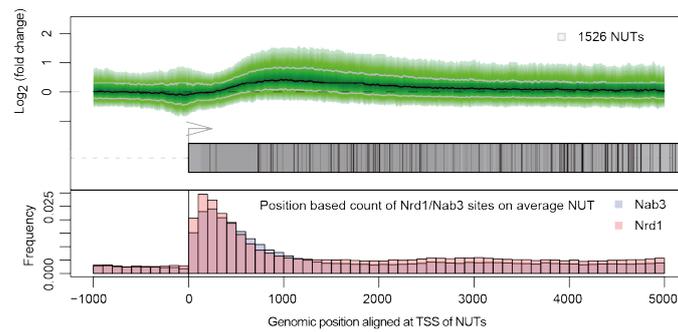
Figure 3.29: Reproducibility assesment of ChIP-Seq and 4tU-Seq (a) Comparison of replicate measurements for ChIP-Seq of the untreated sample. The scatterplot compares read counts of ORF-Ts, SUTs and CUTs. Spearman correlation is 1. (b) Scatterplot of averaged read counts of ORF-Ts, SUTs and CUTs of 4tU-Seq versus ChIP-Seq measurements. Spearman correlation is 0.76.

Figure 3.30 (*following page*): ChIP-seq inferred termination sites reveal termination of ncRNAs before the promoter of down-stream ORF-Ts (a) Scheme illustrating the determination of termination site and Escape Index (EI) from ChIP-Seq data. EIs were calculated as the median fold-change in the transcribed gene body divided by the median fold-change in the proximal TSS region upon nuclear depletion of Nrd1. (b) ncRNA transcription is not terminated upon nuclear depletion of Nrd1. Median log<sub>2</sub> Pol II occupancy fold-change upon nuclear depletion of Nrd1 distribution of Pol II occupancies upon Nrd1 depletion is shown for all NUTs. Transcripts were aligned at the TSS (grey box) and each TTS is depicted by a black line. Nrd1 and Nab3 RNA binding sites as determined by PAR-CLIP are depicted on the bottom and peak within the first 400 bp. (c) Pol II occupancies around the CUT280 locus measured by ChIP-Seq in wild-type conditions (green) and upon nuclear depletion of Nrd1 (blue). The position of sign change in the occupancy fold-change difference profile (red) defines the termination site of the CUT (red vertical line). RNA-binding sites of Nrd1 and Nab3 as determined by PAR-CLIP are shown as green and brown vertical lines over the blue bar at the bottom. (d) Sense NUTs are generally terminated before the promoter of downstream ORF-Ts. The plot shows the frequency of distances from the ncRNA termination site (vertical red line) to the PIC location defined by ChIP-Exo of TFIIB [Rhee and Pugh, 2012].

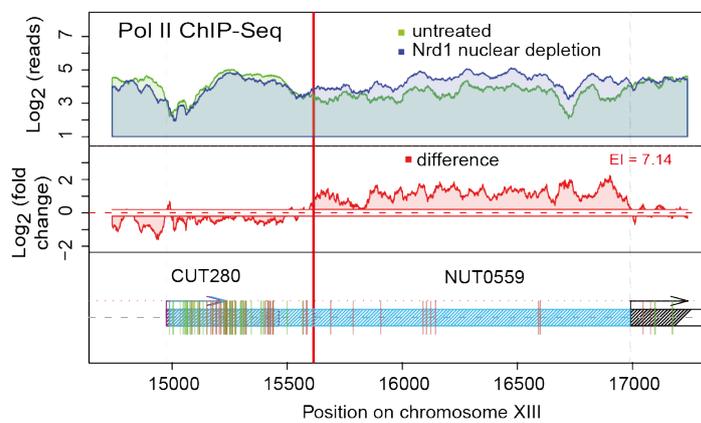
### 3.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis



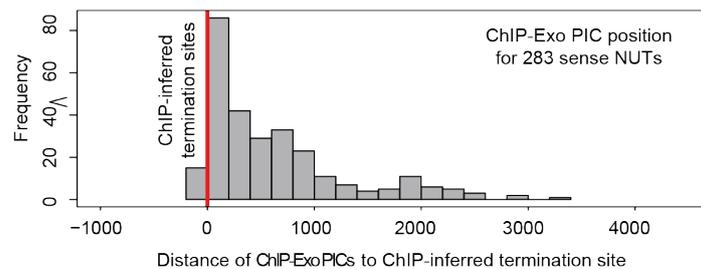
(a)



(b)



(c)



(d)

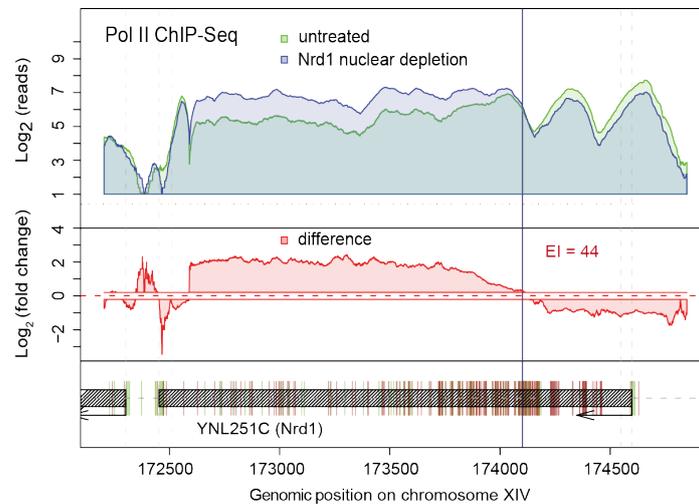
### 3.2.13 Transcription attenuation is rare

Visual inspection of our Pol II ChIP-Seq data at protein-coding genes that are controlled by Nrd1-dependent attenuation [Arigo et al., 2006b, Steinmetz et al., 2006a] revealed an apparent release of Pol II into promoter-proximal regions after nuclear depletion of Nrd1 (Figure 3.31a). To systematically search for genes that are controlled by transcription attenuation, we extended the EI analysis of our ChIP-Seq data to all ORF-Ts. This revealed that transcription attenuation does not generally occur under our experimental conditions (Figure 3.31b).

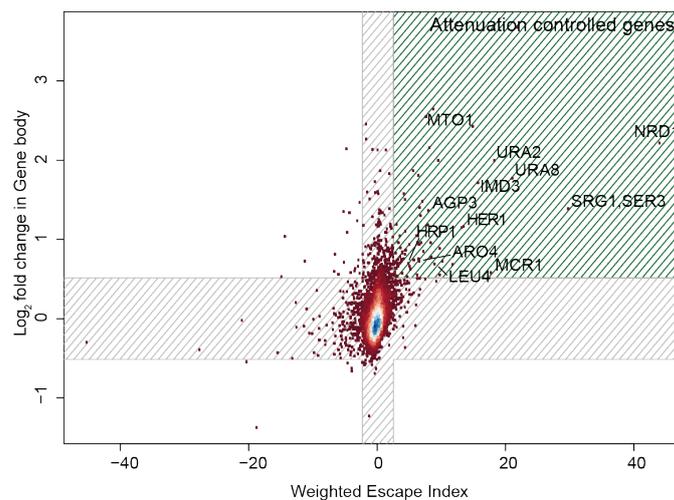
Only 32 ORF-Ts were classified as attenuated genes that fulfilled the following three criteria. First, weighted EIs (Section 2.2.5.4) had to be greater than 2.5 upon nuclear depletion of Nrd1. Second, Pol II occupancy changes in the gene body had to be greater than 1.4-fold. Third, ORF-T transcription had to be up-regulated at least 1.25-fold in 4tU-Seq data (adjusted p-value 0.1). The attenuated genes were generally involved in biosynthetic amino acid and metabolic processes (not shown).

Alignment of the 32 selected ORF-Ts at their TSS showed that the average Pol II occupancy was slightly decreased in the promoter-proximal region after nuclear depletion of Nrd1, likely reflecting a loss of early Pol II termination intermediates (Figure 3.32a). Average Pol II occupancy was however increased from around 400 bp downstream of the TSS (Figure 3.32a), reflecting an increased density of Pol II in promoter-distal regions after attenuation release. Further consistent with attenuation control, PAR-CLIP detected a high density of Nrd1- and Nab3-binding sites in the promoter-proximal RNA region of these 32 genes, which was not observed for other ORF-Ts (Figure 3.32a). We conclude that under optimum growth conditions only few genes are controlled by Nrd1-dependent attenuation, and that the main function of the Nrd1 pathway is to suppress aberrant ncRNA transcription. It remains possible that more genes are under attenuation control during non-optimum growth conditions, such as cell wall stress [Kim and Levin, 2011].

### 3.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis



(a)



(b)

Figure 3.31: Nrd1-dependent transcription attenuation is rare (a) log<sub>2</sub> Pol II reads from ChIP-Seq around the Nrd1 gene locus before (green) and after nuclear depletion of Nrd1 (blue) and calculated log<sub>2</sub> differences in ChIP signal (red). The vertical black line indicates the derived early termination/attenuation site. RNA-binding sites of Nrd1 and Nab3 as determined by PAR-CLIP are shown as green and brown vertical lines at the bottom. (b) Attenuation of mRNA genes upon nuclear depletion of Nrd1 is rare under optimum growth conditions. Only 32 genes show de-attenuation upon nuclear depletion of Nrd1, as indicated by a weighted EI 2.5 and a 1.4-fold change in ChIP-Seq signal (green hatched region).

### 3.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis

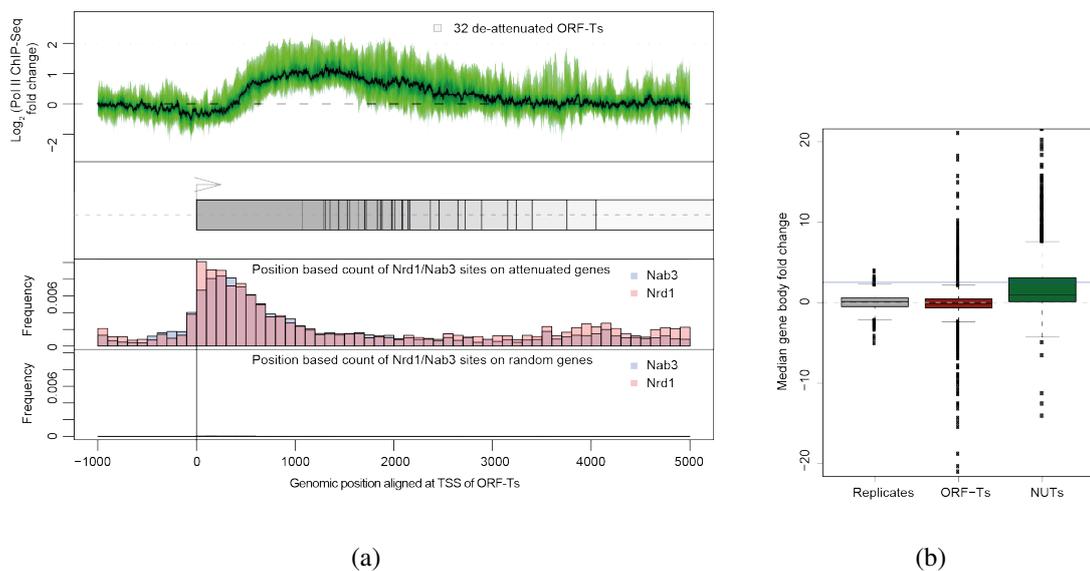


Figure 3.32: De-attenuation leads to Pol II accumulation (a) log<sub>2</sub> Pol II reads from ChIP-Seq around the Nrd1 gene locus before (green) and after nuclear depletion of Nrd1 (blue) and calculated log<sub>2</sub> differences in ChIP signal (red). The vertical black line indicates the derived early termination/attenuation site. RNA-binding sites of Nrd1 and Nab3 as determined by PAR-CLIP are shown as green and brown vertical lines at the bottom. (b) Attenuation of mRNA genes upon nuclear depletion of Nrd1 is rare under optimum growth conditions. Only 32 genes show de-attenuation upon nuclear depletion of Nrd1, as indicated by a weighted EI > 2.5 and a > 1.4-fold change in ChIP-Seq signal (green hatched region).

# Chapter 4

## Discussion

### 4.1 cDTA reveals a mutual feedback loop between mRNA transcription and degradation

A systematic investigation of gene expression requires quantitative monitoring of cellular mRNA metabolism. In particular, a technique is required to quantify absolute mRNA synthesis and decay rates on a genome scale upon genetic perturbation. Here we provide such a technique which we refer to as comparative Dynamic Transcriptome Analysis (cDTA). cDTA is based on non-perturbing metabolic RNA labeling in mutant and wild type budding yeast cells, and the use of fission yeast cells as an internal standard. cDTA is a non-perturbing method for monitoring mRNA turnover and supersedes conventional methods, which require transcription inhibition, resulting in a stress response and perturbation of mRNA metabolism.

cDTA improves our previous DTA protocol [Miller et al., 2011] in several respects. First, cDTA provides reliable estimates of the absolute synthesis and decay rates, thereby allowing for a direct comparison of rates between different yeast strains. Second, cDTA uses 4tU instead of 4sU for RNA labeling, allowing for standard media and abolishing the need for a nucleoside transporter. Third, cDTA requires only two instead of three microarray measurements per rate estimation. As an immediate result, cDTA revealed that *Sp* and *Sc* cells have similar synthesis rates, but *Sp* RNAs have about five-fold longer mRNA half-lives, leading to similar cellular mRNA concentrations despite a different cell volume.

Application of cDTA to *Sc* cells expressing a Pol II point mutant which elongates mRNA slowly *in vitro*, showed that mRNA elongation is a critical determinant for mRNA synthe-

sis in growing cells *in vivo*. It also revealed that cells compensate for low synthesis rates by lowering decay rates, thus stabilizing mRNAs and buffering their levels. Application of cDTA to two mutant strains that lack either one of the two catalytic subunits of the mRNA deadenylase complex Ccr4-Not showed not only the expected defect in mRNA degradation but also a compensatory decrease in mRNA synthesis, also leading to a buffering of mRNA levels. This indicates the existence of a feedback loop that connects mRNA synthesis and degradation, and serves to buffer mRNA levels. These results support published evidence for a global control of mRNA levels depend on cell size [Zhurinsky et al., 2010]. This global control of mRNA levels occurs despite the separation of mRNA synthesis and degradation into nuclear and cytoplasmic compartments.

The mechanisms underlying the synthesis-decay feedback loop and the buffering of mRNA levels are unclear. The feedback loop may be a result of a physical and functional coupling between the various parts of mRNA metabolism. Transcription is coupled to mRNA processing and export [Maniatis and Reed, 2002], and translation is coupled to mRNA degradation [Bregues et al., 2005, Coller and Parker, 2004, Coller and Parker, 2005, Hu et al., 2009]. There is also evidence that nuclear and cytoplasmic mRNA metabolism are linked. The Pol II subcomplex Rpb4/7p shuttles between the nucleus and cytoplasm [Selitrennik et al., 2006], and is involved in transcription [Edwards et al., 1991] and mRNA translation and degradation [Harel-Sharvit et al., 2010, Lotan et al., 2005, Lotan et al., 2007]. The Ccr4-Not complex is involved in mRNA degradation [Tucker et al., 2002], but also in transcription [Collart, 2003, Collart and Timmers, 2004, Kruk et al., 2011, Liu et al., 1998]. From an extension of our kinetic model of mRNA turnover, we propose that the feedback loop is established by a factor that acts as degradation enhancer and transcription inhibitor. It is thus unlikely that factors that act positively on transcription, such as Rpb4/7p and the Ccr4-Not complex, are the feedback factors, although the validity of our model's assumptions remains to be shown.

## 4.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis

We have investigated the global function of Nrd1 using 4tU-Seq, Pol II ChIP-Seq as well as PAR-CLIP and motif analyses of Nrd1 and Nab3. We show that ncRNAs generally originate from NDRs in the yeast genome. Yeast promoters are generally bidirectional, generating divergent ncRNA which originate 150-200 bps upstream of the TSS of the mRNA gene. We show that ncRNA synthesis is generally restricted by Nrd1-dependent termination. A defect in ncRNA transcription termination can lead to genome deregula-

## 4.2 Transcriptome surveillance by selective termination of non-coding RNA synthesis

tion by antisense repression and by transcription interference. We provide evidence that termination of ncRNA transcription is the main function of Nrd1, and attenuation control at mRNA genes is rare. Nrd1 preferentially binds to ncRNAs that frequently contain Nrd1-binding motifs, whereas mRNAs are depleted for these motifs and generally escape Nrd1 action. We conclude that the Nrd1-dependent termination pathway serves as a mechanism for global surveillance of the transcriptome which is based on recognition and removal of polymerases that produce aberrant nascent ncRNA Figure 4.1.

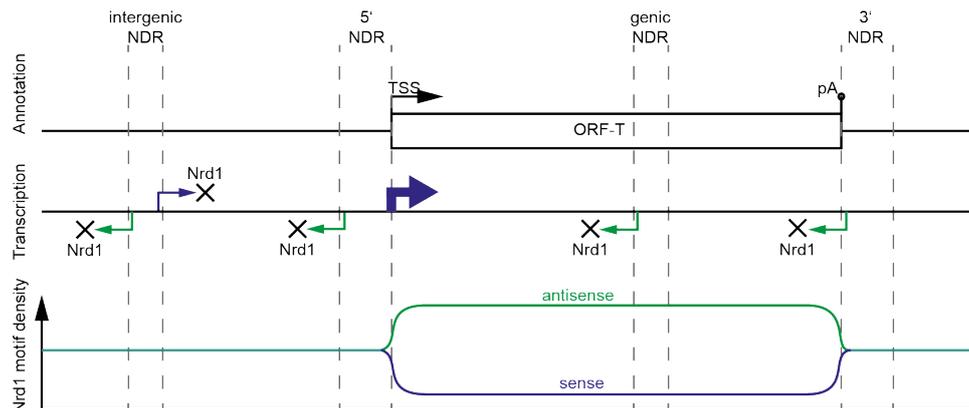


Figure 4.1: Transcription of Pol II generally initiates bidirectionally in NDRs, leading to aberrant ncRNA synthesis. Transcripts that are antisense to ORF-Ts possess a higher density of tetramers with high Nrd1 and Nab3 binding affinities and are preferentially bound by Nrd1 and Nab3, leading to their early termination and rapid degradation. Intergenic transcripts have intermediate levels of Nrd1/Nab binding affinity, are bound by Nrd1 and Nab3 at intermediate levels, and are targeted for early termination. mRNAs originating from ORF-Ts show the lowest density of high-affinity sites, are only weakly bound by Nrd1 and Nab3, and are generally not attenuated by the Nrd1 pathway.

Previous studies that detected divergent RNA transcripts observed bidirectional transcription at about one third of yeast promoters [Neil et al., 2009, Xu et al., 2009]. Here we could observe the bidirectional nature of essentially all yeast promoters apparently because we trapped the short-lived divergent transcript by cross-linking it to its recognition factor Nrd1. Contingent upon the generality of bidirectional promoters, we suggest that transcription directionality is globally achieved by selective termination of divergent ncRNA transcription. Selective termination may be explained by the difference in the occurrence of Nrd1- and Nab3-binding motifs in ncRNA versus mRNA, in particular because Nrd1 and Nab3 can bind cooperatively to RNA [Carroll et al., 2007]. Nrd1 recruitment to early ncRNA transcription complexes may be facilitated by phosphorylation patterns in the C-terminal repeat domain of Pol II [Kubicek et al., 2012, Singh et al., 2009] and by chromatin modifications that are directional [Rando and Chang, 2009, Seila et al., 2008]. The formation of mRNA gene loops further contributes to transcription directionality [Tan-Wong et al., 2012].

Previous studies of ncRNA transcription in yeast used strains with a deletion of the non-essential genes *rrp6* and *xrn1* that encode for a nuclear exosome subunit [Neil et al., 2009, Xu et al., 2009] and a RNA exonuclease [van Dijk et al., 2011], respectively. In these strains, RNA degradation is defective, leading to a stabilization of ncRNAs that would otherwise be rapidly degraded. Our approach of depleting the essential factor Nrd1 provides complementary, more complete insights, because Nrd1 acts upstream of Rrp6 and Xrn1 and its binding to nascent RNA is thought to be the first step in ncRNA transcription termination. Nuclear depletion of Nrd1 gives rise to extended ncRNAs (NUTs) which are normally suppressed by Nrd1-dependent termination and are distinct from previously described ncRNAs because they are on average four times longer. Whereas the deletion of degradation factors does apparently not change ncRNA transcription activity, nuclear depletion of Nrd1 changes global transcription activity and deregulates the transcriptome. This is supported by an absence of significant overlap between genes that are differentially expressed after nuclear depletion of Nrd1 or *rrp6* deletion (not shown), and may explain why the deletion strains  $\Delta rrp6$  and  $\Delta xrn1$  show mild phenotypes, whereas *nrd1* deletion is lethal.

Our data indicate that in yeast it is unavoidable that transcription initiates when the genome is accessible, and that the resulting ncRNA transcription must be suppressed by selective early termination. Other species apparently have similar transcriptome surveillance systems. In *Escherichia coli*, a termination factor-dependent mechanism for suppression of antisense transcription has been described, and proposed to be related to the Nrd1 pathway [Peters et al., 2012]. A mechanism of selective transcription termination was very recently shown to restrict ncRNA transcription from mammalian and mouse bidirectional promoters [Almada et al., 2013, Ntini et al., 2013, Core et al., 2008, Seila et al., 2008]. In this study, termination was shown to be achieved by a different pathway, the pA-dependent pathway, but the general principle for achieving promoter directionality is conserved.

Although transcriptome surveillance suppresses most ncRNA production, some ncRNAs may escape rapid removal and exhibit a function. The overlap of NUTs with XUTs is smaller than for other ncRNA, and 66% of XUTs are antisense to mRNAs and may be involved in gene regulation [van Dijk et al., 2011]. In human cells, the fraction of ncRNAs that serve a cellular function is apparently much higher [Mercer et al., 2009]. It is also likely that the process of ncRNA transcription itself serves a cellular function such as the maintenance of a chromatin state or the enhanced recruitment of polymerase-associated factors for mRNA transcription.

## 4.3 Summary and future perspectives

This thesis covers two topics, global properties of yeast transcription and the selective termination of ncRNAs in *Sc*, which are of broad interest to the scientific community. cDTA was established to detect global changes in mRNA synthesis and degradation rates in yeast. Synthesis and degradation rates could be measured previously [Garcia-Martinez et al., 2004, Shalem et al., 2008, Grigull et al., 2004, Lam et al., 2001] but not on a global scale. Interestingly, global changes in eukaryotic transcription have been neglected until recently [Sun et al., 2012, Lin et al., 2012, Loven et al., 2012]. From studies in our lab it became clear, that *Sc* (and potentially also higher eukaryotes) has evolved mechanism to buffer the overall mRNA levels in the cell [Sun et al., 2012] (Sun et al., accepted in Mol Cell). One of the factors involved in coupling of transcription and decay is Xrn1 (Sun et al., accepted in Mol Cell) [Haimovich et al., 2013]. The Rpb4/7 complex has been postulated to be involved in mRNA synthesis and mRNA degradation coupling for almost a decade by the Choder lab but this hypothesis still needs further validation. To obtain more insights into the coupling mechanism, essential factors of the transcription and degradation pathways need to be tagged and conditionally depleted from either the cell or the nucleus. It should also be possible to inhibit the shuttling of certain factors between the nucleus and the cytoplasm or even to modify the anchor-away technique in order to tether proteins to specific locations upon rapamycin treatment.

Similar to our experiments, the Young lab counted cells and added a certain amount of "spike-ins" to the cells. This revealed that the oncogene *c-Myc* is an "amplifier" of gene expression and induces an overall increase in gene expression. It becomes clear from this, that normalization is crucial and false normalization can lead to the miss-interpretation of results from gene expression studies. It has therefore been proposed that key experiments for gene expression should be done using a normalization method that allows for the detection of global changes in gene expression, such as cDTA.

cDTA makes use of Affymetrix micro-arrays that contain probe sets for *Sc* and *Sp*. However, these microarrays only contain probes for mRNAs and differences in non-coding RNAs, antisense expression or polyadenylation-sites cannot be detected. Second generation sequencing is a superior technique (Section 1.3.1.1) and costs are still dropping, making it ever more attractive. Therefore, cDTA should be taken to the sequencing level, especially when studying higher eukaryotes. Metabolic labeling of RNAs has already been applied in human cells [Doelken et al., 2008, Rabani et al., 2011]. Global changes can be detected by cell counting and spike-ins as shown [Lin et al., 2012], but in this study no metabolic labeling was applied, thus, only steady state levels of transcripts could be detected. A functional cDTA-Seq protocol, that can be used to calculate transcription and

degradation rates, must therefore make use of either spike-ins that have been labeled *in vitro*, or extracted total RNA from an organism that has been metabolically labeled prior to RNA isolation.

In the second part of this thesis we used a combination of state of the art techniques to elucidate the role of Nrd1 in suppression of pervasive transcription. We show that divergent antisense transcripts originate from essentially all yeast promoters. This is in line with findings over the last 5 years, which show that promoters are not generally as restrictive as originally believed [Core et al., 2008, Seila et al., 2008, Xu et al., 2009, Neil et al., 2009, Preker et al., 2008]. However, these findings raised new questions: How can a cell distinguish a functional coding transcript from a non-coding transcript? In yeast this is solved by asymmetric distribution of Nrd1 and Nab3 binding motifs throughout the genome.

The Nrd1-Nab3-Sen1 complex has been intensively studied in the past [Mischo and Proudfoot, 2013]. However, it remains unclear, how the Nrd1 and Nab3 RNA binding event actually translates into a Pol II termination signal. One hypothesis is that Nrd1 and Nab3 binding to the RNA recruits Sen1 to the nascent RNA [Porrua and Libri, 2013b]. Sen1 will then terminate Pol II in a fashion similar to the Rho protein in bacteria [Porrua and Libri, 2013a]. Because all components of the Nrd1, Nab3 and Sen1 trimeric complex are essential, functional studies that disrupt the complex are complicated. Conditional disruption, or tethering of the complex through a modification of the anchor-away technique, would be ideally suited to test the hypothesis of Sen1 recruitment to the nascent RNA via Nrd1 and Nab3.

Very recently, the existence of a mechanism which regulates pervasive transcription at the step of transcription termination and RNA degradation has been described in mammals [Almada et al., 2013, Ntini et al., 2013]. These studies showed that poly(A) signals are asymmetrically distributed around the TSSs of ORF-Ts. These sites reflect effective termination signals for divergent Pol II while Pol II in the sense direction continues transcription [Ntini et al., 2013, Almada et al., 2013]. Additionally, to protect sense Pol II from termination, splice-site-related sequences bound by U1-snRNA can be found downstream of the TSS. Inhibition of U1-snRNP binding leads to similar amounts of transcription termination as for divergent Pol II molecules [Almada et al., 2013]. Although the selective termination of ncRNAs in yeast and mammals is at least partially responsible for transcription directionality, other mechanisms (PIC strength, chromatin modifications and CTD modifications) also affect directionality. Simultaneous disruption of multiple mechanisms that have been proposed to regulate transcription directionality could be used to achieve complete loss of directionality and to model the single contributions of each

mechanism.

Although the mechanism of selective termination seems conserved from yeast to mammals, the proteins are different and it will be interesting to further study the involved proteins, gain mechanistic insights and uncover their evolution. It is possible, that mammals evolved a different mechanism than yeast, because yeast has less conserved poly(A) signals and hardly any splicing. Independent of which proteins are involved in the selective termination process, from an evolutionary point of view, competitive organisms need to adopt mechanisms for selective termination. In principle, a cell cannot know in advance, which DNA sequences will give rise to functional transcripts. If a cell knew in advance which DNA sequences are useful and if transcription was a very stringent process, then how could new mRNAs and especially non-coding, regulatory RNAs emerge? Evolution can only occur randomly, over time. Thus, it makes sense to adopt a very efficient transcription machinery that produces RNA from essentially all accessible DNA and decide afterwards, which transcripts increase fitness. Most ncRNAs are degraded, but in some cases they might have become useful over time, leading to an increased fitness. Further bioinformatic analyses could potentially identify more organisms which show asymmetric distribution of motifs around transcription start sites in the genome supporting this hypothesis. [Almada et al., 2013] show a progressive gain of U1-snRNA sites with gene age, suggesting that suppression of promoter-proximal termination maintains gene function.

Systems biology, although a relatively young discipline has contributed a lot to our understanding of cellular processes. This field will continue to grow and more complex datasets will become available ("1000 Genomes", "ENCODE"). Experiments should be performed under standardized conditions and data sets will encompass different methods of investigation. Strong progress is currently made in the area of global, quantitative proteomics by the Mann and Aebersold labs. A method that allows simultaneous quantification of up to 100 proteins in single cells has recently been developed in the Nolan lab and in Heidelberg, the Bork lab has made strong progress in the field of metagenomics. The integration of different kind of data will be one of the major tasks and bottlenecks in future systems biology. Large projects like the "ENCODE" project produce massive amounts of data. Computer infrastructures are needed, have to be administered, updated and progressively developed in order to allow effective data handling. Computational scientists with professional training will take advantage of pre-existing, accessible data which can be used to address questions that have not been followed up by the authors. However, the exciting part about systems biology, its interdisciplinary nature, will not be lost and computationally based hypotheses will always have to be tested experimentally

and vice versa.

# References

- [Adessi et al., 2000] Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J. J., Mayer, P., and Kawashima, E. (2000). Solid phase dna amplification: characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res, 28(20):E87. 8
- [Almada et al., 2013] Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B., and Sharp, P. A. (2013). Promoter directionality is controlled by u1 snrnp and polyadenylation signals. Nature, 499(7458):360–363. 12, 15, 74, 76, 77
- [Amorim et al., 2010] Amorim, M. J., Cotobal, C., Duncan, C., and Mata, J. (2010). Global coordination of transcriptional control and mrna decay during cellular differentiation. Mol Syst Biol, 6:380. 41
- [Amrani et al., 1997] Amrani, N., Minet, M., Wyers, F., Dufour, M. E., Aggerbeck, L. P., and Lacroute, F. (1997). Pcf11 encodes a third protein component of yeast cleavage and polyadenylation factor i. Mol Cell Biol, 17(3):1102–1109. 5
- [Anders and Huber, 2010] Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol, 11(10):R106. 30, 49, 50
- [Arigo et al., 2006a] Arigo, J. T., Carroll, K. L., Ames, J. M., and Corden, J. L. (2006a). Regulation of yeast nrd1 expression by premature transcription termination. Mol Cell, 21(5):641–51. 13
- [Arigo et al., 2006b] Arigo, J. T., Eyler, D. E., Carroll, K. L., and Corden, J. L. (2006b). Termination of cryptic unstable transcripts is directed by yeast rna-binding proteins nrd1 and nab3. Mol Cell, 23(6):841–51. 12, 68
- [Asin-Cayuela and Gustafsson, 2007] Asin-Cayuela, J. and Gustafsson, C. M. (2007). Mitochondrial transcription and its regulation in mammalian cells. Trends Biochem Sci, 32(3):111–117. 2

## REFERENCES

---

- [Birse et al., 1998] Birse, C. E., Minvielle-Sebastia, L., Lee, B. A., Keller, W., and Proudfoot, N. J. (1998). Coupling termination of transcription to messenger rna maturation in yeast. Science, 280(5361):298–301. 6
- [Blankenberg et al., 2010] Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol, Chapter 19:Unit 19 10 1–21. 29
- [Bolstad et al., 2003] Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics, 19(2):185–193. 11
- [Bourbon et al., 2004] Bourbon, H.-M., Aguilera, A., Ansari, A. Z., Asturias, F. J., Berk, A. J., Bjorklund, S., Blackwell, T. K., Borggreffe, T., Carey, M., Carlson, M., Conaway, J. W., Conaway, R. C., Emmons, S. W., Fondell, J. D., Freedman, L. P., Fukasawa, T., Gustafsson, C. M., Han, M., He, X., Herman, P. K., Hinnebusch, A. G., Holmberg, S., Holstege, F. C., Jaehning, J. A., Kim, Y.-J., Kuras, L., Leutz, A., Lis, J. T., Meisterernest, M., Naar, A. M., Nasmyth, K., Parvin, J. D., Ptashne, M., Reinberg, D., Ronne, H., Sadowski, I., Sakurai, H., Sipiczki, M., Sternberg, P. W., Stillman, D. J., Strich, R., Struhl, K., Svejstrup, J. Q., Tuck, S., Winston, F., Roeder, R. G., and Kornberg, R. D. (2004). A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to rna polymerase ii. Mol Cell, 14(5):553–557. 4
- [Brannan et al., 2012] Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J., and Bentley, D. L. (2012). mrna decapping factors and the exonuclease xrn2 function in widespread premature termination of rna polymerase ii transcription. Mol Cell, 46(3):311–24. 64
- [Bregues et al., 2005] Bregues, M., Teixeira, D., and Parker, R. (2005). Movement of eukaryotic mrnas between polysomes and cytoplasmic processing bodies. Science, 310(5747):486–489. 72
- [Buratowski, 2009] Buratowski, S. (2009). Progression through the rna polymerase ii ctd cycle. Mol Cell, 36(4):541–546. 2, 4, 5, 12
- [Buratowski et al., 1989] Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989). Five intermediate complexes in transcription initiation by rna polymerase ii. Cell, 56(4):549–561. 4

## REFERENCES

---

- [Camblong et al., 2007] Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007). Antisense rna stabilization induces transcriptional gene silencing via histone deacetylation in *s. cerevisiae*. Cell, 131(4):706–17. 60
- [Carroll et al., 2007] Carroll, K. L., Ghirlando, R., Ames, J. M., and Corden, J. L. (2007). Interaction of yeast rna-binding proteins nrd1 and nab3 with rna polymerase ii terminator elements. RNA, 13(3):361–73. 73
- [Carroll et al., 2004] Carroll, K. L., Pradhan, D. A., Granek, J. A., Clarke, N. D., and Corden, J. L. (2004). Identification of cis elements directing termination of yeast non-polyadenylated snorna transcripts. Mol Cell Biol, 24(14):6241–52. 12, 54
- [Cho et al., 1997] Cho, E. J., Takagi, T., Moore, C. R., and Buratowski, S. (1997). mrna capping enzyme is recruited to the transcription complex by phosphorylation of the rna polymerase ii carboxy-terminal domain. Genes Dev, 11(24):3319–3326. 5
- [Churchman and Weissman, 2011] Churchman, L. S. and Weissman, J. S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature, 469(7330):368–73. 10
- [Cleary et al., 2005] Cleary, M. D., Meiering, C. D., Jan, E., Guymon, R., and Boothroyd, J. C. (2005). Biosynthetic labeling of rna with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mrna synthesis and decay. Nat Biotechnol, 23(2):232–237. 10
- [Colin et al., 2011] Colin, J., Libri, D., and Porrua, O. (2011). Cryptic transcription and early termination in the control of gene expression. Genet Res Int, 2011:653494. 62
- [Collart, 2003] Collart, M. A. (2003). Global control of gene expression in yeast by the ccr4-not complex. Gene, 313:1–16. 72
- [Collart and Timmers, 2004] Collart, M. A. and Timmers, H. T. M. (2004). The eukaryotic ccr4-not complex: a regulatory platform integrating mrna metabolism with cellular signaling pathways? Prog Nucleic Acid Res Mol Biol, 77:289–322. 72
- [Coller and Parker, 2004] Coller, J. and Parker, R. (2004). Eukaryotic mrna decapping. Annu Rev Biochem, 73:861–890. 72
- [Coller and Parker, 2005] Coller, J. and Parker, R. (2005). General translational repression by activators of mrna decapping. Cell, 122(6):875–886. 72

## REFERENCES

---

- [Connelly and Manley, 1988] Connelly, S. and Manley, J. L. (1988). A functional mrna polyadenylation signal is required for transcription termination by rna polymerase ii. Genes Dev, 2(4):440–452. 6
- [Corden et al., 1985] Corden, J. L., Cadena, D. L., Ahearn, Jr, J., and Dahmus, M. E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic rna polymerase ii. Proc Natl Acad Sci U S A, 82(23):7934–7938. 2
- [Core et al., 2008] Core, L. J., Waterfall, J. J., and Lis, J. T. (2008). Nascent rna sequencing reveals widespread pausing and divergent initiation at human promoters. Science, 322(5909):1845–8. 10, 11, 74, 76
- [Cramer et al., 2008] Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G. E., Dengl, S., Geiger, S. R., Jasiak, A. J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.-D., Lehmann, E., Leike, K., Sydow, J. F., and Vannini, A. (2008). Structure of eukaryotic rna polymerases. Annu Rev Biophys, 37:337–352. 1, 2
- [Cramer et al., 2001] Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001). Structural basis of transcription: Rna polymerase ii at 2.8 angstrom resolution. Science, 292(5523):1863–1876. 1, 43
- [Creamer et al., 2011] Creamer, T. J., Darby, M. M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S. J., and Corden, J. L. (2011). Transcriptome-wide binding sites for components of the saccharomyces cerevisiae non-poly(a) termination pathway: Nrd1, nab3, and sen1. PLoS Genet, 7(10):e1002329. 12, 13, 32, 54
- [CRICK, 1958] CRICK, F. H. (1958). On protein synthesis. Symp Soc Exp Biol, 12:138–163. 1
- [David et al., 2006] David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W., and Steinmetz, L. M. (2006). A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci U S A, 103(14):5320–5. 11
- [Djebali et al., 2012] Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Roder, M., Kokocinski, F., Abdelhamid, R. F., Alioto, T., Antoshechkin, I., Baer, M. T., Bar, N. S., Batut, P., Bell, K., Bell, I., Chakraborty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca,

## REFERENCES

---

- M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M. J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O. J., Park, E., Persaud, K., Preall, J. B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L. H., Shahab, A., Skancke, J., Suzuki, A. M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S. E., Hannon, G., Giddings, M. C., Ruan, Y., Wold, B., Carninci, P., Guigo, R., and Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, 489(7414):101–8. 11
- [Doelken et al., 2008] Doelken, L., Ruzsics, Z., Raedle, B., Friedel, C. C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., and Koszinowski, U. H. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of rna synthesis and decay. *RNA*, 14(9):1959–1972. 10, 75
- [Dori-Bachash et al., 2011] Dori-Bachash, M., Shema, E., and Tirosh, I. (2011). Coupled evolution of transcription and mrna degradation. *PLoS Biol*, 9(7):e1001106. 38
- [Dressman et al., 2003] Dressman, D., Yan, H., Traverso, G., Kinzler, K. W., and Vogelstein, B. (2003). Transforming single dna molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A*, 100(15):8817–8822. 7
- [Edwards et al., 1991] Edwards, A. M., Kane, C. M., Young, R. A., and Kornberg, R. D. (1991). Two dissociable subunits of yeast rna polymerase ii stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem*, 266(1):71–75. 72
- [Egloff and Murphy, 2008] Egloff, S. and Murphy, S. (2008). Cracking the rna polymerase ii ctd code. *Trends Genet*, 24(6):280–288. 2
- [Eulalio et al., 2007] Eulalio, A., Behm-Ansmant, I., and Izaurralde, E. (2007). P bodies: at the crossroads of post-transcriptional pathways. *Nat Rev Mol Cell Biol*, 8(1):9–22. 10
- [Fan et al., 2008] Fan, X., Lamarre-Vincent, N., Wang, Q., and Struhl, K. (2008). Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments. *Nucleic Acids Res*, 36(19):e125. 28
- [Fedurco et al., 2006] Fedurco, M., Romieu, A., Williams, S., Lawrence, I., and Turcatti, G. (2006). Bta, a novel reagent for dna attachment on glass and efficient generation of solid-phase amplified dna colonies. *Nucleic Acids Res*, 34(3):e22. 8

## REFERENCES

---

- [Fuchs et al., 2009] Fuchs, S. M., Larabee, R. N., and Strahl, B. D. (2009). Protein modifications in transcription elongation. *Biochim Biophys Acta*, 1789(1):26–36. 2
- [Fuda et al., 2009] Fuda, N. J., Ardehali, M. B., and Lis, J. T. (2009). Defining mechanisms that regulate rna polymerase ii transcription in vivo. *Nature*, 461(7261):186–192. 10
- [Garcia-Martinez et al., 2004] Garcia-Martinez, J., Aranda, A., and Perez-Ortin, J. E. (2004). Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell*, 15(2):303–13. 10, 75
- [Ghildiyal and Zamore, 2009] Ghildiyal, M. and Zamore, P. D. (2009). Small silencing rnas: an expanding universe. *Nat Rev Genet*, 10(2):94–108. 2
- [Giardina and Lis, 1993] Giardina, C. and Lis, J. T. (1993). Dna melting on yeast rna polymerase ii promoters. *Science*, 261(5122):759–762. 4
- [Giardine et al., 2005] Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W. J., and Nekrutenko, A. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Res*, 15(10):1451–5. 29
- [Goecks et al., 2010] Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy, T. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*, 11(8):R86. 29
- [Goffeau et al., 1996] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996). Life with 6000 genes. *Science*, 274(5287):546, 563–546, 567. 7
- [Greenleaf et al., 1986] Greenleaf, A. L., Kelly, J. L., and Lehman, I. R. (1986). Yeast rpo41 gene product is required for transcription and maintenance of the mitochondrial genome. *Proc Natl Acad Sci U S A*, 83(10):3391–3394. 2
- [Grigull et al., 2004] Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M. D., and Hughes, T. R. (2004). Genome-wide analysis of mrna stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol*, 24(12):5534–5547. 10, 38, 39, 40, 75
- [Gruenberg et al., 2012] Gruenberg, S., Warfield, L., and Hahn, S. (2012). Architecture of the rna polymerase ii preinitiation complex and mechanism of atp-dependent promoter opening. *Nat Struct Mol Biol*, 19(8):788–796. 4

## REFERENCES

---

- [Gudipati et al., 2012] Gudipati, R. K., Neil, H., Feuerbach, F., Malabat, C., and Jacquier, A. (2012). The yeast *rpl9b* gene is regulated by modulation between two modes of transcription termination. The EMBO journal. 12
- [Guil and Esteller, 2012] Guil, S. and Esteller, M. (2012). Cis-acting noncoding rnas: friends and foes. Nat Struct Mol Biol, 19(11):1068–1075. 1
- [Haag and Pikaard, 2011] Haag, J. R. and Pikaard, C. S. (2011). Multisubunit rna polymerases iv and v: purveyors of non-coding rna for plant gene silencing. Nat Rev Mol Cell Biol, 12(8):483–492. 2
- [Hafner et al., 2010] Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., J., Jungkamp, A. C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., and Tuschl, T. (2010). Transcriptome-wide identification of rna-binding protein and microrna target sites by par-clip. Cell, 141(1):129–41. 9, 54
- [Hahn, 2004] Hahn, S. (2004). Structure and mechanism of the rna polymerase ii transcription machinery. Nat Struct Mol Biol, 11(5):394–403. 4
- [Haimovich et al., 2013] Haimovich, G., Choder, M., Singer, R. H., and Trcek, T. (2013). The fate of the messenger is pre-determined: A new model for regulation of gene expression. Biochim Biophys Acta. 75
- [Hani and Feldmann, 1998] Hani, J. and Feldmann, H. (1998). *trna* genes and retroelements in the yeast genome. Nucleic Acids Res, 26(3):689–696. 11
- [Harel-Sharvit et al., 2010] Harel-Sharvit, L., Eldad, N., Haimovich, G., Barkai, O., Duek, L., and Choder, M. (2010). Rna polymerase ii subunits link transcription and mrna decay to translation. Cell, 143(4):552–63. 10, 72
- [Haruki et al., 2008] Haruki, H., Nishikawa, J., and Laemmli, U. K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol Cell, 31(6):925–32. 14, 48
- [Hazelbaker et al., 2013] Hazelbaker, D. Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2013). Kinetic competition between rna polymerase ii and *sen1*-dependent transcription termination. Mol Cell, 49(1):55–66. 13
- [Hereford and Rosbash, 1977] Hereford, L. M. and Rosbash, M. (1977). Regulation of a set of abundant mrna sequences. Cell, 10(3):463–467. 35, 38

## REFERENCES

---

- [Herr et al., 2005] Herr, A. J., Jensen, M. B., Dalmay, T., and Baulcombe, D. C. (2005). Rna polymerase iv directs silencing of endogenous dna. Science, 308(5718):118–120. 2
- [Hobson et al., 2012] Hobson, D. J., Wei, W., Steinmetz, L. M., and Svejstrup, J. Q. (2012). Rna polymerase ii collision interrupts convergent transcription. Mol Cell, 48(3):365–74. 62
- [Holstege et al., 1997] Holstege, F. C., Fiedler, U., and Timmers, H. T. (1997). Three transitions in the rna polymerase ii transcription complex during initiation. EMBO J, 16(24):7468–7480. 4
- [Holstege et al., 1998] Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., and Young, R. A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell, 95(5):717–28. 10, 11, 38, 39, 40
- [Hongay et al., 2006] Hongay, C. F., Grisafi, P. L., Galitski, T., and Fink, G. R. (2006). Antisense transcription controls cell fate in *saccharomyces cerevisiae*. Cell, 127(4):735–45. 60
- [Honorine et al., 2011] Honorine, R., Mosrin-Huaman, C., Hervouet-Coste, N., Libri, D., and Rahmouni, A. R. (2011). Nuclear mrna quality control in yeast is mediated by nrd1 co-transcriptional recruitment, as revealed by the targeting of rho-induced aberrant transcripts. Nucleic Acids Res, 39(7):2809–20. 12
- [Houseley et al., 2008] Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., and Vogelauer, M. (2008). A ncRNA modulates histone modification and mrna induction in the yeast gal gene cluster. Mol Cell, 32(5):685–95. 60
- [Houseley and Tollervey, 2009] Houseley, J. and Tollervey, D. (2009). The many pathways of rna degradation. Cell, 136(4):763–76. 45
- [Howe et al., 2003] Howe, K. J., Kane, C. M., and Ares, M., J. (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in *saccharomyces cerevisiae*. RNA, 9(8):993–1006. 13
- [Hsin and Manley, 2012] Hsin, J. P. and Manley, J. L. (2012). The rna polymerase ii ctd coordinates transcription and rna processing. Genes Dev, 26(19):2119–37. 12
- [Hu et al., 2009] Hu, W., Sweet, T. J., Chamnongpol, S., Baker, K. E., and Collier, J. (2009). Co-translational mrna decay in *saccharomyces cerevisiae*. Nature, 461(7261):225–229. 72

## REFERENCES

---

- [Hurwitz, 2005] Hurwitz, J. (2005). The discovery of rna polymerase. J Biol Chem, 280(52):42477–42485. 1
- [Jacquier, 2009] Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small rnas. Nat Rev Genet, 10(12):833–44. 12
- [Jimeno-Gonzalez et al., 2010] Jimeno-Gonzalez, S.lez, S., Haaning, L. L., Malagon, F., and Jensen, T. H. (2010). The yeast 5'-3' exonuclease rat1p functions during transcription elongation by rna polymerase ii. Mol Cell, 37(4):580–587. 43
- [Jorgensen et al., 2002] Jorgensen, P., Nishikawa, J. L., Breitzkreutz, B.-J., and Tyers, M. (2002). Systematic identification of pathways that couple cell growth and division in yeast. Science, 297(5580):395–400. 42
- [Kawauchi et al., 2008] Kawauchi, J., Mischo, H., Braglia, P., Rondon, A., and Proudfoot, N. J. (2008). Budding yeast rna polymerases i and ii employ parallel mechanisms of transcriptional termination. Genes Dev, 22(8):1082–1092. 6
- [Kelly and Lehman, 1986] Kelly, J. L. and Lehman, I. R. (1986). Yeast mitochondrial rna polymerase. purification and properties of the catalytic subunit. J Biol Chem, 261(22):10340–10347. 2
- [Kenzelmann et al., 2007] Kenzelmann, M., Maertens, S., Hergenbahn, M., Kueffer, S., Hotz-Wagenblatt, A., Li, L., Wang, S., Ittrich, C., Lemberger, T., Arribas, R., Jonnakuty, S., Hollstein, M. C., Schmid, W., Gretz, N., Groene, H. J., and Schuetz, G. (2007). Microarray analysis of newly synthesized rna in cells and animals. Proc Natl Acad Sci U S A, 104(15):6164–6169. 10
- [Kim et al., 2004a] Kim, H.-J., Jeong, S.-H., Heo, J.-H., Jeong, S.-J., Kim, S.-T., Youn, H.-D., Han, J.-W., Lee, H.-W., and Cho, E.-J. (2004a). mrna capping enzyme activity is coupled to an early transcription elongation. Mol Cell Biol, 24(14):6184–6193. 5
- [Kim and Levin, 2011] Kim, K. Y. and Levin, D. E. (2011). Mpk1 mapk association with the paf1 complex blocks sen1-mediated premature transcription termination. Cell, 144(5):745–56. 13, 68
- [Kim et al., 2004b] Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedeá, E., Greenblatt, J. F., and Buratowski, S. (2004b). The yeast rat1 exonuclease promotes transcription termination by rna polymerase ii. Nature, 432(7016):517–522. 6
- [Kim et al., 2006] Kim, M., Vasiljeva, L., Rando, O. J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct pathways for snorna and mrna termination. Mol Cell, 24(5):723–34. 12

## REFERENCES

---

- [Kishore et al., 2011] Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A quantitative analysis of clip methods for identifying binding sites of rna-binding proteins. Nat Methods, 8(7):559–564. 54
- [Koleske and Young, 1994] Koleske, A. J. and Young, R. A. (1994). An rna polymerase ii holoenzyme responsive to activators. Nature, 368(6470):466–469. 4
- [Komarnitsky et al., 2000] Komarnitsky, P., Cho, E. J., and Buratowski, S. (2000). Different phosphorylated forms of rna polymerase ii and associated mrna processing factors during transcription. Genes Dev, 14(19):2452–2460. 4
- [Konig et al., 2011a] Konig, J., Zarnack, K., Luscombe, N. M., and Ule, J. (2011a). Protein-rna interactions: new genomic technologies and perspectives. Nat Rev Genet, 13(2):77–83. 54
- [Konig et al., 2011b] Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D. J., Luscombe, N. M., and Ule, J. (2011b). iclip–transcriptome-wide mapping of protein-rna interactions with individual nucleotide resolution. J Vis Exp, (50). 9
- [Koonin, 2012] Koonin, E. V. (2012). Does the central dogma still stand? Biol Direct, 7:27. 1
- [Kornberg, 2005] Kornberg, R. D. (2005). Mediator and the mechanism of transcriptional activation. Trends Biochem Sci, 30(5):235–239. 4
- [Kostrewa et al., 2009] Kostrewa, D., Zeller, M. E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). Rna polymerase ii-tfiib structure and mechanism of transcription initiation. Nature, 462(7271):323–330. 4
- [Kruk et al., 2011] Kruk, J. A., Dutta, A., Fu, J., Gilmour, D. S., and Reese, J. C. (2011). The multifunctional ccr4-not complex directly promotes transcription elongation. Genes Dev, 25(6):581–593. 72
- [Kubicek et al., 2012] Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in rnap ii ctd control recruitment of nrd1. Genes Dev, 26(17):1891–1896. 5, 73
- [Kuehner and Brow, 2006] Kuehner, J. N. and Brow, D. A. (2006). Quantitative analysis of in vivo initiator selection by yeast rna polymerase ii supports a scanning model. J Biol Chem, 281(20):14119–14128. 4

## REFERENCES

---

- [Kuehner and Brow, 2008] Kuehner, J. N. and Brow, D. A. (2008). Regulation of a eukaryotic gene by gtp-dependent start site selection and transcription attenuation. Mol Cell, 31(2):201–11. 13
- [Kuo et al., 1996] Kuo, M. H., Brownell, J. E., Sobel, R. E., Ranalli, T. A., Cook, R. G., Edmondson, D. G., Roth, S. Y., and Allis, C. D. (1996). Transcription-linked acetylation by gcn5p of histones h3 and h4 at specific lysines. Nature, 383(6597):269–272. 5
- [Lam et al., 2001] Lam, L. T., Pickeral, O. K., Peng, A. C., Rosenwald, A., Hurt, E. M., Giltneane, J. M., Averett, L. M., Zhao, H., Davis, R. E., Sathyamoorthy, M., Wahl, L. M., Harris, E. D., Mikovits, J. A., Monks, A. P., Hollingshead, M. G., Sausville, E. A., and Staudt, L. M. (2001). Genomic-scale measurement of mrna turnover and the mechanisms of action of the anti-cancer drug flavopiridol. Genome Biol, 2(10):RESEARCH0041. 10, 75
- [Langmead et al., 2009] Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short dna sequences to the human genome. Genome Biol, 10(3):R25. 29
- [Li et al., 2009] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The sequence alignment/map format and samtools. Bioinformatics, 25(16):2078–9. 29
- [Li et al., 1994] Li, Y., Flanagan, P. M., Tschochner, H., and Kornberg, R. D. (1994). Rna polymerase ii initiation factor interactions and transcription start site selection. Science, 263(5148):805–807. 4
- [Lieberman-Aiden et al., 2009] Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., and Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 326(5950):289–293. 9
- [Lin et al., 2012] Lin, C. Y., Loven, J., Rahl, P. B., Paranal, R. M., Burge, C. B., Bradner, J. E., Lee, T. I., and Young, R. A. (2012). Transcriptional amplification in tumor cells with elevated c-myc. Cell, 151(1):56–67. 75
- [Liu et al., 1998] Liu, H. Y., Badarinarayana, V., Audino, D. C., Rappsilber, J., Mann, M., and Denis, C. L. (1998). The not proteins are part of the ccr4 transcriptional complex

## REFERENCES

---

- and affect gene expression both positively and negatively. EMBO J, 17(4):1096–1106. 72
- [Liu et al., 2012] Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. J Biomed Biotechnol, 2012:251364. 7
- [Logan et al., 1987] Logan, J., Falck-Pedersen, E., Darnell, Jr, J., and Shenk, T. (1987). A poly(a) addition site and a downstream termination region are required for efficient cessation of transcription by rna polymerase ii in the mouse beta maj-globin gene. Proc Natl Acad Sci U S A, 84(23):8306–8310. 6
- [Longtine et al., 1998] Longtine, M. S., McKenzie, 3rd, A., Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J. R. (1998). Additional modules for versatile and economical pcr-based gene deletion and modification in *saccharomyces cerevisiae*. Yeast, 14(10):953–961. 22
- [Lotan et al., 2005] Lotan, R., Bar-On, V. G., Harel-Sharvit, L., Duek, L., Melamed, D., and Choder, M. (2005). The rna polymerase ii subunit rpb4p mediates decay of a specific class of mRNAs. Genes Dev, 19(24):3004–3016. 10, 72
- [Lotan et al., 2007] Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G., and Choder, M. (2007). The rpb7p subunit of yeast rna polymerase ii plays roles in the two major cytoplasmic mRNA decay mechanisms. J Cell Biol, 178(7):1133–43. 10, 72
- [Loven et al., 2012] Loven, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., Levens, D. L., Lee, T. I., and Young, R. A. (2012). Revisiting global gene expression analysis. Cell, 151(3):476–482. 10, 75
- [Malagon et al., 2006] Malagon, F., Kireeva, M. L., Shafer, B. K., Lubkowska, L., Kashlev, M., and Strathern, J. N. (2006). Mutations in the *saccharomyces cerevisiae* rpb1 gene conferring hypersensitivity to 6-azauracil. Genetics, 172(4):2201–9. 13, 43
- [Malik and Roeder, 2005] Malik, S. and Roeder, R. G. (2005). Dynamic regulation of pol ii transcription by the mammalian mediator complex. Trends Biochem Sci, 30(5):256–263. 4
- [Mandal et al., 2004] Mandal, S. S., Chu, C., Wada, T., Handa, H., Shatkin, A. J., and Reinberg, D. (2004). Functional interactions of rna-capping enzyme with factors that positively and negatively regulate promoter escape by rna polymerase ii. Proc Natl Acad Sci U S A, 101(20):7572–7577. 5

## REFERENCES

---

- [Mandel et al., 2008] Mandel, C. R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mrna 3'-end processing. Cell Mol Life Sci, 65(7-8):1099–122. 5
- [Maniatis and Reed, 2002] Maniatis, T. and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature, 416(6880):499–506. 72
- [Margaritis and Holstege, 2008] Margaritis, T. and Holstege, F. C. (2008). Poised rna polymerase ii gives pause for thought. Cell, 133(4):581–4. 5
- [Marquardt et al., 2011] Marquardt, S., Hazelbaker, D. Z., and Buratowski, S. (2011). Distinct rna degradation pathways and 3' extensions of yeast non-coding rna species. Transcription, 2(3):145–154. 12
- [Martinez-Rucobo et al., 2011] Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M., and Cramer, P. (2011). Architecture of the rna polymerase-spt4/5 complex and basis of universal transcription processivity. EMBO J, 30(7):1302–1310. 5
- [Mason and Struhl, 2005] Mason, P. B. and Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of rna polymerase ii in vivo. Mol Cell, 17(6):831–40. 13
- [Mavrich et al., 2008] Mavrich, T. N., Ioshikhes, I. P., Venters, B. J., Jiang, C., Tomsho, L. P., Qi, J., Schuster, S. C., Albert, I., and Pugh, B. F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res, 18(7):1073–83. 51
- [Maxam and Gilbert, 1977] Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing dna. Proc Natl Acad Sci U S A, 74(2):560–564. 7
- [Maxon et al., 1994] Maxon, M. E., Goodrich, J. A., and Tjian, R. (1994). Transcription factor iie binds preferentially to rna polymerase iia and recruits tfiih: a model for promoter clearance. Genes Dev, 8(5):515–524. 4
- [Maxwell and Fournier, 1995] Maxwell, E. S. and Fournier, M. J. (1995). The small nucleolar rnas. Annu Rev Biochem, 64:897–934. 1
- [Mayer et al., 2012] Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). Ctd tyrosine phosphorylation impairs termination factor recruitment to rna polymerase ii. Science, 336(6089):1723–1725. 2, 6, 13

## REFERENCES

---

- [Mayer et al., 2010] Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010). Uniform transitions of the general rna polymerase ii transcription complex. Nat Struct Mol Biol, 17(10):1272–8. 5, 27, 62
- [Meinhart and Cramer, 2004] Meinhart, A. and Cramer, P. (2004). Recognition of rna polymerase ii carboxy-terminal domain by 3'-rna-processing factors. Nature, 430(6996):223–226. 6
- [Melvin et al., 1978] Melvin, W. T., Milne, H. B., Slater, A. A., Allen, H. J., and Keir, H. M. (1978). Incorporation of 6-thioguanosine and 4-thiouridine into rna. application to isolation of newly synthesised rna by affinity chromatography. Eur J Biochem, 92(2):373–379. 10
- [Mercer et al., 2009] Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding rnas: insights into functions. Nat Rev Genet, 10(3):155–159. 74
- [Miller et al., 2011] Miller, C., Schwalb, B., Maier, K., Schulz, D., Dumcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dolken, L., Martin, D. E., Tresch, A., and Cramer, P. (2011). Dynamic transcriptome analysis measures rates of mrna synthesis and decay in yeast. Mol Syst Biol, 7:458. 10, 14, 24, 25, 26, 33, 34, 36, 38, 39, 71
- [Minvielle-Sebastia et al., 1998] Minvielle-Sebastia, L., Beyer, K., Krecic, A. M., Hector, R. E., Swanson, M. S., and Keller, W. (1998). Control of cleavage site selection during mrna 3' end formation by a yeast hnrnp. EMBO J, 17(24):7454–7468. 5
- [Minvielle-Sebastia and Keller, 1999] Minvielle-Sebastia, L. and Keller, W. (1999). mrna polyadenylation and its coupling to other rna processing reactions and to transcription. Curr Opin Cell Biol, 11(3):352–357. 5
- [Mischo and Proudfoot, 2013] Mischo, H. E. and Proudfoot, N. J. (2013). Disengaging polymerase: terminating rna polymerase ii transcription in budding yeast. Biochim Biophys Acta, 1829(1):174–85. 2, 5, 6, 12, 13, 76
- [Munchel et al., 2011] Munchel, S. E., Shultzaberger, R. K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mrna turnover reveals gene-specific and system-wide regulation of mrna decay. Mol Biol Cell, 22(15):2787–2795. 33, 35, 40
- [Murray et al., 2012] Murray, S. C., Serra Barros, A., Brown, D. A., Dudek, P., Ayling, J., and Mellor, J. (2012). A pre-initiation complex at the 3'-end of genes drives anti-sense transcription independent of divergent sense transcription. Nucleic Acids Res, 40(6):2432–44. 11, 53

## REFERENCES

---

- [Nechaev and Adelman, 2011] Nechaev, S. and Adelman, K. (2011). Pol ii waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochim Biophys Acta, 1809(1):34–45. 2
- [Neil et al., 2009] Neil, H., Malabat, C., d'Aubenton Carafa, Y., Xu, Z., Steinmetz, L. M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature, 457(7232):1038–42. 1, 11, 12, 51, 73, 74, 76
- [Neumann and Nurse, 2007] Neumann, F. R. and Nurse, P. (2007). Nuclear size control in fission yeast. J Cell Biol, 179(4):593–600. 42
- [Ng et al., 2003] Ng, H. H., Robert, F., Young, R. A., and Struhl, K. (2003). Targeted recruitment of set1 histone methylase by elongating pol ii provides a localized mark and memory of recent transcriptional activity. Mol Cell, 11(3):709–719. 5
- [Nonet et al., 1987] Nonet, M., Scafe, C., Sexton, J., and Young, R. (1987). Eucaryotic rna polymerase conditional mutant that rapidly ceases mrna synthesis. Mol Cell Biol, 7(5):1602–1611. 38
- [Ntini et al., 2013] Ntini, E., Jaervelin, A. I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P. R., Andersen, P. K., Preker, P., Valen, E., Zhao, X., Pelechano, V., Steinmetz, L. M., Sandelin, A., and Jensen, T. H. (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nat Struct Mol Biol. 12, 15, 74, 76
- [Onodera et al., 2005] Onodera, Y., Haag, J. R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C. S. (2005). Plant nuclear rna polymerase iv mediates sirna and dna methylation-dependent heterochromatin formation. Cell, 120(5):613–622. 2
- [Ozsolak et al., 2010] Ozsolak, F., Kapranov, P., Foissac, S., Kim, S. W., Fishilevich, E., Monaghan, A. P., John, B., and Milos, P. M. (2010). Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell, 143(6):1018–1029. 5
- [Parker and Sheth, 2007] Parker, R. and Sheth, U. (2007). P bodies and the control of mrna translation and degradation. Mol Cell, 25(5):635–646. 10
- [Perales and Bentley, 2009] Perales, R. and Bentley, D. (2009). "cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol Cell, 36(2):178–191. 2

## REFERENCES

---

- [Perocchi et al., 2007] Perocchi, F., Xu, Z., Clauder-Muenster, S., and Steinmetz, L. M. (2007). Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin d. Nucleic Acids Res, 35(19):e128. 30
- [Peters et al., 2012] Peters, J. M., Mooney, R. A., Grass, J. A., Jessen, E. D., Tran, F., and Landick, R. (2012). Rho and nusg suppress pervasive antisense transcription in escherichia coli. Genes Dev, 26(23):2621–2633. 74
- [Pikaard et al., 2008] Pikaard, C. S., Haag, J. R., Ream, T., and Wierzbicki, A. T. (2008). Roles of rna polymerase iv in gene silencing. Trends Plant Sci, 13(7):390–397. 2
- [Polz and Cavanaugh, 1998] Polz, M. F. and Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate pcr. Appl Environ Microbiol, 64(10):3724–3730. 8
- [Pontier et al., 2005] Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.-A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct rna polymerases iv in arabidopsis. Genes Dev, 19(17):2030–2040. 2
- [Porrua et al., 2012] Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R. K., Stefl, R., and Libri, D. (2012). In vivo selex reveals novel sequence and structural determinants of nrd1-nab3-sen1-dependent transcription termination. EMBO J, 31(19):3935–48. 12, 54, 55
- [Porrua and Libri, 2013a] Porrua, O. and Libri, D. (2013a). A bacterial-like mechanism for transcription termination by the sen1p helicase in budding yeast. Nat Struct Mol Biol, 20(7):884–891. 13, 76
- [Porrua and Libri, 2013b] Porrua, O. and Libri, D. (2013b). Rna quality control in the nucleus: the angels' share of rna. Biochim Biophys Acta, 1829(6-7):604–611. 12, 76
- [Preker et al., 2008] Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., Schierup, M. H., and Jensen, T. H. (2008). Rna exosome depletion reveals transcription upstream of active human promoters. Science, 322(5909):1851–1854. 76
- [Proudfoot, 2011] Proudfoot, N. J. (2011). Ending the message: poly(a) signals then and now. Genes Dev, 25(17):1770–1782. 5
- [Rabani et al., 2011] Rabani, M., Levin, J. Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., Amit, I., and

## REFERENCES

---

- Regev, A. (2011). Metabolic labeling of rna uncovers principles of rna production and degradation dynamics in mammalian cells. Nat Biotechnol, 29(5):436–42. 10, 26, 49, 75
- [Rando and Chang, 2009] Rando, O. J. and Chang, H. Y. (2009). Genome-wide views of chromatin structure. Annu Rev Biochem, 78:245–71. 73
- [Rhee and Pugh, 2011] Rhee, H. S. and Pugh, B. F. (2011). Comprehensive genome-wide protein-dna interactions detected at single-nucleotide resolution. Cell, 147(6):1408–1419. 9
- [Rhee and Pugh, 2012] Rhee, H. S. and Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature, 483(7389):295–301. 11, 52, 53, 64, 66
- [Ringel et al., 2011] Ringel, R., Sologub, M., Morozov, Y. I., Litonin, D., Cramer, P., and Temiakov, D. (2011). Structure of human mitochondrial rna polymerase. Nature, 478(7368):269–273. 2
- [Roeder and Rutter, 1969] Roeder, R. G. and Rutter, W. J. (1969). Multiple forms of dna-dependent rna polymerase in eukaryotic organisms. Nature, 224(5216):234–237. 1
- [Ronaghi et al., 1996] Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., and Nyren, P. (1996). Real-time dna sequencing using detection of pyrophosphate release. Anal Biochem, 242(1):84–89. 7
- [Rondon et al., 2009] Rondon, A. G., Mischo, H. E., Kawauchi, J., and Proudfoot, N. J. (2009). Fail-safe transcriptional termination for protein-coding genes in *s. cerevisiae*. Mol Cell, 36(1):88–98. 57
- [Sadowski et al., 2003] Sadowski, M., Dichtl, B., Huebner, W., and Keller, W. (2003). Independent functions of yeast pcf11p in pre-mrna 3' end processing and in transcription termination. EMBO J, 22(9):2167–2177. 6
- [Sainsbury et al., 2013] Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing rna polymerase ii-tfiib complex. Nature, 493(7432):437–440. 5
- [Sanger et al., 1977] Sanger, F., Nicklen, S., and Coulson, A. R. (1977). Dna sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A, 74(12):5463–5467. 7

## REFERENCES

---

- [Schena et al., 1995] Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary dna microarray. Science, 270(5235):467–470. 7
- [Schroeder et al., 2000] Schroeder, S. C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing rna polymerase ii. Genes Dev, 14(19):2435–2440. 4, 5
- [Seila et al., 2008] Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., and Sharp, P. A. (2008). Divergent transcription from active promoters. Science, 322(5909):1849–51. 11, 73, 74, 76
- [Seila et al., 2009] Seila, A. C., Core, L. J., Lis, J. T., and Sharp, P. A. (2009). Divergent transcription: a new feature of active promoters. Cell Cycle, 8(16):2557–2564. 5, 12
- [Selitrennik et al., 2006] Selitrennik, M., Duek, L., Lotan, R., and Choder, M. (2006). Nucleocytoplasmic shuttling of the rpb4p and rpb7p subunits of saccharomyces cerevisiae rna polymerase ii by two pathways. Eukaryot Cell, 5(12):2092–103. 72
- [Shalem et al., 2008] Shalem, O., Dahan, O., Levo, M., Martinez, M. R., Furman, I., Segal, E., and Pilpel, Y. (2008). Transient transcriptional responses to stress are generated by opposing effects of mrna production and degradation. Mol Syst Biol, 4:223. 10, 38, 39, 40, 75
- [Singh et al., 2009] Singh, N., Ma, Z., Gemmill, T., Wu, X., Defiglio, H., Rossetini, A., Rabeler, C., Beane, O., Morse, R. H., Palumbo, M. J., and Hanes, S. D. (2009). The ess1 prolyl isomerase is required for transcription termination of small noncoding rnas via the nrd1 pathway. Mol Cell, 36(2):255–66. 73
- [Snyder and Gallagher, 2009] Snyder, M. and Gallagher, J. E. G. (2009). Systems biology from a yeast omics perspective. FEBS Lett, 583(24):3895–3899. 6
- [Steinmetz and Brow, 1996] Steinmetz, E. J. and Brow, D. A. (1996). Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, nrd1, and the putative helicase sen1. Mol Cell Biol, 16(12):6993–7003. 12
- [Steinmetz et al., 2001] Steinmetz, E. J., Conrad, N. K., Brow, D. A., and Corden, J. L. (2001). Rna-binding protein nrd1 directs poly(a)-independent 3'-end formation of rna polymerase ii transcripts. Nature, 413(6853):327–31. 12

## REFERENCES

---

- [Steinmetz et al., 2006a] Steinmetz, E. J., Ng, S. B., Cloute, J. P., and Brow, D. A. (2006a). cis- and trans-acting determinants of transcription termination by yeast rna polymerase ii. Mol Cell Biol, 26(7):2688–96. 68
- [Steinmetz et al., 2006b] Steinmetz, E. J., Warren, C. L., Kuehner, J. N., Panbehi, B., Ansari, A. Z., and Brow, D. A. (2006b). Genome-wide distribution of yeast rna polymerase ii and its control by sen1 helicase. Mol Cell, 24(5):735–46. 13
- [Sun et al., 2012] Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Lariviere, L., Maier, K. C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cdta) reveals mutual feedback between mrna synthesis and degradation. Genome Res. 24, 33, 34, 42, 43, 46, 47, 48, 75
- [Svejstrup et al., 1997] Svejstrup, J. Q., Li, Y., Fellows, J., Gnatt, A., Bjorklund, S., and Kornberg, R. D. (1997). Evidence for a mediator cycle at the initiation of transcription. Proc Natl Acad Sci U S A, 94(12):6075–6078. 4
- [Tan-Wong et al., 2012] Tan-Wong, S. M., Zaugg, J. B., Camblong, J., Xu, Z., Zhang, D. W., Mischo, H. E., Ansari, A. Z., Luscombe, N. M., Steinmetz, L. M., and Proudfoot, N. J. (2012). Gene loops enhance transcriptional directionality. Science, 338(6107):671–5. 11, 73
- [Temin, 1985] Temin, H. M. (1985). Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. Mol Biol Evol, 2(6):455–468. 1
- [Thiebaut et al., 2008] Thiebaut, M., Colin, J., Neil, H., Jacquier, A., Seraphin, B., Lacroute, F., and Libri, D. (2008). Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in *s. cerevisiae*. Mol Cell, 31(5):671–82. 13
- [Thiebaut et al., 2006] Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol Cell, 23(6):853–64. 12
- [Thomas and Chiang, 2006] Thomas, M. C. and Chiang, C.-M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol, 41(3):105–178. 4

## REFERENCES

---

- [Tucker et al., 2002] Tucker, M., Staples, R. R., Valencia-Sanchez, M. A., Muhlrads, D., and Parker, R. (2002). Ccr4p is the catalytic subunit of a ccr4p/pop2p/notp mrna deadenylase complex in *saccharomyces cerevisiae*. EMBO J, 21(6):1427–1436. 45, 72
- [van de Peppel et al., 2003] van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D., and Holstege, F. C. (2003). Monitoring global messenger rna changes in externally controlled microarray experiments. EMBO Rep, 4(4):387–93. 11
- [van Dijk et al., 2011] van Dijk, E. L., Chen, C. L., d'Aubenton Carafa, Y., Gourvenec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Ne, P., Loeillet, S., Nicolas, A., Thermes, C., and Morillon, A. (2011). Xuts are a class of xrn1-sensitive antisense regulatory non-coding rna in yeast. Nature, 475(7354):114–7. 1, 12, 74
- [Vasiljeva and Buratowski, 2006] Vasiljeva, L. and Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of rna polymerase ii transcripts. Mol Cell, 21(2):239–48. 12
- [Vasiljeva et al., 2008] Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The nrd1-nab3-sen1 termination complex interacts with the ser5-phosphorylated rna polymerase ii c-terminal domain. Nat Struct Mol Biol, 15(8):795–804. 5, 12
- [Venters and Pugh, 2009] Venters, B. J. and Pugh, B. F. (2009). How eukaryotic genes are transcribed. Crit Rev Biochem Mol Biol, 44(2-3):117–141. 2, 4, 5
- [Wada et al., 1998] Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., and Handa, H. (1998). Dsif, a novel transcription elongation factor that regulates rna polymerase ii processivity, is composed of human spt4 and spt5 homologs. Genes Dev, 12(3):343–356. 5
- [Wang et al., 2002] Wang, Y., Liu, C. L., Storey, J. D., Tibshirani, R. J., Herschlag, D., and Brown, P. O. (2002). Precision and functional specificity in mrna decay. Proc Natl Acad Sci U S A, 99(9):5860–5865. 11, 38, 39, 40
- [Wei et al., 2011] Wei, W., Pelechano, V., Jarvelin, A. I., and Steinmetz, L. M. (2011). Functional consequences of bidirectional promoters. Trends Genet, 27(7):267–76. 5, 12

## REFERENCES

---

- [Weinmann and Roeder, 1974] Weinmann, R. and Roeder, R. G. (1974). Role of dna-dependent rna polymerase 3 in the transcription of the trna and 5s rna genes. Proc Natl Acad Sci U S A, 71(5):1790–1794. 1
- [Weiss and Gladstone, 1959] Weiss, S. and Gladstone, L. (1959). A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. J. Am. Chem. Soc., 81:4118–4119. 1
- [West et al., 2004] West, S., Gromak, N., and Proudfoot, N. J. (2004). Human 5'  $\rightarrow$  3' exonuclease xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature, 432(7016):522–525. 6
- [Wiederhold and Passmore, 2010] Wiederhold, K. and Passmore, L. A. (2010). Cytoplasmic deadenylation: regulation of mrna fate. Biochem Soc Trans, 38(6):1531–1536. 10
- [Windhager et al., 2012] Windhager, L., Bonfert, T., Burger, K., Ruzsics, Z., Krebs, S., Kaufmann, S., Malterer, G., L'Hernault, A., Schilhabel, M., Schreiber, S., Rosenstiel, P., Zimmer, R., Eick, D., Friedel, C. C., and Dolken, L. (2012). Ultrashort and progressive 4su-tagging reveals key characteristics of rna processing at nucleotide resolution. Genome Res, 22(10):2031–42. 49
- [Wlotzka et al., 2011] Wlotzka, W., Kudla, G., Granneman, S., and Tollervy, D. (2011). The nuclear rna polymerase ii surveillance system targets polymerase iii transcripts. EMBO J, 30(9):1790–1803. 12, 13, 54, 55
- [Wyers et al., 2005] Wyers, F., Rougemaille, M., Badis, G., Rousselle, J. C., Dufour, M. E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., Libri, D., and Jacquier, A. (2005). Cryptic pol ii transcripts are degraded by a nuclear quality control pathway involving a new poly(a) polymerase. Cell, 121(5):725–37. 12
- [Xu et al., 2011] Xu, Z., Wei, W., Gagneur, J., Clauder-Munster, S., Smolik, M., Huber, W., and Steinmetz, L. M. (2011). Antisense expression increases gene expression variability and locus interdependency. Mol Syst Biol, 7:468. 60, 62
- [Xu et al., 2009] Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L. M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature, 457(7232):1033–7. 1, 11, 12, 30, 31, 49, 51, 55, 57, 73, 74, 76
- [Yassour et al., 2010] Yassour, M., Pfiffner, J., Levin, J. Z., Adiconis, X., Gnirke, A., Nusbaum, C., Thompson, D. A., Friedman, N., and Regev, A. (2010). Strand-specific

## REFERENCES

---

- rna sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. Genome Biol, 11(8):R87. 53
- [Yonaha and Proudfoot, 1999] Yonaha, M. and Proudfoot, N. J. (1999). Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. Mol Cell, 3(5):593–600. 6
- [Zeidan and Hart, 2010] Zeidan, Q. and Hart, G. W. (2010). The intersections between o-glcacylation and phosphorylation: implications for multiple signaling pathways. J Cell Sci, 123(Pt 1):13–22. 2
- [Zenklusen et al., 2008] Zenklusen, D., Larson, D. R., and Singer, R. H. (2008). Single-rna counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol, 15(12):1263–1271. 35, 38
- [Zhang et al., 2012a] Zhang, D. W., Mosley, A. L., Ramisetty, S. R., Rodriguez-Molina, J. B., Washburn, M. P., and Ansari, A. Z. (2012a). Ssu72 phosphatase-dependent erasure of phospho-ser7 marks on the rna polymerase ii c-terminal domain is essential for viability and transcription termination. J Biol Chem, 287(11):8541–8551. 5
- [Zhang et al., 2012b] Zhang, D. W., Rodriguez-Molina, J. B., Tietjen, J. R., Nemecek, C. M., and Ansari, A. Z. (2012b). Emerging views on the ctd code. Genet Res Int, 2012:347214. 2, 4, 6
- [Zhang et al., 2005] Zhang, Z., Fu, J., and Gilmour, D. S. (2005). Ctd-dependent dismantling of the rna polymerase ii elongation complex by the pre-mrna 3'-end processing factor, pcf11. Genes Dev, 19(13):1572–1580. 6
- [Zhurinsky et al., 2010] Zhurinsky, J., Leonhard, K., Watt, S., Marguerat, S., Baehler, J., and Nurse, P. (2010). A coordinated global control over cellular transcription. Curr Biol, 20(22):2010–2015. 72
- [Zylber and Penman, 1971] Zylber, E. A. and Penman, S. (1971). Products of rna polymerases in hela cell nuclei. Proc Natl Acad Sci U S A, 68(11):2861–2865. 1

## Abbreviations

4tU	4-thiourcail
bp	base pairs
ChIP	Chromatin immunoprecipitation
cDTA	comparative Dynamic Transcriptome Analysis
CTD	Carboxy terminal domain (of Pol II)
CUTs	Cryptic unstable transcripts
FRB	FKBP12-rapamycin-binding
SUTs	Stable unannotated transcripts
XUTs	Xrn1-sensitive unstable transcripts
NGS	Next generation sequencing
NUTs	Nrd1-unterminated transcripts
OD	Optical density
ORF-T	Transcribed protein coding region
pA	poly(a) site
PAR-CLIP	Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation
PCR	Polymerase chain reaction
<i>Sc</i>	<i>Saccharomyces cerevisiae</i>
<i>Sp</i>	<i>Schizosaccharomyces pombe</i>