



# Genomewide analysis of road-block termination

Tito Candelli

## ► To cite this version:

Tito Candelli. Genomewide analysis of road-block termination. Genetics. Université Paris Saclay (COmUE), 2016. English. NNT: 2016SACLS587 . tel-01674204

**HAL Id: tel-01674204**

**<https://theses.hal.science/tel-01674204>**

Submitted on 2 Jan 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2016SACLS587

THÈSE DE DOCTORAT  
DE L'UNIVERSITÉ PARIS-SACLAY  
PRÉPARÉE À L'UNIVERSITE PARIS-SUD

École doctorale n°577  
Structure et dynamique des systèmes vivants  
Spécialité de doctorat: Sciences de la Vie et de la Santé

par

**Tito Candelli**

GENOMEWIDE ANALYSIS OF ROAD-BLOCK TERMINATION

Thèse présentée et soutenue à l'Institut Jacques Monod, le 12 Décembre 2016.

Composition du Jury :

M. DOMENICO LIBRI	Directeur de recherche Institut Jacques Monod	(Directeur de thèse)
M. ALAIN JACQUIER	Directeur de recherche Institut Pasteur	(Rapporteur)
M. VINCENT GELI	Directeur de Recherche CRCM	(Rapporteur)
M. DANIEL GAUTHERET	Professeur Univ. Paris-sud	(President du Jury)



# Genomewide analysis of road-block termination

## ABSTRACT

Transcription of DNA into RNA intermediates constitutes the first step in gene expression. During the last decade, several studies showed that about 80-90% of the genome is transcribed, and that transcription can initiate almost anywhere. This process—known as pervasive transcription—represents a serious threat to proper gene expression as it has the potential to interfere with not only other transcription events, but any DNA-based process. Selective transcription termination is therefore a mechanism of paramount importance for genome transcriptome stability and correct regulation of gene expression. Here we describe road-block termination, a novel termination mechanism for RNA polymerase II that functions to limit pervasive transcription and buffer the consequences of readthrough transcription at canonical terminators in *S.cerevisiae*. We show that several transcription factors can elicit this termination and that a number of unexpected genomic loci are associated with it. Additionally, we explore the possibility that road-block termination might contribute to specification of replication origins.

## RÉSUMÉ

La transcription de l'ADN en ARN constitue la première étape de l'expression d'un gène. Durant les dix dernières années, plusieurs études ont montré qu'environ 80-90% du génome est transcrit et que la transcription peut démarrer presque partout. Ce phénomène, connu sous le nom de transcription envahissante, représente une menace sérieuse contre l'expression correcte du génome car il peut interférer non seulement avec d'autres événements de transcription mais également avec n'importe quel procédé impliquant l'ADN. Une terminaison sélective est donc un mécanisme de la plus haute importance pour la stabilité du génome et la correcte régulation de l'expression des gènes. Ici nous décrivons la terminaison road-block, un nouveau mécanisme de la terminaison par l'ARN polymérase II, qui a pour fonction de limiter la transcription envahissante et de limiter les conséquences d'une translecture au niveau des sites de terminaison canoniques de *S.cerevisiae*. Nous démontrons également que plusieurs facteurs de transcription peuvent entraîner cette terminaison et que certains sites génomiques y sont associés. De plus, nous explorons également la possibilité que ces terminaisons road-block puissent contribuer à rendre spécifiques les origines de répllication.



# Contents

<b>Table of Contents</b> . . . . .	<b>iv</b>
<b>List of Figures</b> . . . . .	<b>viii</b>
<b>Preface</b> . . . . .	<b>x</b>
<b>I Introduction</b> . . . . .	<b>1</b>
<b>1 Transcription Initiation</b> . . . . .	<b>2</b>
1.1 Spatial Definition: Chromatin Structure and Core Promoter Elements	2
1.2 Temporal Definition: Gene-Specific Transcription Factors . . . . .	4
1.3 PIC Assembly and Promoter Clearance . . . . .	5
<b>2 Transcription Elongation</b> . . . . .	<b>8</b>
2.1 Elongation Through Chromatin . . . . .	9
2.2 Transcriptional Pausing . . . . .	11
2.3 The CTD . . . . .	12
2.3.1 CTD Phosphorylation Dynamics . . . . .	12
2.3.2 Functional Interactions . . . . .	14
<b>3 Transcription Termination</b> . . . . .	<b>16</b>
3.1 The CPF-CF Pathway . . . . .	17
3.1.1 Recruitment and Assembly . . . . .	17
3.1.2 The Allosteric Model . . . . .	18
3.1.3 The Torpedo Model . . . . .	19
3.1.4 A Unified View of CPF-CF Transcription Termination . . .	21

3.2	The NNS Pathway . . . . .	22
3.2.1	The NNS Complex . . . . .	22
3.2.2	The Mechanism of Transcription Termination . . . . .	24
3.2.3	Processing Products of the NNS Pathway . . . . .	27
3.3	Non-Canonical Termination Pathways . . . . .	28
3.3.1	Rnt1-Dependent Termination . . . . .	29
3.3.2	Road-Block Termination . . . . .	30
<b>4</b>	<b>The Transcriptional Landscape of <i>S.cerevisiae</i> . . . . .</b>	<b>32</b>
4.1	Control of pervasive Transcription . . . . .	33
4.2	Classes of Pervasive Transcript . . . . .	33
4.3	Quality Control Pathways . . . . .	36
4.4	Functional Role of Pervasive Transcription . . . . .	37
<b>5</b>	<b>General Regulatory Factors . . . . .</b>	<b>38</b>
5.1	Rap1 . . . . .	38
5.2	Abf1 . . . . .	40
5.3	Reb1 . . . . .	41
5.4	Cbf1 . . . . .	42
5.5	Chromatin Remodeling . . . . .	43
5.5.1	Genome-Wide Effect on Chromatin Structure . . . . .	44
<b>6</b>	<b>Transcription and Replication . . . . .</b>	<b>46</b>
6.1	Replication Origin and Their Specification . . . . .	47
6.1.1	Origin DNA Elements . . . . .	47
6.1.2	Nucleosome Positioning in Origins . . . . .	48
6.1.3	Transcription in Origins . . . . .	49
6.2	Mechanisms of DNA replication . . . . .	50

<b>II</b>	<b>Results and Discussion</b>	<b>53</b>
<b>7</b>	<b>Termination of RNA Polymerase II Through a Road-Block Mechanism . . . . .</b>	<b>54</b>
7.1	Road-Block Termination by Reb1 Restricts Cryptic and Readthrough Transcription . . . . .	55
7.2	Genomewide Analysis of Road-Block Termination . . . . .	87
7.2.1	Introduction . . . . .	88
7.2.2	Results . . . . .	91
7.2.3	Discussion . . . . .	111
7.2.4	Supplementary Figures . . . . .	118
7.3	General Discussion . . . . .	124
7.3.1	Fail-Safe Termination . . . . .	124
7.3.2	Road-Block Termination Promotes Genome Stability . . . .	125
<b>8</b>	<b>The Effect of Endogenous Transcription on Origin Specification</b>	<b>128</b>
8.1	Global Visualization of Transcription Around Replication Origins .	129
8.2	Transcriptional Pausing and Termination Are Associated With Replication Origins . . . . .	130
8.3	Transcription Levels Asymmetrically Affect Origin Efficiency . . . .	133
8.3.1	Licensing Efficiency . . . . .	134
8.3.2	Firing Efficiency . . . . .	136
8.3.3	Timing of Firing . . . . .	138
8.4	Discussion . . . . .	140
8.4.1	Transcription Termination Is a Feature of Replication Origins	141
8.4.2	ACS Orientation Determines the Impact of Transcription on Replication Efficiency . . . . .	141
<b>9</b>	<b>Dynamics of Nrd1-Nab3 RNA Binding <i>in vitro</i> and <i>in vivo</i> . . .</b>	<b>145</b>
9.1	<i>In vitro</i> Selection of RNA Sequences With High Affinity for the Nrd1-Nab3 Heterodimer . . . . .	146
9.2	Arrangement of Binding Sites Influences Heterodimer Affinity <i>in vitro</i>	148

9.3	Comparison of SELEX and <i>in vivo</i> Artificial CUT Selection Unveils Unexpected Dynamics of Nrd1 Binding . . . . .	152
9.4	Nrd1 Binding Sites GUAG and GUAA Possess Different Extended Consensuses . . . . .	154
9.4.1	Confirming the Differences <i>in vivo</i> . . . . .	155
9.5	Discussion . . . . .	157
9.5.1	Site Arrangement and Spacing Influence Binding Affinity and Termination Efficiency . . . . .	158
9.5.2	Comparison of SELEX and <i>in vivo</i> CUT Selection Reveals Clusters of Differentially Enriched Sequences . . . . .	159
<b>Methods . . . . .</b>		<b>160</b>
<b>Glossary . . . . .</b>		<b>173</b>
<b>Bibliography . . . . .</b>		<b>175</b>

# Figures

1.1	Stepwise PIC assembly . . . . .	6
2.1	Mechanism of transcription through chromatin. . . . .	9
2.2	CTD phosphorylation states throughout the transcription cycle. . .	13
3.1	Mechanism of CPF-CF termination . . . . .	20
3.2	Mechanism of NNS termination . . . . .	25
4.1	Classes of transcripts and their fates. . . . .	34
5.1	GRFs binding sites and <i>S.cerevisiae</i> centromere structure . . . . .	39
6.1	ACS consensus and arrangement relative to other origin DNA elements	48
6.2	Stepwise mechanism of DNA replication . . . . .	51
7.1	Rap1-dependent transcripts are isolated from <i>in vivo</i> selection . . .	92
7.2	Northern blot analysis of Rap1 and Reb1 terminated species . . . . .	95
7.3	Examples of Rap1-dependent termination detected <i>in vivo</i> . . . . .	97
7.4	Metagene analysis around sites of Rap1 and Reb1. . . . .	99
7.5	Metagene analysis around poly(A) sites that are within 300bp up- stream of Rap1 sites. . . . .	102
7.6	Examples of CPF-terminated transcripts followed by sites of Rap1 .	103
7.7	Analysis of 4 snoRNAs followed by road-block by CRAC and RNA- seq in several mutant strains. . . . .	104
7.8	RNA-seq analysis of genes followed by Rap1 sites. . . . .	107
7.9	Metagene analysis around several putative road-blocking factors . .	109
7.10	Centromere structure and polymerase occupancy profile . . . . .	110

7.11	metagene analysis of RNAPII occupancy around tRNAs . . . . .	111
8.1	ACS consensus and arrangement relative to other origin DNA elements	129
8.2	Metagene analyses showing polymerase occupancy and termination around replication origins . . . . .	131
8.3	Northern blot analysis of transcription through ARS305 . . . . .	132
8.4	Cartoon of sense and antisense transcription relative to the ACS . .	133
8.5	Boxplots comparing licensing efficiencies in high- and low-transcription populations . . . . .	135
8.6	Boxplots comparing transcription levels in high- and low-firing pop- ulations . . . . .	137
8.7	Boxplots comparing median replication times in high- and low- transcription populations . . . . .	138
8.8	A model of how transcription can affect replication efficiency . . . .	142
8.9	Correlations between transcription and replication efficiency in a NNS defective strain . . . . .	143
9.1	SELEX procedure . . . . .	147
9.2	Enrichment of Nrd1 and Nab3 sites with different arrangements and spacings in the SELEX experiment . . . . .	149
9.3	Schematic of the selection process in the artificial CUT selection . .	150
9.4	Enrichment of Nrd1 and Nab3 sites with different arrangements and spacings in the artificial CUT selection . . . . .	151
9.5	Direct enrichment comparison between SELEX and artificial CUT selection for all 4 nucleotides motifs . . . . .	153
9.6	Comparison of flanking nucleotides between GUAG and GUAA in SELEX and genomic CUT sequences . . . . .	156

# Preface

Transcription of DNA into RNA intermediates constitutes the first step in gene expression. Even minute changes in transcription patterns can upset the balance of many essential cellular constituents, generating a cascade of responses with significant repercussions on every biological process. Because of this massive potential, transcription is one of the most finely regulated events in the cell and according to the Saccharomyces Genome Database (SGD) [27] Gene Ontology annotation, 1231 out of 6691 genes in *S.cerevisiae* (18%) can influence or directly take part in the transcriptional process.

In eukaryotes, three distinct RNA polymerases exist. RNA Polymerase I (RNAPI): responsible for the transcription of Ribosomal RNA (rRNA); RNA Polymerase II (RNAPII): responsible for the transcription of both protein coding genes and many non-coding RNAs; and RNA Polymerase III (RNAPIII): responsible mainly for the transcription of tRNAs and some rRNA. Although products of RNAPI and RNAPIII are by far the most abundant in the cell, RNAPII is tasked with the production of an extremely varied set of transcripts and it is estimated that 80% of the genome is actively transcribed by it [34]. Because of this pervasiveness, transcription by RNAPII must be tightly regulated to ensure its products are viable, as well as to prevent interference with other processes. In this dissertation i will focus on how transcription by RNAPII is controlled—especially through transcription termination—and what its effects are on other DNA-based biological processes.

The first three chapters of the introduction to this work will describe the transcriptional process along its three main steps: initiation, elongation and termination. I will highlight the main molecular determinants that give rise to each phase, as

well as mechanistically characterize the process when appropriate. Because of the relevance for the results that will be presented, I have devoted particular attention to transcription termination and described it in detail. In chapter 4, i will talk about the transcriptional landscape of *S.cerevisiae*; a look into the world of pervasive transcription, along with the mechanisms that control it. I will highlight the different classes of non-coding RNAs transcribed by RNAPII as well as the quality control pathways that ensure their degradation. In connection with the results of this dissertation, chapter 5 will discuss a particular class of transcription factors known as General Regulatory Factors. I will describe these factors in the context of their multiple functions, focusing on their chromatin remodeling capabilities and their function at gene promoters. Finally, in chapter 6 I will consider the process of DNA replication and its interaction with transcription. I will first put replication in its appropriate context by describing the structure of replication origins and the mechanics of the process itself. I will then discuss the available literature in regard to the effect of transcription on replication initiation and origin specification.

In the results part, i will outline three different projects. The first consists of the characterization of road-block termination, a novel termination mechanism for RNAPII. Second, i will explore the interaction between transcription and DNA replication, with particular attention to the effect of transcription on origin usage. Finally, the last chapter will focus on NNS termination and how the components of the NNS complex contact their cognate binding sites in different contexts.

The results presented here were obtained using *S.cerevisiae* as a model organism. Therefore, the ensemble of data cited in this manuscript refers to this organism unless otherwise stated.



“MATHEMATICS, RIGHTLY VIEWED,  
POSSESSES NOT ONLY TRUTH, BUT SUPREME BEAUTY  
—A BEAUTY COLD AND AUSTERE,  
WITHOUT THE GORGEOUS TRAPPINGS OF PAINTING OR MUSIC.”  
— BERTRAND RUSSELL

“THE ENDING ISN’T MORE IMPORTANT  
THAN ANY OF THE MOMENTS THAT LED TO IT.”  
— DR. ROSALENE, *To the Moon*

“SOMETIMES IT IS MORE IMPORTANT TO TAKE IN THE SPECTACULAR  
THAN TO WORRY ABOUT THE PRESSING BUSINESS OF STAYING ALIVE.”  
— DOUGLAS DORST, *S.*

## Part I

# Introduction

# 1

## Transcription Initiation

Initiation is the first step in any transcription event. It therefore needs to be accurate in when and where it occurs. Transcription initiation fundamentally relies on the assembly of the Pre-Initiation Complex (PIC) (a super-complex 1.5 megadaltons in size containing RNAPII [47]) on DNA. The assembly of such complex is spatially defined by two elements: chromatin structure and core promoter elements. Both contribute to limit the amount of spurious transcription by ensuring robust assembly of the PIC only in promoter regions. In addition to spatial regulation, timing and intensity of transcription initiation must also be controlled. Specific promoters can be finely tuned by the binding of gene-specific transcription factors, that can act as either enhancers or repressors; modulating initiation efficiency either constitutively or in response to environmental effects. Finally, when the PIC is fully assembled, it eventually escapes the promoter and enter productive elongation.

### 1.1 Spatial Definition: Chromatin Structure and Core Promoter Elements

Chromatin is a higher order structure that forms when DNA wraps around histones, proteins that can efficiently arrange loose DNA into compact structures. The

simplest unit of chromatin consists of 140 nucleotides of DNA tightly wrapped around a histone, forming a nucleosome. The organization of the genome around nucleosome units has a multitude of consequences, not least of which is to sterically prevent DNA binding proteins from accessing their substrate. As transcription relies on assembly of RNAPII and the PIC on DNA to complete its initial phase, nucleosomes pose a considerable barrier to efficient initiation [52, 81]. The insulation of DNA by nucleosomes has been harnessed by the cell and made into a regulatory mechanism that can spatially define where transcription initiates, as transcription factors must bind DNA for it to occur. To favor transcription factor binding to DNA in promoter regions, the latter are always associated with an Nucleosome Free Regions (NFR), an area of the genome where nucleosomes are depleted, leaving naked DNA available for binding. Although certain sequence elements can passively discourage nucleosome association, several complexes actively mediate depletion of nucleosomes from promoter regions, such as SWI1/SNF and the closely related RSC complex. These complexes can be recruited in two ways: through sequence specificity [8, 90] or through recruitment by gene-specific transcription factors such as Reb1p, Abf1p, and Rap1p to promoter regions [8, 53, 66, 172].

While chromatin defines the position of transcription initiation, core promoter elements provide specificity for many early-acting general transcription factors. A number of promoter elements were identified in metazoans, where they have been shown to regulate position and intensity of transcription initiation; but although a Transcription Start Site (TSS) consensus was recently defined [112], no sequence was found to be universally required for transcription initiation [19]. In *S.cerevisiae* promoter elements remain poorly characterized and seem to lack the majority of the sequence elements found in their metazoan counterparts. The major element known to bring about the assembly of the PIC in *S.cerevisiae* is the TATA box. This very short consensus sequence, TATAWAWR [9], is present in about 15% of yeast genes [83] and is recognized by the TATA Binding Protein (TBP), an essential factor for PIC assembly. At these promoters, TBP binds DNA as part of the Spt-Ada-Gcn5-Acetyl transferase (SAGA) complex, changing the conformation of DNA and

priming the promoter for assembly of other general transcription factors. TATA-dependent promoters, however, are not the only type of promoter in *S.cerevisiae*. The majority of yeast promoters (85-90% ) are known as TATA-less and require binding of the TFIID complex in lieu of SAGA [155]. Curiously, TBP, along with a number of other shared subunits and co-factors, is also contained in the TFIID complex, but it was recently shown that, in this context, its binding activity is not required for gene activation [83]. TFIID and SAGA have largely overlapping roles in activating gene expression, however, the predominant activity of the two complexes can be associated with functional differences. While TFIID generally dominates over house-keeping genes that do not require regulation, SAGA—and as a consequence TBP binding—has a larger effect over highly regulated and stress-inducible genes [76]. The binding of either complex represents the first step towards assembly of other general transcription factors into the PIC.

## 1.2 Temporal Definition: Gene-Specific Transcription Factors

While nucleosome positioning and core promoter elements define where transcription should initiate, they do not generally actively regulate it on their own. In the cell, many genes need to be activated in response to specific conditions or external stimuli. These regulated genes are generally inactive and become actively transcribed only when the conditions of their activation are met. The main mechanism that enables these transcriptional switches is the presence of gene-specific transcription factors. These DNA-binding proteins specifically target promoter regions, modulating their activity in response to a large number of conditions. Gene-specific transcription factors can activate—or repress—transcription in a variety of ways: activation can occur by binding DNA and recruiting NFR-generating complexes, or otherwise facilitating PIC assembly, and even by relocating chromatin to the nuclear periphery [21]. Alternatively, transcription factors can constitutively repress their target genes and selectively lose the DNA-binding capability under certain conditions, such as

the presence of a ligand.

Genome-wide studies on transcription factor organization highlighted the combinatorial potential that emerges when several transcription factors interact with the same promoters [65]. Regulation of a single promoter by several distinct transcription factors can exploit their different requirements—qualitative or quantitative—to force the emergence of complex regulatory logic.

### 1.3 PIC Assembly and Promoter Clearance

Assembly of the Pre-Initiation Complex starts with the binding of either TFIID or SAGA to promoter DNA. The presence of TBP in these complexes modifies the structure of DNA, allowing the step-wise recruitment of several general transcription factors and of RNAPII [For review see 162].

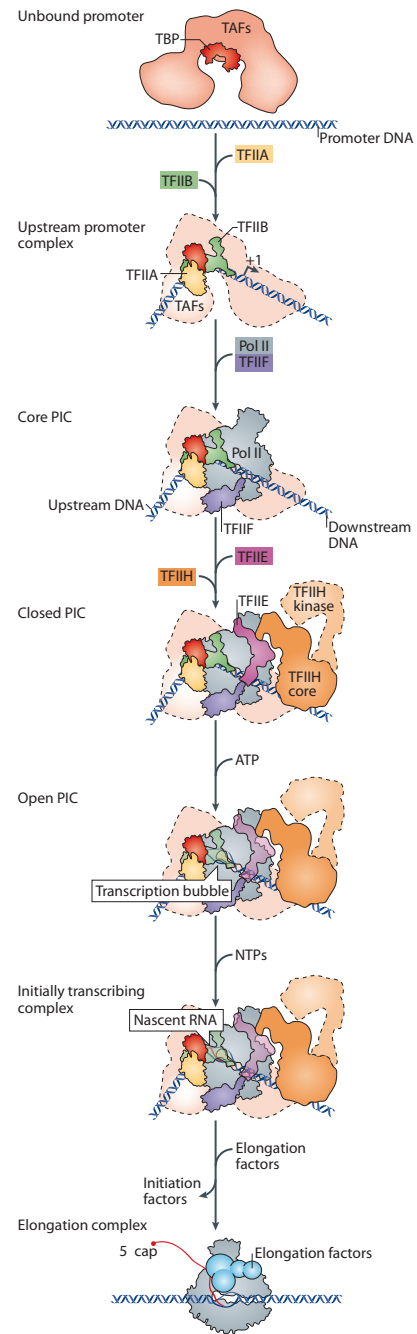
TFIIA and TFIIB are the first factors to make contact with TBP, stabilizing its interaction with DNA. However, while TFIIA simply acts as an auxiliary factor and is dispensable [78], TFIIB is required for RNAPII recruitment [17]. The presence of TFIIB acts as a platform for TFIIF docking. The addition of TFIIF has the double effect of recruiting RNAPII (RNAPII is bound to TFIIF when in free form [147]) and of further stabilizing the whole PIC. Despite the inclusion of RNAPII in the forming PIC, at this stage promoter DNA is firmly wound-up in a double helix and therefore the ternary complex<sup>1</sup> required for transcription cannot yet form. TFIIIE and TFIIH are recruited to the PIC to solve this problem. TFIIIE acts as a bridge between RNAPII and TFIIH, who contains an ATPase module and is able to unwind promoter DNA [72]. This will eventually contribute to DNA melting and the formation of the open PIC, a structural variant that precedes the shift into elongation.

---

<sup>1</sup>The ternary complex is defined as the three-way interaction between DNA, RNA, and RNAPII that forms within transcribing polymerases

The order of stepwise assembly of general transcription factors into a functional PIC was first discovered *in vitro* [15]. *In vivo*, however, there is evidence for the activity of the mediator complex in providing additional assembly pathways [46]. Mediator is a large and flexible protein complex that can interact with virtually every general transcription factor and with RNAPII. It is known for its fundamental role in transducing regulatory signals from gene-specific transcription factors to the polymerase. Without Mediator, the PIC can drive basal transcription levels, but its activity cannot be modulated in response to external factors. Studies have implicated mediator in the recruitment of TFIIE and TFIIH independently of RNAPII, providing alternative ways to assemble the complete PIC. Additionally, interactions between RNAPII and mediator were found to be required for transcription *in vivo* [171].

After the assembly of the PIC and the Mediator complex on the promoter, RNAPII relies on TFIIH to relax DNA and physically separate the two strands, creating what is referred to as the transcription bubble. Studies in human report that once the bubble first opens, it spans about 7 nucleotides. It then extends forward, allowing the process of transcription to begin. Polymerases at this stage, however, have to contend with the fact that the RNA-DNA hybrid is



**Figure 1.1:** Stepwise assembly of general transcription factors and RNAPII on a promoter. adapted from [162].

too short to be stable. According to *in vitro* studies, forming a sufficiently long—and therefore stable—hybrid requires several rounds of abortive initiation, where the small RNA is displaced from the template and released. When the RNA-DNA hybrid reaches a length of about 10 nucleotides, the upstream half of the bubble, which now spans 17-18 nucleotides, collapses, suddenly closing [71]. This event marks the detachment of what will eventually become the elongation complex from the scaffold of general transcription factors that is going to be retained at the promoter [134].



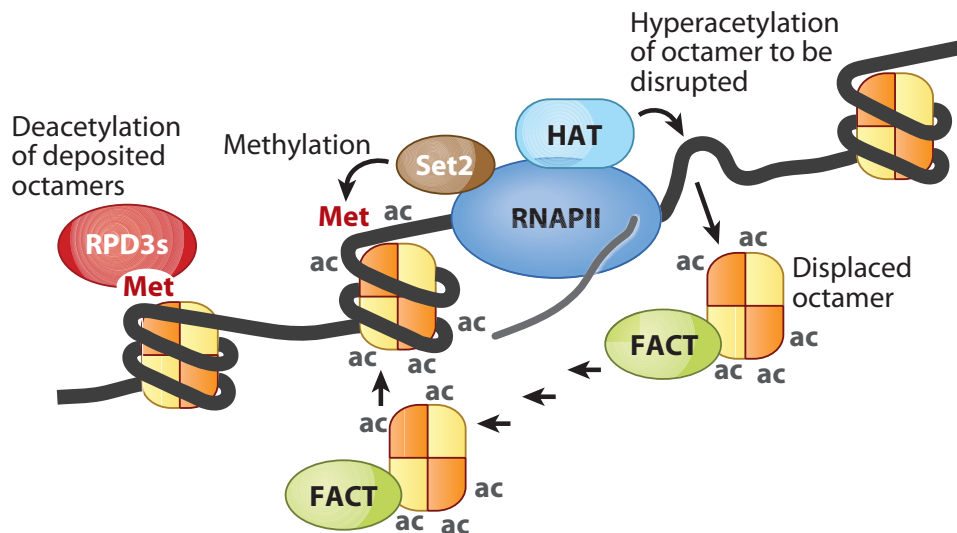
# 2

## Transcription Elongation

After escaping the PIC, RNAPII enters the phase of productive elongation. During this phase, the polymerase travels along DNA, catalyzing the addition of nucleotides to the growing RNA molecule that is being synthesized. The simple synthesis of RNA, however, is not enough to qualify a mature transcript. Several essential processing steps take place during transcription elongation and contribute to the production of fully formed transcripts. Among these, the addition of the 5' cap, addition of a poly(A) tail, and formation of an export-competent transcript all rely on the presence of RNAPII and the Transcription Elongation Complex (TEC) in order to be carried out properly. The precise composition of the TEC is poorly understood. However, as RNAPII progresses through the transcription unit, several complexes and co-factors are known to dynamically associate with it in order to enact the various maturation steps. Transcription elongation is therefore a highly regulated activity that coordinates several different processes to produce mature transcripts. This dynamic regulation is enacted by the cell through several distinct mechanisms, such as the phosphorylation of the C-Terminal Domain (CTD) and the modification of histones. These very same regulation mechanisms—along with important regulatory sequences—will eventually mark the end of transcription elongation and the transition to transcription termination.

## 2.1 Elongation Through Chromatin

Chromatin represents an extremely repressive barrier to any kind of DNA based process. As I briefly touched upon in previous sections, chromatin components—histones—need to be actively dislodged from promoter regions in order to allow the Pre-Initiation Complex to assemble. Elongating RNAPII faces very similar problems, as in order to synthesize the RNA, it has to move through an array of nucleosomes without losing contact with DNA. Although *in vitro* evidence has shown that RNAPII can effectively elongate through a single nucleosome [104]—possibly due to spontaneous disassembly and reassembly of nucleosomes, a process that was recently shown to happen every few seconds [86]—the elongation complex alone is not enough to mediate transcription through multiple nucleosomes.



**Figure 2.1:** Overview of the main actors in the mechanism of transcription through chromatin. Nucleosomes are destabilized through acetylation and chaperoned away—either partially or completely—by FACT and other complexes. Addition of methyl groups to histone tails allows the recruitment of Histone De-Acetylases (HDACs) and the restoration of chromatin structure. adapted from [168].

The TEC can overcome this problem by enlisting the help of several histone chaperones and chromatin remodeling complexes, as well as by exploiting post translational

modifications of histones (Fig: 2.1). The current model for transcription through nucleosomes posits that, depending on the intensity of transcription, histones can either be completely removed from DNA, or be partially destabilized as to allow RNAPII to more easily transcribe through them [94]. The most notable actors in this phase are Histone Acetyl-Transferases (HATs) such as Gcn5 and the FACT (Facilitates chromatin transcription) complex [For review see 150]. HATs are posited to travel with the polymerase, depositing an acetyl group on histone tails. This has the consequence of destabilizing inter-nucleosome interactions as well as lowering the affinity for DNA by increasing negative charges, resulting in a more relaxed chromatin structure and more unstable nucleosomes. Once histones are acetylated, FACT—also traveling with the polymerase—destabilizes the H2A-H2B dimer<sup>1</sup>, removing it and facilitating transcription through the remaining incomplete nucleosome structure.

We saw how, in order to efficiently elongate, RNAPII needs to destabilize the chromatin structure. However, in the long run, this destabilization can have negative effects, as it results in a more relaxed structure that can potentially give rise to intragenic transcription initiation. In order to prevent this phenomenon, the composition, modifications, and overall structure of nucleosomes must be reset after the passage of RNAPII. Specific histone chaperones such as Spt6, together with methyl-transferases and HDACs, are involved in this process. First, Spt6 and other histone chaperones reconstruct a complete histone in the wake of transcribing RNAPII. Subsequently, methyl-transferases such as Set2 methylate lysine 36 on histone H3. Although this modification—unlike acetylation—has no structural consequences on the organization of nucleosomes, it can act as a platform for recruitment of HDACs. The RPD3 complex has high affinity for H3K36 methylation and is recruited immediately after the passage of RNAPII in order to remove the acetyl groups from histones and thus reset the structure of chromatin.

---

<sup>1</sup>Two of the four core components of a histone. Histones are composed of two H2A-H2B dimers and one H3-H4 tetramer arranged in a symmetrical structure.

## 2.2 Transcriptional Pausing

Nucleosomes do not represent the only obstacle to productive elongation. A number of events can potentially prevent RNAPII from elongating forward, such as DNA damage, misincorporation of a nucleotide, or collision with another DNA-bound protein. This causes RNAPII to temporarily stop, a phenomenon known as transcriptional pausing. One of the mechanisms used by the polymerase to resolve pausing is called backtracking.

During backtracking, RNAPII moves backwards, retracing its steps. how much the polymerase is able to backtrack *in vivo* is a matter of debate, early *in vitro* studies reported up to 100 nucleotides, while more recent crystal structures concluded that this figure can range from 4-5 up to 12-15 nucleotides [28]. This backwards movement causes part of the already synthesized RNA to slide forward into a channel connected to the outside of the complex. Presence of RNA into the channel promotes the binding of TFIIS<sup>2</sup> to the complex [28]. This stimulates the intrinsic endonucleolytic activity of RNAPII, which results in cleavage of the extruding RNA and realignment of the 3' end of the nascent transcript with the catalytic site of the polymerase. At this point, RNAPII has effectively reset its position, having moved back and gotten rid of the extra segment of RNA. It can therefore restart its forward translocation and resume the normal catalytic activity.

In some cases, the backtracking process cannot resolve pausing. As transcriptional pausing becomes more prolonged, it progressively evolves towards another state called transcriptional arrest. An arrested polymerase necessitates the intervention of specific factors in order to restart elongation or be removed from the DNA template. For example, when DNA damage causes RNAPII to arrest, a number of sequentially acting E3 ubiquitin ligases are required to attach ubiquitin chains on the polymerase and lead to its degradation, allowing DNA repair factors to access the damage. The current model posits that Rsp5 acts to monoubiquitylate RNAPII, while

---

<sup>2</sup>Also known as *Dst1*

the Elc1/Cul3 complex elongates that chain, resulting in the recruitment of the proteasome and disassembly of the elongation complex through degradation of the polymerase [10].

## 2.3 The CTD

RNAPII and the elongation complex are fundamental elements in coordinating many of the co-transcriptional processes that contribute to the maturation of the nascent RNA. The main subunit of the polymerase, Rpb1<sup>3</sup>, possesses a structure that allows it to dynamically recruit all the necessary factors and complexes in a timely fashion: The CTD. The CTD is an unstructured C-terminal domain composed, in *S.cerevisiae*, of 26 repeats of the heptapeptide YSPTSPS<sup>4</sup>. This cluster of repeats can be differentially phosphorylated in different phases of transcription elongation, acting as a dynamically changing interaction surface for different co-factors.

### 2.3.1 CTD Phosphorylation Dynamics

The CTD heptapeptide contains a high number of phosphorylatable residues. Out of the 7 aminoacids, 5 can support the addition of a phosphate group: TYR<sub>1</sub>, SER<sub>2</sub>, THR<sub>4</sub>, SER<sub>5</sub>, and SER<sub>7</sub>. The combinatorial phosphorylation of SER<sub>2</sub> and SER<sub>5</sub>, however, provides the majority of the better known functional contribution to transcription elongation and it was recently shown that phospho-groups at these two residues are more abundant than on any of the other residues [176].

Unlike SER<sub>2</sub> and SER<sub>5</sub>, our understanding of the consequences of TYR<sub>1</sub>, THR<sub>4</sub>, and SER<sub>7</sub> phosphorylation is still limited. In vertebrates, THR<sub>4</sub> has been implicated in the processing—but not transcription—of histone genes [74], while SER<sub>7</sub> was shown to recruit the CTD phosphatase Rpap2 specifically to Small Nuclear RNAs

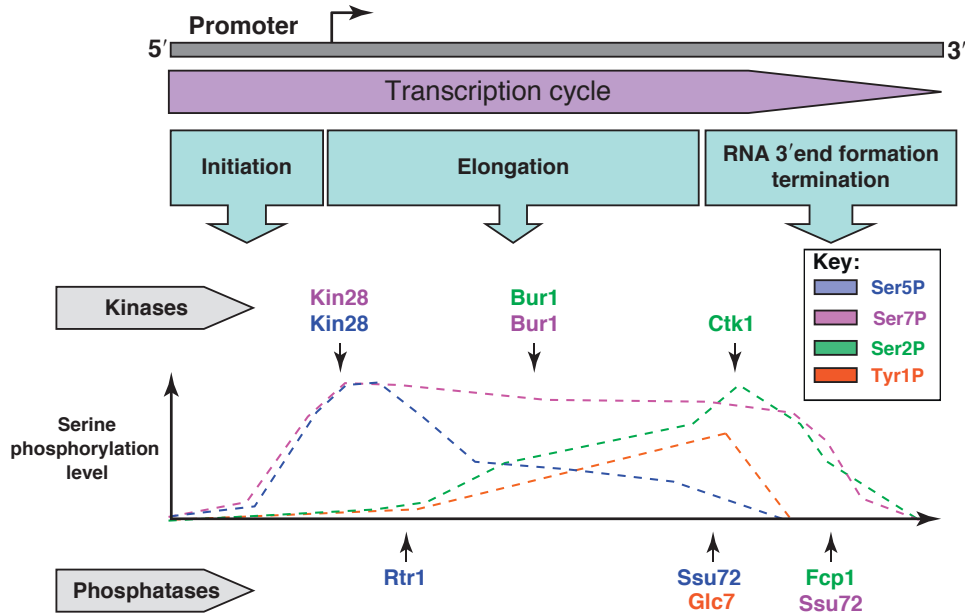
---

<sup>3</sup>Also known as Rpo21

<sup>4</sup>TYR<sub>1</sub>-SER<sub>2</sub>-PRO<sub>3</sub>-THR<sub>4</sub>-SER<sub>5</sub>-PRO<sub>6</sub>-SER<sub>7</sub> in expanded nomenclature

(snRNAs) genes [44]. Recent studies in *S.cerevisiae* found that phosphorylation of TYR<sub>1</sub> impairs recruitment of specific termination factors [119, 165], however no role is known for TYR<sub>1</sub> in transcription elongation.

In light of this, in the following paragraphs I will focus mainly on the mechanisms and effects of SER<sub>2</sub> and SER<sub>5</sub> phosphorylation.



**Figure 2.2:** General view of SER<sub>2</sub>, SER<sub>5</sub>, SER<sub>7</sub>, and TYR<sub>1</sub> phosphorylation along the transcription cycle, kinases and phosphatases involved in CTD modification are represented immediately above and below the graph. The two main phosphorylation states, SER<sub>2</sub> and SER<sub>5</sub>, are dominant at the 3' and 5' respectively, reflecting their functional roles in the termination and early elongation phases of transcription. SER<sub>7</sub> is consistently present throughout the transcription cycle, but its functional impact in yeast remains elusive. Adapted from [45].

During the initiation phase of transcription, the CTD of RNAPII starts off unphosphorylated (Fig: 2.2). When the PIC is fully assembled, Kin28, a catalytic subunit of the general transcription factor TFIIF, phosphorylates the CTD heptapeptide on SER<sub>5</sub>. In *S.cerevisiae*, the CTD remains mostly SER<sub>5</sub> phosphorylated for the first 450 nucleotides of transcription elongation [120]. After this point the combined action of the SER<sub>5</sub>-phosphatase Rtr1 [77, 126] and the SER<sub>2</sub>-kinases Ctk1 [144]

make SER<sub>2</sub> the most prominent mark<sup>5</sup>. Despite phosphorylation of SER<sub>2</sub> reaching saturation about 600 nucleotides from the TSS [120], SER<sub>5</sub> phosphorylation is still present on many repeats, resulting in the presence of a double phosphorylation pattern with important functional consequences (see below). Only Towards the 3' end of the gene the action of CTD phosphatase Ssu72 completely abrogates the SER<sub>5</sub>-P mark, leaving SER<sub>2</sub>-P as the only active mark. Finally, additional activity of the Fcp1 phosphatase results in the removal of most phospho-marks from the CTD, readying the polymerase for another round of transcription.

### 2.3.2 Functional Interactions

As I outlined above, the transcription cycle follows specific patterns of CTD phosphorylation: unphosphorylated CTD is recruited to promoter regions, SER<sub>5</sub>-P dominates during early elongation and gradually makes way for SER<sub>2</sub>-P, which is the dominant mark in the later stages of transcription. Each of these stages comes with the potential to interact with numerous co-factors and provides modularity to the elongation complex.

The unphosphorylated state of free-form RNAPII CTD allows the polymerase to interact with the mediator complex; an interaction that is thought to contribute to the recruitment of RNAPII to active promoters. Once the PIC is assembled, the polymerase needs to escape the promoter and leave the Pre-Initiation Complex behind. The modifications that take place at this stage, namely SER<sub>5</sub> phosphorylation, are thought to disrupt the interaction between RNAPII and mediator—thereby allowing promoter clearance—although evidence remains inconclusive [35, 170].

The presence of SER<sub>5</sub> mark during early elongation has two direct consequences: it stimulates capping of the nascent transcript through recruitment of the capping enzymes [160], and it has the potential to promote early transcription termination

---

<sup>5</sup>It is interesting to note that the phosphorylation state of RNAPII CTD is independent of transcript length, but exclusively depends on the amount of nucleotides from the TSS. This will have important implications for the termination of non-coding transcripts.

through the recruitment of the Nrd1-Nab3-Sen1 (NNS) complex [186]. While capping is ubiquitous and required to prevent premature degradation of the transcript, early termination is a quality control mechanism that requires (in addition to SER<sub>5</sub>-P) the presence of specific sequence elements on the nascent transcript and will be described in detail in chapter 3.

Studies in mammals have reported that the CTD is required for splicing to occur properly [161]. In particular heptapeptides containing SER<sub>2</sub> phosphorylation are known to recruit several splicing factors [62]. Recent studies in *S.cerevisiae* show differential phosphorylation patterns in intronless and intron-containing genes, hinting at a possible functional interaction between splicing and CTD phosphorylation also in yeast [122].

Towards the end of the transcription cycle, SER<sub>2</sub>-P becomes the most prominent mark. This phase sees the recruitment of a number of different actors. Chromatin remodelers and histone modifying complexes such as Set2 and Spt6 are recruited through the CTD, making sure that the structure of nucleosomes is maintained [16].

Finally, 3' end processing, termination, and export are all affected by the CTD. binding of components of the cleavage and polyadenylation complex such as Pcf11 and Rtt103 stimulates the termination of transcription and the processing of the transcript 3' end (such as poly(A) tail addition), while recruitment of export factors such as Yra1 direct a rapid and efficient export to the cytoplasm [96].



# 3

## Transcription Termination

After its synthesis and maturation are complete, the nascent RNA molecule must be released from the DNA template, and the elongation complex must be disassembled and its components recycled. In *S.cerevisiae*, transcription termination is enacted by several widely different mechanisms. Two predominating pathways terminate the vast majority of transcripts generated by RNA Polymerase II: the Cleavage and Polyadenylation Factor/Cleavage Factor I (CPF-CF) pathway and the Nrd1-Nab3-Sen1 (NNS) pathway. Both these mechanisms rely on short sequences on the nascent RNA—coupled with specific modifications on the CTD of RNAPII—to recruit specific factors and enact the disassembly of the elongation complex and the release of the transcript in the nucleus. Moreover, both transcription termination mechanisms are strictly intertwined with some steps of 3' end processing and maturation, influencing the fate of the transcript after termination.

In addition to the two main pathways cited above, several non-canonical termination mechanism will be described. These mechanisms are dedicated to the termination of specific RNA species, or can act as backups when the main pathways fail.

## 3.1 The CPF-CF Pathway

The CPF-CF pathway was the first termination mechanism described in *S.cerevisiae* because of its association with the termination of protein-coding genes<sup>1</sup>. CPF-CF termination is unique as it results in cleavage of the nascent RNA before termination occurs. The site of cleavage is specified through sequence elements present on the nascent RNA and plays an important role in kickstarting the termination reaction.

The main actor of this termination mechanism is the CPF-CF complex, a large assembly of modular sub-complexes that act in concert to execute all the required steps. This complexity makes CPF-CF the most reliable, efficient, and precise termination mechanism in *S.cerevisiae*.

### 3.1.1 Recruitment and Assembly

Recruitment and initial assembly of the CPF-CF complex onto the nascent RNA is promoted by two mechanisms: interaction with specific sequences elements, and interaction with the polymerase CTD.

A key component of the CPF-CF complex, Pcf11, contains a peptide sequence able to recognize the CTD. This CTD Interaction Domain (CID) is able to specifically recognize the SER<sub>2</sub>-phosphorylated version of the heptapeptide. Given the nature of this CTD modification—which is confined to the later stages of transcription—density of the CPF-CF complex around the polymerase is selectively increased where the complex is more likely to be needed for termination (i.e. at the 3' end of transcription units), facilitating the eventual binding of CPF-CF to the sequence elements on the nascent RNA.

Unlike in human, where the cleavage site is defined by a single highly conserved hexanucleotide sequence on the nascent RNA, Yeast CPF-CF complex recognizes a

---

<sup>1</sup>Its activity can extend to certain kinds of non-coding transcripts as well (see section 4.2 for details)

number of degenerate short sequences. Two sub-complexes of CPF-CF, Cleavage Factor 1A (CF1A) and Cleavage Factor 1B (CF1B), are responsible for the recognition of these sequences. In particular, Rna15 and Hrp1 (components of CF1A and CF1B respectively) directly bind the nascent RNA. Associated factors Rna14 and Pcf11 contribute to the assembly of the whole complex by interacting with RNAPII and forming a scaffold that serves to tether the catalytic portion of the CPF-CF complex to the cleavage site.

The bulk of the catalytic activity of the CPF-CF complex is contained in the Cleavage and Polyadenylation Factor (CPF) sub-complex. CPF directly contacts the cleavage site with its Ysh1 subunit and is responsible for the cleavage of the nascent RNA, one of the events that is thought to kickstart the termination reaction. CPF also coordinates the polyadenylation reaction through the subunits Yth1 and Fip1. These factors recruit and tether the poly(A) polymerase Pap1 to the complex, which will begin catalyzing the addition of a poly(A) tail after the transcript has been cleaved.

Despite the wealth of knowledge available on the mechanics of CPF-CF recruitment and assembly, some controversy still surrounds the actual termination mechanism. Two main models describing the termination reaction exist in the literature, the allosteric model and the torpedo model.

### **3.1.2 The Allosteric Model**

After cleavage and release of the RNA, the elongation complex has successfully accomplished its job in the transcriptional process and is ready to be disassembled. The allosteric model is one of the two main mechanistic models that describes the process by which the TEC is removed from the DNA template.

The allosteric model argues that cleavage of the RNA is a dispensable signal, and that termination can happen independently of this step. It posits that after transcription of the cleavage site, RNAPII loses a lot of factors that qualify the

elongation complex as such. Loss of these “anti-termination” factors—components of the elongation complex that would prevent termination from occurring—would trigger conformational changes, destabilize the polymerase, and allow components of the CPF-CF complex itself to elicit the disassembly of RNAPII from the template.

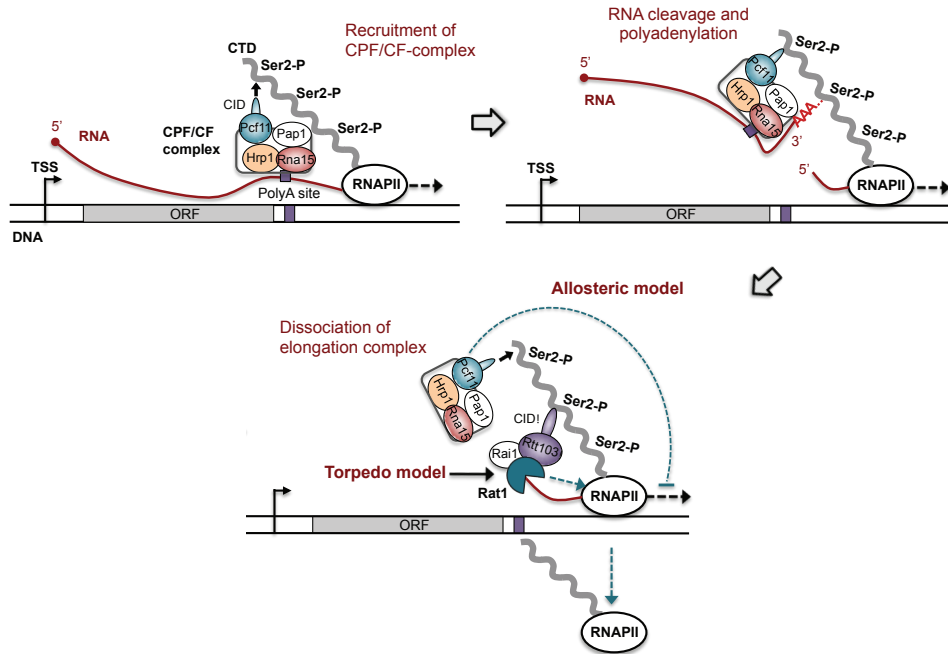
Several studies support this model. RNAPII was shown to lose a number of associated elongation factors after reaching the 3' end [87]. In addition, the component of the CPF-CF complex Pcf11 was shown to be able to terminate the polymerase *in vitro* by binding the nascent RNA and the SER<sub>2</sub>-phosphorylated moiety of RNAPII [205]. Ulterior support to this last study was provided by the same authors two years later, when they discovered that Pcf11 is able to perform the same feat in drosophila [206]. Finally, a very recent study was able to reconstitute transcription termination in an *in vitro* system in the absence of cleavage [204].

### 3.1.3 The Torpedo Model

According to the torpedo model, cleavage represents the main termination signal for the CPF-CF complex, as it leaves an uncapped 5'-P on the transcript associated with the still transcribing elongation complex. These unprotected 5' is the substrate of 5'→3' exonucleases, a class of enzymes that are known to progressively degrade RNA polypeptides. The 5'→3' exonuclease Rat1 was discovered to be associated with the CPF-CF complex and is thought to attack the 5' moiety of the RNAPII-associated transcript, starting a processivity race with RNAPII. Upon winning the race, Rat1 would destabilize the structure of the ternary complex within the polymerase, causing it to break apart and detach from the DNA template.

There are several lines of evidence that support this model for CPF-CF transcription termination. Both Rat1 and its human homologue Xrn2 exhibit termination defects in model cases when mutated [88, 190]. Furthermore, Rat1 and its co-factor Rtt103 were found to be strongly associated with the 3' end of genes and in physical association with the CPF-CF complex [88, 107], supporting the idea of a functional

recruitment to zones of active transcription termination. Homology studies found that homologues of Rtt103 in both humans and *C.elegans* have roles in transcription termination [31, 125]. Finally, recent mechanistic studies *in vivo* have demonstrated the kinetic competition between Rat1 and the elongation complex. By employing mutant polymerases that elongate faster or slower than the wild type version, the authors were able to show that slower polymerases result in earlier termination, consistent with the notion that Rat1 needs to physically catch up with the polymerase in order to elicit termination [54].



**Figure 3.1:** Overview of the main mechanistic step that lead to CPF-CF termination. The complex is recruited thanks to CTD phosphorylation and binding sites on the RNA. The transcript is then cleaved and the elongation complex terminated in accordance with the torpedo or allosteric model.

At the same time, several reports argue against the torpedo model as sole effector of transcription termination. *In vitro* studies were unable to reproduce the termination effect observed *in vivo* using only Rat1 [37]. More recent ventures re-attempted the *in vitro* approach with limited success [136], but managed to demonstrate that Rat1 is able to terminate polymerases that are destabilized by nucleotide misincorporation.

Several additional mechanistic studies showed that the exonucleolytic activity of Rat1 is unable to mediate the release of the polymerase from the template [107, 138]. Moreover, termination defects caused by Rat1 mutants were not associated with stabilization of the RNAPII-associated transcript, arguing against the model.

### 3.1.4 A Unified View of CPF-CF Transcription Termination

As evidence for and against the two models piles up, a unified view that combines elements of both torpedo and allosteric model is taking shape. While the effect of Rat1 on transcription termination (of at least some transcripts) is established, its role as main effector of CPF-CF termination has been repeatedly called into question. Several studies have now described interdependencies between Rat1 and other subunits of the CPF-CF complex—notably Pcf11—and the perceived nature of Rat1 is shifting towards that of a molecular effector that is integrated into a larger system. The proof of principle that termination is possible without cleavage has been recently provided—albeit *in vitro* [204]—and presence of Rat1 has been convincingly shown to facilitate termination [54], arguing for a model that integrates these two mechanisms.

## 3.2 The NNS Pathway

NNS dependent transcription termination is the second of the main termination mechanisms in *S.cerevisiae*. It is involved in the termination of Small Nuclear RNAs (snRNAs), Small Nucleolar RNAs (snoRNAs) and a number of other non-functional non-coding RNAs. It sets itself apart from CPF-CF termination in a number of ways. First and foremost, it relies on a completely different—and much smaller—set of proteins: the two RNA binding proteins Nrd1 and Nab3 [30], together with the helicase Sen1. Because of the different molecular effectors, the termination mechanism—although still not fully elucidated—is appreciably different.

The NNS complex also distinguishes itself because of the different fate imposed on the RNA released: instead of being exported to the cytoplasm after polyadenylation, the transcripts released are subjected to the activity of degradation enzymes [185]. To this end the NNS complex recruits both the nuclear exosome and a specific set of 3' end processing factors known as TRAMP (Trf4/Air2/Mtr4p Polyadenylation), which drives polyadenylation and stimulates degradation [82, 185].

NNS termination operates mainly on non-coding RNAs and is generally restricted to the early stages of transcription elongation. Despite not being directly involved in the termination of protein-coding genes, it can play a role in the regulation of gene expression by acting as an attenuator (i.e. terminating some transcription events, preventing them from producing functional RNAs) [5]. Examples of this phenomenon include the IMD2 or URA2 genes [79]. Alternatively, NNS was shown to terminate transcription of non-coding RNAs whose transcription is involved in regulation [180].

### 3.2.1 The NNS Complex

The main molecular effectors of the NNS complex are the three protein Nab3, Nrd1, and Sen1.

**Nab3** This factor was originally identified as a polyadenylated RNA binding protein. Nab3 contains several structural domains: a conserved RNA Recognition Motif (RRM) that can contact specific sequence elements on the nascent RNA, a region necessary for the interaction with Nrd1, and an essential Glutamine/Proline region at the C-terminus.

Biochemical experiments have shown that Nab3 forms a stable heterodimer with Nrd1 and contacts the RNA as such [30]. In addition, the structure of the RRM has been solved, revealing the structural basis for the preference of the sequence UCUUG [106]. Finally, its Glutamine/Proline region—despite being generally unstructured—can assemble into amyloid structures [187].

**Nrd1** Identified as part of the “nuclear pre-mRNA downregulation” family of proteins, Nrd1 is the most abundant of the three members of the complex. Its main features consist of an RRM structure that allows it to contact the nascent RNA, a CTD interaction domain (CID) that mediates the interaction with RNAPII (see below) and a Nab3 interaction motif that allows it to form a stable heterodimer.

Nrd1’s RRM was shown *in vivo* to contact the consensus sequence GTA[A/G] [174]. Recent *in vitro* studies, however, have shown that several other G-rich and A-rich sequences could be bound equally well [7], although the *in vivo* relevance of these studies remains to be demonstrated.

In addition to the RNA, Nrd1 can contact RNAPII through its CID [92, 186]. Although dispensable for cell viability, the CTD-CID interaction is required for efficient termination.

Curiously, Nrd1 also contains a Glutamine/Proline region at the C-terminus, similarly to Nab3. Deletion of this region shows no growth or termination defects, but is synthetic lethal if combined with other aphenotypic mutations on Nab3 [our unpublished data]. The functional implications of these genetic interactions are still unknown.



**Sen1** This extremely large (253kDa) and very low abundance (125 molecules per cell) protein is the only member of the NNS complex to have enzymatic activity [173]. Sen1 was characterized as a helicase of the SFI superfamily and is very closely related to Upf1, a member of the Non-sense Mediated mRNA Decay (NMD) pathway in the cytoplasm. Unlike its close relative, Sen1 possesses a nuclear localization signal and acts in the nucleus, where it can physically interact with the other members of the NNS complex Nrd1 and Nab3.

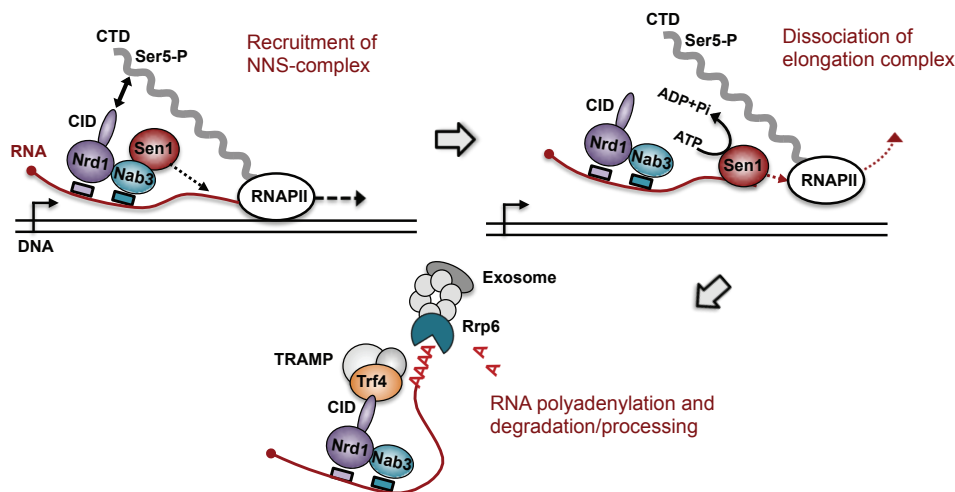
Structurally, Sen1 contains a helicase domain able to hydrolyze ATP and a large N terminal domain. The helicase domain was recently purified in E.coli and biochemically analyzed, revealing binding affinity for both DNA and RNA, but a slower translocation rate on RNA [175]. Moreover, its ATPase activity was shown to be necessary for termination *in vitro* [143]. The N-terminal region of Sen1 was implicated in the interaction with the RNAPII, as well as other factors such as Rnt1 and Rad2, but the implications of the latter interactions remain obscure.

### 3.2.2 The Mechanism of Transcription Termination

As in the case of CPF-CF, the NNS complex is recruited to the region of termination through two distinct mechanisms that cooperate to maximize efficiency: the CTD of the polymerase [186] and specific sequence elements on the nascent RNA [30]. Within the NNS complex, Nrd1 and Nab3 are the major interactors of these elements, providing specificity and ensuring that Sen1—believed to be the molecular effector of NNS termination—is recruited only in the appropriate circumstances [143].

The CTD of RNAPII is contacted by the CID domain of Nrd1. This domain preferentially recognizes the SER<sub>5</sub>-phosphorylated variant, which is the prevalent CTD phosphorylation state in the first 500-600 nucleotides of transcription. This preference confers to the NNS complex a high degree of specificity for terminating transcription in the early stages of elongation. According to the current model for NNS termination, the interaction with the CTD occurs prior to RNA binding, and

facilitates recognition of sequence elements on the nascent transcript. Presence of SER<sub>5</sub>-P CTD was shown to be a pre-requisite for efficient termination, as placing high efficiency NNS binding sites at the end of long transcription units—where the levels of SER<sub>5</sub>-P would be completely supplanted by SER<sub>2</sub>-P—does not result in termination [63].



**Figure 3.2:** Main stages of NNS-dependent termination. the NNS complex is recruited thanks to SER<sub>5</sub>-phosphorylated CTD and sequence elements on the transcript. Termination is elicited by Sen1, presumably by translocating along the transcript. Finally, the exosome is recruited to the transcript and the transcript is either trimmed or completely degraded.

Recruitment of Nrd1 to the CTD, however necessary, is not sufficient to trigger termination. The Nrd1-Nab3 heterodimer must also contact the nascent RNA through the RRM domains of the two subunits. Original studies have investigated the sequence elements that drive NNS termination, pinpointing two core consensus: UCUU as the main binding site for Nab3, and GUA[A/G] as the main site for Nrd1 [23]. More recent investigations redefined these consensus and identified new sequence elements that can increase termination efficiency when in proximity of canonical binding sites. Use of an *in vivo* SELEX (Systematic Evolution of Ligands by Exponential enrichment) strategy allowed to extend the core consensus sequences for both Nrd1 and Nab3 with nucleotides that proved critical for binding [142]. In addition, AU-rich sequences found downstream of Nrd1 sites were shown to play a

role in increasing both termination efficiency and recruitment of Nrd1 [142]. Similar conclusions have been reached by *in vivo* crosslinking studies [195].

Despite the efforts expended in identifying sequence elements that could univocally lead to NNS termination, a lot of ambiguity remains on what constitutes an NNS terminator *in vivo*. While presence of Nrd1-Nab3 binding sites is required, no consistent pattern emerges in number, spacing, or quality of Nrd1/Nab3 sites at known NNS termination sites. *In vitro* studies on model cases have identified some features of heterodimer binding. For example, mutation of Nab3 binding sites proved to be more deleterious to heterodimer recruitment than mutation of Nrd1 sites [22]. Moreover, multiple heterodimers were found to bind the same RNA sequence, possibly cooperatively [22]. It remains impossible, however, to generalize these results beyond the few sequences tested. While the NNS complex could simply rely on a high number of low affinity sites to reach an occupancy threshold, it remains possible that several unseen elements play a role in qualifying NNS terminators, influencing the quantity and quality of Nrd1 and Nab3 binding sites necessary for an efficient termination.

When the Nrd1-Nab3 heterodimer is bound to the nascent RNA, the molecular effector of NNS termination, the helicase Sen1, is recruited to the complex. Studies have shown that Sen1 is strictly required to terminate transcription, but the mechanism through which this happens is not clear. Significant advances in the understanding of this phenomenon came from use of an *in vitro* transcription termination system [143]. In this context, Sen1 alone was found to be sufficient to disassemble the elongation complex. Termination was shown to occur preferentially at sites of pausing and to require both the interaction of Sen1 with the nascent transcript and ATPase activity. It is unclear whether ATP-dependent translocation of Sen1 on the nascent RNA is required for termination. However, results from an *in vivo* study suggest the existence of a kinetic competition between transcription elongation and Sen1 translocation on the RNA. The authors investigated the effect of the speed of transcription on NNS termination, showing that faster transcription results in longer

NNS-terminated transcripts, while slower transcription produces shorter transcripts and is able to suppress mutations on Sen1 [69]. Taken together, these results support a model where, akin to the bacterial termination factor Rho, Sen1 would contact the nascent transcript and translocate in a 5' to 3' direction, eliciting termination upon catching up with the polymerase.

### 3.2.3 Processing Products of the NNS Pathway

The process of NNS termination is strictly connected with 3' end processing or degradation mediated by the nuclear exosome, a multiprotein complex endowed with exonuclease activity [185]. The exosome plays a major role in nuclear RNA quality control, degrading aberrant transcripts, a number of non-functional non-coding RNAs, and trimming the precursors of functional small non-coding RNAs such as sn/snoRNA [for review see 85]. The exosome is composed of six non-catalytic subunits arranged in a ring-like structure, together with three cap subunits that can bind RNA. The catalytic activity of this complex is dependent on two active 3'→5' exonuclease, Dis3 and Rrp6. Dis3 associates with the ring on the opposite side of the three cap subunits, and degrades RNAs that are threaded through the cap proteins and into the ring [111]. The exosome is present throughout the nucleus and in the cytoplasm. However, only the nuclear version can associate with the other exonuclease, Rrp6, whose activity is known to regulate the levels of many NNS targets.

Recruitment of the exosome to NNS targets takes place via one of the exosome's co-factors: the TRAMP complex. TRAMP (for Trf4/Air2/Mtr4p Polyadenylation) is a nuclear complex composed of the poly(A) polymerase Trf4, the RNA-binding protein Air2 and the helicase Mtr4. Trf4 is the core subunit of the complex, to which both Air2 and Mtr4 bind independently. It possesses poly(A) polymerase activity, but unlike Pap1—the canonical poly(A) polymerase associated with the CPF-CF complex—it can only add tails in a distributive manner. Trf4 is also the factor responsible for the coordination between the NNS complex and the nuclear

exosome. A recent study showed that Trf4 contacts Nrd1 through a small motif called Nrd1 Interaction Motif (NIM). The NIM on Trf4 mimics SER<sub>5</sub>-P CTD and can therefore compete with the CTD of RNAPII for the interaction with the CID (CTD interaction domain) on Nrd1. The interaction of Nrd1's CID with the CTD and Trf4 are mutually exclusive. These findings have suggested a model whereby TRAMP is recruited to the RNA when the CID of Nrd1 is freed from the CTD of the polymerase [182], allowing the coordination of events going from termination to the handover of the transcript to TRAMP and the exosome.

As a co-factor of the exosome, TRAMP is able to both recruit and stimulate its activity. Addition of a poly(A) tail to the terminated transcript is thought to provide an unstructured platform that can be easily be threaded through the non-catalytic subunits of the exosome. However, TRAMP has been known to stimulate exosome activity even independently of poly(A) polymerase activity [182].

By virtue of the tight connection between NNS and TRAMP, NNS-terminated transcripts are usually subject to rapid degradation. SnoRNAs and snRNAs constitute notable exceptions, in that they are heavily structured functional non-coding transcripts that are recruited to the exosome, but undergo only trimming of their 3' ends instead of complete degradation. This is thought to occur thanks to the presence of secondary structure and additional proteins binding the RNA, preventing the transcript from being entirely threaded through the exosome [123].

### 3.3 Non-Canonical Termination Pathways

CPF-CF- and NNS-dependent termination seemingly account for the vast majority of RNAPII transcription termination events in the cell. Several additional mechanisms, however, can terminate transcription in *S.cerevisiae*. These non-canonical termination pathways are generally thought to elicit termination of particular RNA species, but can also act as fail-safe pathways in restricting readthrough transcription [60].

### 3.3.1 Rnt1-Dependent Termination

The yeast Rnase III homologue Rnt1 is an enzyme that binds and cleaves double-stranded RNA stem-loops at a defined recognition site. Rnt1's known function in the cell is that of cleaving polycistronic rRNAs and snoRNAs transcripts, promoting their subsequent trimming and processing by the exosome [60]. Recently, Rnt1 binding sites have been identified downstream of a number of genes and its cleavage activity has been implicated in transcription termination.

Studies on the model gene NPL3 have shown that deletion of Rnt1 leads to transcriptional readthrough and can even mediate the production of dicistronic transcripts [59]. Rat1, the mediator of the CPF-CF termination according to the torpedo model, was found to be also required for proper termination by Rnt1. This led to a model where Rnt1 cleaves a stem-loop that forms downstream of the CPF-CF cleavage site, generating a non-polyadenylated transcript, and leaving an uncapped 5' on the nascent transcript. This free 5'-OH is a substrate for exonuclease Rat1, and transcription termination is thought to occur with a mechanism akin to the CPF-CF torpedo model, with Rnt1 as the cleaving agent instead of the CPF complex [59, 157].

The termination mechanism is usually very intimately connected with 3' end processing and with the fate of the transcripts it produces. The case of Rnt1-dependent termination, however, is peculiar in this respect. Use of *in vivo* reporter systems showed that, in the absence of a polyadenylation site, Rnt1-dependent transcripts are unstable and supposedly targeted by TRAMP and the exosome [59]. However, addition of a cryptic polyadenylation site close to the Rnt1 binding site in the same system results in increased transcript stability that is Pap1-dependent. This suggests that depending on its environment, Rnt1 can either stimulate the usage of a nearby Polyadenylation site or produce transcripts that are targeted for degradation [157].

### 3.3.2 Road-Block Termination

Road-block termination represents another non-canonical mechanism that can mediate transcription termination. Road-block was first observed as a termination mechanism for RNAPI, where a DNA binding factor acts as a physical obstacle for the polymerase. The polymerase is thought to stall at the DNA binding site and eventually dissociate from the template through unclear mechanisms [98, 99].

When the mechanism was first described, *in vitro* work had shown that transcription factor Reb1 was able to pause all three yeast RNA polymerases [98]. Later studies from the same authors confirmed that the DNA binding site for Reb1 was coincident with sites of RNAPI transcription termination *in vivo* [149]. Combination of these experiences led to a model where Reb1 is binding DNA and terminating RNA polymerase I at specific rDNA loci. It was only in 2012 that a Reb1 paralogue—Nsi1, who binds the same consensus sequence as Reb1—was implicated as the true *in vivo* effector of RNAPI termination, while Reb1 was proven to not have a role [151].

I have participated to a study of the laboratory showing that Reb1 is the effector of roadblock transcription termination for RNA polymerase II *in vivo*. This study will be described in the results section.





# 4

## The Transcriptional Landscape of *S.cerevisiae*

The rise of microarrays and next generation sequencing techniques has made the exploration of the transcriptome possible. Early application of tiling arrays to the transcriptome of *S.cerevisiae* showed that, in addition to protein coding genes and a multitude of functional non-coding RNAs, the genome is pervasively transcribed and RNA molecules can arise from many unannotated regions [34, 128, 197]. There are multiple possible reasons for this phenomenon. Studies have shown that yeast promoters, despite showing directionality, can fire bidirectionally and give rise to non-functional RNAs [128, 197]. Additionally, transcription usually arises in poorly chromatinized areas of the genome, pointing to the possibility that the genome might provide a low barrier to transcription initiation outside regions of high nucleosome occupancy. These factors contribute to the widespread occurrence of transcription outside of annotated regions, which is usually referred to as pervasive transcription and contributes to the generation of large quantities of non-coding (mostly non-functional) RNAs.

## 4.1 Control of pervasive Transcription

Pervasive transcription represents a non-negligible fraction of all RNAPII transcription. Therefore, it has the potential to interfere with other physiological events and needs to be carefully regulated. Control of pervasive transcription occurs on two levels: First, RNAPII that initiates spuriously need to be rapidly terminated, in order to avoid interference with other processes on DNA; second, the resulting transcripts need to be efficiently degraded, to prevent accumulation of toxic species.

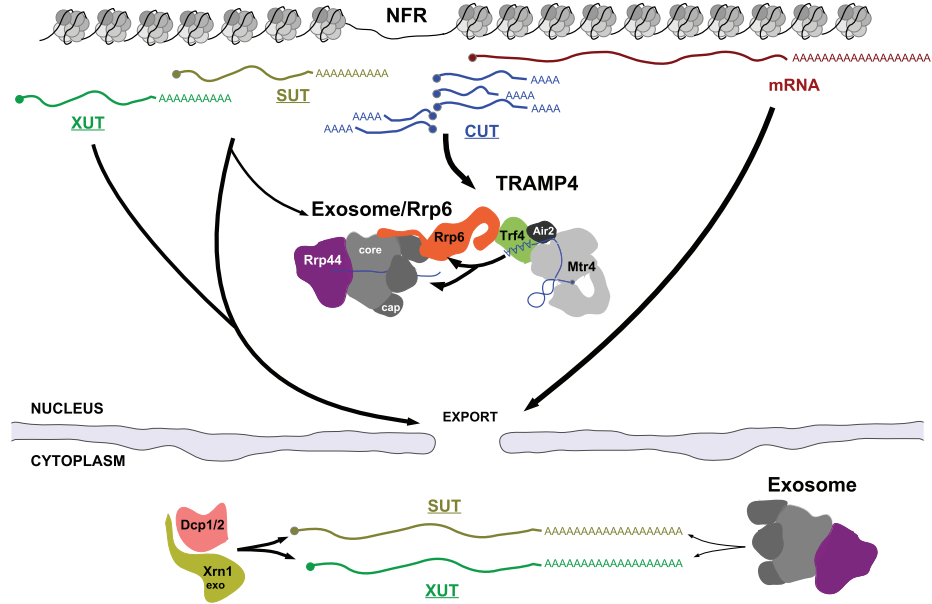
The NNS complex is the main termination pathway involved in control of pervasive transcription [5, 179]. Binding sites for Nrd1 and Nab3 are frequently enriched in areas where pervasive transcription occurs, such as antisense to coding RNAs and in intergenic regions [179]. Acting early in the transcription cycle, NNS is an effective tool to block such transcription events before they can do damage. Despite the major role of NNS, CPF-CF, as well as some non-canonical termination pathways, have been implicated in termination of pervasive transcription [29, 114, 183].

Once termination has occurred, transcripts are released into the nucleus. These RNA species do not possess coding potential and might be deleterious to the cell if accumulated in sufficient quantities. In order to prevent such accumulation, the cell evolved RNA quality control systems that can degrade spurious and aberrant transcripts. These decay pathways can be directly connected to termination and 3' processing, as in the case of NNS and the TRAMP-Exosome [179], or recognize specific features that mark non-functional transcripts, such as poor coding potential.

## 4.2 Classes of Pervasive Transcript

Because of their rapid turnover, the majority of pervasive transcripts are difficult to detect in wild type cells. Several studies found that deletion of certain elements of RNA quality control would affect the stability of only a subset of pervasive transcripts, making them appear in transcriptome analyses [183, 196]. Over time, it

became obvious that several classes exist, each responding differently to inactivation of specific quality control pathways. The following classes, therefore, represent sets of transcripts sharing one or more features that make them more susceptible to specific branches of quality control.



**Figure 4.1:** The cellular fate of pervasive transcripts. Different classes of non-coding RNAs are represented at the top. Black arrows indicate the fate of the transcript after transcription, either immediate degradation via the exosome/TRAMP quality control pathway, or export to the cytoplasm. Here, cytoplasmic quality control is shown.

**CUTs** The first—and most abundant—class of pervasive transcripts to be described, Cryptic Unstable Transcripts (CUTs) were identified in a strain missing the exosome co-factor Rrp6 [196]. CUTs are short transcripts (400-800 bp) originating from intergenic regions and bidirectional promoters. They can often be detected in the antisense direction to protein coding genes and their transcription can sometimes contribute to gene regulation [6].

CUTs are terminated by the NNS pathway [5]. This greatly facilitates their turnover, which occurs exclusively in the nucleus. After transcription termination has occurred, CUTs are contacted by TRAMP and handed over to the nuclear exosome, resulting

in their rapid degradation [179].

**SUTs** Unlike CUTs, Stable Untranslated Transcripts (SUTs) are detectable in wild type cells [34]. This difference is due to the termination mechanism that characterizes these transcripts. While CUTs are terminated early by the NNS pathway, SUTs are longer and terminate through the CPF-CF pathway [114]. This difference in termination implies that SUTs can more easily escape the nucleus and be exported into the cytoplasm. It should be noted that a large portion of SUTs is partially affected by exosome mutations, suggesting that multiple termination mechanisms might contribute to the generation of these transcripts.

Despite being exported to the cytoplasm, SUTs have poor coding potential and are targeted by specific quality control pathways in this compartment (see below) [112].

**XUTs** Very close to SUTs, Xrn1-dependent Unstable Transcripts (XUTs) have essentially the same characteristics. They are terminated by the CPF-CF pathway and rapidly exported to the cytoplasm [183]. However, while the turnover rate of SUTs is sufficiently slow to allow their detection in wild type cells, XUTs are more susceptible to cytoplasmic decay pathways, and therefore require deletion of Xrn1—the main molecular effector of cytoplasmic RNA degradation—to become visible in transcriptome analyses [183].

**NUTs** Largely overlapping with CUTs, Nrd1-dependent Unterminated Transcripts (NUTs) are defined as transcripts that gain stability when NNS termination is impaired [166]. Normally, these transcripts are rapidly degraded by the nuclear exosome. However, when NNS termination is impaired, they gain in length and stability, becoming detectable.

**RUTs** Only recently identified as a new class of pervasive transcripts, Reb1-dependent Unstable Transcripts (RUTs) are transcripts subjected to road-block

termination by the transcription factor Reb1 and subsequently degraded by the nuclear exosome [29].

### 4.3 Quality Control Pathways

RNA quality control eliminates aberrant and pervasive transcripts through degradation. Several multisubunit complexes located throughout the cell carry out this function through use of endo- and exo-nuclease activities. Targeting of transcripts to these complexes (i.e. marking for degradation) can occur through several means: it can be directly connected to the termination mechanism used to release the transcript, as in the case of NNS termination, or it can depend on certain features of the RNA, such as presence of a premature stop codon.

The exosome is known to act in both the nucleus and the cytoplasm. Its catalytic activity depends on the subunit Dis3, which possesses 3' to 5' exonuclease and endonuclease activity. In the nucleus, the exosome is associated with two specific co-factors: a second 3' to 5' exonuclease called Rrp6, and a polyadenylation complex called TRAMP [82]. While Rrp6 significantly contributes to RNA degradation through its exonuclease activity, TRAMP stimulates the activity of the exosome through addition of short poly(A) tails and other, less clear means [64, 80]. This ensemble of factors makes the exosome the foremost quality control agent in the nucleus. In the cytoplasm, the exosome is not found in complex with Rrp6 or TRAMP and has only a minor role in RNA degradation [for review see 181].

In the cytoplasm, RNA degradation is mainly enforced by the 5' to 3' exonuclease Xrn1. Several decay pathways can lead to degradation by Xrn1 (and to some extent the cytoplasmic exosome): Non-sense Mediated mRNA Decay (NMD), triggered by the presence of a premature stop codon; No-Go Decay (NGD), triggered by lack of a translation start codon; and No-Stop Decay (NSD), caused by lack of a stop codon. These pathways target transcripts that do not possess the typical features of mRNAs, stopping potentially toxic elements from being translated [for review see

73]. Xrn1 is known to target pervasive transcripts with poor coding potential, such as SUTs and XUTs, providing a backup system that can deal with those RNAs that manage to escape the nuclear quality control [112].

## 4.4 Functional Role of Pervasive Transcription

The question of whether pervasive transcripts in yeast possess any functional activity remains unclear. While the act of pervasive transcription has been associated with regulatory events on multiple occasions, very little is known about the function of the transcripts themselves.

For instance, SER3 expression is known to be regulated by the upstream transcription unit SRG1—producing a non-coding RNA—through a mechanism of transcriptional interference [116]. This phenomenon occurs when an elongating polymerase invades a promoter, thereby reducing the efficiency of transcription initiation. Similarly, the PHO84 gene seems to be regulated by an antisense transcript that runs along the whole gene, reaching the promoter and downregulating expression [24]. In both these cases, repression is mediated by a modification of the chromatin state of the promoter, which prevents assembly of the Pre-Initiation Complex. Stabilization of the transcript, however, did not in any way affect the repression.

Other regulation mechanisms involve NNS termination and a conditional generation of CUTs. Several nucleotide biosynthesis genes (URA2, URA8, IMD2 among others) can initiate transcription from two regions separated by an NNS terminator sequence [79, 178]. Only transcription from the downstream TSS results in productive elongation, while transcription starting from the upstream TSS results in early termination and degradation of the transcript. It has been shown that nucleotide availability modulates TSS selection, and said genes are properly expressed only when specific nucleotide concentrations are low.

# 5

## General Regulatory Factors

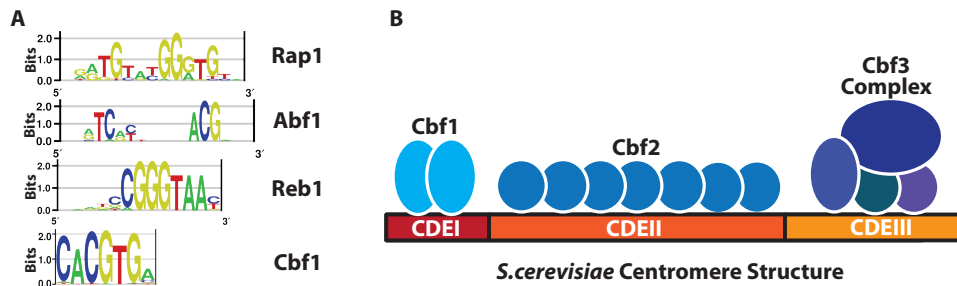
General Regulatory Factors (GRF) are a subset of abundant, widespread, and multi-functional DNA-binding proteins involved in several aspects of chromosomal function. In addition to their role as transcriptional activators, GRF are involved in transcriptional silencing, telomere maintenance, and centromere function.

The proteins defined as GRF are, among others, Rap1, Reb1, Abf1 and Cbf1 [39]. GRFs are a functionally and structurally heterogeneous group of proteins. However, they have the capability of activating transcription through specific binding in promoter regions and modification of the chromatin structure. Through this mechanism, GRF are known to regulate a substantial number of genes.

In this section, I will describe in brief the specific roles of each GRF and subsequently focus on their transcriptional activity.

### 5.1 Rap1

The essential transcription factor Rap1 is probably the best characterized GRF and it has a multitude of functions. Rap1 has a strong preference for the specific DNA



**Figure 5.1:** **A:** Sequence logos representing the main binding sites for the four GRFs Rap1, Abf1, Reb1, and Cbf1. **B:** structure of the centromere in *S.cerevisiae* and its main interactors, A Cbf1 dimer is stably bound to CDEI.

element shown in figure 5.1A. This binding is mediated by two large DNA binding domains very similar to those of the human oncogene Myb [154].

Rap1 is the main transcriptional activator of ribosomal protein (RP) genes, controlling the expression of about 90% of these species [124]. This regulation is enacted through multiple pathways. First, Rap1 recruits a number of ancillary transcription factors: Fhl1, Ifh1, Sfp1, and Hmo1. Together they modify the structure of chromatin and stimulate transcription [152]. Second, Rap1 is able to independently recruit TFIIA and TFIID to the promoter of RP genes, accelerating the rate of PIC formation at these loci [135].

In addition to its activator capabilities, Rap1 works as an active silencer of transcription. During vegetative growth, the mating type loci of *S.cerevisiae* are transcriptionally inactive. Their silencing is mediated by binding of Rap1, Orc1, and another GRF, Abf1. These proteins are able to recruit Sir1, Sir2, Sir3, and Sir4, which mediate the spread of heterochromatin over the HML and HMR loci, preventing transcription initiation [95]. The transcriptional repressor activity of Rap1 has also been reported for RP genes under conditions of nutrient starvation, but in these conditions the silencing mechanism remains unclear [152].

Lastly Rap1 has been implicated in the maintenance of telomeres [108]. In this context, Rap1 is part of a complex named Telosome together with Rif1 and Rif2. The telosome forms a protective cap around telomere sequences and is required for



different aspects of telomere homeostasis such as telomere length regulation, inhibition of end resection, protection from fusion and inhibition of untimely activation of the DNA damage checkpoint [for review see 189]. Recent genome-wide studies identified Rap1 binding sites both at telomeres and RP genes, showing that these two classes of binding sites are distinct [154]. Somewhat consistent with this notion, another study showed how Rap1 possesses two binding modes. According to the authors, Rap1 can either bind a single site with high efficiency making use of both its Myb-like DNA binding domains, or it can bind more degenerate sequences with lower affinity using only one domain, but forming higher stoichiometry complexes [48]. However, whether these two binding modes have functional consequences is unknown.

Rap1, together with other GRF such as Reb1 and Abf1, has been shown to have a role as an insulator (i.e. preventing the spread of heterochromatic silencing), and is thought to act in this capacity at the mating type loci [55].

## 5.2 Abf1

Both Structurally and functionally close to Rap1, Abf1 is another essential factor implicated in numerous processes. Abf1 binds the split DNA site shown in figure 5.1A, which is known to regulate hundreds of promoters.

While the vast majority of RP genes are regulated by Rap1, a cohort representing 10% of the total is under the control of Abf1 [36]. A recent study investigated the mechanism of Abf1-dependent RP gene regulation, showing that Abf1 is found in association with Fhl1 and Ifh1, but has a lower occupancy on the promoter relative to Rap1 [50]. Abf1-dependent regulation of RP genes seems to possess distinct features from the canonical Rap1 regulation. Under nutrient starvation, Abf1 was observed to be more stably associated with the promoter and this resulted in a severe downregulation of gene expression. The authors speculated that stable association of Abf1 with DNA could mediate transcriptional silencing, while a more dynamic

interaction could mediate activation [51].

Akin to Rap1, Abf1 is known to act in silencing at the mating type loci, as well as an insulator in sub-telomeric regions [110].

Abf1 is present in a number of autonomous replicating sequences (hence the name, ARS Binding Factor 1). These regions of the genome are essential to the process of DNA replication and act as its starting points. The C-terminal region of Abf1 was found to enhance replicative activity independently of the transcription activation domain [194]. In addition, replication factors have been shown to increase Abf1 DNA-binding activity [49]. Despite these data, however, Abf1's mechanism of action at replication origins has never been fully elucidated.

Lastly, Abf1 is implicated in the activity of the global genome nucleotide excision repair mechanism (GG-NER). Abf1 was shown to form a stable complex with Rad7 and Rad16, two essential proteins for GG-NER activity [148]. Additionally, impairing Abf1 DNA binding results in UV-sensitive yeast. The Rad7-Rad16-Abf1 complex is known to generate superhelical torsion in DNA [201], and Abf1 is thought to provide specificity to the complex through its DNA binding activity [200].

### 5.3 Reb1

Reb1 was first identified as a ribosomal DNA (rDNA) enhancer binding protein, where it acts in stimulating transcription of ribosomal RNA [141]. Reb1 tightly binds the consensus reported in figure 5.1A with a bipartite myb-like DNA binding domain. Functionally, it acts to promote transcription of about 600 genes and it was implicated as an insulator in sub-telomeric regions. The homologue of Reb1 in *S.pombe* has been extensively studied as a DNA replication termination factor, as it is able to stall replication forks [2]. The implication in this process in *S.cerevisiae*, however, is still unproven.

Reb1 was mistakenly believed to be the effector of RNAPII transcription termination [100]. This notion, however, was dispelled when it was shown that Nsi1, a related protein that binds the same consensus on DNA, was the true molecular effector of RNAPII termination [151]. Interestingly, Reb1 is now implicated in the termination of RNAPII through the same road-block mechanism with which it was thought to terminate RNAPII [29].

## 5.4 Cbfl

Cbfl is the only GRF thus far to not possess a myb-like DNA binding domain. Instead, it is a member of the helix-loop-helix family of DNA binding factors and specifically binds the consensus represented in figure 5.1A. Cbfl is mostly known for its activity as a structural element in centromeres, but can stimulate transcription of a limited number of genes [121].

In *S.cerevisiae*, centromeres are short (120 nucleotides) DNA sequences coated with proteins that mediate assembly of the kinetochore and proper chromosome segregation during mitosis (Fig. 5.1B). Structurally, the centromere sequence is divided into three centromere DNA element (CDE): CDEI, CDEII, and CDEIII. Cbfl is the main binder of CDEI, a region of the centromere known to be important, but not essential for chromosome segregation [131].

Additionally, Cbfl is known to form a complex with transcription factors Met4 and Met28. Through this complex, Cbfl is able to target Met4 to genes involved in Sulphur metabolism [84]. In addition to bringing Met4 to the promoter of MET genes, Cbfl is also known to modify the structure of chromatin at MET and other gene loci through a still unknown mechanism.

GRF	DNA-binding	Chromatin Remodeling	Functions
Rap1	bipartite Myb-like	multiple [152]	RP genes activation, silencing of mating type loci, telomere maintenance
Abf1	Myb-like	RSC	RP genes activation, silencing of mating type loci, insulator, stimulator of DNA replication
Reb1	bipartite Myb-like	RSC	terminator of RNAPII, insulator
Cbf1	Helix-turn-helix	unknown	part of centromeres, transcriptional activator

**Table 5.1:** summary of GRF functions, binding sites and associated chromatin remodeling system.

## 5.5 Chromatin Remodeling

A common feature of GRFs is the capability of altering the local chromatin structure in the vicinity of their binding sites. This mechanism is used to clear nucleosomes from promoters and thus stimulate transcription. As a general rule, GRFs are considered “obligate synergizers”: they can weakly stimulate transcription on their own, but achieve a much greater effect when another weak activator binds the same promoter. The chromatin remodeling activity of GRFs is therefore thought to act as a force multiplier, allowing normally weakly binding transcriptional activators, who would not be able to bind a more chromatinized template, to be stably bound on DNA [18, 26, 140].

Although all GRF described possess some level of chromatin remodeling activity, it is unclear whether this stems from use of a common system or multiple independent pathways. Studies implicated Reb1 and Abf1 in connection with the RSC complex [66]. To prove this point, the authors depleted Abf1 and Reb1, which resulted in a shrinkage of several NFRs. Subsequent depletion of the catalytic subunit of the RSC complex, Sth1, showed that NFRs regulated by Reb1 and Abf1 are also regulated by

RSC. In addition, the authors inserted a Reb1 site within an ORF and observed that an NFR could form depending on the presence of both Reb1 and of Sth1. These findings were confirmed by more recent investigations [93], which found a large overlap between promoters regulated by Reb1 and Abf1 and promoters regulated by RSC. The same study, however, discovered a number of promoters where NFRs are generated in a Reb1- and Abf1-dependent manner, but independently of RSC, arguing for a more complex regulation mechanism.

### 5.5.1 Genome-Wide Effect on Chromatin Structure

The stereotypical view of eukaryotic promoters is characterized by well-positioned +1 and -1 nucleosomes surrounding a 150+ stretch of poorly chromatinized DNA. This notion was challenged by a recent study that showed the existence of nucleosomal particles inside a large number of promoter NFRs [93]. These fragile nucleosomes (FN) are particularly sensitive to the amount of micrococcal nuclease (MNase) used to reveal nucleosome positioning in a genome-wide manner, and therefore went undetected until now. Analysis of the distribution of GRF binding sites inside promoters showed that fragile nucleosomes are significantly associated with GRF binding. Additionally, the GRF-associated chromatin remodeling complex RSC was implicated in the process, and insertion of GRF binding sites in previously unaffected promoters was shown to induce fragile nucleosome formation. For Reb1 and Abf1, the GRF binding site seems to coincide with the position of the fragile nucleosome, suggesting a kinetic competition between histones and GRFs. In the case of Rap1, however, the situation is less clear. The binding site was detected upstream of the fragile nucleosome, and often entailed the presence of two, not one, of these unstable particles. How such large NFR is generated and how fragile nucleosomes are maintained within it is still unknown.

Another study recently investigated the effect of GRFs Rap1 and Abf1 on genome-wide chromatin assembly [57]. While chromatin remodeling activity has been (expectedly) detected at directly regulated promoters, the two GRFs were shown

to affect—albeit to a lesser extent—the chromatin structure of thousands of genes. Using thermosensitive mutants of Rap1 and Abf1 the authors analyzed genome-wide nucleosome occupancy. Analysis of these datasets led to the conclusion that a modest but significant change in nucleosome disposition was occurring at a number of loci that were not described as regulated by either Rap1 or Abf1. Upon further analysis, these promoters were found to be enriched in low affinity or degenerate Rap1 and Abf1 sites. This suggests that even low affinity binding of GRFs can contribute to the regulation of gene expression through a chromatin remodeling activity, and underscores the idea of GRFs as force multipliers—or enabler of transcription—on a much larger scale than previously thought.

# 6

## Transcription and Replication

DNA replication is the biological process that duplicates a cell's genetic information so that, upon division, daughter cells can inherit a full copy of the genome. Replication starts at loci named replication origins. The replisome begins to assemble at these loci during the G1 phase and, upon transition to S-phase, splits into two replicative forks that elongate in opposite directions (Fig. 6.2). This process needs to be tightly regulated, as over- or under-replication can lead to severe genome instability.

While presence of one replication origin is generally sufficient to duplicate prokaryotic genomes, eukaryotic ones are often too large and require multiple origins to be replicated in a timely fashion (i.e. within the confines of S-phase). The genome of *S.cerevisiae* contains 410 confirmed origins (also called Autonomously Replicating Sequences (ARS)) [169], but not all of them are used every time the genome is replicated. Although studies on replication initiation detected discernable patterns in origin specification, studies on single cells have shown that origin selection is not entirely deterministic, but rather a stochastic process [32, 137]. This notion raises the question of which elements (either intrinsic to the replication process, or independent of it) can influence origin specification.

In this chapter I will describe what qualifies a replication origin and explore the mechanisms of origin specification in *S.cerevisiae*, with particular emphasis on the controversial relationship between transcription and DNA replication.

## 6.1 Replication Origin and Their Specification

Replication origins are cis-acting DNA elements upon which the replisome can assemble and start the replicative process. Because of the stochastic nature of their usage, origins are unlike other cis-acting elements. They are collectively required for cell viability, but individually dispensable and redundant [12, 38]. This plasticity lessens the selective pressure for any particular origin, as long as the replication process as a whole remains efficient.

Several elements can influence the likelihood that an origin will be used to start the replication process. Among them, some are intrinsic to the sequence of each origin, like the affinity for replication factors. Some, however, can be heavily influenced by factors external to the replicative process, such as nucleosome deposition and transcription.

### 6.1.1 Origin DNA Elements

Origins in *S.cerevisiae* are usually small (100-150 bp), preferentially intergenic, and AT-rich sequences [146]. Origin-specific motifs are degenerate and generally not conserved. Despite this heterogeneity, several common consensus sequences were identified as promoters of origin activity and classified as A and B elements<sup>1</sup> (Fig. 6.1).

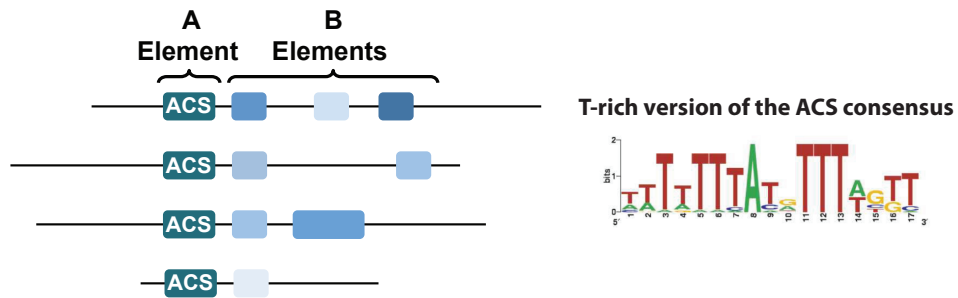
**A element** The only essential sequence element, the A element is also called ARS Consensus Sequence (ACS). The ACS is a non-palindromic 11 bp consensus

---

<sup>1</sup>It should be noted that C elements were also described by Celniker and colleagues [25]. However, evidence for the relevance of these motifs *in vivo* is lacking and they will not be discussed here.



(see Fig. 6.1) [25, 133] that is the binding site of a protein complex called Origin Recognition Complex (ORC). Binding of this complex to the origin represents the first step in the replicative process [40].



**Figure 6.1:** Cartoon showing the most typical arrangement of sequence elements within origins. the ACS is required, while several B elements contribute to origin specification downstream of the T-rich strand of the ACS.

**B elements** A family of motifs with little sequence conservation, B elements are mainly AT-rich and always map downstream of the T-rich strand of the ACS (Fig. 6.1) [146]. While a match to the A element consensus is found in every origin [25], no individual B element is universally required for origin activity [105, 113]. Collectively, however, B elements constitute a requirement for proper origin activity.

B elements were originally thought to facilitate DNA unwinding due to their AT-richness [75]. Subsequent studies, however, revealed that some B elements are playing a more active role, contributing to the recruitment of the replisome [192].

### 6.1.2 Nucleosome Positioning in Origins

Sequence elements are not enough to qualify an active origin. More than 10,000 matches for the ACS exist in the genome of *S.cerevisiae*, however, only 400 replication origins were identified. Moreover, it has been reported that some origins are able to efficiently drive replication of a plasmid, but are rarely used *in vivo* [129, 163]. Investigation of this context-dependent activity showed that nucleosome positioning

plays a crucial role in origin activity and that functional origins *in vivo* are always associated with nucleosome free regions. In support of this notion, experiments forcing nucleosome assembly within the A or B elements resulted in abrogation of replisome assembly [103].

The ACS itself was speculated to be able to drive nucleosome positioning, a notion supported by *in vivo* studies [11, 43]. In these reports, the authors show that ACS sequences contained in active origins are surrounded by NFR, although formation of the latter is per se not sufficient to specify an origin as non-functional ACS are also associated with low nucleosome occupancy, albeit to a lesser extent.

Lastly, binding sites for transcription factors were detected within origins and, although their function remains somewhat unclear, are speculated to contribute to the maintenance of NFR within the origin [40, 156]. For example, the transcription factor Abf1 (ARS Binding Factor) is often found near origins and is known to recruit chromatin remodeling complexes to deplete nucleosomes at promoter regions. Studies on these origins showed that deleting Abf1 binding sites results in loss of activity, but replacing the sites with those of other transcription factors associated with NFR generation, such as Rap1, retains the replication activity [113].

### 6.1.3 Transcription in Origins

Multiple factors can affect the efficiency of an origin. For example, sequence elements can strongly contribute by affecting either ORC binding. However, extrinsic factors such as nucleosome positioning can epistatically affect origin activity by occluding said elements [109]. This raises the question: what other extrinsic processes can impact the initiation of replication? Several studies have investigated the effects of transcription on origin activity, but the results in the literature are controversial. While it is generally agreed upon that transcription has a deleterious effect on origin activity [132, 177], a substantial amount of evidence exists to argue that presence of RNAPII within origins can enhance their activity [58, 198].

Tanaka and colleagues originally investigated the problem by analyzing ARS1, an origin that partially overlaps with the gene TRP1. They observed that changing the endogenous promoter of TRP1 to a stronger one led to significant loss of origin activity [177]. A later study found that high transcriptional output across origins increases their sensitivity to ORC mutants (i.e. transcription increases loss of activity in the context of ORC mutations). The authors proposed a model according to which susceptibility to ORC mutants strictly depends on the intrinsic properties of each origin (e.g. sequence elements) but can be affected by extrinsic elements such as transcription [132].

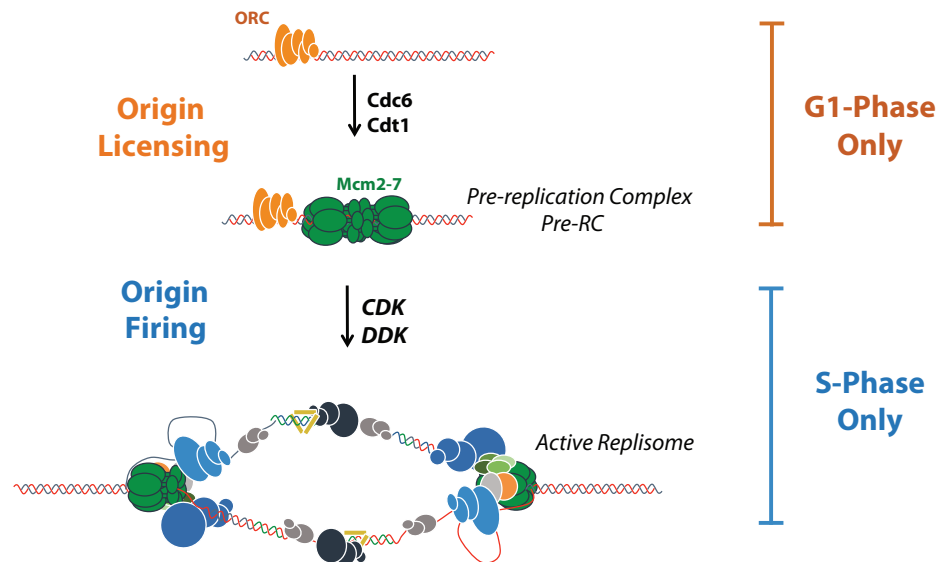
In seeming contradiction, several other studies demonstrated how RNAPII is capable of enhancing origin activity through its presence in the vicinity of the origin. In particular, the CTD of RNAPII was shown to interact with subunits of the replicative helicase MCM2-7 in both xenopus and human [198]. Studies in yeast corroborated this result by showing that not only tethering of RNAPII CTD at replication origins can enhance activity, but also that cells with a shortened CTD (10 heptapeptide motifs in place of 26) show increased plasmid loss rates [58]. Lastly, a recent study investigated origins associated with rDNA loci and concluded that RNAPII molecules participate to ORC binding to origins through their ser2-Phosphorylated CTD [118].

Although there are compelling arguments on both sides, mechanistic details are still lacking and current models cannot yet account for these discrepancies.

## 6.2 Mechanisms of DNA replication

Because of the importance of proper DNA replication for genome stability, its mechanism of action must ensure that the entirety of the genome is duplicated once and only once. In order to achieve this result, replication occurs in two discrete steps [41]. The first step occurs exclusively in the G1-phase, and is called origin licensing. During this step the six subunits Origin Recognition Complex (ORC)

binds the ACS and associates with two ancillary factors named Cdc6 and Cdt1. When interacting, these three licensing factors are able to recruit multiple pairs of inactive replicative helicases around double stranded DNA, forming the pre-replication complex (pre-RC) (Fig 6.2) [42, 56, 153, 158, 167]. Replicative helicases are multisubunit complexes composed of minichromosome maintenance proteins (MCM2-7). In S-phase, MCM2-7 will serve both as platforms for the assembly of other replisome components and as driving force for replicative fork elongation. It is important to note that during the first step of replication, only a subset of all origins are licensed. This subset is influenced by specific origin properties (e.g. strength of the ORC binding site, nucleosome occupancy etc.), but not deterministically chosen.



**Figure 6.2:** DNA replication takes place in two distinct steps. First ORC is required to bind to replication origins during the G1-phase and recruit the Pre-replication complex. Second, upon entry in S-phase, the pre-RC (among others) is phosphorylated and the full replisome can assemble and eventually fire.

The end of the G1-phase and the beginning of the S-phase marks the end of the licensing step and the beginning of the second step of DNA replication: the activation step. During this step, dormant pre-replication complexes at licensed origins activate, assemble into the active replisome and eventually fire. In order to prevent re-licensing of an already activated origin—and therefore avoid the risk of

firing the same origin twice, leading to re-replication—the ORC complex is inhibited through phosphorylation and MCM2-7 are rapidly depleted from the nucleus before this step begins [130].

Entry into S-phase coincides with a cascade of phosphorylation signals that activates the CDK and DDK kinase complexes. These cell cycle specific enzymes phosphorylate the pre-replication complex and are thought to induce structural rearrangements that allow assembly of the complete replisome. First the replicative helicases form the Cdc45-MCM-GINS (CMG) complex [3, 127]. Subsequently, DNA polymerases pol  $\delta$  and  $\epsilon$  join the forming replisome. This allows the complete replisome to form and marks the start of fork elongation.

## Part II

# Results and Discussion

# 7

## Termination of RNA Polymerase II Through a Road-Block Mechanism

In section 3.3.2 I described how road-block termination is an effective tool in transcription termination of RNAPI at rDNA loci. At the beginning of my doctoral studies, the laboratory had found that road-block can also serve as a termination mechanism for RNA polymerase II, and that one of the molecular effectors of this phenomenon was the general regulatory factor Reb1. A major part of my thesis work was therefore dedicated to exploring the notion of road-block applied to RNA polymerase II through the use of genome-wide techniques.

This work led to the identification of other effectors of road-block termination and the better characterization of the genome-wide extent of this pathway. The work is summarized in two manuscripts presented below. The first describes road-block termination elicited by Reb1 and mechanistically characterizes the pathway. The second identifies other effectors of road-block and further explores its genome-wide extent.

## **7.1 Road-Block Termination by Reb1 Restricts Cryptic and Readthrough Transcription**

In this work, I focused on the genome-wide characterization of road-block termination. Previous work from the group had identified specific hallmarks of road-block in synthetic sequences, such as the accumulation of RNAPII about 15-20 nucleotides upstream of Reb1 binding sites. I used genome-wide RNAPII occupancy datasets to probe Reb1 binding sites on the genome and determine whether they are associated with polymerase pausing. In order to achieve this, I devised an algorithm able to identify peaks of polymerase pausing and their position relative to Reb1 binding sites. Additionally, I analyzed a set of synthetic sequences known to elicit Reb1-dependent road-block termination in order to expand the binding consensus for Reb1.



# Roadblock Termination by Reb1p Restricts Cryptic and Readthrough Transcription

Jessie Colin,<sup>1,4</sup> Tito Candelli,<sup>1,4</sup> Odil Porrua,<sup>1,4</sup> Jocelyne Boulay,<sup>1</sup> Chenchen Zhu,<sup>2</sup> François Lacroute,<sup>1</sup> Lars M. Steinmetz,<sup>2,3</sup> and Domenico Libri<sup>1,4,\*</sup>

<sup>1</sup>Centre de Génétique Moléculaire, CNRS UPR3404, 91190 Gif sur Yvette, France

<sup>2</sup>European Molecular Biology Laboratory (EMBL), Genome Biology Unit, 69117 Heidelberg, Germany

<sup>3</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>4</sup>Present address: Institut Jacques Monod, CNRS, UMR 7592, Univ Paris Diderot, Sorbonne Paris Cité, F-75205 Paris, France

\*Correspondence: [libri.domenico@ijm.univ-paris-diderot.fr](mailto:libri.domenico@ijm.univ-paris-diderot.fr)

<http://dx.doi.org/10.1016/j.molcel.2014.10.026>

## SUMMARY

Widely transcribed compact genomes must cope with the major challenge of frequent overlapping or concurrent transcription events. Efficient and timely transcription termination is crucial to control pervasive transcription and prevent transcriptional interference. In yeast, transcription termination of RNA polymerase II (RNAPII) occurs via two possible pathways that both require recognition of termination signals on nascent RNA by specific factors. We describe here an additional mechanism of transcription termination for RNAPII and demonstrate its biological significance. We show that the transcriptional activator Reb1p bound to DNA is a roadblock for RNAPII, which pauses and is ubiquitinated, thus triggering termination. Reb1p-dependent termination generates a class of cryptic transcripts that are degraded in the nucleus by the exosome. We also observed transcriptional interference between neighboring genes in the absence of Reb1p. This work demonstrates the importance of roadblock termination for controlling pervasive transcription and preventing transcription through gene regulatory regions.

## INTRODUCTION

In *S. cerevisiae*, two main transcription termination pathways have been described for RNA polymerase II (RNAPII) (for a recent review, see Kuehner et al., 2011). The first pathway involves the cleavage and polyadenylation (CPF/CF) complex and is mainly devoted to the transcription of mRNA coding genes. The CPF/CF complex is recruited at the 3' end of genes by the interaction with the carboxy-terminal domain (CTD) of the largest RNAPII subunit and by the recognition of specific signals within the nascent RNA. After cotranscriptional cleavage of the nascent transcript, the polymerase is released by a mechanism that is still not fully understood and might involve exonucleolytic digestion of the RNAPII-associated RNA. This mode of termination and

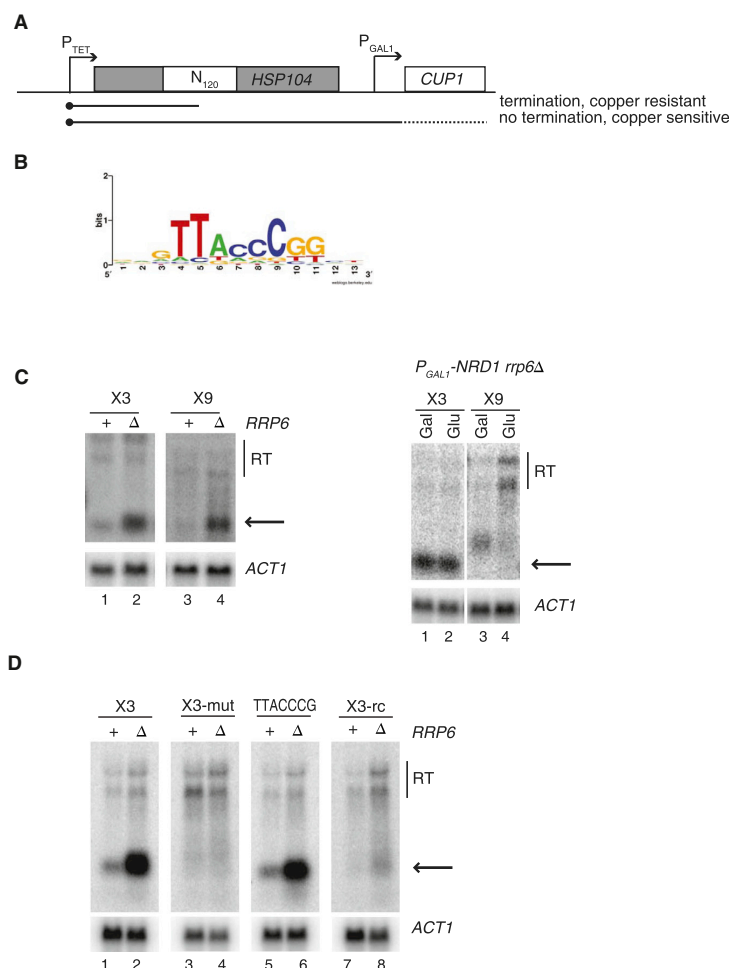
3' end processing produces stable transcripts that are exported to the cytosol for translation.

The second pathway involves the Nrd1p-Nab3p-Sen1p (NNS) complex and was first described for termination of snRNAs/snoRNAs (Steinmetz et al., 2001). This mode of termination is triggered by the recognition of specific motifs on the nascent RNA by the RNA-binding proteins Nrd1p and Nab3p (Creamer et al., 2011; Gudipati et al., 2008; Porrua et al., 2012; Wlotzka et al., 2011), which precedes release of the polymerase by the Sen1p helicase (Kuehner et al., 2011; Porrua and Libri, 2013; Hazelbaker et al., 2013). One distinctive feature of NNS-dependent termination is that the released transcript is polyadenylated by a different poly(A) polymerase, Trf4p, a subunit of the TRAMP4 (Trf4p-Air2p-Mtr4p) complex (Wyers et al., 2005). TRAMP4 stimulates degradation by the nuclear exosome, a multimeric enzyme containing two catalytic subunits, Rrp6p and Dis3p.

One important role of the NNS complex is to control pervasive transcription, i.e., the widespread occurrence of spurious transcription events defining noncanonical transcription units (Schulz et al., 2013; Steinmetz et al., 2006). Pervasive transcription is potentially harmful, as it can interfere with transcription of canonical genes and generate toxic noncoding RNA molecules. The NNS complex terminates transcription of cryptic unstable transcripts (CUTs), the major products of pervasive transcription, and targets these RNAs for degradation by the nuclear exosome (Arigo et al., 2006; Thiebaut et al., 2006; Schulz et al., 2013).

Similar to canonical gene transcription, cryptic transcription generally originates from nucleosome-free regions (NFRs; Neil et al., 2009; Xu et al., 2009), and the Reb1p protein plays an important role in the positioning of NFRs. Reb1p contains a bipartite, Myb-like DNA binding domain and was originally described as a transcriptional activator for RNAPII and RNA polymerase I (RNAPI) transcription (Brandl and Struhl, 1990; Kulkens et al., 1992). It has been shown to preferentially bind intergenic regions, and ectopically inserting a Reb1p binding site within an ORF is sufficient to induce formation of an NFR. Reb1p-dependent nucleosome depletion depends on the RSC chromatin remodeling complex and its catalytic subunit, the Sth1p ATPase, which are likely targeted by Reb1p and other factors to sites of NFR formation (Hartley and Madhani, 2009).

Although Reb1p has been reported to be involved in transcription termination of rDNA transcription by RNAPI, recent studies have challenged this notion, showing that a Reb1p-related protein,



**Figure 1. Selection of Terminators from Naïve Sequences**

(A) Scheme of the reporter used to select terminators. The random sequence of 120 nt (white box) and the *HSP104* coding sequence (gray boxes) are flanked by the tetracycline repressible (*P<sub>TET</sub>*) and the *GAL1* promoter (*P<sub>GAL1</sub>*). Transcripts expected in the event of termination or transcriptional read-through are indicated as solid lines. Two read-through species are produced, one ending at a cryptic terminator at the end of *P<sub>GAL1</sub>* and a second at the end of the *CUP1* gene (indicated by a dotted line).

(B) Consensus logo obtained from 83 selected terminators. See also Figure S1A.

(C) Northern blot analysis of transcripts derived from clone X3 or, as a control, from an artificial NNS-dependent terminator (X9; Porrua et al., 2012). Analysis was performed in an *rrp6Δ* background (left panels) or upon metabolic depletion of Nrd1p by growth in glucose of *P<sub>GAL1</sub>-NRD1 rrp6Δ* cells (right panels). An arrow indicates the position of the RNAs derived from termination at the selected sequence. RT, readthrough transcripts. See also Figure S1B.

(D) The selected motif is necessary and sufficient for termination. Northern blot analysis of wild-type or *rrp6Δ* strains harboring reporters containing either the clone X3, its mutated version (X3-mut), the motif alone (TTACCCGG), or clone X3 containing the reverse complement of the selected motif (X3-rc).

additional tool adopted by the cell to tame pervasive transcription and ensure the stability of the transcriptome.

## RESULTS

### Selection of Terminators from Naïve Sequences

We used an *in vivo* SELEX approach combined with a reporter system in

which transcription termination prevents the silencing of a downstream promoter by transcriptional interference to identify sequence motifs inducing RNAPII transcription termination (Figure 1A; Porrua et al., 2012). Briefly, a pool of random sequences was cloned between the Tet-repressible promoter (Garí et al., 1997) and the *GAL1* promoter, driving expression of the *CUP1* gene, which confers copper resistance to yeast. In the presence of termination signals between the two promoters, transcriptional interference is prevented and the *P<sub>GAL1</sub>-CUP1* unit is active, which allows selecting terminators from the pool of random sequences based on the copper resistance readout. We isolated several hundred sequences, the majority of which induced transcription termination in an NNS complex-dependent manner (Porrua et al., 2012). However, we also selected a large fraction of sequences (roughly 15%) containing a well-conserved, nonpalindromic sequence motif. Statistical analysis of the nucleotide variations in the selected pool

Nsi1p, is the rDNA termination factor (Kawauchi et al., 2008; Reiter et al., 2012). RNAPII termination is thought to occur by a roadblock mechanism whereby Nsi1p binds its cognate site within the rDNA terminator to induce polymerase release. Although roadblock termination likely occurs for RNAPII, it has not been demonstrated for RNAPII, which predominantly utilizes mechanisms linked to the recognition of signals on the nascent RNA.

Here we show in yeast that Reb1p can terminate RNAPII transcription by roadblock-induced pausing followed by ubiquitination of the stalled polymerase. Reb1p-dependent termination occurs at several sites of cryptic transcription in the genome and produces unstable transcripts that, like CUTs, are degraded by the TRAMP-exosome pathway. We also provide evidence that Reb1p-dependent termination functions as a fail-safe mechanism neutralizing transcriptional leakage from adjacent genes, thereby attenuating transcriptional interference. These data reveal an important role of Reb1p as an “NFR guard,” thus defining an

generated the highly conserved core motif shown in Figure 1B and a more relaxed flanking nucleotides context (see Figure S1A available online). Copper resistance was indeed due to the occurrence of transcription termination, because short transcripts ending within the selected regions could be detected by northern blot analysis (Figures 1C and S1B, clones X3, X18, and X31). Similar to CUTs, these RNAs are unstable in a wild-type strain and are strongly stabilized in *rrp6Δ* degradation-defective strains (Figures 1C and S1B, left panels). As expected, termination is not dependent on the NNS complex, because metabolic depletion of Nrd1p in a *P<sub>GAL1</sub>-NRD1* strain did not affect termination at these sequences but impaired termination at a control, NNS-dependent clone (X9, Figures 1C and S1B, right panels).

Mutation of the conserved motif led to the loss of copper-resistant growth (data not shown) and the disappearance of the short transcripts, while longer, readthrough RNAs terminating at downstream sites increased (Figure 1D, lanes 3 and 4). Moreover, insertion of this motif alone in a coding region terminated transcription as efficiently as the full-length terminator (Figure 1D, lanes 5 and 6). Interestingly, the termination motif was found in the selected clones preferentially in one orientation, and replacement of this sequence by its reverse complement almost fully abolished termination (Figure 1D, lanes 7 and 8), suggesting directional specificity for termination, at least in the context of our system (see Discussion).

From these experiments we conclude that the selected motif is necessary and sufficient to induce efficient transcription termination.

### The Transcription Factor Reb1p Is Involved in RNAPII Transcription Termination

The motif identified is the putative binding site of two DNA-binding proteins: Reb1p and Nsi1p/Ydr026c (Harbison et al., 2004). These two proteins belong to the family of Myb-like DNA-binding factors. Reb1p functions in transcription activation and is required for the proper positioning of NFRs (Hartley and Madhani, 2009). Nsi1p is required for transcription termination of rDNA genes and is expressed to lower levels than Reb1p (432 versus 7510 molecules per cell; Ghaemmamghami et al., 2003).

To assess if either one of these two factors is required for inducing transcription termination at the selected motif, we constructed yeast strains containing either a deletion of *NSI1* or *REB1* under control of the *GAL1* glucose-repressible promoter (*REB1* is essential). Since these two factors could have redundant functions in termination, we also constructed strains carrying both modifications.

As shown in Figure 2A, deletion of *NSI1* has no effect on termination (compare lanes 1 and 2 to lanes 5 and 6). However, metabolic depletion of Reb1p induced a clear loss of the short unstable transcript to the profit of a longer readthrough RNA (Figures 2A and S1C), which is diagnostic of a termination defect. Depletion of Reb1p in an *nsi1Δ* genetic background did not further increase readthrough, excluding a possible redundant implication of Nsi1p. From these results we conclude that Reb1p is required for RNAPII transcription termination at these selected terminators.

### Reb1p-Dependent Termination Produces Unstable Transcripts that Are Polyadenylated by Trf4p

Our results indicate that, akin to NNS-dependent termination, Reb1p-dependent termination leads to the production of unstable transcripts. Therefore we investigated whether degradation of these RNAs follows the same pathway as CUTs. Analysis of the poly(A)<sup>+</sup> and poly(A)<sup>−</sup> fractions as well as RNaseH/oligo dT cleavage indicates that a significant fraction of RNAs stabilized in *rrp6Δ* cells are polyadenylated (Figures 2B and S1D). The polyadenylated fraction is strongly reduced in *rrp6Δtrf4Δ* cells, indicating that Trf4p, presumably within the TRAMP complex, is mainly responsible for adding poly(A) tails to these transcripts (Figure 2B, compare lanes 5 and 6 and lanes 8 and 9; Figure S1D, lanes 3 and 4 and lanes 5 and 6). Similarly to CUTs, Trf4p contributes to efficient degradation, because in *trf4Δ rrp6Δ* cells the steady-state level of these transcripts was considerably higher than in single *rrp6Δ* mutant cells (Figure 2B, compare lane 4 to lane 7; Figure S1D, compare lanes 3 and 4 and lanes 5 and 6) and stabilization was also clearly observed in single *trf4Δ* mutant (data not shown). Finally, the core exosome/Dis3p also contributes to degradation, because stabilization of the short RNAs was observed in a catalytic *dis3-exo<sup>−</sup>* mutant, which was further increased in double *rrp6Δ dis3-*defective cells (data not shown).

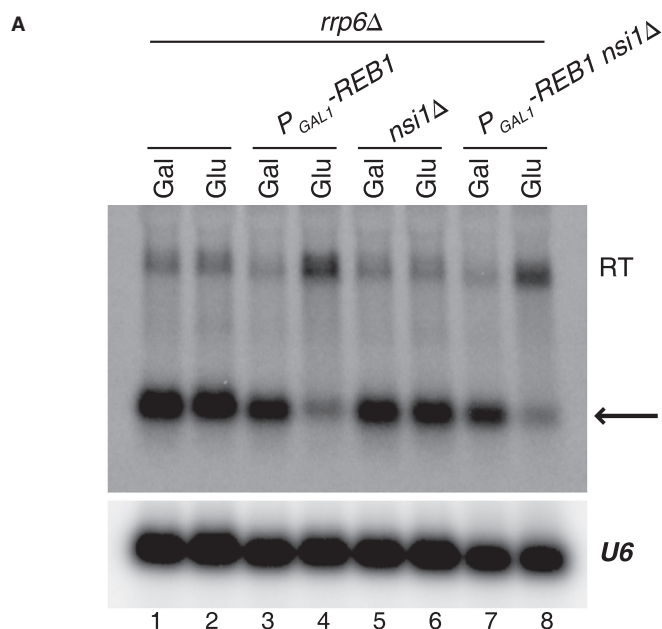
Poly(A)<sup>−</sup> RNAs were also consistently detected (Figure S1D, lanes 1–3), a fraction of which is also present in a wild-type strain, unlike poly(A)<sup>+</sup> species (Figure S1D, lanes 1 and 2). These stable, nonadenylated RNAs most likely represent nascent transcripts that are protected by RNA polymerase from the exosome (see below).

From these experiments, we conclude that transcripts derived from Reb1p-dependent termination events are unstable because they are polyadenylated by Trf4p and degraded by the Rrp6p and Dis3p exosome subunits in wild-type cells. Considering the similar nature of these RNAs and CUTs, we dubbed them Reb1p-dependent unstable transcripts, or RUTs.

### Reb1p Terminates Transcription by DNA Binding Rather Than RNA Binding

The two major pathways of RNAPII transcription termination rely on the essential recognition of the nascent RNA by termination factors (either the CPF/CF or the NNS complex). However, Reb1p is a DNA-binding protein, suggesting that termination might occur via a different mechanism. We first assessed Reb1p binding to the RNA version of its DNA binding site. As shown in Figure 3A, recombinant Reb1p bound efficiently double-stranded DNA with an affinity around 70 nM. However, recognition of the single-stranded DNA or RNA version of the same site was very inefficient, most likely reflecting unspecific binding at the highest concentrations of Reb1p used.

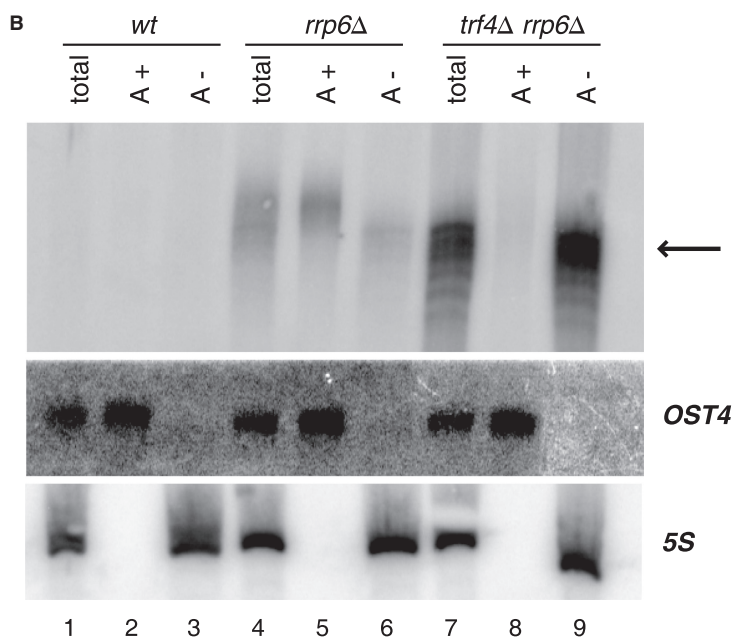
This suggests that binding to the DNA induces transcription termination by a roadblock mechanism. One prediction of such a model is that the Reb1p binding site should not be transcribed. Consistently, mapping by RACE the 3' ends of transcripts produced from different clones selected showed that Reb1p-dependent termination always occur 11–13 nt before the Reb1 site (Figure S2A). This size is compatible with the physical distance between the leading edge of the polymerase (touching

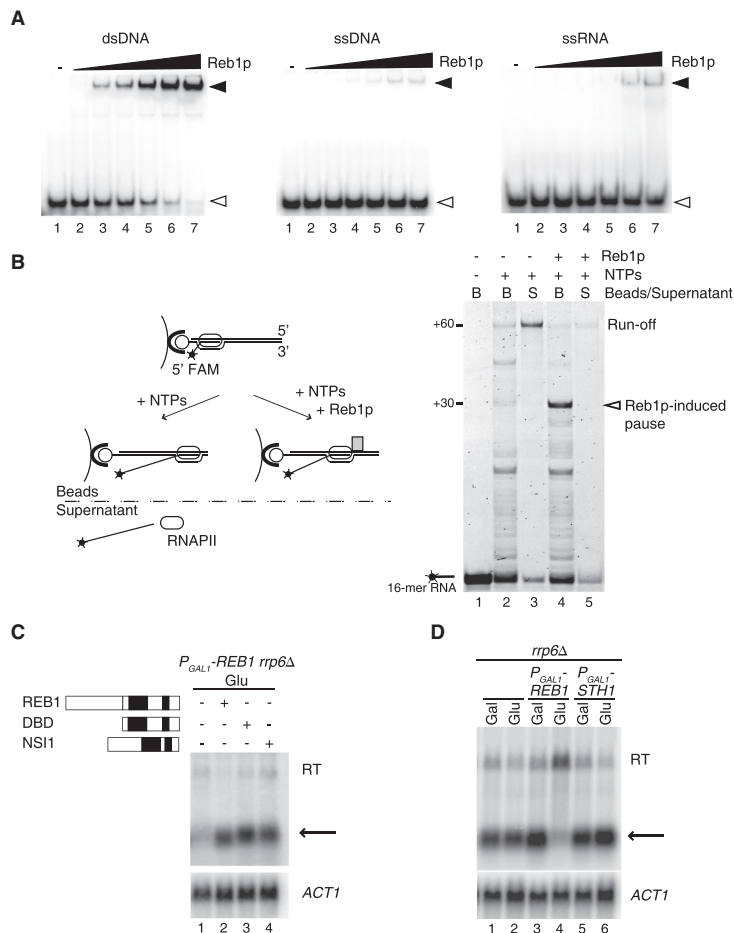


**Figure 2. Reb1p Is Required for Transcription Termination at the Selected Terminators**

(A) Northern blot analysis of transcripts derived from clone X3 upon metabolic depletion of Reb1p, in the presence or absence of Nsi1p as indicated. Labeling is as in Figure 1C.

(B) Reb1p terminated transcripts are polyadenylated by Trf4p. Northern blot analysis of RNAs generated by Reb1p-dependent termination and affinity-selected on oligo-dT magnetic beads. The total poly(A)<sup>+</sup> and poly(A)<sup>-</sup> fractions were analyzed as indicated. Note that poly(A)<sup>-</sup> RNAs are not visible in this experiment in the wild-type strain due to the shorter exposure of the blot. See also Figures S1C and S1D.





**Figure 3. Binding of Reb1p to the DNA Is Sufficient to Elicit Termination**

(A) Recombinant Reb1 (rReb1) binds dsDNA but not ssDNA or RNA. EMSA analysis was carried out using rReb1p and probes of same sequences in a dsDNA, ssDNA, or RNA form containing a Reb1p binding site. The complex formed is indicated by a black arrowhead. Unbound probe is indicated by a white arrowhead. Concentration of rReb1p used was 12.5, 25, 50, 100, 200, and 300 nM (respectively, lanes 2, 3, 4, 5, 6, and 7).

(B) rReb1p induces transcriptional pausing in vitro. In vitro transcription termination assay using immobilized templates and highly purified RNAPII. A scheme of the experiment is shown on the left. Reb1p-dependent pausing is indicated by the accumulation of a transcript that extends up to 12 nt upstream of the Reb1 site but remains associated to the immobilized template (B) and is not released in the supernatant (S). Reb1p is indicated by a gray box.

(C) DNA binding by Reb1p or Nsi1p is sufficient to induce termination in vivo. Northern blot analysis of RNAs extracted from  $P_{GAL1}$ -*REB1* *rrp6Δ* cells containing the X3 reporter. Endogenous Reb1p was depleted by growth in glucose in the presence of wild-type Reb1p, a truncated form of Reb1p containing only the DNA binding domain (DBD) or Nsi1p under control of the *REB1* promoter. Black boxes represent conserved region in Reb1p and Nsi1p.

(D) Northern blot analysis of RNAs extracted from cells depleted for Reb1p or Sth1p and containing the X3 reporter. See also Figure S2.

### Reb1p-Dependent Termination Does Not Require NFR Formation

Reb1p is known to activate transcription by recruiting the RSC chromatin remodeling complex, which, in turn, contributes to the positioning of the NFR (Hartley and Madhani, 2009). Our results

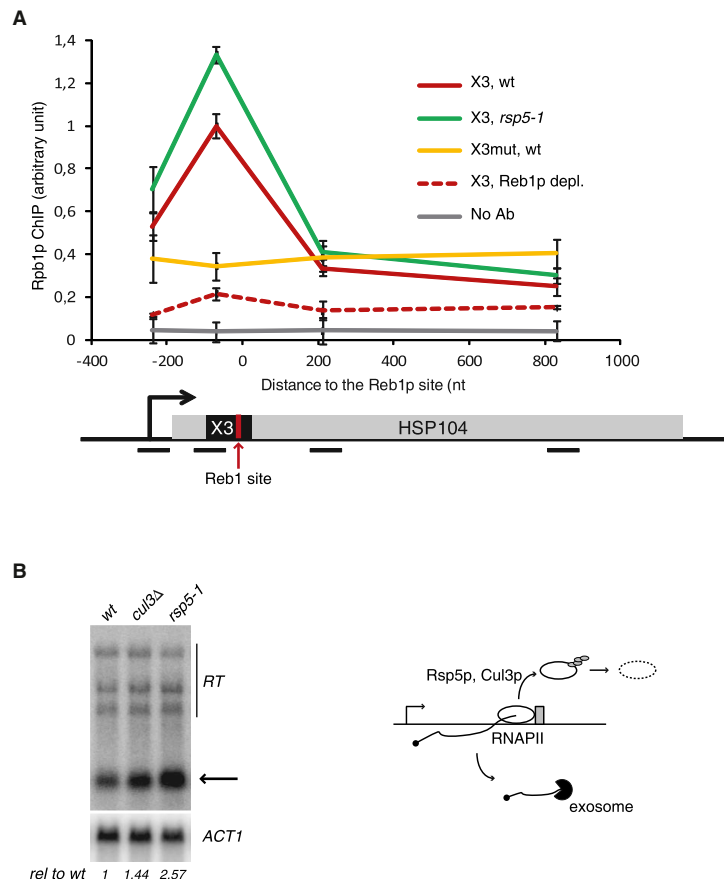
from in vitro transcription assays do not rule out the possibility that in vivo DNA binding might be necessary but not sufficient for Reb1p-dependent termination and that formation of an NFR might also be required. To address this question, we generated a version of Reb1p truncated for the first 336 aa (Reb1-DBD), which retains full ability to bind DNA, as verified by EMSA (Figure S2C), but cannot activate transcription (Figure 6C). We also considered that Nsi1p, which binds the same sequence as Reb1p but is only similar in the DNA binding domain, might also terminate transcription when expressed to the same levels as Reb1p. We ectopically expressed Reb1-DBD and Nsi1p under control of the *REB1* promoter, in a strain containing the endogenous *REB1* driven by the glucose-repressible *GAL1* promoter ( $P_{GAL1}$ -*REB1*). Metabolic depletion of endogenous Reb1p in the presence of Reb1-DBD or Nsi1p indicated that neither protein supports viability (Figure S2B) nor can either activate transcription of Reb1p-dependent genes (Figure 6C; data not shown).

the DNA bound Reb1p) and the catalytic center of the enzyme (where the 3' OH of the nascent RNA is positioned) (Saeki and Svejstrup, 2009). To substantiate these results, we performed in vitro transcription experiments using purified RNAPII and recombinant Reb1p. In this system, an elongation complex was assembled on an immobilized template that contains a Reb1p binding site (Porrua and Libri, 2013) and allowed to transcribe in the presence or absence of rReb1p. Consistent with in vivo observations, in the presence of Reb1p the polymerase pauses roughly 12 nt upstream of the Reb1 site, as indicated by the accumulation of a transcript that remains associated to the template (Figure 3B).

These findings strongly support the notion that the first step in Reb1p-dependent termination is the collision between RNAPII and Reb1p. However, DNA-bound Reb1p alone cannot provoke the release of the polymerase in vitro, suggesting that a mechanism must exist in vivo for dismantling the elongation complex (see below).

from in vitro transcription assays do not rule out the possibility that in vivo DNA binding might be necessary but not sufficient for Reb1p-dependent termination and that formation of an NFR might also be required. To address this question, we generated a version of Reb1p truncated for the first 336 aa (Reb1-DBD), which retains full ability to bind DNA, as verified by EMSA (Figure S2C), but cannot activate transcription (Figure 6C). We also considered that Nsi1p, which binds the same sequence as Reb1p but is only similar in the DNA binding domain, might also terminate transcription when expressed to the same levels as Reb1p. We ectopically expressed Reb1-DBD and Nsi1p under control of the *REB1* promoter, in a strain containing the endogenous *REB1* driven by the glucose-repressible *GAL1* promoter ( $P_{GAL1}$ -*REB1*). Metabolic depletion of endogenous Reb1p in the presence of Reb1-DBD or Nsi1p indicated that neither protein supports viability (Figure S2B) nor can either activate transcription of Reb1p-dependent genes (Figure 6C; data not shown).





**Figure 4. Reb1p-Dependent Termination Involves Transcriptional Pausing In Vivo**

(A) Rpb1p ChIP occupancy on a reporter construct containing the X3 or X3mut sequence, containing the wild-type or the mutated Reb1 site, respectively. The experiment was also performed in *rsp5-1* cells as indicated. A strain metabolically depleted for Reb1p (*P<sub>GAL1</sub>-REB1*, growth in glucose) was included as a control. Average of three experiments; error bars represent standard deviation. A scheme of the construct with the position of the Reb1 site and the positions of the amplified regions is shown on the bottom. See also Figure S3.

(B) Northern blot analysis of RNAs extracted from *rsp5-1* or *cul3Δ* cells containing the X3 reporter. Defects in the clearance of RNAPII paused at the Reb1 site lead to the increased detection of nascent RNAs (indicated by an arrow). Quantification of the nascent RNA levels relative to *ACT1* and to the wild-type for this gel is shown below the panel. From multiple independent analyses, the levels in *rsp5-1* cells relative to wild-type are  $2.68 \pm 0.12$  (average and standard deviation from three experiments). A scheme of the experiment is shown on the right.

#### The Mechanism of RNAPII Release in Reb1p-Dependent Termination

The in vitro transcription experiments reported above indicate that Reb1p induces RNAPII pausing but cannot dismantle an elongation complex in this minimal experimental set up. This suggests the existence of a mechanism that ensures the release of RNAPII roadblocked by Reb1p.

We first assessed whether RNAPII pausing could also be detected in vivo

around Reb1 sites by chromatin immunoprecipitation (ChIP). To this end we measured RNAPII occupancy at the X3 selected terminator and its mutated version (X3mut). As shown in Figure 4A, the presence of a functional Reb1 site induced a dramatic increase in RNAPII occupancy, which was fully abolished by metabolic depletion of Reb1p. Binding of Reb1p to its site in vivo was verified by ChIP (Figure S3). These findings indicate that Reb1p bound to the DNA induces RNAPII pausing in vivo, which is consistent with the detection of nonadenylated, nascent transcripts associated with paused polymerases (Figures 2B and S1D).

The persistence of paused polymerase is thought to be detrimental for several cellular functions, and it has been shown that polymerases stalled upstream of DNA damage are released by a ubiquitination/degradation mechanism. This implicates mono-ubiquitination by Rsp5p of the largest subunit of RNAPII, Rpb1p, which induces ubiquitin chain extension by the E1c1p-Cul3p complex and subsequent proteasomal degradation (Harreman et al., 2009, and references therein). We surmised that this pathway might also be required for the release of polymerases

Both factors, however, efficiently induced transcription termination at the Reb1p-dependent terminator (Figure 3C, lanes 3 and 4).

To further support the notion that formation of an NFR is not required for termination, we turned to the RSC complex, which is the main effector of Reb1p in NFR formation. Sth1p is the catalytic subunit of the RSC, and its depletion strongly affects Reb1p-dependent NFR formation genome-wide (Hartley and Madhani, 2009). Consistent with this notion, while metabolic depletion of Sth1p using a glucose repressible promoter strongly affected the activation of Reb1p target genes (see below, Figures S5A and 6C), it had no effect on Reb1p-dependent termination (Figure 3D). This strongly suggests that Reb1p-dependent termination does not depend on the ability of this factor to promote nucleosome remodeling.

Together our results indicate that DNA binding is necessary and sufficient, in vivo, to promote Reb1p-mediated transcription termination, and strongly support the notion that termination is triggered by the occurrence of a roadblock to transcription.

paused by a roadblock and assessed Reb1p-dependent termination into an *rsp5-1* or a *cul3Δ* strains, both defective for this ubiquitination pathway. Failure to ubiquitinate Rpb1p is expected to result in increased persistency of the polymerase at the site of Reb1p-dependent pause. This was indeed verified by RNAPII ChIP in *rsp5-1* cells (Figure 4A). This increase was not due to higher Reb1p levels or binding to the DNA, because Reb1p occupancy at the pause site was even slightly lower in *rsp5-1* cells compared to the wild-type (Figure S3).

Increased levels of roadblocked polymerase in *rsp5* or *cul3* mutants are expected to result in increased levels of nonadenylated, nascent RNA. Consistently, we observed a significant increase of short RNAs ending at the site of roadblock in *rsp5-1* and *cul3Δ* cells transformed with the reporter (Figure 4B). These RNAs are stable, because they are detected in cells proficient for nuclear degradation (Figure 4B), and are nonadenylated (data not shown), as expected for nascent transcripts.

From these experiments we conclude that RNAPII roadblocked at sites of Reb1p binding is released by the Rsp5p-Cul3p ubiquitination pathway, akin to the removal of RNAPII paused at sites of DNA damage.

### Reb1p Terminates RNAPII Transcription at Natural Sites

The experiments reported above demonstrate the proof of principle of Reb1p-dependent termination and its mechanism, but not its occurrence in natural cases. To assess the biological relevance of this termination pathway, we analyzed by tiling arrays the transcriptome of cells transiently depleted of Reb1p, using the *P<sub>GAL1</sub>-REB1* strain, with or without the *rrp6Δ* deficiency. These data were crossed to the known genome-wide distribution of Reb1p, as determined by high-resolution ChIP-seq (Rhee and Pugh, 2011) and the position of Reb1p binding sites, and to the distribution of unstable transcripts (Neil et al., 2009; Gudipati et al., 2012). Several cases of Reb1p-dependent termination were indeed identified for natural transcripts, most of which are unstable (Figure S4; data not shown). A characteristic feature of these transcripts, revealed by 3' end SAGE genome-wide analyses (Neil et al., 2009) and our analyses using the artificial terminators (e.g., see Figures S1 and S2), is their well-defined 3' end compared to canonical CUTs, which is due to the different mechanism of termination.

We validated the occurrence of Reb1p-dependent termination for a few model cases (Figures 5 and S5). For instance, an unstable short transcript (*uATP5*) can be detected upstream of the *ATP5* gene, terminating a few nucleotides before a Reb1 site. Transcription termination of *uATP5* was Reb1p dependent, because depletion of the latter strongly affected the levels of the short RNA to the profit of a longer species, terminating at a downstream site. Note that the longer species is itself unstable because it is only visible in an *rrp6Δ* context and at levels that do not fully compensate for the loss of the short transcript. As for other cases reported below, a fraction of the longer RNAs might also be degraded in the cytoplasm.

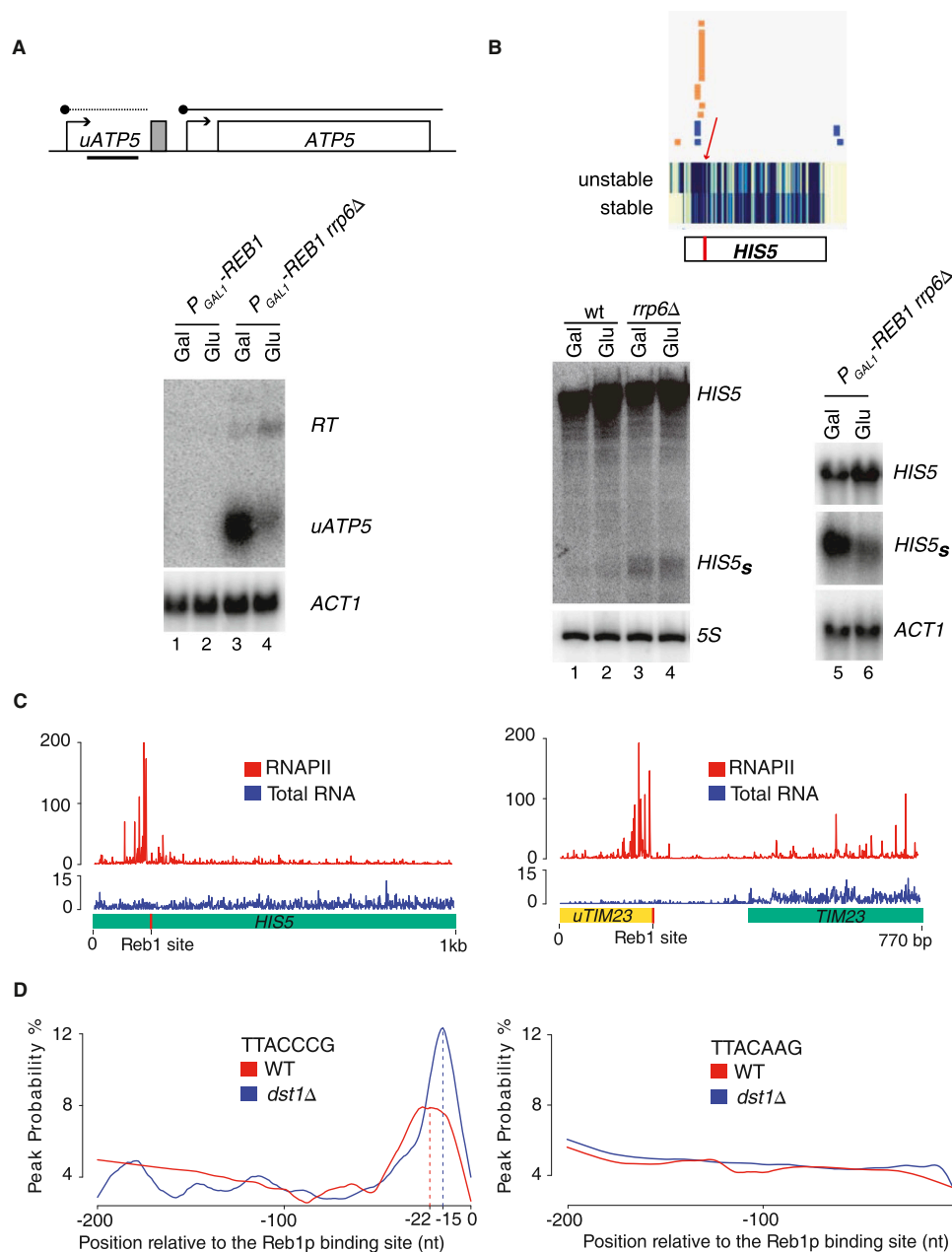
Most Reb1 sites are located in intergenic regions. The case of *HIS5* is peculiar, since this gene contains two Reb1 sites, one upstream and one within the coding region. A short unstable transcript terminating upstream of the internal site can be detected by SAGE and northern blot analysis (Figure 5B; Neil et al.,

2009). Assessing the Reb1p dependency of the short transcript is complicated by the fact that the *HIS5* gene is strongly Reb1p dependent and depletion of Reb1p or the RSC catalytic subunit Sth1p strongly affects *HIS5* expression (Figure S5A). Therefore we ectopically expressed *HIS5* under control of a heterologous promoter and showed that depletion of Reb1p strongly reduced early termination at the Reb1p internal site to the profit of the full-length transcript (Figure 5B, lanes 5 and 6). Whether this unusual organization implies the occurrence of regulated attenuation is unclear so far and is a matter for future studies.

We have shown above that RNAPII pauses in vivo in a Reb1p-dependent manner at an artificial terminator. We explored whether RNAPII pausing could be detected at natural Reb1p termination sites using the high-resolution genome-wide distribution of RNAPII defined by native elongating transcript sequencing (NET-Seq; Churchman and Weissman, 2011). As shown in Figure 5C, prominent RNAPII pausing peaks are precisely located a few nucleotides before the Reb1p binding site at the *HIS5* gene and the RUT upstream of the *TIM23* gene (Figure S5B). Pausing is not expected to be detected at all sites of termination, because accumulation of the signal depends on a balance between transcription levels upstream of the pause site and the rate of clearance of the paused polymerase. For instance, at *uATP5* RUT, increased RNAPII signals can be observed upstream of the Reb1 site, but the overall signal is too low to unambiguously identify a peak as in the case of *HIS5* or *uTIM23* (data not shown). To generalize this finding, we combined data from all Reb1 sites in the genome and plotted the frequency of RNAPII peaks that are at least two standard deviations above the mean at each position in a 200 nt region upstream of the Reb1 site. The distribution of the peak frequency indicates a significantly increased RNAPII pausing 14–25 nt before the average Reb1 site, but not at sites containing single-nucleotide mutations known to strongly affect Reb1p binding (Figure 5D), which is fully compatible with in vitro experiments.

Elongation pausing is known to induce backtracking of RNAPII, during which the 3' end of the nascent RNA is displaced from the catalytic site of the enzyme. The RNA is cleaved by the coordinated action of the polymerase and the TFIIS elongation factor to correctly position the 3' OH of the molecule and resume elongation (Reines et al., 1992; Izban and Luse, 1992). In a wild-type strain, the 3' end of the nascent RNA at the region of pausing is determined by cleavage, which can occur at multiple, closely spaced positions during backtracking; when cleavage is inhibited (as in the absence of TFIIS), the 3' end of the nascent RNA generally coincides with the last position occupied by the polymerase before backtracking. To confirm the occurrence of pausing at the average Reb1 site, we repeated the same analysis in a strain deleted for TFIIS (*dst1Δ*). As shown in Figure 5D, the RNAPII peak was still observed in the mutant, but with a distribution slightly displaced toward the Reb1 site and less dispersed relative to a wild-type strain, consistent with the notion of pausing and backtracking at sites of roadblock.

Together, these results demonstrate the existence of natural sites of Reb1p-dependent transcription termination in the yeast genome, which occurs via a roadblock mechanism.



**Figure 5. Reb1p Terminates Transcription of Natural ncRNAs, "RUTs"**  
 (A) Northern blot analysis of the unstable transcript located upstream of *ATP5* (*uATP5*). A scheme of the genomic region is shown on the top; the Reb1 site (gray box) is located a few nucleotides downstream of the mapped 3' end of the transcript.  
 (B) Stable and unstable RNAs species derived from the *HIS5* locus. (Upper panel) Tiling array heatmap and SAGE tags analyses (Neil et al., 2009). The ends of individual RNA species mapped by 3' end SAGE are indicated by orange (unstable fraction) or blue (stable fraction) squares. The position of the Reb1 site in *HIS5* (158 nt downstream of the ATG) is indicated by a red arrow and a red bar in the scheme. (Lower panels) Northern blot analyses revealing a short unstable transcript  
 (legend continued on next page)



### Functional Cooperation between Termination Pathways to Control Pervasive Transcription

Because Reb1p occupancy is prominent in intergenic regions, in many cases a putative site of Reb1p-dependent termination overlaps with the canonical terminator of a gene, suggesting that different pathways may concur to ensure efficient termination either by acting synergistically or sequentially as reciprocal fail-safe mechanisms.

The YSY6 locus is one example of an alternative pattern of main and backup termination mechanisms. This gene (Figure S6A) contains a Reb1 site that coincides with the end of the transcript, still transcription termination is mainly dictated by the CPF/CF, as production of YSY6 RNA is sensitive to mutation of Rna14p (an essential CF subunit; Figure S6B) and insensitive to Reb1p depletion (Figure 6A). However, a low level of transcriptional readthrough naturally occurs that is neutralized by the occurrence of Reb1p-dependent termination. Indeed, upon depletion of Reb1p a longer transcript is produced that is exported and degraded in the cytoplasm by the nonsense-mediated decay (NMD) pathway, as shown by the marked accumulation of this species when NMD is impaired by the *upf1*  $\Delta$  mutation (Figures 6A and S6C). A similar organization is found at the YDL233w locus, where Reb1p-dependent termination acts as a fail-safe mechanism to neutralize leakage from the main CPF/CF termination. In this case, the readthrough transcripts that are not terminated at the Reb1p-dependent site (i.e., when Reb1p is depleted) are degraded both by the exosome and the NMD pathway because they can only be detected in a *dis3-exo<sup>-</sup> upf1*  $\Delta$  context (Figure 6B, lanes 5–8).

In the cases described above, depletion of Reb1p leads to a downregulation of the downstream genes (*DEM1* and *OST4* for YSY6 and YDL233w respectively; Figures 6, S6A, and S6D). One possibility is that increased upstream readthrough due to impairment of the backup termination pathway silences the downstream promoter by transcriptional interference. However, it is also possible that transcriptional activation of the downstream gene depends on Reb1p via the RSC pathway. To distinguish between the two possibilities, we took advantage of the finding that the Reb1-DBD can terminate but is unable to activate transcription (Figure 3C). The expression of *DEM1* and *OST4* was analyzed after depletion of Reb1p in the presence of Reb1-DBD, which should prevent transcriptional interference but should not allow activation. As shown in Figure 6C, the Reb1-DBD could only poorly restore expression of *DEM1*, indicating that the latter critically depends on the activation function of Reb1p. Consistently, expression of *DEM1* was markedly affected by depletion of Sth1p, the catalytic subunit of the RSC (Figure 6C). Therefore, whether transcription interference occurs at the *DEM1* locus in

the absence of Reb1p cannot be reliably determined. In contrast, *OST4* was dependent on Reb1p for activation only to a lesser extent, as shown by the poor effect of Sth1p depletion (Figure 6D). Importantly, expression of the Reb1-DBD alone was sufficient to partially restore *OST4* expression, supporting the notion that silencing of *OST4* upon Reb1p depletion is mainly due to transcriptional interference.

The paradigmatic case of *OST4* suggests that Reb1p-dependent termination could play a general role as a fail-safe mechanism preventing transcriptional interference to occur at contiguous genes. One important prediction of this model is that Reb1p-dependent termination should become essential under conditions where CPF/CF termination is also defective. To test this prediction, we generated a hypomorphic, temperature-sensitive allele of Reb1p containing a mutation in its DNA binding domain, *reb1-ts1*. This mutation only slightly affects growth in an otherwise wild-type context, presumably because DNA binding is only partially affected. However, when combined with the CPF/CF *ma14-3* mutation that is mildly defective at permissive temperature (our unpublished data; Libri et al., 2002), the *reb1-ts1* mutation becomes detrimental for growth (Figure S6E). Importantly, the growth defect of the double mutant can be partially suppressed by expressing the DNA binding domain of Reb1p alone or by Nsi1p, which bind the same sequence and terminate transcription but lack the activation function of Reb1p (Figure S6F). This indicates that the synthetic growth defect is due to loss of fail-safe Reb1p-dependent termination in a context partially defective for CPF/CF termination.

Together these findings support the notion that Reb1p has a dual role in the control of gene expression. On one side, it induces the correct positioning of NFRs, which is essential for transcription activation; on the other side, it plays an important role in “protecting” promoter regions from the deleterious effect of even modest readthrough transcription from neighboring transcription units.

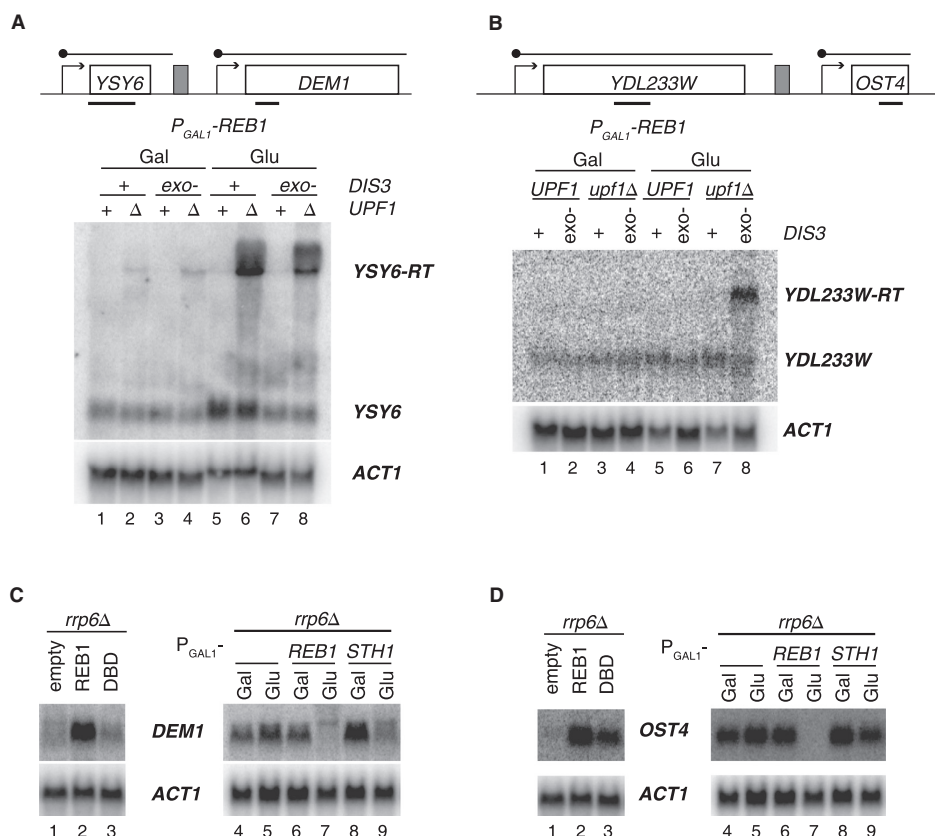
### DISCUSSION

Because the yeast genome is compact and extensively transcribed, efficient transcription termination is very important for robust and stable gene expression. In this study we describe an additional pathway for transcription termination that depends on the DNA binding factor Reb1p, a factor hitherto implicated in the activation of transcription (see Figure 7 for a model). These data open up an additional perspective on the insulation of transcriptional units, the control of pervasive transcription, and the role of DNA binding proteins in these processes.

(HIS5<sub>2</sub>) derived from roadblock termination at the Reb1 site in an *rrp6*  $\Delta$  strain (left panel, lanes 3 and 4). (Right) RNAs derived from *HIS5* expressed under the control of a heterologous promoter (*P<sub>TE1</sub>*) in a *P<sub>GAL1</sub>-REB1 rrp6*  $\Delta$  strain. See also Figure S5A.

(C) RNAPII occupancy revealed by NET-Seq analysis (red histograms) at *HIS5* and *TIM23* (data from Churchman and Weissman, 2011). Total RNA signal in the same regions is shown in blue as a control. An unstable transcript (*uTIM23*) terminated by Reb1p is located immediately upstream of *TIM23*. The position of the Reb1 site is indicated by a red bar in the scheme. See also Figure S5B.

(D) Metasite analysis of RNAPII pausing revealed by NET-Seq (Churchman and Weissman, 2011) upstream of Reb1 sites in the genome. The plots show the frequency of polymerase pausing peaks in a 200 nt window preceding a generic Reb1 site (TTACCCG) or a mutated sequence that cannot bind Reb1p (TTACAAG), calculated over all the sites in the genome aligned at position 0. The analysis was performed in a wild-type strain (red) or a TFIIS mutant (*dst1*  $\Delta$ , blue). See also Figure S7.



**Figure 6. Reb1p-Dependent Termination Functions as a Fail-Safe Mechanism**

(A and B). Reb1p-dependent termination functions as a fail-safe mechanism for YSY6 and YDL233W mRNA termination. Northern blot analysis of YSY6 and YDL233W mRNAs under the indicated conditions and strains. Readthrough transcripts are only detected upon Reb1p depletion when NMD is impaired in *upf1Δ* cells (YSY6) or when both the exosome and NMD are affected (YDL233W).

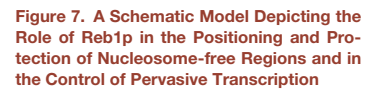
(C) Northern blot analysis of *DEM1* expression. *DEM1* expression depends on Reb1p and Sth1p and is not restored by expression of the Reb1p DNA binding domain.

(D) Northern blot analysis of *OST4* expression. In the absence of Reb1p, expression of *OST4* can be restored with the reestablishment of fail-safe termination upon expression of the Reb1p DNA binding domain alone. The position of the Reb1 site is indicated by a gray box. See also Figure S6.

### Roadblock Termination Is a General and Conserved Mechanism

Roadblock termination has been shown to occur for transcription of rDNA genes by RNAPII in several species, including yeast. Termination is triggered by the collision between the transcribing polymerase and a DNA-bound factor that is TTF1 in the mammalian system and most likely Nsi1p in yeast (Evers and Grummt, 1995; Evers et al., 1995; Reiter et al., 2012). However, the sensitivity to roadblocks is not a specific feature of RNAPII, and the three RNA polymerases can be paused in vitro by the DNA-bound Reb1p (Lang et al., 1994; this study). The inability to overcome a DNA-bound obstacle might therefore be an intrinsic property of RNA polymerases and possibly an ancient mechanism of termination, which is underscored by the occurrence of

roadblock termination in bacteria (Belitsky and Sonenshein, 2011; Pavco and Steege, 1990). We show here that roadblock termination also occurs in vivo for RNAPII. This indicates that in spite of the existence of elongation factors and chromatin-remodeling factors to allow transcription through nucleosomes in vivo, the cell apparently lacks specific mechanisms to efficiently remove nonhistone, DNA-bound proteins in front of transcribing RNA polymerases. Reb1p-dependent roadblock termination is also very likely to be conserved in other species. Reb1 sites are similarly excluded from regions of active elongation in several Hemiascomycetes species (*S. paradoxus*, *K. lactis*, *S. mikatae*, *S. bayanus*, and *C. glabrata*) ( $p < 10^{-4}$ ) to an extent that is not likely to be due only to the role of Reb1p in NFR formation. For comparison, the distribution of Cbf1p, a



clearance of the paused polymerase) might more specifically characterize roadblock termination.

An interesting feature of Reb1p-induced roadblock is its directionality. It was originally demonstrated that only when bound to a site in the “G-rich” orien-

tation relative to the direction of transcription can Reb1p roadblock RNAPII in vitro (Lang and Reeder, 1993; Lang et al., 1994). Although we did not observe in vitro a similar directionality with RNAPII (data not shown), one orientation of the Reb1 site was markedly preferred in our selection experiment, and its reverse orientation could not terminate transcription in our reporter system (Figure 1D), suggesting directionality. Interestingly enough, the preferred site orientation for roadblocking RNAPII is the “C-rich,” i.e., opposite to that for RNAPI. However, natural sites of roadblock termination were observed upstream of the Reb1 site in both orientations (e.g., the site upstream *OST4* is “G-rich”), and RNAPII pausing was also detected by “metastate” analysis when the Reb1 site is in the “G-rich” orientation (Figure S7A). This suggests that the geometry of the RNAPII-Reb1p collision affects the efficiency of the roadblock. The “G-rich” orientation would be sufficient to roadblock the less robust RNAPII elongation in vitro and the generally low frequency transcription in vivo (Zenklusen et al., 2008). Still, it does not suffice to arrest transcription driven by the strong Tet-repressible promoter in our more stringent selection system, because high-frequency firing might efficiently outcompete Reb1p binding.

Differently from RNAPII (Lang et al., 1994; Mason et al., 1997a, 1997b), release of roadblocked RNAPII does not require a release sequence element because the isolated 8 nt Reb1 site is sufficient for termination (Figure 1D). However, we cannot exclude that, in addition to the Rsp5p and elongin-cullin ubiquitination system, other factors might contribute to the release of paused RNAPII, as proposed for the Rat1p exonuclease and the Sen1p helicase in RNAPII termination (Kawauchi et al., 2008). Nevertheless, we have shown that the CPF/CF- and the NNS-dependent termination pathways are dispensable for at least a few tested cases of Reb1p-dependent termination (data not shown; Figures 1C and S1B).

It remains unclear whether ubiquitination of roadblocked RNAPII is followed by degradation of the enzyme, in analogy with what was demonstrated for RNAPII stalled upstream of DNA damage (Somesh et al., 2005; Verma et al., 2011; Wilson et al., 2013). Destruction of one polymerase molecule per termination event might not look economical, but advantages are

These considerations beg the question of what restricts roadblock termination for the cell to cope with a wealth of DNA binding activities that could potentially interfere with transcription. It is likely that not all DNA-bound proteins elicit RNAPII pausing and termination. For instance, we observed that binding of Gal4p does not elicit termination of upstream transcription regardless of growth conditions (data not shown). On the other side, binding to the DNA of the *E. coli* Lac repressor was found to prevent expression of SV40 large T antigen in mammals, suggesting RNAPII roadblocking (Deuschle et al., 1990). Similarly, a short sequence motif that is a binding site for a DNA binding factor in the adenovirus major late promoter has been suggested to prevent elongation (Connolly and Manley, 1989). Roadblock termination must depend on a complex interplay of several competing events: first, nucleosomes might occlude binding sites for several factors; second, transcription itself might prevent the interaction of potential roadblockers with the DNA, either by sterically hindering binding sites or by inducing a chromatin structure that limits the exposure of these sites. The ability of a DNA binding factor to counteract these competing events, in terms of abundance and affinity for the DNA, conceivably qualifies a roadblocker, and Reb1p might occupy a pre-eminent position in this respect.

RNAPII pausing and the consequent protection of nonadenylated RNAs are readily observed at sites of roadblock termination using NET-Seq (Figure 5; data from Churchman and Weissman, 2011) and PAR-CLIP (Creamer et al., 2011; data not shown) data. Although pausing has also been proposed to occur as a prerequisite of CPF/CF- and NNS-dependent termination, the same techniques do not detect significant RNAPII accumulation at these sites, suggesting that localized RNAPII pausing (or slow

probably to be found in the overall low evolutionary cost of the system, which requires very limited sequence information (typically 8–10 nt) and uses enzymatic pathways selected for other purposes. This pathway appears to be well adapted for neutralizing low levels of transcription at “sensitive” locations such as NFRs and in any instances where the benefits of compact termination signals might justify its costs in terms of polymerase losses (e.g., in the case of regions “crowded” with transcription units).

### Functional Significance of Roadblock Termination

Because Reb1p-dependent termination leads to the production of unstable transcripts, it is a reasonable assumption that its significance relates more to the control of transcription than to the generation of functional RNA molecules. The bulk of cryptic transcription originates in NFRs, where Reb1p is preferentially localized, and it is very possible that pervasive transcription events are neutralized early by Reb1p (or other roadblockers) within the NFR from which they originate, and escape detection because of their small size. We suggest that Reb1p and roadblock termination in general might contribute to suppress promoter bidirectionality.

Importantly, Reb1p has also a role in preventing transcriptional interference, sometimes as part of a fail-safe termination pathway (Figure 7). Regions where transcription originates in NFRs are extremely sensitive to “invasions” from neighboring polymerases. Because elongating transcription induces a chromatin state that is repressive for initiation (for a review, see Jensen et al., 2013), even low levels of transcription through promoters suffice to inhibit initiation (Bumgarner et al., 2012; Castelnovo et al., 2013; van Werven et al., 2012). Thus, isolating regions of transcription initiation from elongating polymerases is a major challenge, and the preferential localization of Reb1p in intergenic regions might be economically exploited to sustain such a double role as a transcriptional activator and an NFR “guard.” The protection of the *OST4* gene from interfering readthrough transcription is paradigmatic in this regard. The importance of enforcing termination is underscored by the existence of fail-safe pathways described for other genes, such as the one triggered by cleavage of the nascent RNA by the endoribonuclease Rnt1p (Ghazal et al., 2009; Rondón et al., 2009).

To estimate the extent of Reb1p-dependent termination in the yeast genome, we can propose a minimal figure based on the occurrence of robust RNAPII pausing at sites of termination (Figures 4, 5, and S7). When we considered all the putative Reb1 sites in the yeast genome based on sequence alone, roughly 12% show an RNAPII pausing signal upstream (Figures S5D and S7A). However, when we restricted the analysis to the subset of sites with the highest actual Reb1p occupancy (MacIsaac et al., 2006), this figure approaches 40% (Figure S7B). This could still be an underestimation, because sites of low transcriptional activity and cases of fail-safe termination are likely overlooked.

Finally, many cases of roadblock pausing/termination might translate into regulation, not only at the level of transcription but also splicing and 3' end processing. Altering the abundance or the binding affinity of DNA binding factors might affect pausing at sensitive locations, repress or generate regulatory transcription, or more directly attenuate the expression of genes. The case of *HIS5* described here might be paradigmatic in this

respect, and investigating the regulatory potential of its unusual architecture is an exciting future perspective.

### EXPERIMENTAL PROCEDURES

#### Yeast and Constructs

Yeast strains and oligonucleotides used in this study are listed in Supplemental Experimental Procedures. Plasmids were generally constructed using homologous recombination in yeast with standard procedures and are listed in Supplemental Experimental Procedures. Standard molecular biology analyses including transcriptome analyses are described in Supplemental Experimental Procedures. In vivo selection of terminators has been extensively described elsewhere (Porrua et al., 2012). The statistical analysis of the selected pool of sequence was performed as described in Supplemental Experimental Procedures.

#### Genome-wide Analyses of Reb1p-Induced Polymerase Pausing

Reb1p-induced RNAPII pausing was analyzed using the NET-seq data sets (Churchman and Weissman, 2011) in wild-type and *dst1Δ* context. A total of 791 occurrences of the Reb1p core binding site (TTACCCG) were identified in the yeast genome (SGD R62).

To analyze RNAPII pausing at each of these sites, we searched for local peaks of polymerase occupancy in an upstream window of 200 nt. A peak was defined as a read value that is higher than the mean plus two times the standard deviation calculated over all the nonzero read values in the 200 nt window. All the windows containing fewer than three nonzero read values were excluded from this calculation, resulting in a total of 555 (wild-type) and 604 (*dst1Δ*) processed sites for the “TTACCCG” sequence. The same analysis was performed for polymerases transcribing toward the reverse complement of the site (CGGGTAA, 534 occurrences for the wild-type and 594 for the *dst1Δ* data set), and at a mutated site (TTACaaG) that does not bind Reb1p as a control. In order to make each metasite analysis comparable, the number of peaks occurring at any given position in the 200 nt window was divided by the total number of nonzero read values at that position, which represents peak frequency. Note that only the occurrence of a peak and not its height was taken into account. To obtain the plots shown in Figures 5 and S7, a smoothing of the data was performed using the “supsmu” R function.

#### ACCESSION NUMBERS

The whole tiling array set of data has been deposited into ArrayExpress under accession number E-MTAB-2241.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and Supplemental Experimental Procedures and can be found with this article at <http://dx.doi.org/10.1016/j.molcel.2014.10.026>.

#### AUTHOR CONTRIBUTIONS

J.C. designed and performed experiments and wrote the paper; O.P. designed and performed experiments; T.C. designed and performed bioinformatics analysis; O.P. and T.C. contributed equally to this work. J.B., C.Z., and F.L. performed experiments; L.M.S. designed experiments; D.L. designed experiments and wrote the paper.

#### ACKNOWLEDGMENTS

We would like to thank J. Svejstrup for the kind gift of strains, and F. Pugh for allowing access to unpublished data; Libri lab members for fruitful discussions; and J. Svejstrup, N. Proudfoot, and F. Feuerbach for critical reading of this manuscript. This work was supported by the CNRS (D.L.), the Danish National Research Foundation (D.L.), the Agence Nationale pour la Recherche (ANR, ANR-08-Blan-0038-01 and ANR-12-BSV8-0014-01 to D.L.), the Fondation pour la Recherche Médicale (FRM, programme Equipes 2013 to D.L.), the Fondation Bettencourt-Schueller (prix Coup d'Élan 2009 to D.L.), the National

Institutes of Health (L.M.S.), and the Deutsche Forschungsgemeinschaft (L.M.S.). This study was technically supported by the EMBL Genomics Core Facility. O.P. was supported by fellowships from the EMBO and the FRM. This research was carried out within the scope of the Associated European Laboratory LEA "Laboratory of Nuclear RNA Metabolism."

Received: January 30, 2014

Revised: June 9, 2014

Accepted: October 29, 2014

Published: December 4, 2014

## REFERENCES

- Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell* 23, 841–851.
- Belitsky, B.R., and Sonenshein, A.L. (2011). Roadblock repression of transcription by *Bacillus subtilis* CodY. *J. Mol. Biol.* 411, 729–743.
- Brandl, C.J., and Struhl, K. (1990). A nucleosome-positioning sequence is required for GCN4 to activate transcription in the absence of a TATA element. *Mol. Cell. Biol.* 10, 4256–4265.
- Bumgarner, S.L., Neuert, G., Voight, B.F., Symbor-Nagraska, A., Grisafi, P., van Oudenaarden, A., and Fink, G.R. (2012). Single-cell analysis reveals that noncoding RNAs contribute to clonal heterogeneity by modulating transcription factor recruitment. *Mol. Cell* 45, 470–482.
- Castellano, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F., and Zenklusen, D. (2013). Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat. Struct. Mol. Biol.* 20, 851–858.
- Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368–373.
- Connelly, S., and Manley, J.L. (1989). RNA polymerase II transcription termination is mediated specifically by protein binding to a CCAAT box sequence. *Mol. Cell. Biol.* 9, 5254–5259.
- 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, e1002329.
- Deuschle, U., Hipskind, R.A., and Bujard, H. (1990). RNA polymerase II transcription blocked by *Escherichia coli* lac repressor. *Science* 248, 480–483.
- Evers, R., and Grummt, I. (1995). Molecular coevolution of mammalian ribosomal gene terminator sequences and the transcription termination factor TTF-I. *Proc. Natl. Acad. Sci. USA* 92, 5827–5831.
- Evers, R., Smid, A., Rudloff, U., Lottspeich, F., and Grummt, I. (1995). Different domains of the murine RNA polymerase I-specific termination factor mTTF-I serve distinct functions in transcription termination. *EMBO J.* 14, 1248–1256.
- Gari, E., Piedrafitra, L., Aldea, M., and Herrero, E. (1997). A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast* 13, 837–848.
- Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. *Nature* 425, 737–741.
- Ghazal, G., Gagnon, J., Jacques, P.-E., Landry, J.-R., Robert, F., and Elela, S.A. (2009). Yeast RNase III triggers polyadenylation-independent transcription termination. *Mol. Cell* 36, 99–109.
- Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat. Struct. Mol. Biol.* 15, 786–794.
- Gudipati, R.K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L.M., Jacquier, A., and Libri, D. (2012). Extensive degradation of RNA precursors by the exosome in wild-type cells. *Mol. Cell* 48, 409–421.
- Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., MacIsaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.-B., Reynolds, D.B., Yoo, J., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99–104.
- Harreman, M., Taschner, M., Sigurdsson, S., Anindya, R., Reid, J., Somesh, B., Kong, S.E., Banks, C.A.S., Conaway, R.C., Conaway, J.W., and Sveistrup, J.Q. (2009). Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. *Proc. Natl. Acad. Sci. USA* 106, 20705–20710.
- Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. *Cell* 137, 445–458.
- 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, 55–66.
- Izban, M.G., and Luse, D.S. (1992). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3' → 5' direction in the presence of elongation factor SII. *Genes Dev.* 6, 1342–1356.
- Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. *Mol. Cell* 52, 473–484.
- 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, 1082–1092.
- Kent, N.A., Eibert, S.M., and Mellor, J. (2004). Cbf1p is required for chromatin remodeling at promoter-proximal CACGTG motifs in yeast. *J. Biol. Chem.* 279, 27116–27123.
- Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol.* 12, 283–294.
- Kulkens, T., van der Sande, C.A., Dekker, A.F., van Heerikhuizen, H., and Planta, R.J. (1992). A system to study transcription by yeast RNA polymerase I within the chromosomal context: functional analysis of the ribosomal DNA enhancer and the RBP1/REB1 binding sites. *EMBO J.* 11, 4665–4674.
- Lang, W.H., and Reeder, R.H. (1993). The REB1 site is an essential component of a terminator for RNA polymerase I in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 13, 649–658.
- Lang, W.H., Morrow, B.E., Ju, Q., Warner, J.R., and Reeder, R.H. (1994). A model for transcription termination by RNA polymerase I. *Cell* 79, 527–534.
- Libri, D., Dower, K., Boulay, J., Thomsen, R., Rosbash, M., and Jensen, T.H. (2002). Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. *Mol. Cell. Biol.* 22, 8254–8266.
- MacIsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* 7, 113.
- Mason, S.W., Sander, E.E., and Grummt, I. (1997a). Identification of a transcript release activity acting on ternary transcription complexes containing murine RNA polymerase I. *EMBO J.* 16, 163–172.
- Mason, S.W., Wallisch, M., and Grummt, I. (1997b). RNA polymerase I transcription termination: similar mechanisms are employed by yeast and mammals. *J. Mol. Biol.* 268, 229–234.
- 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, 1038–1042.
- Pavco, P.A., and Steege, D.A. (1990). Elongation by *Escherichia coli* RNA polymerase is blocked in vitro by a site-specific DNA binding protein. *J. Biol. Chem.* 265, 9960–9969.
- Porrua, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat. Struct. Mol. Biol.* 20, 884–891.
- 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, 3935–3948.
- Reines, D., Ghanouni, P., Li, Q.Q., and Mote, J., Jr. (1992). The RNA polymerase II elongation complex. Factor-dependent transcription elongation involves nascent RNA cleavage. *J. Biol. Chem.* 267, 15516–15522.
- Reiter, A., Hamperl, S., Seitz, H., Merkl, P., Perez-Fernandez, J., Williams, L., Gerber, J., Németh, A., Léger, I., Gadal, O., et al. (2012). The Reb1-homologue



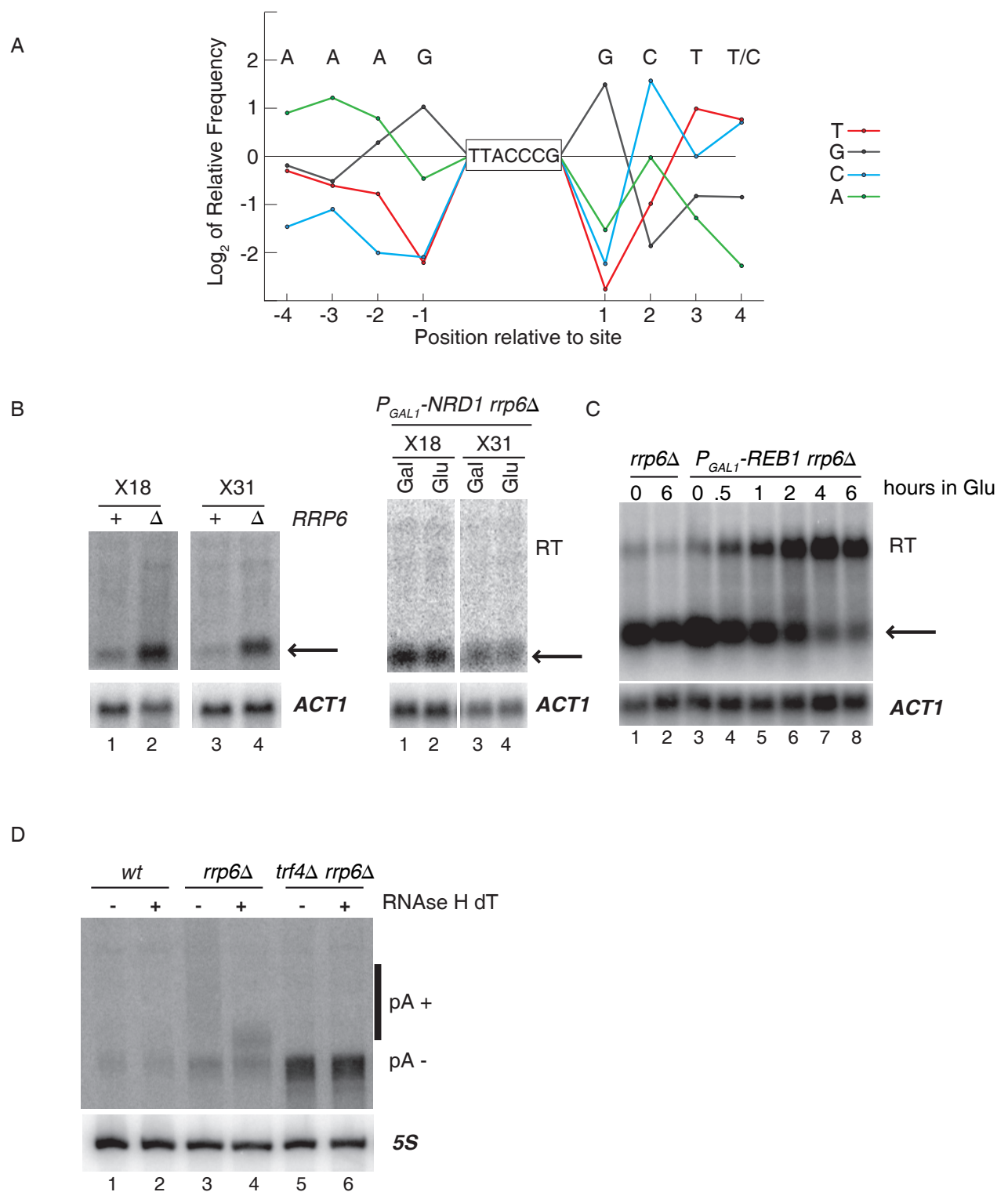
- Ydr026c/Nsi1 is required for efficient RNA polymerase I termination in yeast. *EMBO J.* 31, 3480–3493.
- Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 147, 1408–1419.
- Rondón, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Mol. Cell* 36, 88–98.
- Saeki, H., and Svejstrup, J.Q. (2009). Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. *Mol. Cell* 35, 191–205.
- Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* 155, 1075–1087.
- Somesh, B.P., Reid, J., Liu, W.-F., Sogaard, T.M.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2005). Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. *Cell* 121, 913–923.
- 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, 327–331.
- Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol. Cell* 24, 735–746.
- 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, 853–864.
- van Werven, F.J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., Primig, M., and Amon, A. (2012). Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* 150, 1170–1181.
- Verma, R., Oania, R., Fang, R., Smith, G.T., and Deshaies, R.J. (2011). Cdc48/p97 mediates UV-dependent turnover of RNA Pol II. *Mol. Cell* 41, 82–92.
- Wilson, M.D., Harreman, M., and Svejstrup, J.Q. (2013). Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. *Biochim. Biophys. Acta* 1829, 151–157.
- Wlotzka, W., Kudla, G., Granneman, S., and Tollervy, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J.* 30, 1790–1803.
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnauld, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121, 725–737.
- Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033–1037.
- Zenkhusen, 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, 1263–1271.

Molecular Cell, Volume 56

## **Supplemental Information**

### **Roadblock Termination by Reb1p Restricts Cryptic and Readthrough Transcription**

Jessie Colin, Tito Candelli, Odil Porrua, Jocelyne Boulay, Chenchen Zhu,  
François Lacroute, Lars M. Steinmetz, and Domenico Libri

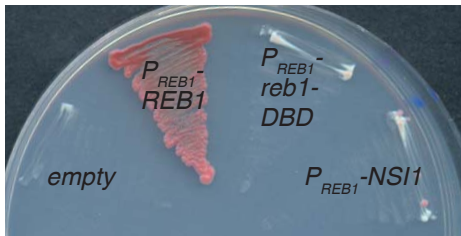




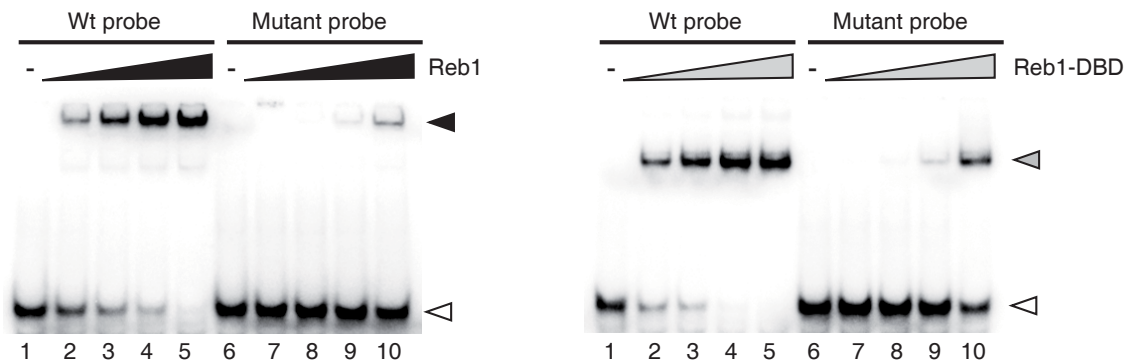
A

		Distance to Reb1p site (nt)
HSP::Reb1	TATTCTAGCTGC ↓ CTTCAATGAAACGGTTACCCGG	15
X3	ATTCGCAAGGACT ↓ ATAGTTCATTTAGTTACCCGG	14
X18	GATACCGGTCACCG ↓ CATTGCAAAAGATTACCCGG	13
X20	TATGGCTTTGGTGAG ↓ AGCAAGGAAGGTTACCCGG	12
X25	ACGATAAACGCTG ↓ CTGGTCGTTTACGTTACCCGG	14
X28	ATTCGCAAGGACTAT ↓ AGTTCATTTAGTTACCCGG	12
X31	TCGGTAGGATGAAT ↓ AGGTCAGAAAAGTTACCCGG	13

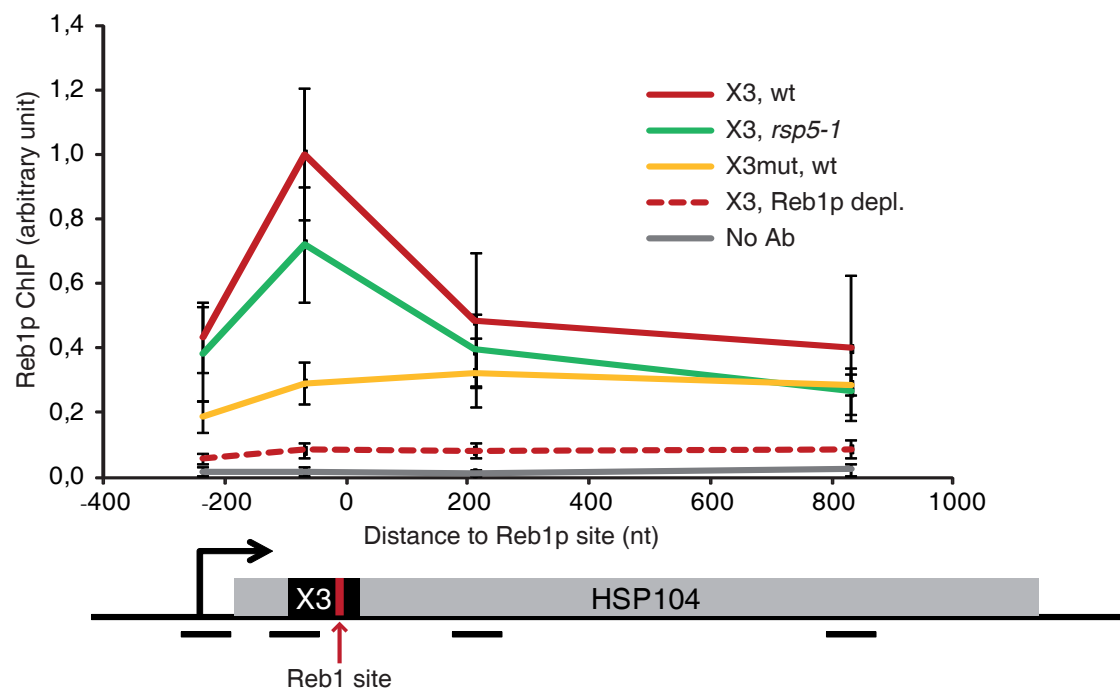
B

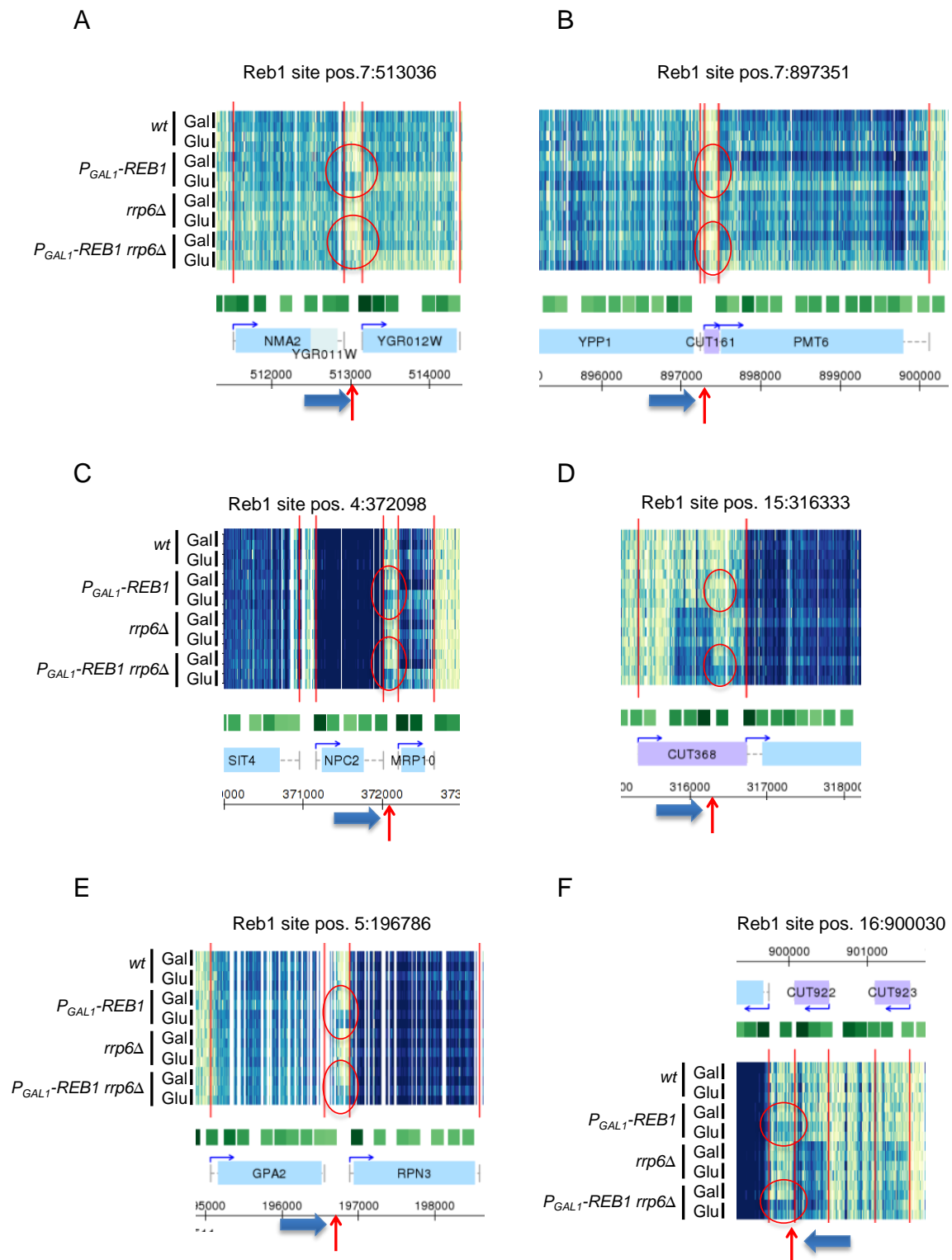


C



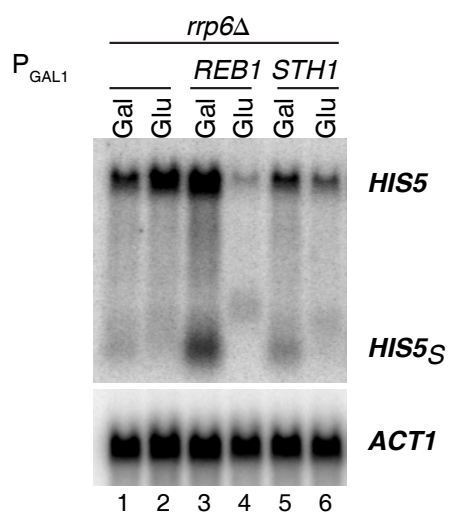
Wt probe 5' -ATGATGACCTAGAT**GTTACCCGG**GTAAAGAGCCCCATTAT-3'  
 Mutant probe 5' -ATGATGACCTAGAT**accgaagcc**GTAAAGAGCCCCATTAT-3'



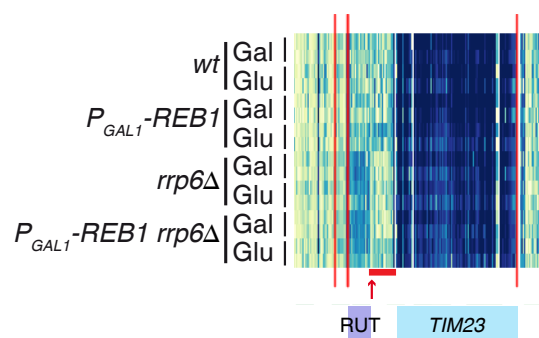


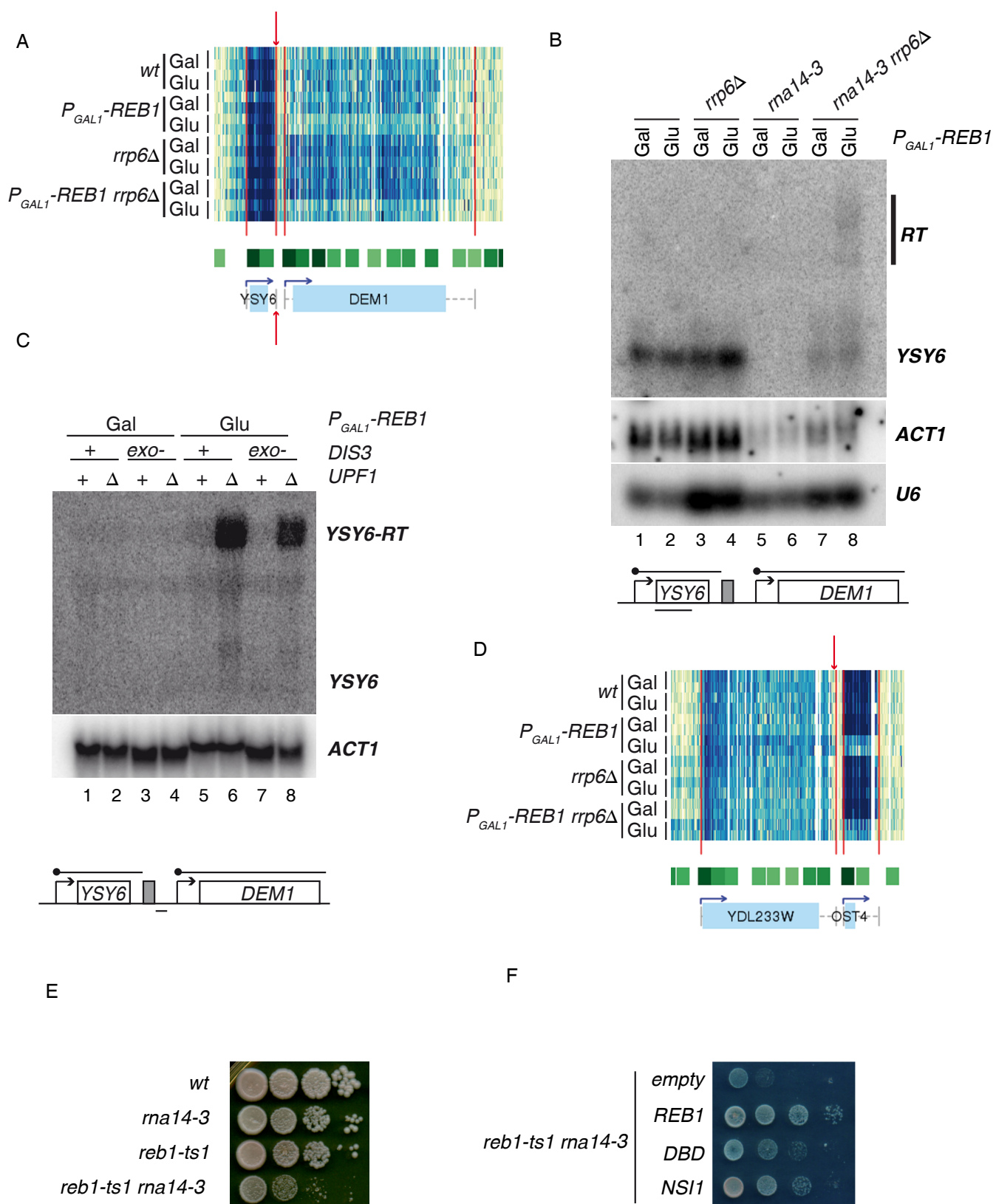
Colin et al., Figure S4, related to figures 5-6

A



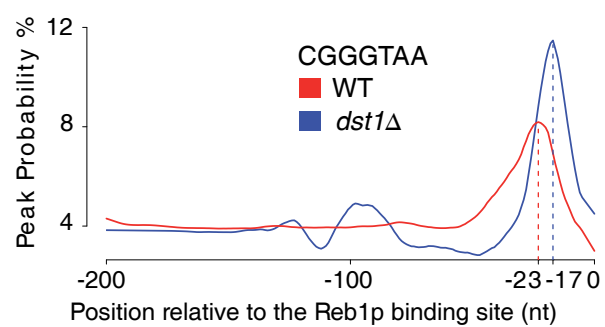
B



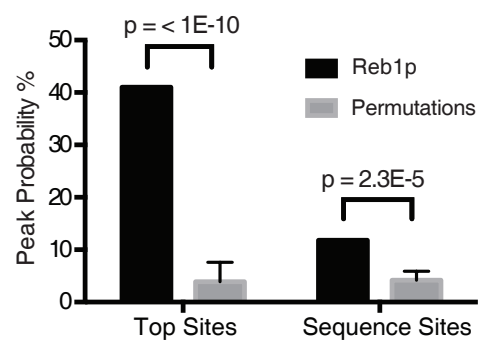


Colin et al. Figure S6, related to figure 6

A



B



### Supplemental Figure Legends

Figure S1. Reb1p induces RNAPII transcription termination. Related to Figures 1 and 2. **A.** Statistical analysis of the nucleotides flanking the Reb1p site in the selected terminators. For every position flanking the core Reb1p site, the plot indicates the  $\log_2$  of the ratio between the nucleotide frequencies in the selected and the non-selected pools. **B.** Northern-blot analysis of RNAs generated from two additional selected clones (X18 and X31). Similarly to clone X3 (Figure 1C), termination at these sequences is independent of the NNS pathway (right panels) and generates unstable transcripts that are stabilized in *rrp6* $\Delta$  cells (indicated by an arrow, left panel). **C.** Time course of Reb1p depletion in cells containing the Reb1p-dependent terminator X3. Labeling as in Figure 1. **D.** Northern-blot analysis of the polyadenylation status of transcripts generated by termination at the X3 sequence. Poly(A) tails were degraded by RNase H/oligo dT treatment as indicated. The smear of polyadenylated species in *rrp6* $\Delta$  cells (indicated by a solid bar) disappears when the gene coding for the poly(A) polymerase Trf4p is also deleted (compare lanes 3 and 5) and collapses to species with short poly(A) tails that cannot be further degraded by RNase H/oligo dT treatment (compare lanes 3 and 4). The band present in the wild type strain is non-adenylated because is insensitive to oligo dT/RNaseH digestion and cannot be detected after oligo-dT selection (Figure 2B).

Figure S2. Related to Figure 3 **A.** RACE mapping of 3' ends of transcripts generated by Reb1p-dependent termination within several independent selected sequences. Termination always occurs 12 to 15 nucleotides before the Reb1p-binding site, independently of the sequence. Clone HSP::Reb1 contains only the Reb1p binding site inserted in the *HSP104* coding sequence. The mapped 3' ends are indicated by an arrow and the Reb1p binding site is represented in bold. **B.** The DNA binding domain of Reb1p is not sufficient to support growth. *P<sub>GAL1</sub>-REB1* cells containing Reb1p, the DNA-binding domain of Reb1p or Nsi1p, under control of the *REB1* promoter are grown on glucose plates to deplete endogenous Reb1p. Complementation only occurs with wild type Reb1p. **C.** Electrophoretic mobility shift assay (EMSA) using either rReb1p or its DNA binding domain. Both proteins bind efficiently dsDNA carrying a Reb1p binding site but not its mutated version.

Figure S3. Reb1p occupancy does not increase in *rsp5-1* cells. Related to Figure 4. Reb1p occupancy determined by ChIP on a template carrying the X3 sequence in wild type or *rsp5-1* cells. As in figure 4,

the experiment was also performed with a clone containing a mutated Reb1p site and after Reb1p depletion as a control.

Figure S4. Related to Figures 5 and 6. Heatmap of additional examples of natural transcripts terminated by the Reb1p-dependent pathway. For all these examples, an extended read-through transcript (red oval) is readily detected upon Reb1p depletion. Reb1 site is indicated by a red arrow.

Figure S5. Related to Figure 5C and 6A **A.** Reb1p and the RSC are required for the activation of *HIS5*. Northern-blot analysis of *HIS5* RNAs. Depletion of either Reb1p or Sth1p affects expression of full length *HIS5*. **B.** Heatmap of the RNA signals derived from the *TIM23* locus revealed by tiling arrays. A Reb1p-dependent cryptic unstable transcript (*uTIM23* RUT) is detected upstream of *TIM23* in an *rrp6Δ* strain. Upon Reb1p depletion in glucose-containing medium, an extended transcript can be detected (extension marked by a red bar). The position of the Reb1p binding site is indicated by a red arrow.

Figure S6. Related to Figure 6 **A.** Heatmap of transcripts derived from the *YSY6/DEM1* locus showing that depletion of Reb1p induces a downregulation of *DEM1*. The position of the Reb1 site is indicated by a red arrow. Note that under these conditions the elongated transcripts generated from read-through at the *YSY6* locus (presumably *YSY6-DEM1* chimeric RNAs) cannot be detected because degraded by the NMD pathway in the cytoplasm (see part B and figure 6B). **B.** Reb1p-dependent termination functions as a fail-safe mechanism at the *YSY6* gene. When canonical CPF/CF termination is impaired in an *rna14-3* mutant (compare lanes 1-4 to 5-8), read-through transcripts cannot be observed even upon impairment of nuclear degradation in an *rrp6Δ* background (lanes 5 and 7). However, longer and heterogeneous read-through species (RT, indicated by a black bar) are clearly visible when Reb1p-dependent termination is also impaired by metabolic depletion of Reb1p (lane 8). **C.** Same blot as in figure 6B, hybridized with a probe specific for the extended *YSY6-DEM1* chimeric species (RT). The position of the probe used is shown on the scheme. **D.** Heatmap of transcripts derived from the *YDL233W/OST4* locus. Upon depletion of Reb1p, readthrough occurs at the *YDL233W* gene (not markedly visible in these conditions, see figure 6C) and the downstream gene *OST4* is downregulated. The Reb1p binding site is indicated by a red arrow. **E.** The integrity of Reb1p function is required for normal growth when the CPF/CF termination pathway is also partially impaired. Both the *reb1-ts1* and *rna14-3* alleles are hypomorphic and grow relatively well at the permissive temperature. The double mutant is strongly impaired. **F.** The function of Reb1p in termination is required for growth. *reb1-ts1 rna14-3* cells were transformed with plasmids expressing wild type Reb1p, a truncated form of Reb1p



containing only the DNA binding domain or Nsi1p. The DNA binding domain of Reb1p is sufficient to improve growth of *reb1-ts1 ma14-3* cells.

Figure S7. A. Metasite analysis of RNAPII pausing upstream of Reb1p binding sites in the genome. Related to Figure 5E. The plots show the profile of the frequencies of polymerase pausing peaks in the 200 positions preceding the reverse complement of the Reb1 site used in figure 5E. Analysis and labeling as in figure 5E. B. Statistical analysis of the extent of RNAPII pausing at the best 109 sites showing the highest Reb1p occupancy (top sites) or at all the sites in the genome (sequence sites) irrespective of occupancy. The RNAPII peak frequency at the nucleotide corresponding to the maximum of the profile (position -17 relative to the site) is plotted and compared with the distribution of values obtained from randomly sampling equivalent number of sites from the genome. Error bars in the random distribution correspond to the standard deviation generated by the simulation.

## **Supplemental Experimental Procedures**

### **RNA analysis**

Northern blot analyses were performed with standard procedures, using 5% acrylamide/7.5M urea or 1.2% agarose/0.67% formaldehyde gels. RNAs were transferred to Hybond N+ membrane (GE Healthcare) and probed with 5' end-labeled oligonucleotides or PCR fragments labeled by random priming (Megaprime kit, GE Healthcare). Hybridizations were performed in UltraHyb or UltraHyb-Oligo (Ambion) commercial buffers at 42°C. Analysis of the polyadenylation status of transcripts was performed either by cleaving the poly(A) tail with RNaseH (Invitrogen) and oligo(dT) or by affinity selection of poly(A)<sup>+</sup> species with oligo(dT)-Dynabeads (Invitrogen) according to the manufacturer instructions. 3' RACE were performed using a commercial kit (Ambion) according to the manufacturer instructions.

### **Chromatin Immunoprecipitation**

Chromatin immunoprecipitation was performed as previously described (Rougemaille et al., 2008). Immunoprecipitations were performed with anti-HA (F-7) or anti-Rpb1(Y-80) antibodies from Santa Cruz. Data presented are the average of at least three biological replicates and error bars represent standard deviations.

### **Analyses of the selected pool of sequences**

Sequences containing Reb1p-dependent terminators were first identified by visual detection of the Reb1p binding site in a low number of clones generated by manual sequencing. A larger pool was subsequently assembled from large scale sequencing data. The sequence logo was generated with Weblogo (<http://weblogo.berkeley.edu/logo.cgi>) using 83 sequences previously aligned with Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

Nucleotides flanking the core Reb1p binding sites were statistically analyzed using a subset of the selected sequences containing only the most highly represented clones (i.e. each clone representing at least 2% of the selected pool) and the total pool of naive sequences (1,431,308 unique sequences) to estimate background frequencies. Selected sequences were purged to prevent biases due to single nucleotide variants of the same clone (Thomas-Chollier et al., 2011). A total of 1025 sequences were present in the Reb1p binding sites enriched pool. The analysis was performed by extracting all the occurrences of the consensus core Reb1p binding site "TTACCCG", plus 4 nt on each side, from both the naive and the enriched pool (1828 and 45 sequences respectively). For each flanking position, we

plotted in figure S1A the log<sub>2</sub> ratio of the frequency of each nucleotide in the enriched versus the naive pool.

#### ***In vitro* analysis of Reb1p nucleic acid binding and function in termination**

rReb1 and rReb1-DBD were produced and purified as previously described (Porrua and Libri, 2013). Electrophoretic mobility shift assays (EMSA) were performed using recombinant proteins and 5'-labelled probes of same sequence (5'-ATGATGACCTAGATGTTACCCGGGTAAAGAGCCCCATTAT-3') in double stranded DNA, single stranded DNA, or RNA forms. The mutant probe has the sequence (5'-ATGATGACCTAGATaccgaagccGTAAAGAGCCCCATTAT-3'). *In vitro* transcription termination assays were performed essentially as previously described (Porrua and Libri, 2013).

#### **Transcriptome analyses by tiling microarrays**

RNAs for tiling arrays hybridizations were prepared from strains containing the *P<sub>GAL1</sub>-REB1* construct (or the endogenous *REB1* gene as a control) after 2 hours of growth in glucose to minimize indirect effects. At this time point Reb1p depletion induces a clear termination defect (Figure S1C), but no marked effects on growth can be observed. Hybridizations and analyses were performed as previously described (Xu et al., 2011). Briefly, total RNA was treated with RNase-free DNaseI using Turbo DNA-free kit (Ambion). For first-strand cDNA synthesis, 20 µg of total RNA was mixed with 1.72 µg of random hexamers, 0.034 µg of oligo(dT) primer and incubated at 70°C for 10 min followed by 10 min at 25°C, then transferred on ice. The synthesis included 2,000 units of SuperScript II Reverse Transcriptase, 50 mM TrisHCl, 75 mM KCl, 3 mM MgCl<sub>2</sub>, 0.01 M DTT, dNTP + dUTP mix (0.5 mM for dCTP, dATP and dGTP; 0.4 mM for dTTP and 0.1 mM for dUTP, Invitrogen), 20 µg/mL actinomycin D in a total volume of 105 µL. The reaction was carried out in 0.2 mL tubes in a thermal cycler with the following thermal profile: 25°C for 10 min, 37°C for 30 min, 42°C for 30 min followed by 10 min at 70° for heat inactivation and 4°C on hold. Samples were then subjected to RNase treatment of 20 min at 37°C (30 units RNase H, Epicentre, 60 units of RNase Cocktail, Ambion). First-strand cDNA was purified using the MinElute PCR purification kit (Qiagen) and 5 µg were fragmented and labeled using the GeneChip WT Terminal labeling kit (Affymetrix) according to manufacturer's protocol. The labeled cDNA samples were denatured in a volume of 300 µl containing 50 pM control oligonucleotide B2 (Affymetrix) and Hybridization mix (GeneChip Hybridization, Wash and Stain kit, Affymetrix) of which 250 µl were hybridized per array (*S. cerevisiae* yeast tiling array, Affymetrix, PN 520055). Hybridizations were carried out at 45°C for 16 h with 60 rpm rotation. The staining was carried out using the GeneChip Hybridization, Wash and Stain kit with fluidics protocol FS450\_0001 in an Affymetrix Fluidics station.

Table of strains used in this study

strain	genotype	Source/reference
W303	<i>ura3-1, ade2-1, his3-11,5, trp1-1, leu2-3,112, can1-100</i>	(Thomas and Rothstein, 1989)
DL671 BMA64	<i>ura3-1, ade2-1, his3-11,5, trp1Δ, leu2-3,112, can1-100</i>	(Baudin et al., 1993)
DLY678	as BMA64, MATa, <i>trf4::KAN<sup>R</sup>, rrp6::URA3</i>	
DLY815	as BMA64, MATa, <i>rrp6::KAN<sup>R</sup></i>	
DLY885	as W303, MATa, <i>HIS::P<sub>GAL1</sub>-NRD1, rrp6::KAN<sup>R</sup></i>	(Thiebaut et al., 2006)
DLY1582	as W303, MATa, <i>hsp104::LEU CUP1::KAN<sup>R</sup></i>	
DLY1583	as W303, <i>hsp104::LEU cup1::KAN<sup>R</sup> rrp6::KAN<sup>R</sup></i>	
DLY1650	as W303, <i>HIS::P<sub>GAL1</sub>-NRD1, rrp6::KAN<sup>R</sup></i>	
DLY1905	as BMA64, <i>reb1-ts1</i>	This study
DLY1912	as BMA64, MATa, <i>HIS::P<sub>GAL1</sub>-HA-REB1</i>	This study
DLY1914	as BMA64, MATa, <i>HIS::P<sub>GAL1</sub>-HA-REB1 rrp6::KAN<sup>R</sup></i>	This study
DLY1968	as BMA64, <i>nsi1::URA3kl rrp6::KAN<sup>R</sup></i>	This study
DLY1970	as BMA64, <i>HIS::P<sub>GAL1</sub>-HA-REB1 nsi1::URA3kl, rrp6::KAN<sup>R</sup></i>	This study
DLY1986	<i>reb1-ts1 ma14-3</i>	This study
DLY2109	<i>HIS::P<sub>GAL1</sub>-HA-REB1 ma14-3</i>	This study
DLY2110	<i>HIS::P<sub>GAL1</sub>-HA-REB1, ma14-3, rrp6::KAN<sup>R</sup></i>	This study
DLY2210	as BMA64, MATa, <i>HIS::P<sub>GAL1</sub>-HA-REB1, dis3-D551N-protA::TRP1kl</i>	This study
DLY2213	as BMA64, <i>HIS::P<sub>GAL1</sub>-HA-REB1, his5::LEU2cg, rrp6::URA3</i>	This study
DLY2240	as W303, <i>cul3::TRP</i>	
DLY2241	W303	
DLY2242	as W303, <i>rsp5-1::HIS</i>	(Harreman et al., 2009)
DLY2275	<i>HIS::P<sub>GAL1</sub>-HA-REB1, rsp5-1::HIS</i>	This study
DLY2304	<i>HIS::P<sub>GAL1</sub>-HA-REB1, upf1::TAP::TRP1kl</i>	This study
DLY2305	<i>HIS::P<sub>GAL1</sub>-HA-REB1, upf1::TAP::TRP1kl, dis3-D551N-protA::TRP1kl</i>	This study
DLY2351	<i>HIS::P<sub>GAL1</sub>-HA-STH1, rrp6::KAN<sup>R</sup></i>	This study

Table of plasmids used in this study

pDL431	pCM190(TRP1)- <i>P<sub>TET</sub></i> -HSP104- <b>X3</b> -HSP104- <i>P<sub>GAL1</sub></i> -LACZ
pDL438	pCM190(TRP1)- <i>P<sub>TET</sub></i> -HSP104-“ <b>GTTACCCGG</b> ”-HSP104- <i>P<sub>GAL1</sub></i> -LACZ
pDL457	pCM190(URA3)- <i>P<sub>TET</sub></i> -HSP104- <b>X9</b> -HSP104- <i>P<sub>GAL1</sub></i> -CUP1
pDL459	pCM190(URA3)- <i>P<sub>TET</sub></i> -HSP104- <b>X18</b> -HSP104- <i>P<sub>GAL1</sub></i> -CUP1
pDL460	pCM190(URA3)- <i>P<sub>TET</sub></i> -HSP104- <b>X20</b> -HSP104- <i>P<sub>GAL1</sub></i> -CUP1
pDL641	pCM190(URA3)- <i>P<sub>TET</sub></i> -HSP104- <b>X25</b> -HSP104- <i>P<sub>GAL1</sub></i> -CUP1
pDL642	pCM190(URA3)- <i>P<sub>TET</sub></i> -HSP104- <b>X28</b> -HSP104- <i>P<sub>GAL1</sub></i> -CUP1
pDL643	pCM190(URA3)- <i>P<sub>TET</sub></i> -HSP104- <b>X31</b> -HSP104- <i>P<sub>GAL1</sub></i> -CUP1
pDL513	pCM188
pDL537	pCM188- <i>P<sub>REB1</sub></i> - <b>REB1</b>
pDL538	pCM188- <i>P<sub>REB1</sub></i> - <b>reb1-DBD</b>
pDL539	pCM188- <i>P<sub>REB1</sub></i> - <b>NS1</b>
pDL550	pCM190(URA3)- <i>P<sub>TET</sub></i> -HSP104-“ <b>GTTACCCGG</b> ”-HSP104- <i>P<sub>GAL1</sub></i> -CUP1
pDL551	pCM190(URA3)- <i>P<sub>TET</sub></i> -HSP104- <b>X3mut</b> -HSP104- <i>P<sub>GAL1</sub></i> -CUP1
pDL552	pCM190(URA3)- <i>P<sub>TET</sub></i> -HSP104- <b>X3rc</b> -HSP104- <i>P<sub>GAL1</sub></i> -CUP1
pDL665	pCM190- <b>HIS5</b>
pDL572	pCM188- <i>P<sub>TET</sub></i> -HSP104- <b>X3</b> -HSP104- <i>T<sub>CYC1</sub></i>
pDL573	pCM188- <i>P<sub>TET</sub></i> -HSP104- <b>X3mut</b> -HSP104- <i>T<sub>CYC1</sub></i>

Table of oligonucleotides used in this study

name	Sequence 5'-3'	Gene	use
DL190	TTGAGCCAACGTCAAAATCGTTAGAGCCCTTCTGTAAATT GCGTTTGGTCGTTTCAT	<i>HSP104</i>	probe PNK
DL275	ATCTCTTTGTAAAACGGTTCATCC	<i>U6</i>	probe PNK
DL377	ATGTTCCCAGGTATTGCCGA	<i>ACT1</i>	PCR (Klenow probe)
DL378	ACACTTGTGGTGAACGATAG	<i>ACT1</i>	
DL751	TTTCCCAGAGCAGCATGACT	<i>CUP1</i>	probe PNK
DL1198	GCGGAGATAACTCCAAGTTAT	<i>HSP104</i>	qPCR
DL1199	TAGAGCAAACAATATATGGTC	<i>HSP104</i>	qPCR
DL1202	AAGGACGACGCTGCTAACATC	<i>HSP104</i>	qPCR
DL1203	AGATCTATATTCGTTATTGGT	<i>HSP104</i>	qPCR
DL1359	CCTTATACATTAGGTCTTT	<i>P<sub>TET</sub></i>	qPCR
DL1360	ATCCCCGAATTGATCCGG	<i>P<sub>TET</sub></i>	qPCR & RNase H
DL2812	GAATGCGGTGACCGGTATCG	<i>HSP104-Reb1-BS</i>	qPCR
DL2813	CCTATACATATTCTAGCTGCC	<i>HSP104-Reb1-BS</i>	qPCR
DL1566	AGTTGATCGGACGGGAAAC	<i>5S</i>	probe PNK
DL2321	TGCTAAACCAGCAGTACGTGTGTGCCG	<i>HIS5</i>	PCR (Klenow probe)
DL2322	GGTCTTGCTCAATTCAACTGGAG	<i>HIS5</i>	
DL2323	ATAGCAGGGTGACAGGGTCG	<i>YSY6</i>	PCR (Klenow probe)
DL2324	ACACCACCACTACGAGAAG	<i>YSY6</i>	
DL2325	TGCAAAGGACCAAAGACAGC	<i>uATP5</i>	PCR (Klenow probe)
DL2326	CCAGCACCGTCTCTCGAAAG	<i>uATP5</i>	
DL2370	CGGCACACACGTAAGTCTGG	<i>HIS5</i>	RNase H
DL2668	CGTCGTGGTGCGAGGAGAGATACCGCTGTTGCGTCGCCT CTTTGCGTGCAAGGGC	<i>YDL233W</i>	probe PNK
DL2669	GGAGACATGGTGGAGTCAACAGCATGG	<i>OST4</i>	probe PNK
DL2840	TGGGCCGGGCACTAATCAAC	<i>DEM1</i>	PCR (Klenow probe)
DL2841	TCGCTCAACAATGCTCCTTCC	<i>DEM1</i>	
DL2910	CGCATCCATAAATGACAACGCGGGTAATACTGCC	<i>YSY6-RT</i>	probe PNK
DL2503	GGCTAACGCGTGGTGCGAATGACCAGGCCAGTATGGAAA TTGATGATGACCTAGATTATAGTTCATTGTGTT <b>ACCCGG</b> AC TCAGAGCC	Non-template strand of IVT substrate containing a Reb1 binding site (bold)	
DL2504	GGCTCTGAGT <b>CCGGGT</b> AACACAATGAAGTATAATCTAGGT CATCATCAATTTCCATACTGGCCTGGTCATTGCACACAG CGT	Template strand of IVT substrate containing a Reb1 binding site (bold)	
DL2492	UGCAUUUCGACCAGGC	Fluorescently 5' labeled RNA oligonucleotide for promoter- independent assembly of elongation complex (IVT)	

### Supplemental References

- Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 21, 3329–3330.
- Harreman, M., Taschner, M., Sigurdsson, S., Anindya, R., Reid, J., Somesh, B., Kong, S.E., Banks, C.A.S., Conaway, R.C., Conaway, J.W., et al. (2009). Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 20705–20710.
- Porrua, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat. Struct. Mol. Biol.* 20, 884–891.
- Rougemaille, M., Dieppois, G., Kisseleva-Romanova, E., Gudipati, R.K., Lemoine, S., Blugeon, C., Boulay, J., Jensen, T.H., Stutz, F., Devaux, F., et al. (2008). THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association. *Cell* 135, 308–321.
- 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, 853–864.
- Thomas, B.J., and Rothstein, R. (1989). Elevated recombination rates in transcriptionally active DNA. *Cell* 56, 619–630.

## 7.2 Genomewide Analysis of Road-Block Termination

Following the publication of Colin *et al.*, I focused on the possibility that other DNA binding proteins might be effectors of road-block termination *in vivo*. Besides Reb1, Rap1 had also been identified as a possible road-blocking factor from previous experiments. This led me to investigate the family of General Regulatory Factors, which resulted in the identification of several other possible road-blocking factors belonging to this class, such as Abf1. Using a combination of already published datasets and newly generated data, I applied meta-gene analyses and other computational techniques to identify genomic loci associated with polymerase pausing and termination events.

In this work, Jean-Baptiste Briand and Jessie Colin performed northern blot, EMSA (Electro Mobility Shift Assay), and RT-qPCR experiments. Drice Challal performed all the CRAC experiments. Jessie Colin performed the RNA-seq experiments. I processed all datasets and performed all the computational analyses.



### 7.2.1 Introduction

The compact genome of *S.cerevisiae* is covered by several machineries that need to be temporally and spatially coordinated to limit interferences that might affect robust reading and perpetuation of the genetic information. Transcription itself best exemplifies the complexity of the genomic landscape. Transcription initiation occurs frequently in regions and direction that largely overrun the annotations of genes with an assigned function [34, 197]. This is believed to be due to a leaky control on initiation and to the general of bi-directionality of promoters, which is also generally conserved in evolution. Transcription units largely overlap in both sense and antisense direction, and although RNA polymerases II (RNAPII) only seldom collide [202], the chromatin marks associated with ongoing transcription persist, and are susceptible to considerably impact concurrent transcription events. Overlapping transcription has also a large potential for regulation of gene expression, and is sometimes controlled and tamed to the need of the cell [116].

The pervasive nature of transcription brings about two main potentially perturbing elements: the first is the presence of transcribing RNA polymerases which might directly affect other DNA-related events; the second is the production of many non-coding RNA molecules that might titrate RNA-binding factors and indirectly affect gene expression. Cells possess tools to control both, by terminating “spurious” transcription events and degrade a large fraction of the RNA produced. In this perspective, transcription termination and RNA degradation, besides being devoted to the production of functional RNAs, additionally qualify as quality control mechanisms [for review see 181].

In yeast, two main pathways of termination exist. The first is operated by a complex called the Cleavage and Polyadenylation Factor-Cleavage Factor (CPF-CF) and is used to arrest transcription of mRNA coding genes. The CPF-CF complex recognizes signals on the nascent RNA and cleaves it, producing a 5' fragment that is polyadenylated by the Pap1 poly(A) polymerase and exported to the cytoplasm. The 3'-fragment still associated to the transcribing polymerase is recognized and

degraded by a 5'→3' exonuclease, Rat1, which contributes to dismantling the elongation complex by a much discussed but still unclear mechanism [88]. The CPF-CF is also believed to be directly involved in termination by allosterically modifying the properties of the transcription elongation complex [205]. The second canonical pathway is dependent on the NNS (Nrd1-Nab3-Sen1) complex and was traditionally associated to the production of sn- and snoRNAs [89]. Nrd1 and Nab3 bind the RNA at short motifs containing a well-conserved 4-5 nucleotides core [23] and are thought to recruit Sen1 that translocates on the nascent RNA to release the polymerase by a mechanism that remains unclear [143]. Peculiar to this pathway is the treatment of the RNA released, that is polyadenylated by a different poly(A) polymerase, Trf4, functioning within the TRAMP4 (Trf4-Air2-Mtrf4-Polyadenylation) complex, and trimmed to its mature size in the nucleus by the exosome, a large multisubunit complex that is endowed with 3' to 5' exonuclease activities [185].

A large share of the transcripts produced by pervasive transcription do not code for proteins and to what extent these RNAs have specific functions remains matter of debate. They are sorted in classes, generally defined by the pathways associated to their metabolism. CUTs (Cryptic Unstable transcripts) have been first described based on their extreme instability [196]. These RNAs derive from transcription events terminated by the NNS pathway and are degraded to completion by the TRAMP-exosome pathway. When NNS termination is defective, elongated forms of CUTs are produced that are presumably terminated downstream by the CPF-CF pathway because they are insensitive to nuclear, exosomal degradation. These elongated forms of CUTs have been more recently named NUTs (Nrd1 Unterminated Transcripts) [166]. Some of the non-coding RNAs produced by pervasive transcription are sufficiently stable to be detected in wild type cells (SUTs, stable unannotated transcripts [34]) or are degraded in the cytoplasm by the nonsense-mediated decay (NMD) and Xrn1 pathways (XUTs, Xrn1-sensitive Unstable Transcripts [183]). Finally, some are only detected in particular physiological conditions (MUTs, meiotic unannotated transcripts [102]).

We have recently described an additional pathway of transcription termination that depends on the DNA-binding protein Reb1 and that was dubbed road-block (RB) termination [29]. The elongating polymerase was shown to pause upstream of DNA-bound Reb1, which provokes its release by a mechanism that involves its ubiquitylation and presumably degradation. The isolated binding site of Reb1 was shown to be sufficient for eliciting termination when inserted in regions of active elongation, indicating that additional sequence elements are not required for efficient RB termination. Because, akin to CUTs, the RNAs released are polyadenylated by TRAMP and degraded by the nuclear exosome, these transcripts were dubbed RUTs (Reb1-dependent Unstable Transcripts).

In this report we demonstrate that several DNA-binding factors or complexes are able to terminate transcription by a RB mechanism. We generated high-resolution data on the distribution of RNAPII upon depletion of RB factors to address the significance and extension of RB termination at the genomewide scale. We demonstrate that prominent peaks of roadblocked polymerases accumulate in intergenic regions immediately downstream of canonical terminators, indicating the significant occurrence of transcriptional readthrough in wild type cells. Akin to the leaky control on transcription initiation, the constitutive failure to terminate efficiently generates an additional level of pervasive transcription that has the potential to strongly affect the function of downstream regulatory regions or other DNA associated events. We show that RB and canonical termination pathways are not dependent on each other. High resolution analyses of RNAPII occupancy upon affecting either RB or CPF-CF and NNS termination indicates that RB is unlikely to partake in canonical termination and, conversely, that NNS and CPF-CF pathways are unlikely to be involved in RB termination. Rather, RB termination plays an important quality control role in limiting pervasive transcription events due to termination failure. The faculty of DNA associated factors to alter the processivity of elongation complexes, and the widespread occurrence of these factors defines a large potential in shaping and regulating the transcriptome. We propose that road-block termination constitutes an additional, general level of control on

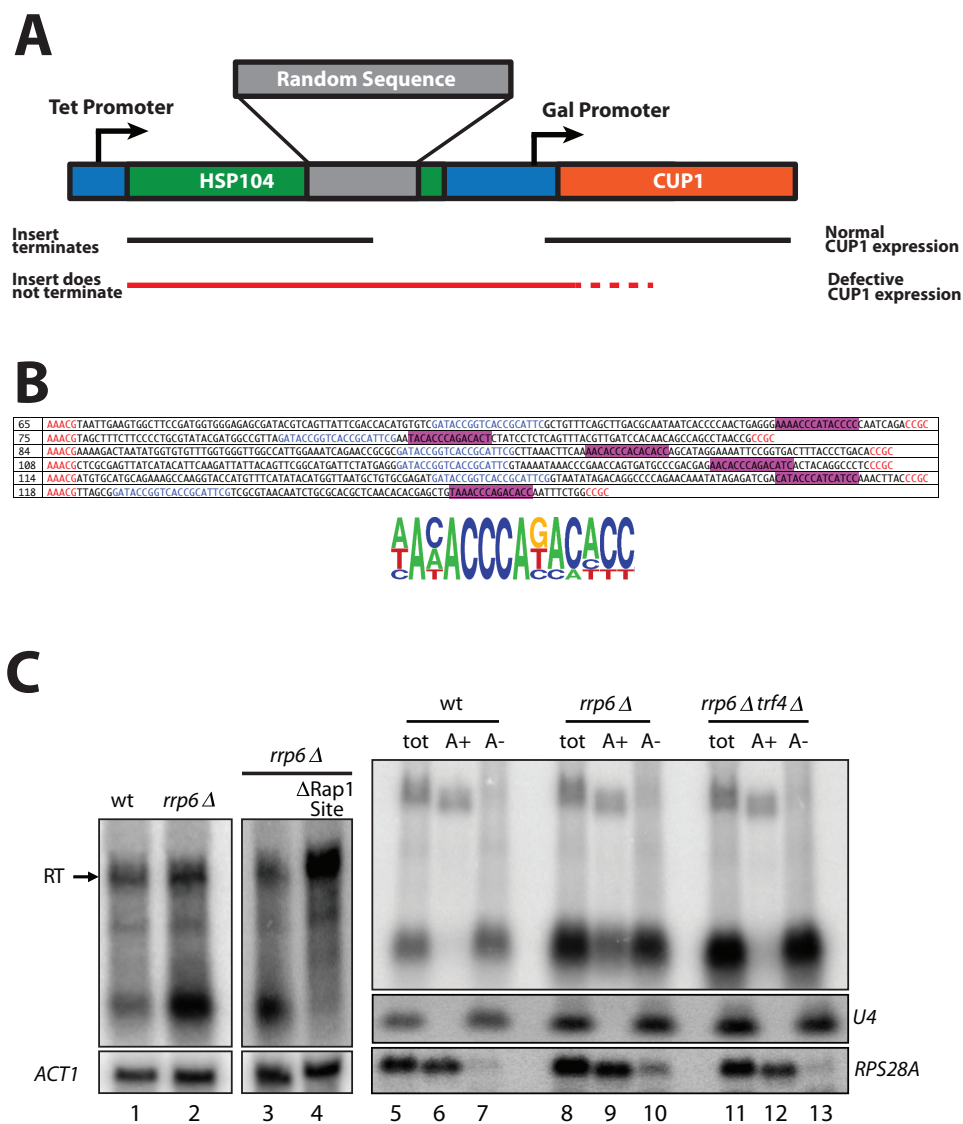
transcription that operates at the post-initiation level by altering the efficiency and extent of RNAPII elongation.

## 7.2.2 Results

### ***In Vivo* Selection Reveals Rap1-Dependent Transcription Termination**

We have previously described a procedure to select transcription terminators from pools of naïve sequences [142]. Briefly, test sequences are inserted within a transcription unit driven by the tetracycline-repressible (TetP) promoter, roughly 200nt downstream of the transcription start site. A second promoter from the GAL1 gene is inserted downstream and drives expression of a selectable marker, CUP1, the expression of which is required for yeast growth in copper-containing medium. In the absence of a terminator in the test sequences, transcription driven from TetP silences the GAL1 promoter by transcription interference and prevents CUP1 expression, which leads to copper-sensitivity. When the test sequence induces termination, the CUP1 gene is expressed and yeasts grow on copper-containing plates (Fig. 7.1A). Using this system we selected terminators from a pool of sequences containing a stretch of 120 random nucleotides. We selected many sequences inducing termination via the NNS pathway and via the Reb1-dependent road-block pathway. We also selected sequences that do not belong to either class, some of which contain a motif resembling a Rap1 binding site (Figure 7.1B). Rap1 recognizes its site via a Myb-like DNA-binding domain and is involved in many DNA-associated processes, including telomere maintenance and gene expression. Rap1 is also strongly associated to the positioning and formation of nucleosome free regions (NFR).

It has previously been shown that the presence of a Rap1 binding site can induce RNAPII stalling in a model Ty1 retrotransposon construct [199]. In this study, the occurrence of Rap1-dependent transcription termination was ruled out based on the analysis of the transcripts produced in the presence of the Rap1 site. These



**Figure 7.1:** **A:** Schematic representation of the reporter system used to select Rap1-terminated transcripts. **B:** Sample of selected sequences containing Rap1 sites, the identified consensus is represented. **C:** Northern blot analysis of several species derived from the reporter system.

RNAs were non-adenylated and insensitive to nuclear degradation, and therefore assumed to be nascent RNAs associated to the stalled polymerase. Moreover, it was not demonstrated that stalling is dependent on the integrity of Rap1 or its binding to the DNA. The stalling model would hardly be compatible with our results, because only loss of polymerases – and therefore termination – is expected to prevent transcription interference. We therefore assessed whether the presence of the selected site would induce Rap1-dependent transcription termination. We first demonstrated that the Rap1 binding site is necessary and sufficient to prevent transcription interference at the GAL1 promoter. Indeed, mutation of the site in the context of a selected clone prevented yeast growth on copper, while insertion of the site in a fragment of the coding region of the HSP104 gene was sufficient to induce copper resistance (data not shown).

These results were confirmed by direct analysis of the RNA produced. To assess whether the transcripts released undergo nuclear degradation, we analyzed the RNAs in both a wt and degradation-defective *rrp6* $\Delta$  and *trf4* $\Delta$  strains. As shown in figure 7.1C, a short RNA is produced when a selected terminator is present in the reporter construct. For all of the terminators analyzed, the size of this RNA is 13-17 nt shorter than the distance between the transcription start site and the Rap1 site (data not shown), suggesting that stalling or release of the polymerase occurs upstream of the site, which is consistent with a road-block mechanism.

The transcripts produced are strongly sensitive to degradation, as indicated by their marked steady state increase when the analysis is performed in a  $\Delta$ *rrp6* exosome mutant (Figure 7.1C, lanes 1-2). This indicates that these transcripts cannot solely correspond to polymerase-associated nascent RNAs but rather that they are released upon transcription termination. The short transcript disappears to the profit of a longer, read-through product when the Rap1 site is deleted (compare lanes 3 and 4). The bulk of the transcripts released and degraded appears to be non-adenylated (Figure 7.1C, compare lanes 7, 10 and 13), although a fraction is polyadenylated by Trf4 (compare lanes 9 and 12). The transcripts that are detected in a wild type

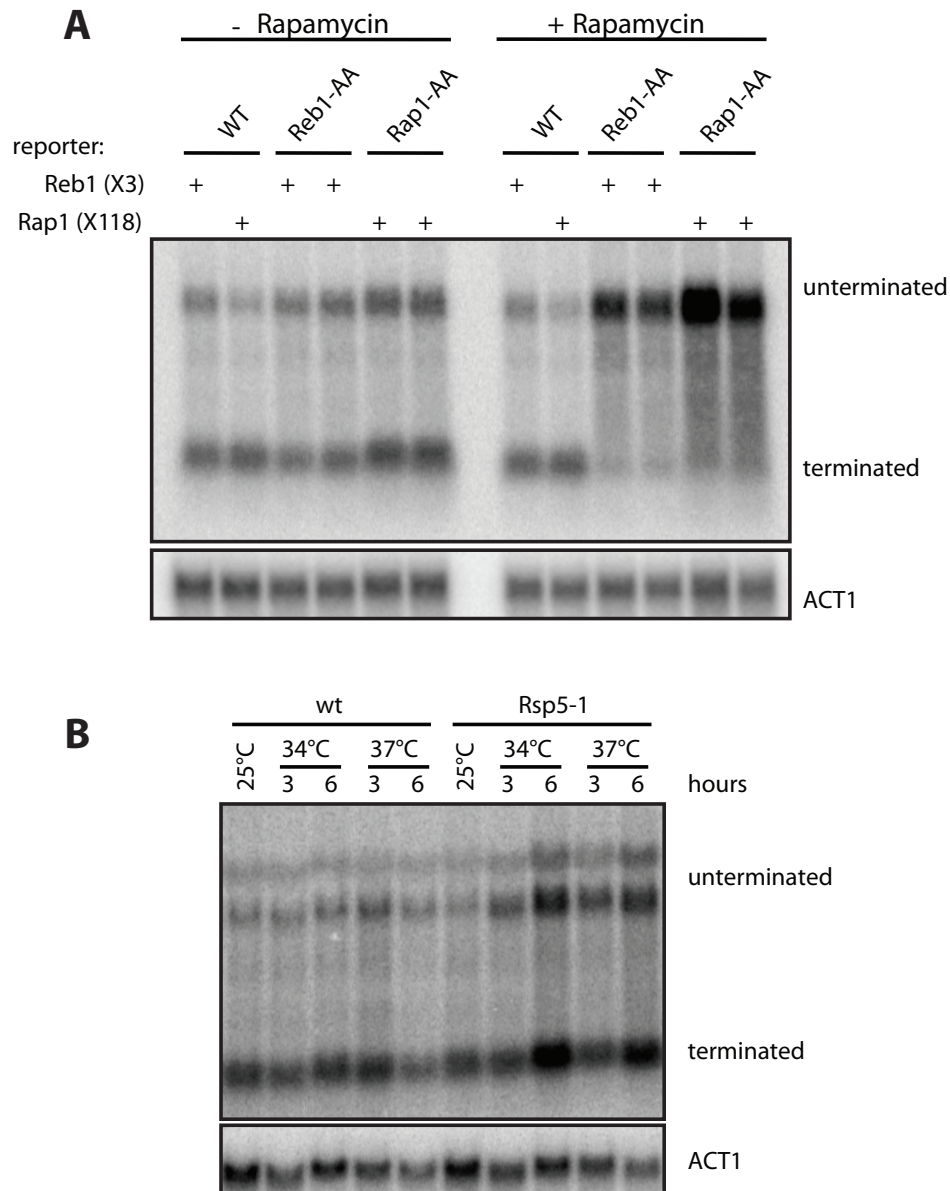
strain are non-adenylated (lanes 5-7) and might correspond to nascent RNAs that are protected from degradation because of their association with the polymerase.

The dependency on the Rap1 site strongly suggests, but does not prove that Rap1 is involved in termination. Indeed, termination might occur via other pathways, e.g. as a result of the recognition of partially or fully overlapping termination signals at the Rap1 site. To prove the Rap1 dependency, we transiently depleted this essential factor with the anchor away strategy and analyzed the transcripts produced. As shown in figure 7.2A, the levels of the short RNA derived from the reporter construct are markedly decreased in the absence of Rap1, to the profit of a longer species earmarking termination at a downstream site. From this result we conclude that Rap1 is necessary to induce termination at the selected sites.

Finally, we have previously shown that release of the road-blocked polymerase from the DNA template occurs following its ubiquitylation that depends on the Rsp5 ubiquitin ligase. When the elongation complex is dismantled, the RNA released is polyadenylated and degraded rapidly; conversely, the persistence of roadblocked RNAPII on the DNA template following mutation of Rsp5 leads to an increase of the nascent, non-adenylated transcript that can be detected in a wild type strain [29]. Northern blot analysis confirmed the expected increase in the levels of nascent RNAs when the Rap1-roadblocked polymerase is less efficiently removed in a thermosensitive *rsp5-1* mutant strain (Fig. 7.2B).

This finding is also substantiated by the observation that recombinant Rap1 binds very efficiently the double stranded DNA but not the RNA or single stranded DNA version of its site (supplementary figure 7.1).

Together, these results demonstrate that transcription termination occurs by a road-block mechanism at sites bound by Rap1.



**Figure 7.2:** **A:** Northern blot analysis in presence or absence of Reb1 and Rap1 of specific sequences containing Reb1 or Rap1 terminators. **B:** Northern blot analysis of Rap1-terminated transcripts in a strain containing thermosensitive ubiquitin ligase Rsp5.

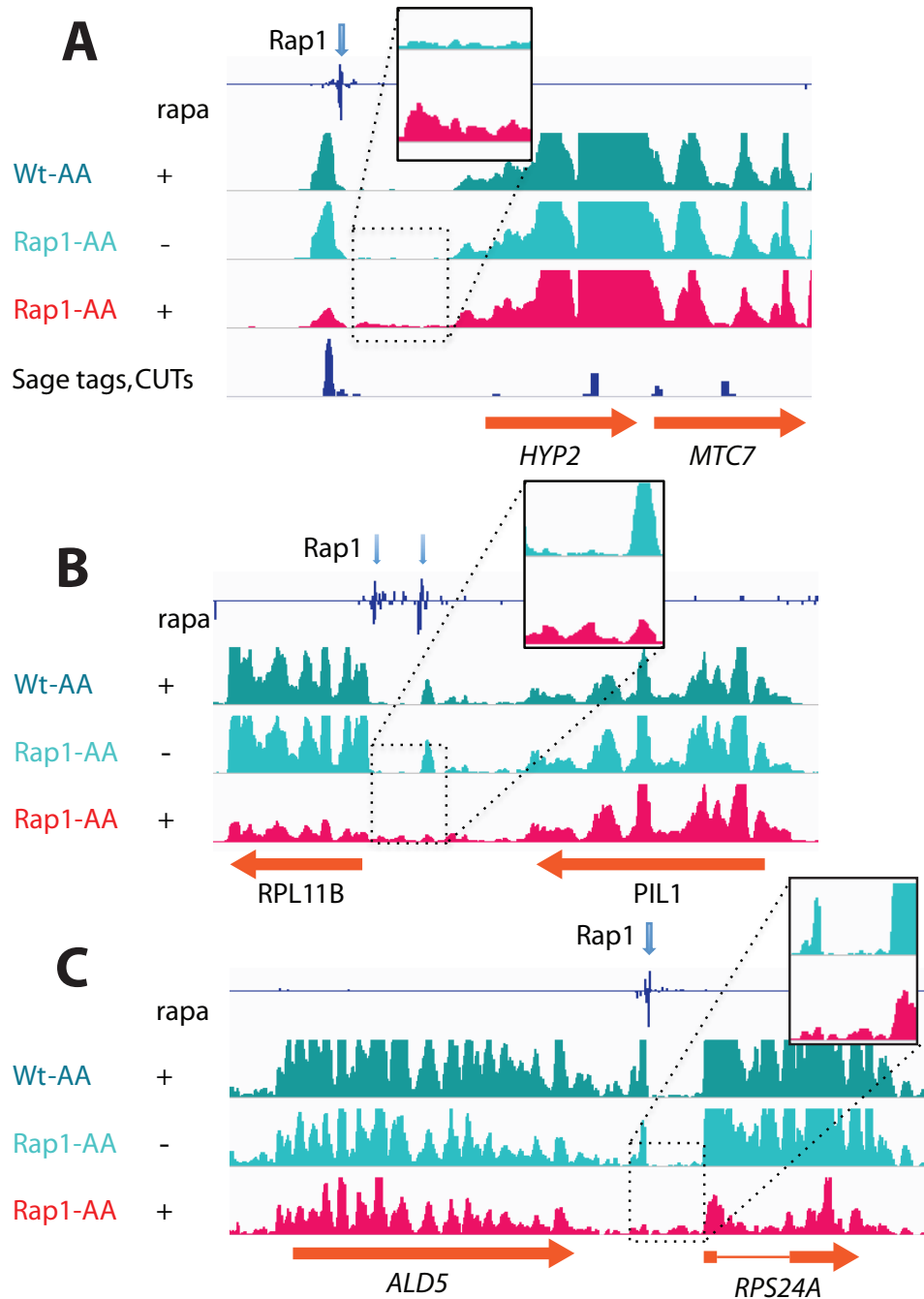


## Rap1-Dependent Transcription Termination in the *S.cerevisiae* Genome

These results constitute the proof-of-principle that transcription termination can occur in a Rap1-dependent manner, but do not prove that it occurs significantly in the *S.cerevisiae* genome. A hallmark of road-block termination is the accumulation of RNAPII immediately upstream of the site of road-block, due to polymerase pausing.

We therefore assessed whether RNAPII pausing can be observed in the *S.cerevisiae* genome at Rap1 sites and whether pausing would be dependent on Rap1. To this end, we analyzed the RNAPII distribution in a wild type and a Rap1 anchor away (Rap1-AA) strain by a modified crosslinking and cDNA analysis (CRAC) method [13, 61]. By this approach, the position of the polymerase is directly inferred by sequencing the nascent transcript associated to the largest subunit of the enzyme after *in vivo* UV crosslinking [122]. Consistent with the notion that the signals obtained genuinely represent nascent and not mature transcripts, intronic regions where largely covered in the RNAPII CRAC dataset but not in the sequencing of mature, total RNAs (supplementary figure 7.2).

In the Rap1-AA strain, Rap1 is rapidly and efficiently depleted from the nucleus upon addition of rapamycin [67]. Notable examples of sites of Rap1-dependent road-block sites are shown in figure 7.3. Two Rap1 binding site are present upstream of the HYP2 gene and constitute a prominent site of Rap1 localization as detected by several techniques [93, 154]. CRAC analysis reveals a prominent accumulation of the RNAPII signal immediately upstream of the Rap1 sites, indicating pausing. The occurrence of termination is demonstrated by the existence of a non-annotated unstable transcript ending in correspondence of the RNAPII peak, revealed by microarray analysis [128] and by a cluster of 3'-end SAGE. RNAPII pausing and termination were Rap1-dependent, because depletion of Rap1 led to a strong reduction in the RNAPII peak and to the appearance of a readthrough signal downstream of the site (Fig. 3A, inset). Finally, insertion of the two Rap1 sites in



**Figure 7.3:** Examples of Rap1-dependent termination detected *in vivo* through the CRAC technique. RNAPII occupancy signal accumulates upstream of Rap1 binding sites in wild type. This occupancy peak is reduced upon the addition of rapamycin to the Rap1-AA strain.

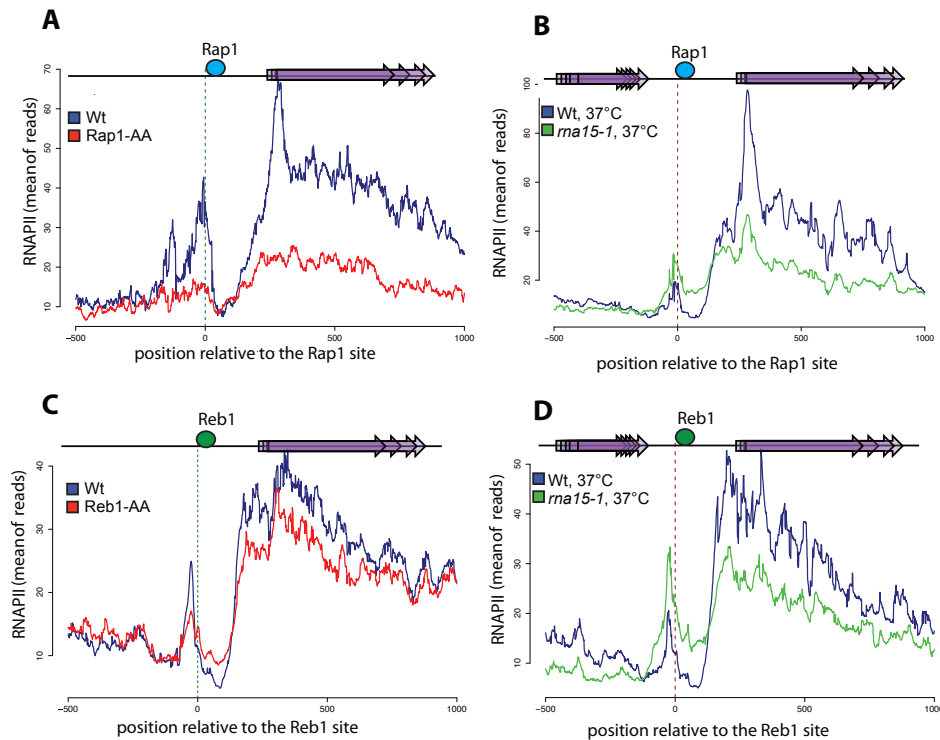
the heterologous context of our reporter system induced Rap1-dependent termination and led to the production of an unstable RNA (supplementary figure 7.3).

Two other examples are shown in figure 7.3B-C. In these cases, the Rap1 occupancy site is located between two tandem genes and the accumulation of RNAPII is most likely due to transcription events reading through the upstream terminator (see below). In both cases, depletion of Rap1 leads to abrogation of the peak and increased RNAPII signals downstream of the site (Fig. 7.3B-C, insets).

To extend these results to a genomewide perspective we profiled the average distribution of the RNAPII CRAC signal around aligned sites of Rap1 occupancy found in promoter regions.

Rap1 is required for the strong expression of ribosomal protein (RP) genes, and is often positioned in nucleosome free regions (NFRs) upstream of these genes. Consistently, a major peak of RNAPII occupancy is observed downstream of the aligned Rap1 binding sites, corresponding to the occurrence of transcription initiation within a relatively short window (Fig. 7.4A and B). Importantly, however, a significant peak demonstrating RNAPII pausing is also observed upstream of Rap1 binding, which is associated to the occurrence of transcription termination in the same region (see below). Importantly, sequestering Rap1 out of the nucleus led to a significant decrease in the RNAPII pausing peak demonstrating that Rap1 dependent road-block occurs at many sites of Rap1 binding in the genome.

Similar RNAPII CRAC analyses were also performed upon Reb1 depletion (Fig. 7.4C and D). Peaks of RNAPII pausing were readily observed at individual sites of Reb1 occupancy that disappeared upon Reb1 depletion (supplementary figure 7.4 and data not show). Because Reb1 is also required for the expression of many genes, profiling RNAPII distribution around aligned sites of Reb1 occupancy revealed a similar transcription initiation peak as for Rap1. Importantly, a prominent peak indicating RNAPII pausing was also observed upstream of Reb1 that strongly decreased upon sequestering Reb1 out of the nucleus. Overall, these results demonstrate



**Figure 7.4:** Metagene analysis around sites of Rap1 and Reb1. **A:** Global view of polymerase occupancy around sites of Rap1 residing in promoter regions. in Rap1-AA, the pause associated with the site decreases significantly. **B:** Global view of polymerase occupancy around sites of Rap1 that follow a CPF-terminated transcript. Upon inactivation of *rna15*, the readthrough from the upstream transcripts accumulates in front of the road-block site. **C and D:** The same analyses were performed with Reb1 sites.

the significant occurrence of Rap1- and Reb1-dependent, road-block transcription termination in *S.cerevisiae*.

## Widespread Redundancy in Transcription Termination

In the compact *S.cerevisiae* genome, efficient and timely release of the elongation complex is essential to prevent interference between contiguous transcription units. Whether CPF termination is inherently highly efficient or enforced by redundant mechanism remains unclear. Many sites of Reb1 and Rap1 occupancy are located in intergenic regions, downstream of genes terminated by the CPF pathway. If sig-

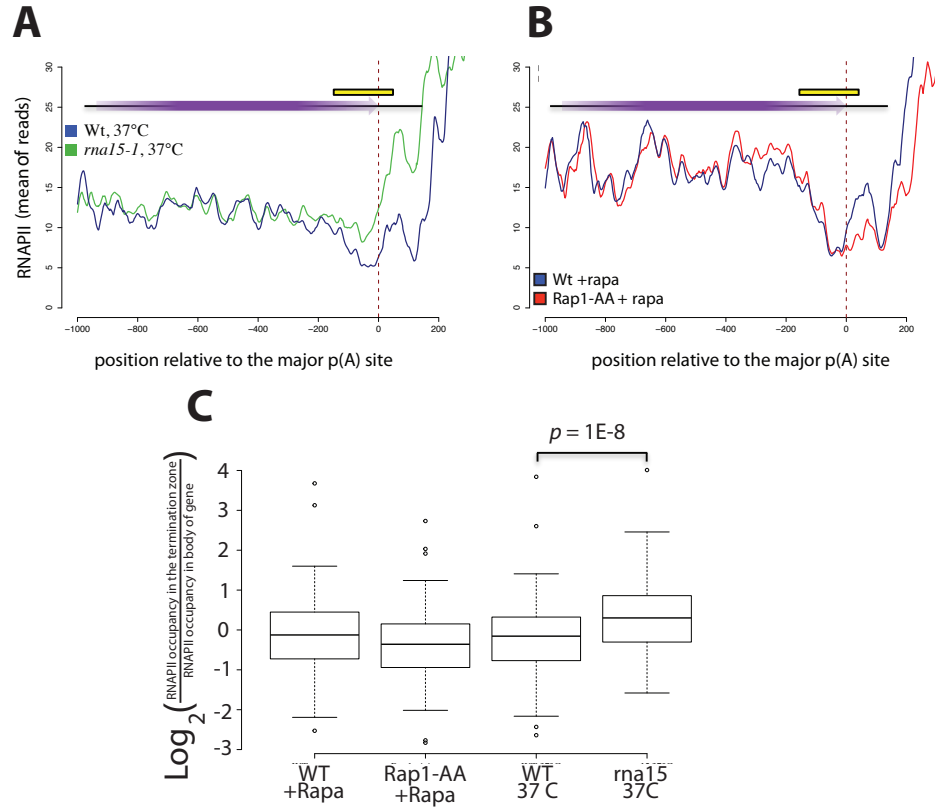
nificant transcriptional read through occurs at these CPF terminators, polymerases are expected to be roadblocked at downstream sites of Reb1 and Rap1 occupancy, as also suggested in the cases of PIL1 and ALD5 (Fig. 7.3B-C). We therefore restricted our metasite analyses to Reb1 and Rap1 occupancy sites located within 300nt downstream of mRNA-coding genes. In these conditions, only polymerases escaping termination (if any) are expected to contribute to the metaprofile observed. As shown in figure 7.4B and D, transcriptional road-block is clearly observed in the wild type strain at sites of Rap1 and Reb1 occupancy downstream of canonical CPF terminators. To prove that roadblocked polymerases indeed originate from readthrough at upstream terminators and not from spurious initiation between terminators and the road-block sites, we also performed a parallel RNAPII CRAC analysis using a thermosensitive *rna15-1* allele, which impairs CPF termination. A prominent increase in the road-block peak was clearly observed upon impairing CPF termination in the *rna15-1* mutant, consistent with the notion that the flux that alimments roadblocked polymerases originates from upstream transcription units and increases when upstream termination is defective. As a control, we profiled RNAPII distribution at the same set of genes using published PAR-CLIP data obtained upon nuclear depletion of Nrd1 [164], an essential actor of NNS termination that is not involved in termination of mRNA coding genes. In these conditions we did not observe an increase in the road-block peak (supplementary figure 7.5 and data not shown) confirming that roadblocked polymerases originate from upstream, CPF-dependent genes. Although less prominent, road-block was also observed at sites of Abf1 occupancy downstream of CPF terminators, which increased, as for Rap1 and Reb1, when termination was impaired in an *rna15-1* mutant (supplementary figure 7.6).

Overall, these results demonstrate the widespread occurrence of significant levels of transcription readthrough at CPF terminators in strains that are proficient for transcription termination. This results in the constitutive accumulation of roadblocked polymerases at sites of Rap1 and Reb1 occupancy (and many additional sites in the genome, see below).

## Roadblock is Not Part of the CPF Termination Mechanisms

Although these results strongly suggest that road-block termination acts to neutralize transcriptional readthrough downstream of CPF-dependent terminators, it cannot be excluded that roadblocking the polymerase is an important requirement for efficient termination at the upstream canonical sites. For instance, it can be envisioned that pausing induced by the road-block favors chasing of the polymerase by Rat1. In this perspective, it is expected that reducing the road-block should affect the efficiency of termination at the upstream sites. To address this possibility, we investigated whether increased transcriptional readthrough could be observed at CPF terminators in the absence (or strong reduction) of the downstream road-block. We analyzed the level of polymerase in the region immediately downstream of CPF terminators in Reb1 or Rap1 anchor away strains upon nuclear depletion of either factor. Three examples of CPF-dependent genes with a downstream road-block are shown in supplementary figures 7.7. In all occurrences, transcription termination occurred efficiently at the CPF sites even in the absence of the road-block as witnessed by a very similar RNAPII signal at and downstream of the termination region.

To generalize these observations, we first compared the RNAPII metaprofiles in regions of CPF termination upstream of a Rap1 binding sites in the presence and absence of the road-block factor. To this end we aligned for each gene the strongest site of poly(A) addition as defined by TIF-seq analyses [139], trusting that this will allow a sufficiently precise approximation of the average termination region. A decrease in the average RNAPII signal was observed in wild type cells in this region, confirming the progressive occurrence of termination. Depletion of Rap1 had no impact on the distribution of the RNAPII signal that declined in the termination region very similarly to the wild type indicating identical efficiencies of termination at the upstream CPF sites (Fig. 7.5B). As a control, a termination defect could clearly be observed in rna15-1 cells at the non-permissive temperature (Fig. 7.5A). Similar results were obtained for the set of CPF-dependent genes upstream of a Reb1-dependent road-block (data not shown).

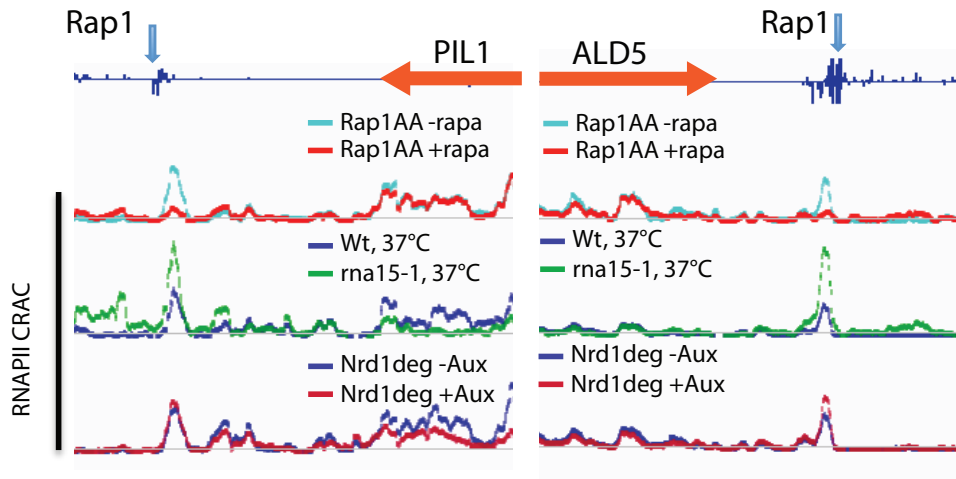


**Figure 7.5:** Metagenome analysis around poly(A) sites that are within 300bp upstream of Rap1 sites. **A:** Comparison of RNAPII average profile in wild type and *rna15-1* at non-permissive temperature and **B:** the same analysis performed in wt or Rap1-AA strains. **C:** Comparison between ratios of RNAPII signal in the termination zone divided by RNAPII signal in the body of the gene. While a significant difference is detectable between *rna15-1* and wt at non-permissive temperature, no such difference is detectable when Rap1-AA is compared with wt.

To substantiate these results we calculated the fractional level of readthrough for each CPF-dependent gene upstream of a Rap1- dependent road-block by dividing the density of reads in the termination region by the density in the gene body. The distribution of the values obtained is strongly affected by the *rna15-1* mutation, as expected for a bona fide termination defect ( $p=1E-8$ ), but not by the absence of Rap1, demonstrating that the road-block does not significantly impact CPF termination (Fig. 7.5C).

## Roadblock and NNS-Dependent Termination

While this work was in progress another study suggested that road-block- and NNS-dependent termination are functionally linked, notably that: i) road-block is part of the mechanism of snoRNA termination and ii) that roadblocked polymerases are released by the NNS pathway. This study relied on the analysis of transcripts produced in different mutant conditions and on the published distribution of polymerases in wt and Nrd1 anchor away strains [159]. We undertook to revisit this important question using our high-resolution RNAPII CRAC in cells defective for the CPF, NNS and road-block pathways.

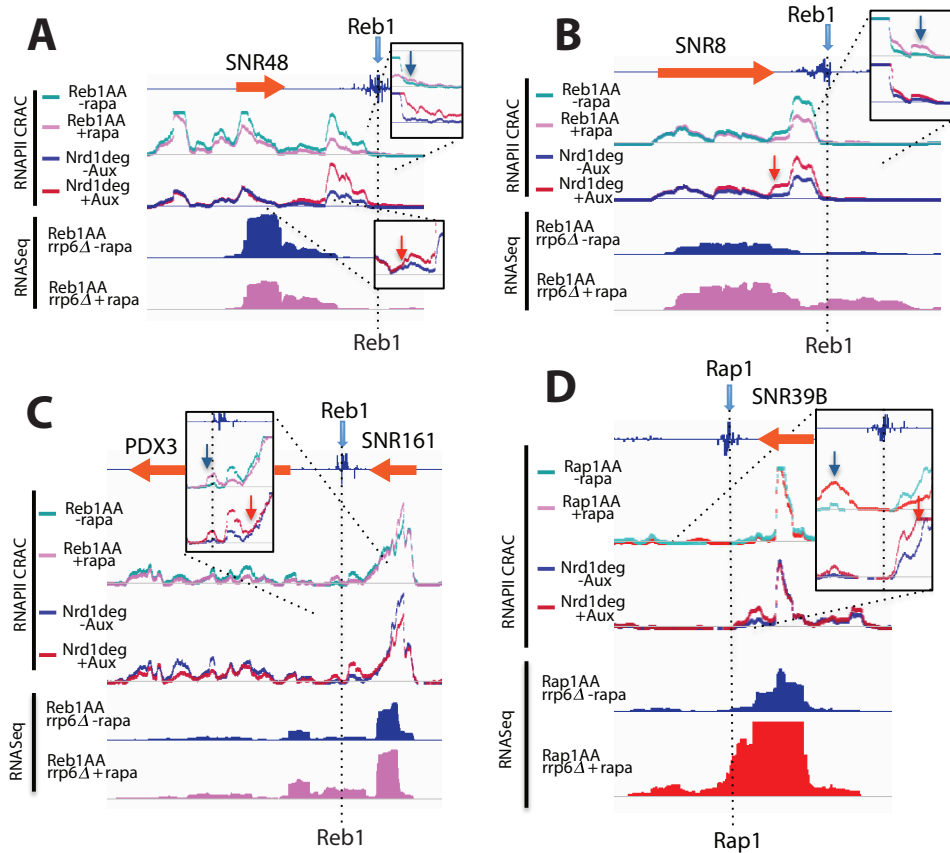


**Figure 7.6:** Examples of CPF-terminated transcripts followed by sites of road-block in the context of Rap1-AA, rna15-1 and Nrd1-degron strains. the position of Rap1 sites and annotation of the transcripts is displayed at the top.

Roadblock peaks have been shown to increase in strains defective for NNS termination, which was interpreted as evidence of defective clearing of roadblocked polymerases when NNS termination is impaired [159]. An alternative interpretation, which we favor, is that when NNS termination is defective, polymerases that do not terminate at primary NNS termination sites accumulate downstream at road-block sites. Consistent with this notion is the finding that the level of roadblocked polymerase is not sensitive to NNS termination at road-block sites preceded by CPF terminators (supplementary figure 7.5). Two such examples are shown in figure



7.6. The RNAPII road-block peak increases considerably when CPF termination is impaired at the PIL1 and ALD5 loci but is unaffected by depletion of NRD1. Identical results were obtained at these loci when Sen1 was depleted (data not shown). Conversely, depletion of Nrd1 leads to an increase of the road-block peaks at Reb1 and Rap1 sites located downstream of NNS terminators (supplementary figure 7.8). Together, these results indicate that roadblocked polymerases are not generally released by the NNS pathway.



**Figure 7.7:** Analysis of 4 snoRNAs followed by road-block by CRAC and RNA-seq in several mutant strains. Position of the Reb1 or Rap1 site, and annotation of the transcripts is represented at the top of each image. Insets are enlargements of the highlighted area.

In a small number of snoRNAs the road-block is located very near to the NNS terminator and it is possible that it contributes to the formation of a functional RNA. We analyzed the polymerase profile around four of these snoRNAs for which a Reb1-

(SNR161, SNR8, SNR48) or Rap1-dependent (SNR39B) road-block peak of variable intensity was observed in the termination region. Depletion of Nrd1 led to a clear increase of the road-block peak, as expected. However, a clear increase in RNAPII occupancy was also observed between the NNS terminator and the region of the road-block (figure 7.7, red arrowheads), indicating the existence of a readthrough at the primary terminator that feeds the flow of polymerases accumulating later at the road-block. This was clearly visible at the SNR8 and SNR48 loci, where the road-block is slightly more distal (Fig. 7.7, panels A-B), but also observed at SNR161 and SNR39B where the signal due to the readthrough increase somewhat merged with the road-block peak (Fig. 7.7 C-D).

Conversely, no evidence of readthrough could be observed at the primary termination site when the road-block factor (Rap1 or Reb1) was depleted, which only led to the expected decrease in the road-block peak. A small but clearly visible readthrough extended downstream of the road-block (Fig. 7.7, blue arrowheads), most likely due to the release of polymerases accumulating at the failsafe site.

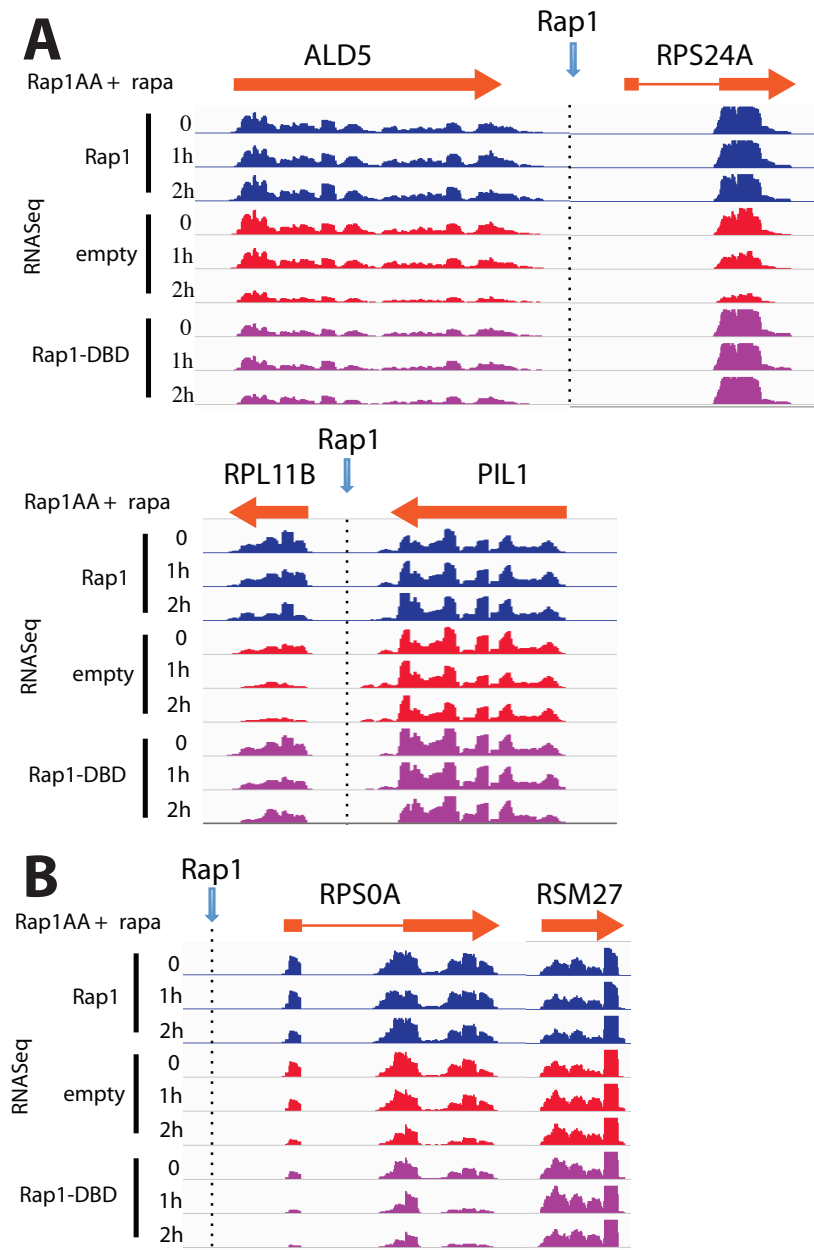
We also analyzed the RNAs produced in the absence of the road-block (Fig. 7.7). We depleted Reb1 or Rap1 in an *rrp6Δ* strain, which allowed visualizing the primary product of transcription that is stabilized in this genetic context. Interestingly, in spite of the overall low level of polymerases going through the road-block site in the absence of Reb1 or Rap1, a significant increase in the amount of pre-snoRNA was generally observed (Fig. 7.7, see the RNAseq profiles at SNR8, SNR161 and SNR39B loci; data not shown), suggesting that transcription events terminating downstream of the road-block produce transcripts that are generally more stable than those produced by transcription terminating at the primary (NNS) or secondary (road-block) terminator. The levels of the mature snoRNAs were generally increased or unchanged in the absence of the road-block (data not shown), although the general stability of these forms prevents from drawing strong conclusions in these transient depletion experiments.

These experiments strongly suggest that the absence of the road-block does not

prevent the production of functional snoRNAs but allows the production of stable transcripts derived from a low level of readthrough transcription at the primary NNS terminator. Importantly they strongly support the notion that road-block termination functions as a fail-safe mechanism for both the CPF and the NNS pathways.

## Functional Importance of Fail-Safe Transcription Termination

As shown in figure 7.8, depletion of Rap1 strongly downregulates transcription of RPL11B and RPS24A. These genes are positioned downstream of a Rap1-dependent road-block where polymerases derived from upstream transcription accumulate. Removal of the road-block allows the progression of these polymerases into the downstream promoters (Fig. 7.8), which might be silenced by transcriptional interference. However, it is also possible that Rap1 directly promotes transcription activation of these genes, independently of its role in roadblocking upstream polymerases. To distinguish between these (non exclusive) possibilities we investigated whether maintaining the sole “protective” function of Rap1 would be sufficient to restore expression of the downstream genes. To this end we depleted Rap1 in cells expressing the well-characterized DNA binding domain of Rap1 (Rap1-DBD, aa. 358-601), which is not expected to activate transcription. As a control, we also expressed the wild type Rap1 or an empty plasmid upon endogenous Rap1 depletion, and analyzed the RNA produces by RNAseq. Prior RT-qPCR analyses demonstrated that expression of Rap1-DBD is sufficient to restore the road-block upstream of HYP2 (supplementary figure 7.9 and data not shown). The overall impact on the transcriptome of Rap1-DBD will be extensively discussed elsewhere (Challal et al., in preparation), but the RPL11B and RPS24A RNA profiles, together with RNAs derived from neighboring genes as a control, are shown in figure 8. Consistent with the RNAPII CRAC data, expression of RPL11B and RPS24A is markedly affected by the depletion of endogenous Rap1 and restored by the concomitant expression of wt Rap1. Importantly, expression of the DNA binding domain alone of Rap1 is sufficient to restore RPL11B and RPS24A to wild type levels. This is not due to



**Figure 7.8:** RNA-seq analysis of genes followed by Rap1 sites. **A:** Strains depleted for Rap1 were rescued with different plasmids, one expressing the full length Rap1, one expressing only the DNA-binding domain of Rap1, and an empty plasmid. The transcript downstream of the Rap1 site is downregulated in absence of the full length protein, but expression is rescued by the presence of the DNA binding Domain of Rap1. **B:** Rap1 site without an upstream feature. The same downregulation is detected in presence of both the empty plasmid and Rap1-DBD, thus proving that Rap1-DBD cannot activate transcription.

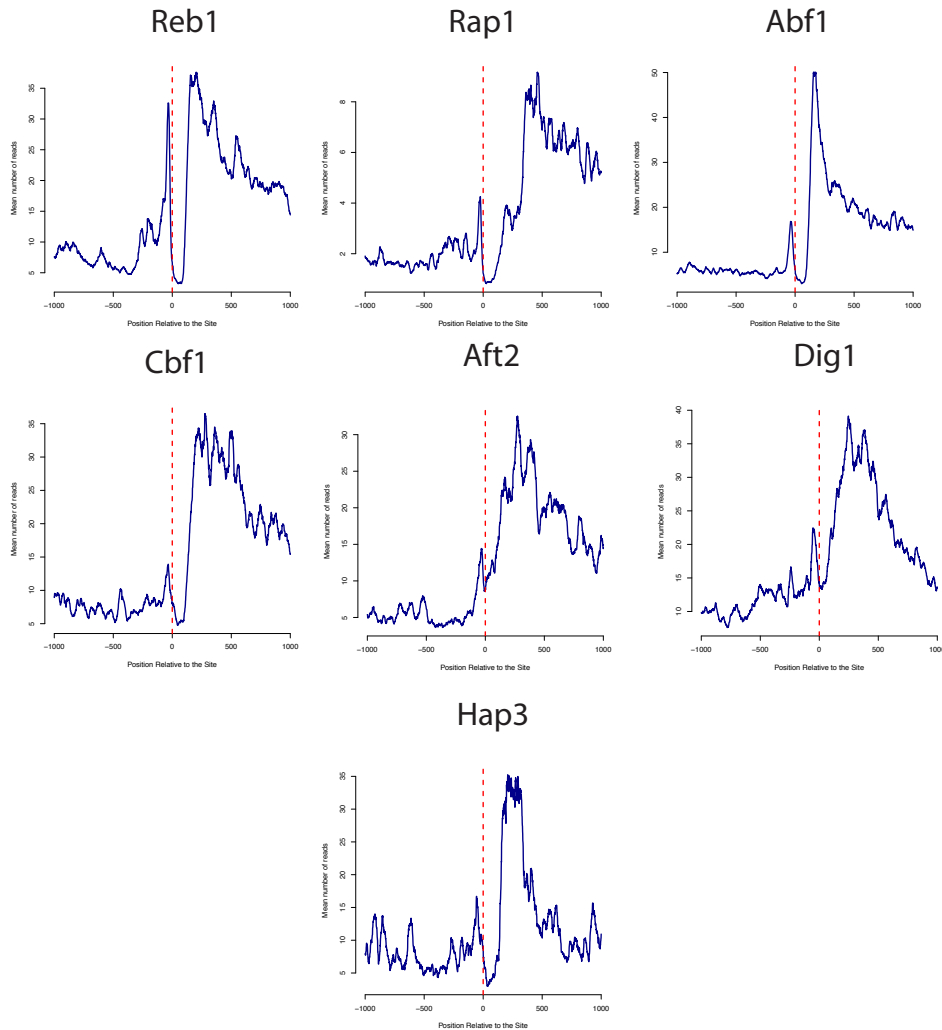
a generalized ability of Rap1-DBD to activate Rap1 target genes as demonstrated by the failure of Rap1-DBD to restore expression of RPS0A (Fig. 7.8B) or RPL29 (data not shown).

Together these results support the notion that the constitutive readthrough at CPF (and possibly NNS) terminators can be sufficient for silencing downstream genes, underscoring the importance of the protective action of road-block factors.

### **Extensive Road-Block Termination in the *S.cerevisiae* genome**

In the light of the results shown here on Rap1 and Reb1, we undertook to assess more generally the occurrence of road-block termination at sites of occupancy for DNA-binding proteins or complexes. For a more stringent and sensitive meta-analysis, we plotted the median level of polymerase occupancy at each position before a given site, which better reflects changes in the whole distribution of occupancy values. Indeed, the appearance of a peak at a given position is more stringently linked to changes that affect the whole distribution of values and less dependent on the contribution of extreme values. Consistently, a prominent and specific peak of polymerase pausing was observed by this method immediately upstream of many transcription factors (Fig. 7.9). Roadblock occurs at a variable distance between 20 and 40 nucleotides upstream of the protein binding site, likely reflecting the topology of the collision between polymerase and the DNA-bound factor or complex of factors.

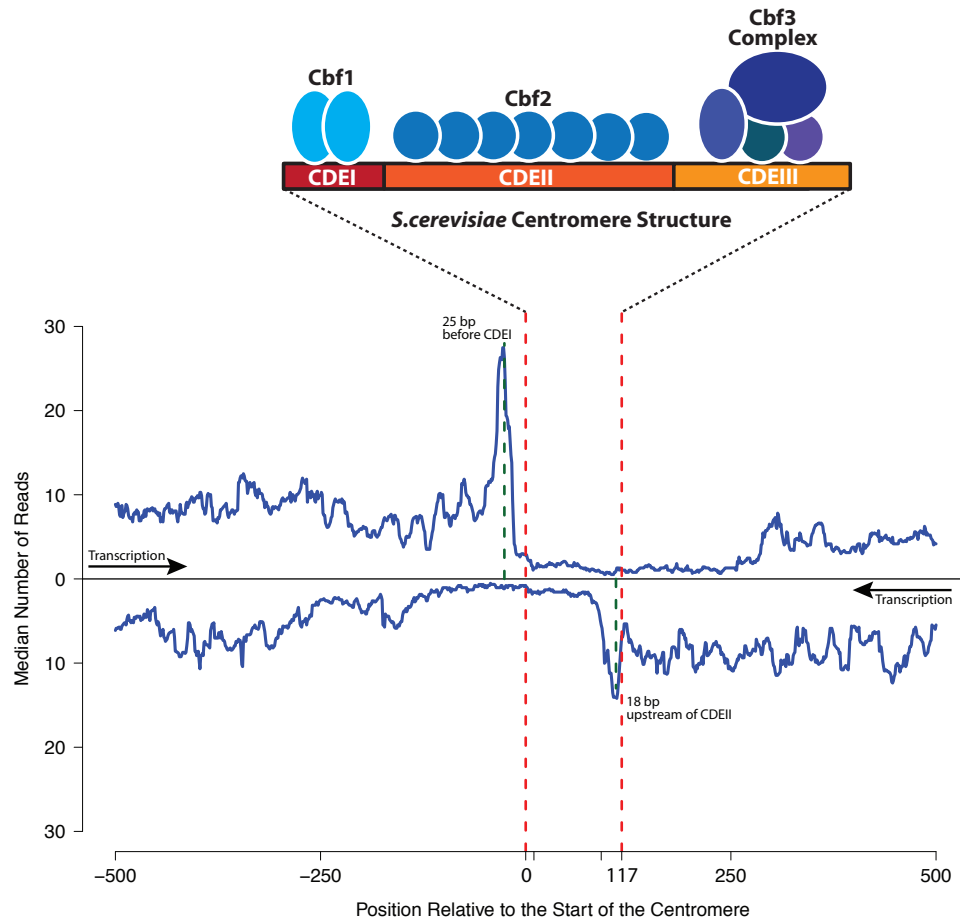
We also sought evidence of road-block termination at other sites where RNAPII transcription might collide with DNA-associated events. Prominent levels of road-block termination were observed at centromeres and tRNAs. In *S.cerevisiae*, centromeres are defined by a set of short, conserved sequence elements located in a 125nt region. These sequences, CDEI, CDEII and CDEIII (Fig. 7.10) are specifically bound by DNA binding complexes that overall constitute the kinetochore, required for the attachment of the chromosomes to microtubules during cell division [for review see 97]. The analysis of the RNAPII metaprofile around centromeres clearly indicates



**Figure 7.9:** Metagene analyses performed around several putative road-blocking factors.

a prominent level of road-block when centromeres are aligned using the external CDEI or the CDEIII sequence motifs, bound respectively by Cbf1 and the Cbf3 complex.

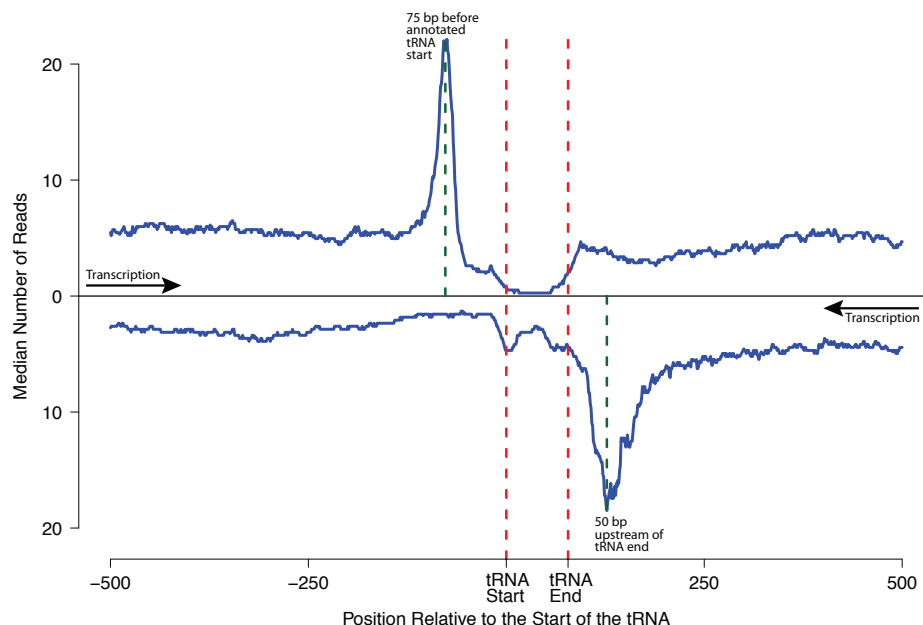
Prominent levels of road-block termination also occur at tRNAs. This was previously observed in the 5'-end of a model tRNA, where road-block was attributed to the binding of the RNAPIII factor TFIIB [91]. We extended this finding genomewide, by showing RNAPIIs piling up at position -75 from the tRNA start site, corresponding



**Figure 7.10:** Schematic representation of the structure of the centromere and the RNAPII occupancy profile around it. The top and bottom of the graph represent the two strands, the direction of transcription is indicated by arrows.

to a road-block induced by TFIIB bound at position -50 from the start site. Importantly, however, we also observed a prominent road-block antisense to the tRNA. RNAPII strongly accumulates about 50 nucleotides upstream of the annotated end of the tRNA (Fig. 7.11). Because RNAPIII transcription is not known to depend on factors bound to the DNA, we infer that roadblocking occurs from the collision of RNAPII with RNAPIII, presumably paused at the termination signal or persistently occupying the tRNA transcription region.

Together, these data demonstrate that road-block termination is not restricted to Reb1 or Rap1 binding sites, but also occurs at many different locations in the yeast



**Figure 7.11:** metagene analysis of RNAPII occupancy around tRNAs. The top and bottom of the graph represent the two strands, the direction of transcription is indicated by arrows.

genome. Aside from operating a quality control mechanism on the efficiency of termination at canonical sites, road-block pausing of polymerases has potential for gene regulation and coordination of transcription with other DNA related events

### 7.2.3 Discussion

In a previous study we have described an additional pathway whereby transcription termination occurs when the elongation complex encounters the factor Reb1 bound to the DNA. Based on a few model cases we have proposed that road-block termination by Reb1 limits pervasive transcription and functions to “protect” promoters regions from “invading” polymerases. The general validity of these concepts was, however, not addressed in this early study. Here we extend these concepts to a genomewide perspective and to other factors, providing a general view of the impact and functional significance of road-block termination in the *S.cerevisiae* genome.



We first demonstrated that Rap1, a DNA binding factors that has roles in transcription activation, gene silencing and telomere homeostasy, is also a road-block factor. An earlier study showed that the fortuitous introduction of a Rap1 site in a Ty1 retrotransposon led to RNAPII stalling and repression of gene expression. Based on the analysis of the RNA produced, which was non-adenylated and insensitive to exosome degradation, it was concluded that termination of transcription did not occur in these conditions [199]. We show that road-block termination occurs upstream of Rap1, leading to the production of RNAs that are polyadenylated by Trf4 and degraded for a large part by the nuclear exosome. We also detected non-adenylated RNAs, which most likely represents the nascent RNA associated to the polymerase that pauses before termination. Importantly, nuclear depletion of Rap1 prevents termination, indicating that the protein – and not the presence of termination signals overlapping its binding site – is responsible for ending transcription. Failure to detect the polyadenylated fraction for technical reasons in the study by Yarrington *et al.* might account for the discrepancies; alternatively, termination might not occur in the Ty1 retrotransposon model for unknown reasons.

## The Mechanism of Roadblock Termination

Similar to what previously shown for Reb1 [29], release of the polymerase stalled upstream of the road-block occurs, at least partially, as a consequence of its ubiquitylation by Rsp5 and presumably degradation. Thus, this pathway is not restricted to Reb1-dependent termination and presumably extends to all cases of road-block, in addition to events of pausing that cannot be resolved in a more “conservative” manner as previously demonstrated for polymerases encountering a DNA damage [193]. Using high resolution RNAPII occupancy data we observed very sharp peaks of stalling at the road-block sites, which is hardly compatible with more than one polymerase roadblocked, on average, at a time. This indicates that the clearance due to the Rsp5 pathway is as efficient as the building up of the peak, at least at steady state.

It has been recently proposed that the NNS pathway is required for releasing roadblocked polymerases. This claim was essentially funded on the observation that i) Nrd1 and Nab3 binding sites are frequently present at sites of road-block and ii) that peaks of polymerase stalling increase upon depletion of Nrd1 from the nucleus, which was taken as evidence that the clearance of roadblocked polymerases is affected when NNS termination is defective. Data shown in this report and in our previous study on road-block termination are not compatible with this model. The strongest counterevidence is that the insertion of an isolated Reb1 [29] or Rap1 site (this report) in a segment of the HSP104 gene lacking NNS termination signals, is sufficient for efficient RB termination. Moreover, these or similar constructs have been shown to be largely insensitive to depletion of Nrd1 and Nab3 [29, data not shown].

This notion also holds for the natural cases of road-block termination in the *S.cerevisiae* genome. We show that the cumulative road-block peak significantly increases upon depletion of Nrd1 only when considering road-block sites downstream of NNS, but not CPF-CF terminators (Fig. 7.6, supplementary figure 7.8). Conversely, when the RB sites are downstream of CPF-CF terminators, mutation in the CPF-CF complex (but not Nrd1 depletion) increase the levels of roadblocked polymerases (figure 7.4B and D, supplementary figure 7.4). This suggests that alterations in the NNS (or CPF-CF) complexes do not affect the clearance of roadblocked polymerases, but their further accumulation upon failure to terminate at upstream NNS- (or CPF-CF) dependent genes.

Thus, we favor a model according to which road-block termination operates independently of the NNS and CPF-CF pathways and does not allow recycling of the polymerase for further steps of transcription but leads to its degradation, together with the RNA that is produced. Such a disruptive mechanism might look uneconomical, but the concept is analogous to the seemingly useless transcription of many non-functional RNAs that are degraded rapidly after production by processing or quality control mechanisms. The energetic balance might still be favorable if the

evolutionary cost of developing highly efficient, error-proof machineries is taken into consideration. In this respect, the genomewide analyses reported here strongly suggest that road-block termination is unlikely to be devoted to the generation of functional molecules, but rather to controlling a relatively low fraction of polymerases that might significantly affect the efficiency or robustness of neighboring processes.

## Functional Significance of Roadblock Termination

We have previously proposed that in a few model cases, Reb1-dependent road-block termination functions to neutralize transcription events that failed to undergo termination at upstream genes. The question addressed here is to what extent this is general, i.e. does significant transcription readthrough occur at CPF and NNS terminators genomewide and in cells that are proficient for termination. We show here that prominent roadblocks at Reb1 and Rap1 sites can be fed by polymerases that escape upstream CPF-CF and NNS-dependent termination, demonstrating the occurrence of constitutive readthrough at canonical terminators. Because we show that many DNA binding factors can road-block, to different extents, the elongation complex, polymerases overlooking canonical termination signals run into “bumpy” roads that limit their progression in intergenic regions, where they could interfere with transcription initiation or other cellular processes. The genomewide analysis of RNAPII distribution in mutants of the CPF pathway show that in most instances, readthrough polymerases accumulate in the adjacent intergenic regions, which is fully consistent with this notion [unpublished results, or Challal et al., in preparation].

A large wealth of evidence exists demonstrating that pervasive transcription is generated to a large extent by the leaky control of chromatin on initiation. This is particularly important for restricting the inherent bi-directionality of promoters and directing preferential initiation towards functional genes. Mutation of many chromatin remodelers or modifiers further weakens such a repressive control on

initiation [115]. Leakiness in transcription termination is functionally analogous to the limited control of chromatin on initiation, in terms of the generation of pervasive transcription. In both cases, this allows transcription elongation in regions that are not necessarily producing functional transcripts and is susceptible to affect regulation of neighboring genes or other DNA-related processes, which requires its control *a posteriori* by quality control pathways. In both cases, additional exposure of genomic information by transcription might confer evolutionary advantages.

An appropriate level of readthrough transcription might allow the option of generating new and longer genes, for instance when the extended transcripts evolve to fuse contiguous ORFs or to generate polypeptide extensions to an existing factor. The regulatory potential of transcription readthrough should also not be neglected: modulation of termination efficiency might allow coordinating expression of tandem genes. Such a modulation might more easily apply over a flexible basal system whereby the efficiency of termination is not plateaued out.

### **Relationships Between Road-Block and the Main Pathways of Termination in *S.cerevisiae***

We considered the possibility that pausing induced by the road-block could favor upstream termination, by slowing down the progression of the polymerase and allowing catching up by “pursuing” enzymes like Sen1 or Rat1. We reasoned that, should the model be correct, the progressive loss of polymerases due to termination upstream of the road-block is expected to be affected when the latter is removed or strongly diminished. However, the average profile of polymerases on aligned termination regions for CPF-dependent genes upstream of Reb1 or Rap1 roadblocks did not support this notion, and rather showed that termination occurred upstream with equal efficiency in the presence or absence of the road-block. Removing the latter has therefore the sole effect of allowing further progression of polymerases that have failed to terminate at the primary site.

We also favor the notion that fail-safe termination is the main function of this

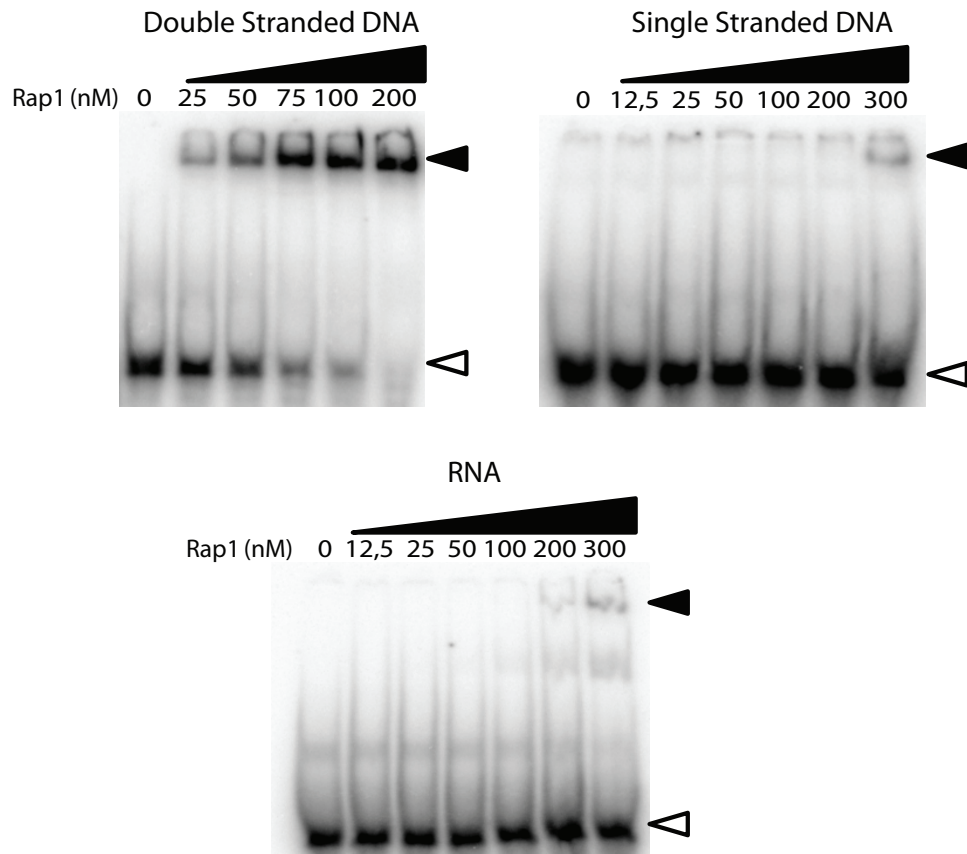
pathway for sn/snoRNAs genes. The strong expression of many of these genes might more strictly require fail-safe termination to protect downstream features, which possibly explains the frequent occurrence of RB sites downstream of sn/snoRNA genes [159]. Analogously to what observed for CPF termination, we found that upon Nrd1 depletion the downstream RB peak increases, consistent with the notion that the RB peak is fed by polymerases that fail to terminate at the primary NNS transcription termination site (TTS). Moreover, when the RB factor is depleted we could not detect termination failure at the primary TTS, supporting the notion that the RB is not generally required for NNS termination. We did observe a significant accumulation of extended RNAs upon depletion of the road-block. However, because the RB only prevents progression of the relatively low fraction of polymerases that have escaped primary termination, it is unlikely that transcription going through the RB site fully accounts for the relatively high levels of extended RNAs observed, for instance at SNR8 and SNR39B. Rather, we favor the notion that the longer RNAs produced are stabler than the pre-snoRNA and accumulate because they escape exosomal degradation that ensues from NNS or RB termination. Whether release of polymerases at the secondary, RB termination also generates precursors that could be trimmed down to the mature snoRNA is unclear; however, we never observed a decrease in the levels of mature RNA when the RB was depleted, arguing against a significant contribution of these transcripts to the mature forms.

We show that many proteins that bind the DNA are able to road-block the RNA polymerase, suggesting that transcriptional activity might be modeled to a large extent by non-histone proteins bound to the DNA. Besides transcription units, other features are “protected” by the RB, including tRNA and centromeres, for which we show evidence in this report, and replication origins [Candelli et al., in preparation, see chapter 8]. The case of tRNAs is possibly anomalous as we observe the occurrence of RB also at the 3’ end, where the specific presence of DNA-binding factors has not been described. The specific topology of these transcription units that are possibly circularized for a more efficient transcription re-initiation, or the general high persistence of RNAPIII at these sites might account for the

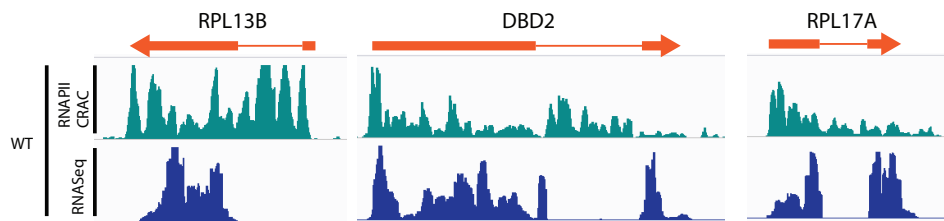
inability of RNAPII to traverse these regions. Besides preventing interferences with RNAPIII transcription, the strong barriers provided by tRNAs might constitute major insulating elements for the protection of transcription units or other sensitive genomic features.

The extent and the properties of road-block termination in the *S.cerevisiae* genome suggest that significant regulation of gene expression and other DNA-related processes might occur as a result of the modulation of RNAPII progression, which might also apply to larger metazoan genomes.

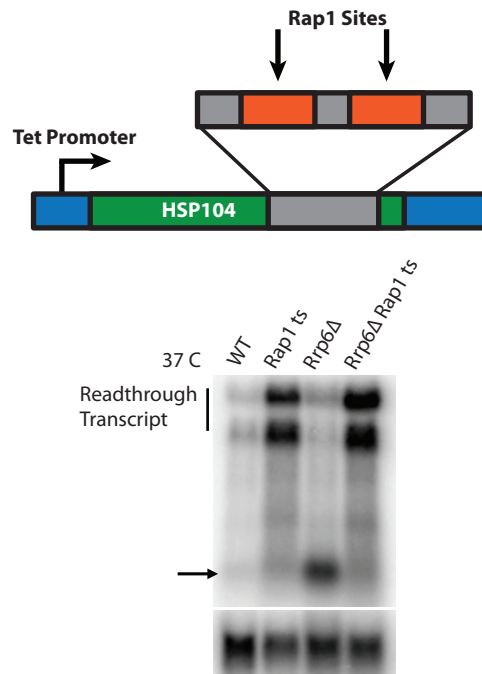
## 7.2.4 Supplementary Figures



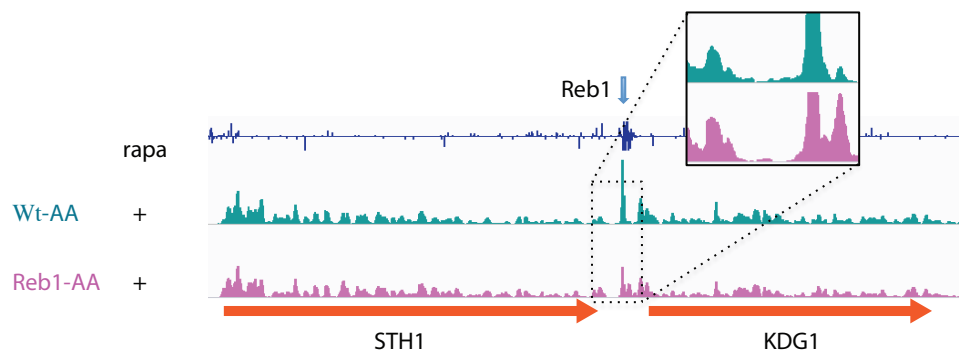
**Supplementary Figure 7.1:** EMSA analysis of increasing concentrations of Rap1 with several species of nucleic acids: single strand DNA, double strand DNA, and RNA.



**Supplementary Figure 7.2:** Comparison between CRAC and RNA-seq signal at several loci. CRAC profiles are characterized by signal within introns.

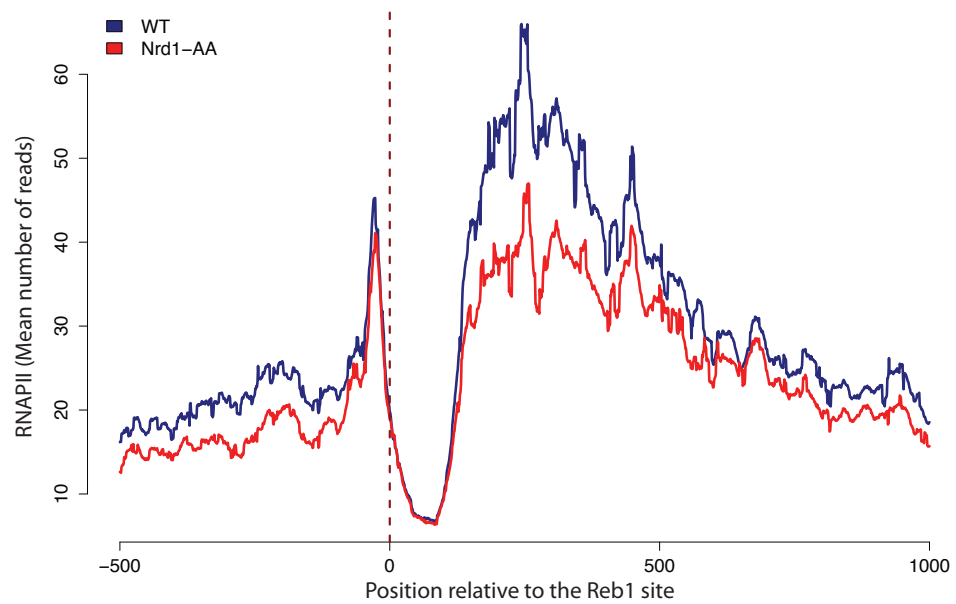


**Supplementary Figure 7.3:** Northern blot analysis of the insertion of two Rap1 sites within the reporter system. A short Rrp6- and Rap1-dependent transcript is present and disappears at non-permissive temperature in a Rap1 thermosensitive strain.

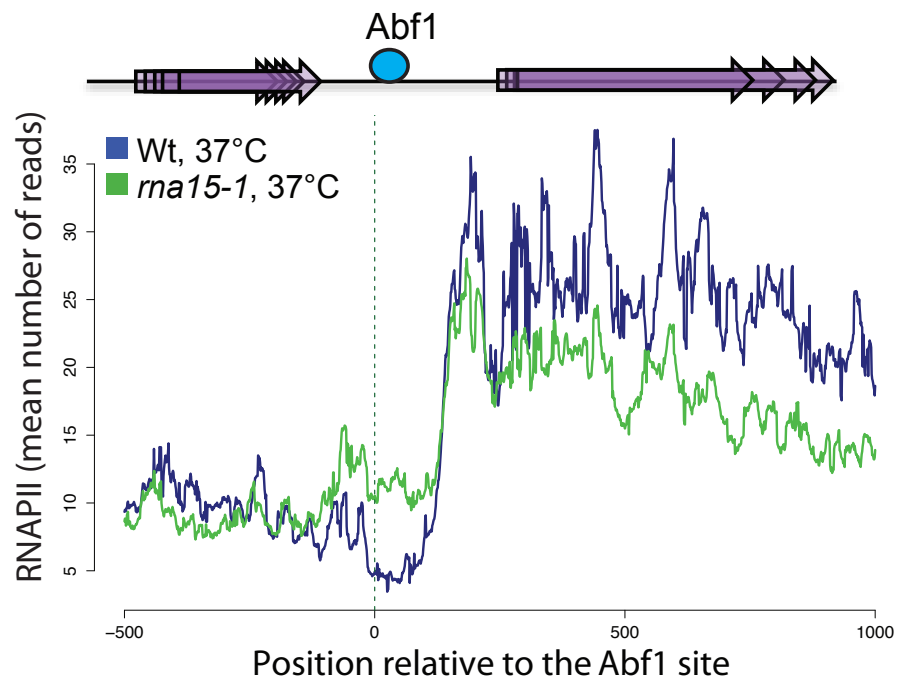


**Supplementary Figure 7.4:** example of Reb1-mediated road-block. CRAC signal is shown in WT-AA and Reb1-AA. RNAPII signal accumulates in proximity of Reb1 site, but this accumulation is diminished in Reb1-AA

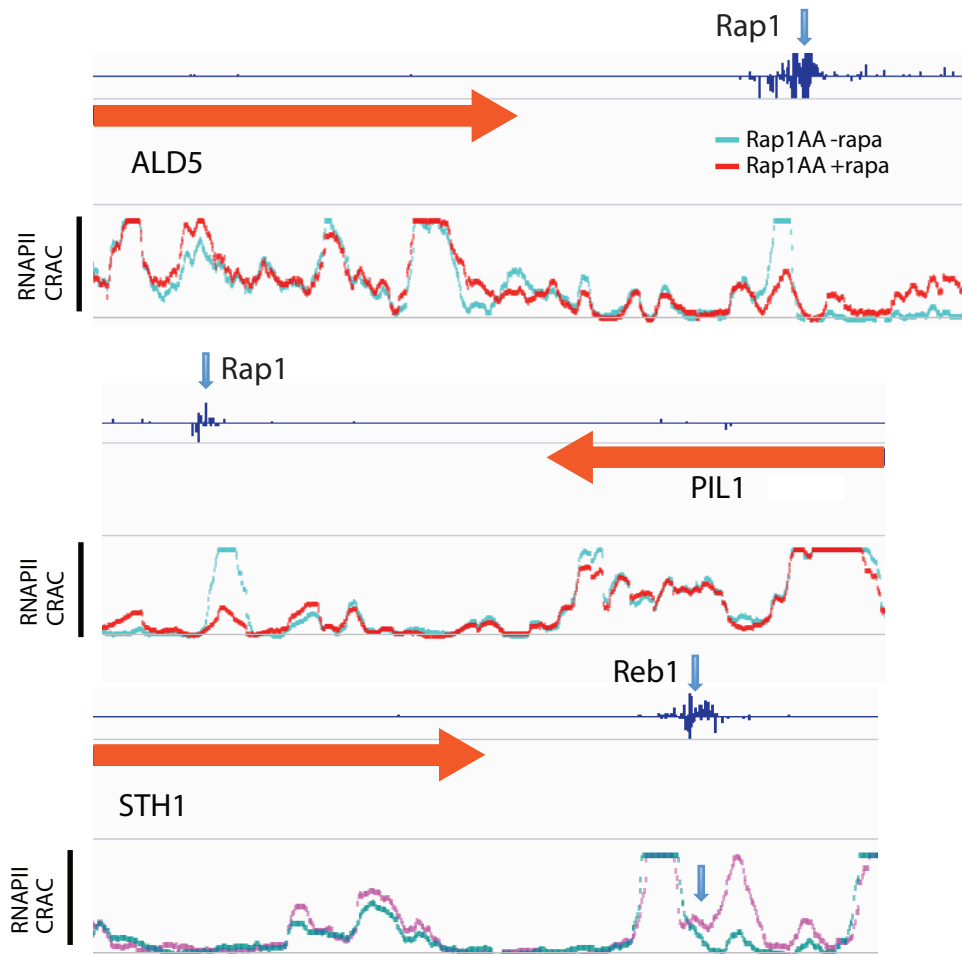




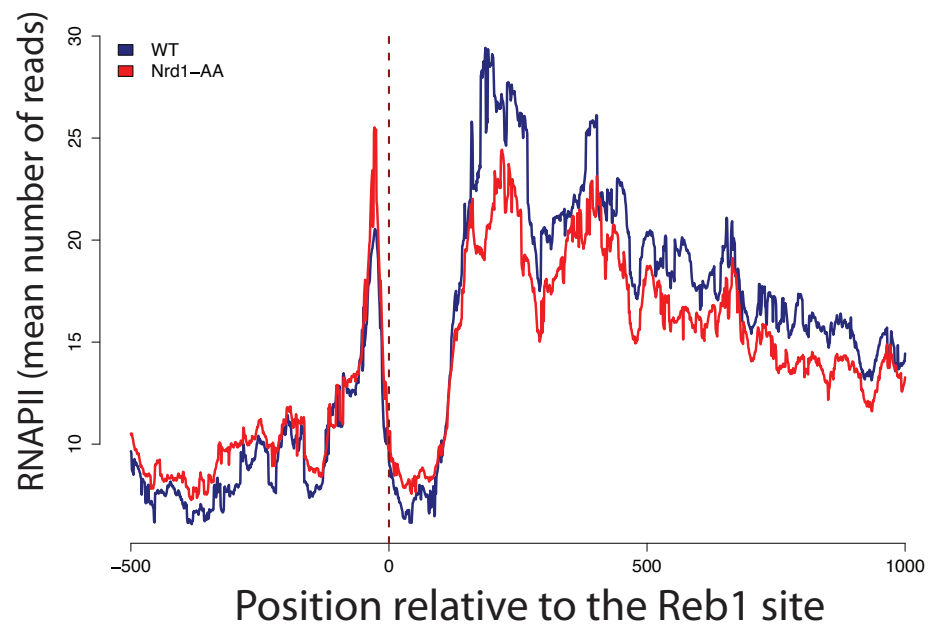
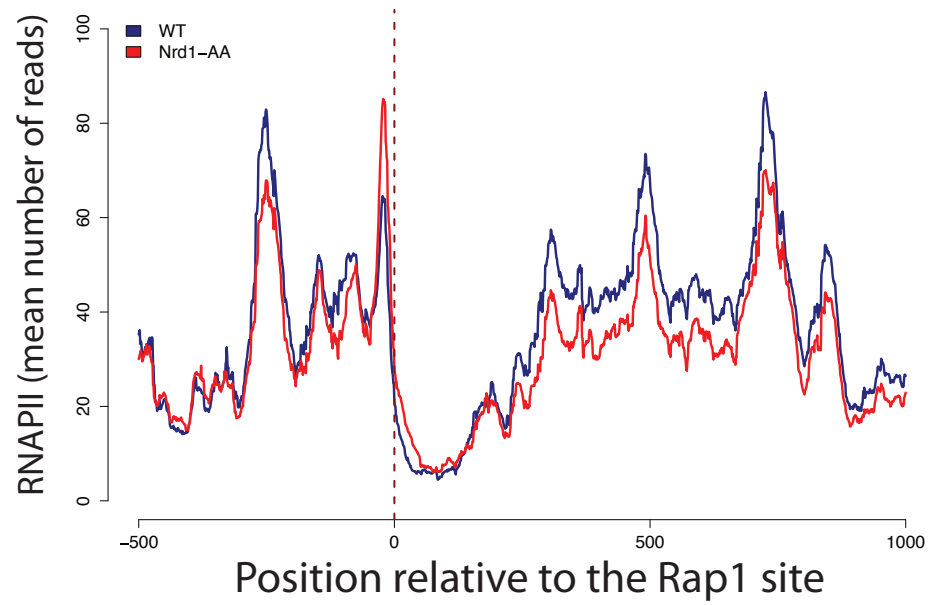
**Supplementary Figure 7.5:** metagene analysis of RNAPII PAR-CLIP signal around Reb1 sites preceded by CPF-terminated transcripts. This analysis was carried out both in a wild type and Nrd1-AA strain.



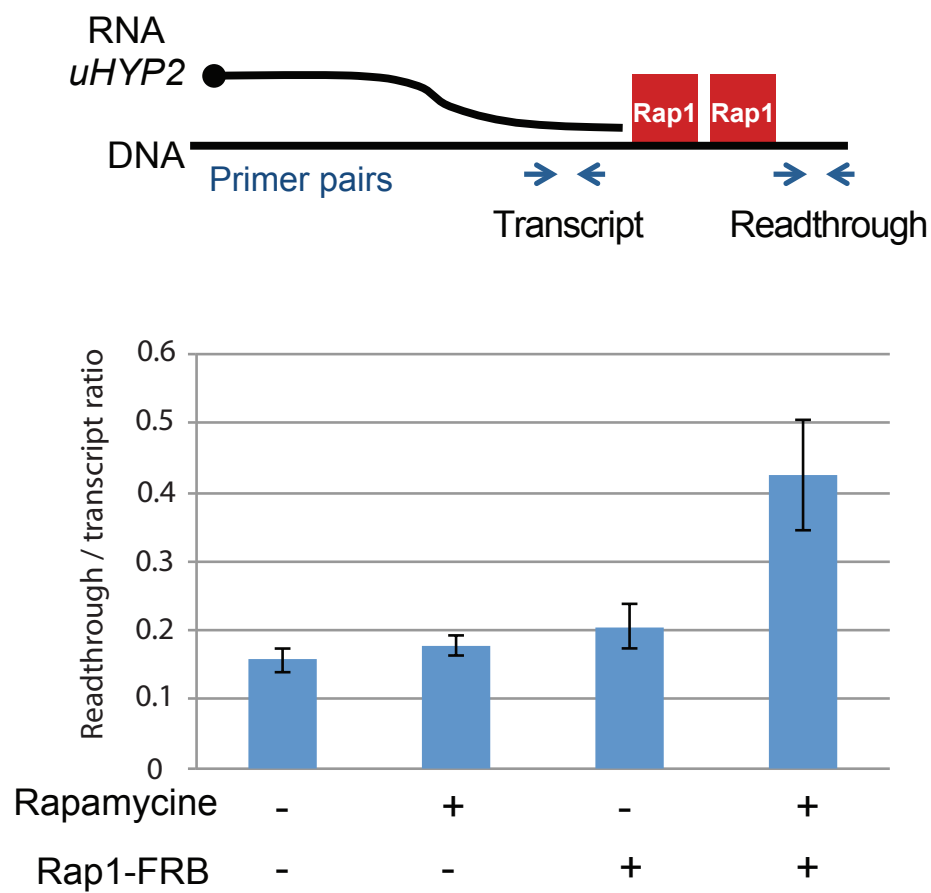
**Supplementary Figure 7.6:** Metagene analysis of RNAPII CRAC around binding sites of Abf1 carried out in a wild type and *rna15-1* strain at non-permissive temperature



**Supplementary Figure 7.7:** Comparison between RNAPII CRAC performed in Rap1-AA + and - rapa at three Rap1 sites downstream of CPF-terminated features. Despite depletion of Rap1, no change in the efficiency of CPF termination can be detected.



**Supplementary Figure 7.8:** Metagene analyses of wild type and Nrd1-AA RNAPII PAR-CLIP signal around Rap1 and Reb1 sites preceded by NNS-terminated transcripts.



**Supplementary Figure 7.9:** RT-qPCR analysis of Rap1 dependent termination upstream of the HYP2 gene. The ration between qPCR signal after and before the Rap1 sites increases significantly upon removal or Rap1.

## 7.3 General Discussion

In these two manuscripts we describe a novel non-canonical termination pathway for RNA polymerase II. General regulatory factors Reb1 and Rap1—and possibly other genomic features such as centromeres, tRNAs, and binding sites for the transcription factor Abf1—were shown to stall RNAPII, prevent elongation and result in transcription termination. Road-block termination was shown to be an extensively used mechanism to terminate polymerases that escape canonical termination pathways. However, road-block termination is able to act independently of other termination mechanisms and has no effect on their efficiency.

### 7.3.1 Fail-Safe Termination

General regulatory factors are a family of transcription factors that regulate a substantial amount of genes in *S.cerevisiae* (10-15% [154]) We showed that three members of this family, Reb1, Rap1, and Abf1, are bona fide road-block terminators in addition to their activator roles. Because road-block termination results in the production of unstable transcripts, we speculate that its functional relevance concerns more the control of pervasive transcription rather than the production of functional RNAs. Indeed, a number of GRF binding sites were found to be associated with termination of CUTs or other non-functional transcripts. Moreover, we show that sites of road-block in proximity of canonical CPF terminators still display accumulation of RNAPII, suggesting that constitutive readthrough at CPF terminators is a major source of road-block dependent transcripts. This evidence is consistent with a model where road-block would serve as a fail-safe termination mechanism to prevent transcriptional readthrough (or other spurious transcription events) from invading promoter regions. This notion is particularly relevant in yeast, where, due to the compact nature of the genome, unchecked transcriptional readthrough is very likely to interfere with other biological processes.

### 7.3.2 Road-Block Termination Promotes Genome Stability

In addition to transcription factors, we identified several other genomic loci that were associated with strong polymerase pausing. Although we provided no formal proof that these loci represent true transcription termination sites, the presence of several hallmarks of road-block termination (position and shape of RNAPII pausing peaks and presence of RNA 3' ends) supports this hypothesis. Both centromeres and tRNA genes displayed localized increases of polymerase occupancy at their borders, suggesting that the protective role that road-block termination has at promoter regions could extend to other loci that are sensitive to transcriptional interference.

Strong transcription through a centromeric region leads to loss of the parent chromosome [4] in the following mitotic cycles. It is therefore possible that even physiological amounts of readthrough or pervasive transcription could negatively impact the efficiency of the processes associated with centromeres. We observe strong polymerase pausing in the vicinity of Cbfl binding sites within the centromere, and speculate that presence of this protein on DNA could be an effector of road-block. Interestingly, deletion of Cbfl is not lethal but is associated with chromosomal instability [20], although whether this is due to increase in transcriptional interference or loss of other centromeric-specific functions elicited by Cbfl remains unclear.

In addition to centromeres, we find strong polymerase pausing associated with both ends of tRNA genes. Earlier studies provided the proof of principle that tRNA genes could be involved in preventing transcriptional interference [91], citing the presence of TFIIB as a requisite for the effect. In this study we detected strong polymerase pausing in close proximity of TFIIB binding sites in about 70 % of all tRNA genes, suggesting that RNAPII initiation complex provides a strong barrier to transcription elongation. In addition to this putative road-block, we found that polymerases were also stalling in the vicinity of RNAPII termination site. Although we do not know what the cause for this accumulation is, we speculate that a head-to-head collision between RNAPII and terminating RNAPII might prevent elongation. Alternatively, a gene-looping mechanism could result in a

similar molecular phenotype by physically linking the RNAPIII termination site with the initiation complex. Overall, it is tempting to speculate that the high rate of transcription of RNAPIII genes makes them particularly susceptible to transcriptional interference, and therefore resulted in the presence of protective mechanisms to prevent it.





# 8

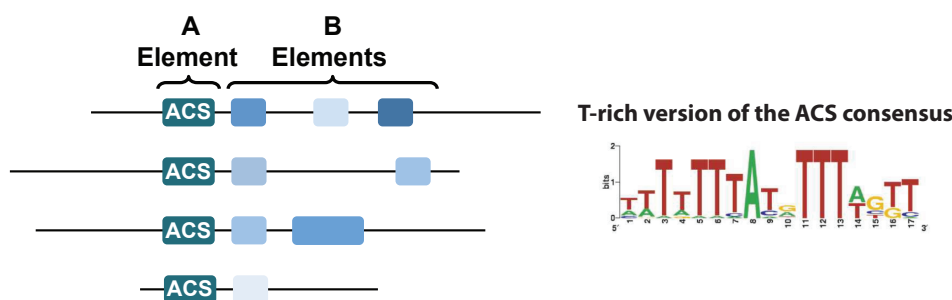
## The Effect of Endogenous Transcription on Origin Specification

In chapter 6, I discussed the mechanisms of DNA replication and how extrinsic factors, such as chromatin structure, can affect origin efficiency. Despite many years of study, the effect of transcription on the activity of replication origins remains a controversial topic, as evidence exists for both a negative and a positive role. Here we show that physiological levels of transcription have a negative impact on origin activity and that mechanisms exist to limit the interference between transcription and replication initiation.

In this work I applied metagene analyses to several published datasets in order to obtain a global view of polymerase occupancy and termination events around origins. I then used correlative techniques to explore the relationships between transcription levels and different stages of replication initiation. Northern blot analyses were performed by Julien Gros in the lab. Preliminary data regarding origin efficiency in *Nrd1* defective strains (see discussion, section 8.4) was generated by our collaborator Julien Soudet, I performed the computational analyses.

## 8.1 Global Visualization of Transcription Around Replication Origins

In order to have a global view of transcription around replication origins, we decided to produce aggregate plots using published polymerase occupancy datasets [164] around sites of replication initiation. This powerful technique allows to visualize average transcription levels at a single nucleotide resolution across any number of genomic loci, but requires a common feature along which all loci can be aligned (e.g. an annotated feature or a sequence element, such as a TSS or a transcription factor binding site). Because of the nature of replication origins, we chose the ACS—the only sequence element absolutely required for origin activity—as our common feature.



**Figure 8.1:** Cartoon showing the most typical arrangement of sequence elements within origins. the ACS is required, while several B elements contribute to origin specification downstream of the T-rich strand of the ACS.

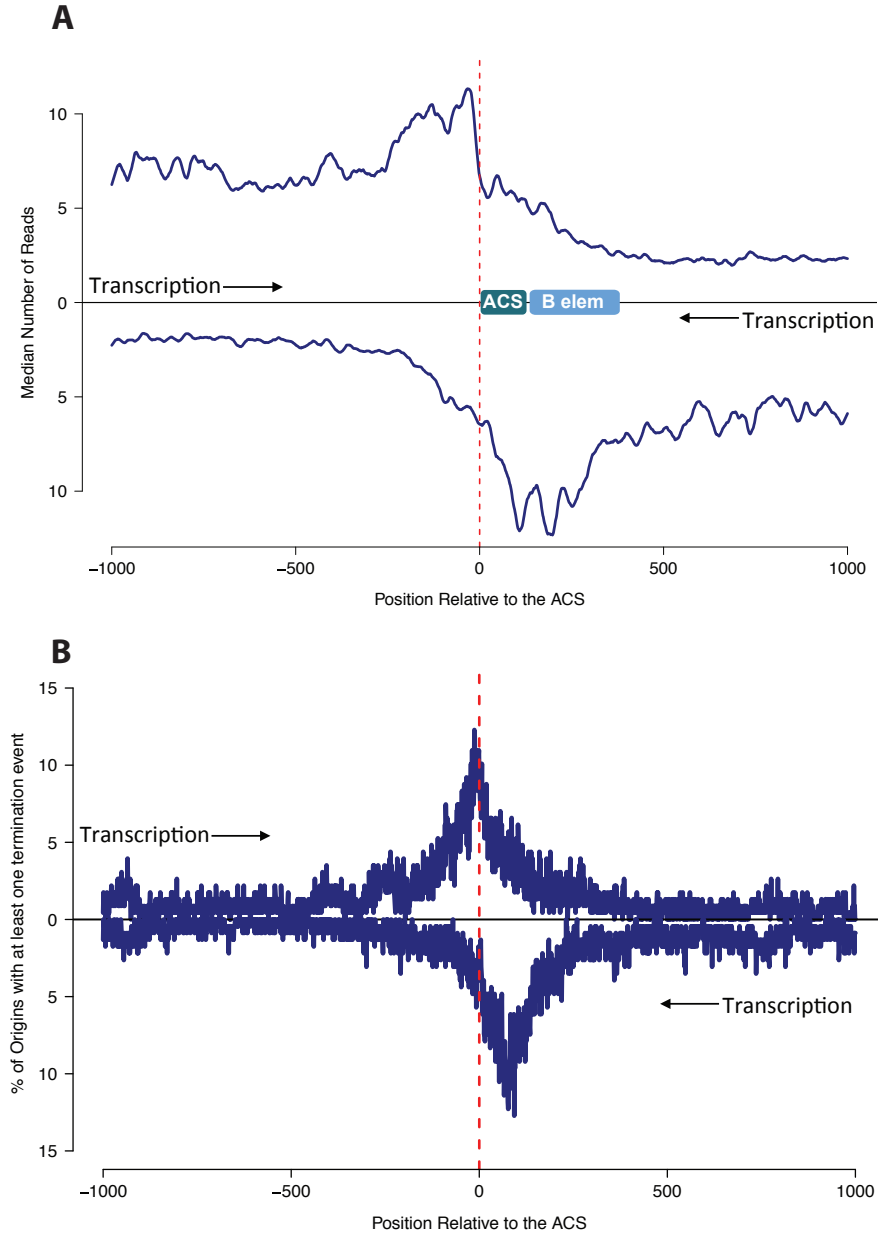
The ACS is an AT-rich sequence element that is bound by the ORC complex and acts as the assembly site for the pre-replication complex. Because of its non-palindromic sequence, we could appropriately distinguish between transcription along the T-rich or A-rich versions of the ACS consensus. Additionally, the orientation of the ACS determines the location of other important sequence elements of the origin, the B elements (Fig. 8.1). This allowed us to not only be strand-specific with respect to the ACS, but with respect to the whole structure of the origin.

## 8.2 Transcriptional Pausing and Termination Are Associated With Replication Origins

The meta-site analysis of RNAPII occupancy around replication origins is shown in figure 8.2A. The top part of the plot represents transcription along the T-rich strand of the ACS, while the bottom part represents transcription along the A-rich strand of the ACS. To obtain this plot, we used a set of origins for which the ACS was annotated [133]. Because transcription in origins is generally low, we restricted our analysis to those surrounded by convergent or tandem genes in order to have a more distinct signal.

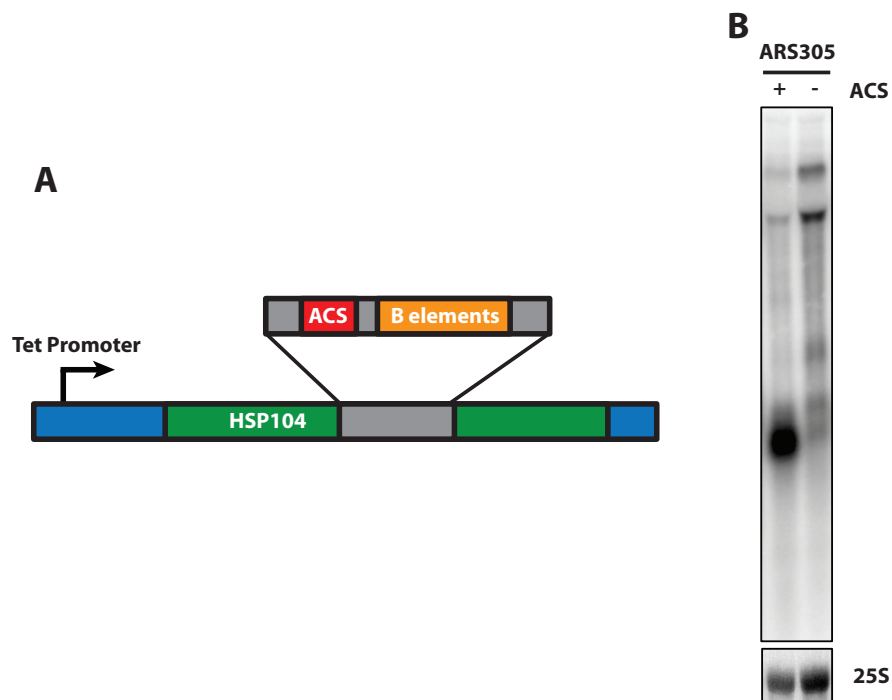
We detect an increase in polymerase occupancy, relative to the incoming average, in the vicinity of the ACS for both along the T- and A-rich strands of the ACS. However, while transcription on the T-rich strand shows a distinct occupancy peak about 20-30 nucleotides before the ACS, transcription on the A-rich strand displays multiple peaks that reside 110-130 nucleotides away from it. Additionally, both occupancy increases—especially the one of the T-rich strand—are characterized by a steep drop in signal after their peak. Because of this sharp drop, we reasoned that the peaks could represent polymerase pausing caused by transcription termination. We therefore generated an aggregate plot across all origins, showing the location of termination events (as defined by the production of RNA 3'-ends [191]) relative to the position of the ACS (fig 8.2B). We detected a substantial number of termination events in the vicinity of the ACS. Moreover, the asymmetry that we highlighted between the two strands with respect to polymerase occupancy is maintained in this analysis. The peak of termination events in the T-rich strand resides about 20 nucleotides from the ACS, while in the A-rich strand this peak is shifted 100 nucleotides away from the ACS. These results show that RNAPII accumulates around replication origins, and this accumulation coincides with transcription termination events.

Although we did not know which pathway was responsible for the termination events



**Figure 8.2:** **A:** Metagenome analysis performed on a polymerase occupancy dataset [164]. profiles represent the average levels of transcription across origins surrounded by convergent and tandem genes. The top part of the plot represents transcription along the T-rich strand of the ACS, while the bottom part represents transcription along the A-rich strand of the ACS. **B:** Plot representing the percentage of assayed origins with at least one termination event at any given position.

we detected around origins, we identified several hallmarks of road-block termination. Polymerase pausing is coincident with precise termination and, at least in the case of the T-rich strand, is positioned 20 nucleotides before the binding site of a DNA binding factor. This led us to speculate that termination on the T-rich strand is caused by a road-block dependent on ORC.



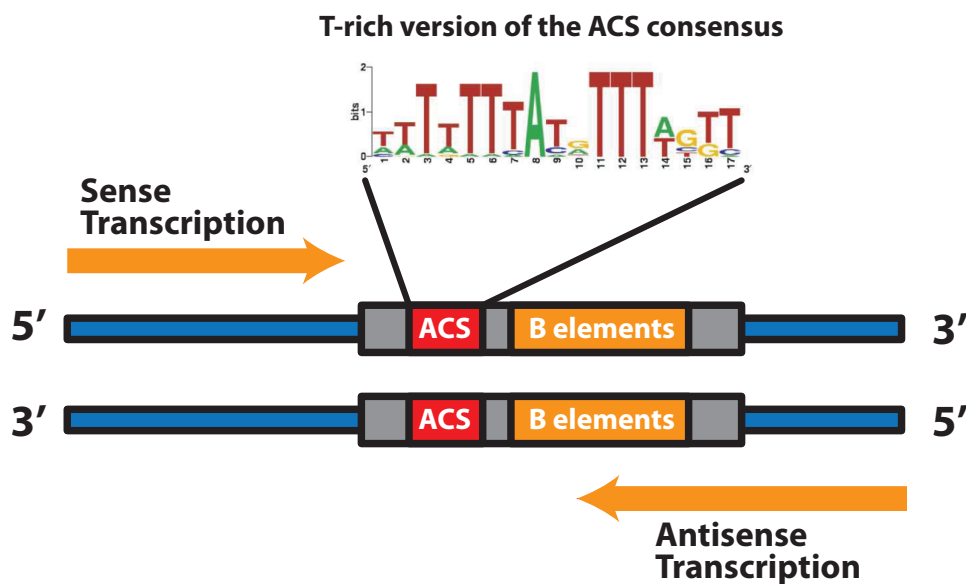
**Figure 8.3: A:** In our reporter system, a tet promoter directs transcription of a fragment of the HSP104 gene, within which ARS305 is embedded with or without the ACS. **B:** Northern blot analysis of the reporter system shows the presence of a short transcript that disappear upon ACS deletion.

In order to test this hypothesis, Julien Gros, post-doc in the lab, performed northern blots using a reporter system. In this system, the sequence of interest is embedded in a fragment of the HSP104 gene, whose expression is then driven by a strong promoter (Fig 8.3A). We tested the sequence of origin ARS305 carrying the deletion of the ACS sequence. Figure 8.3B shows species generated by transcription through the T-rich strand of the ACS in presence or absence of the ACS sequence itself. Strong signal for a short transcript is detectable when the ACS is present, while in its absence, the short transcript disappears to the profit of a longer species. Although

we cannot formally exclude that sequence elements are playing a role in transcription termination of the short species, this results argues in favor of termination by a road-block mechanism.

### 8.3 Transcription Levels Asymmetrically Affect Origin Efficiency

Road-block is known to act as fail-safe termination to protect promoter regions and other loci from transcriptional interference [29]. We reasoned that termination enacted by ORC could have a similar role by protecting the B elements, which are known to aid in pre-RC assembly [192]. Because of the arrangement of ACS and B



**Figure 8.4:** Schematic representation of “sense” and “antisense” transcription relative to the structure of the origin. While sense transcription can be blocked by ORC before reaching the B elements, antisense transcription has no such impairments.

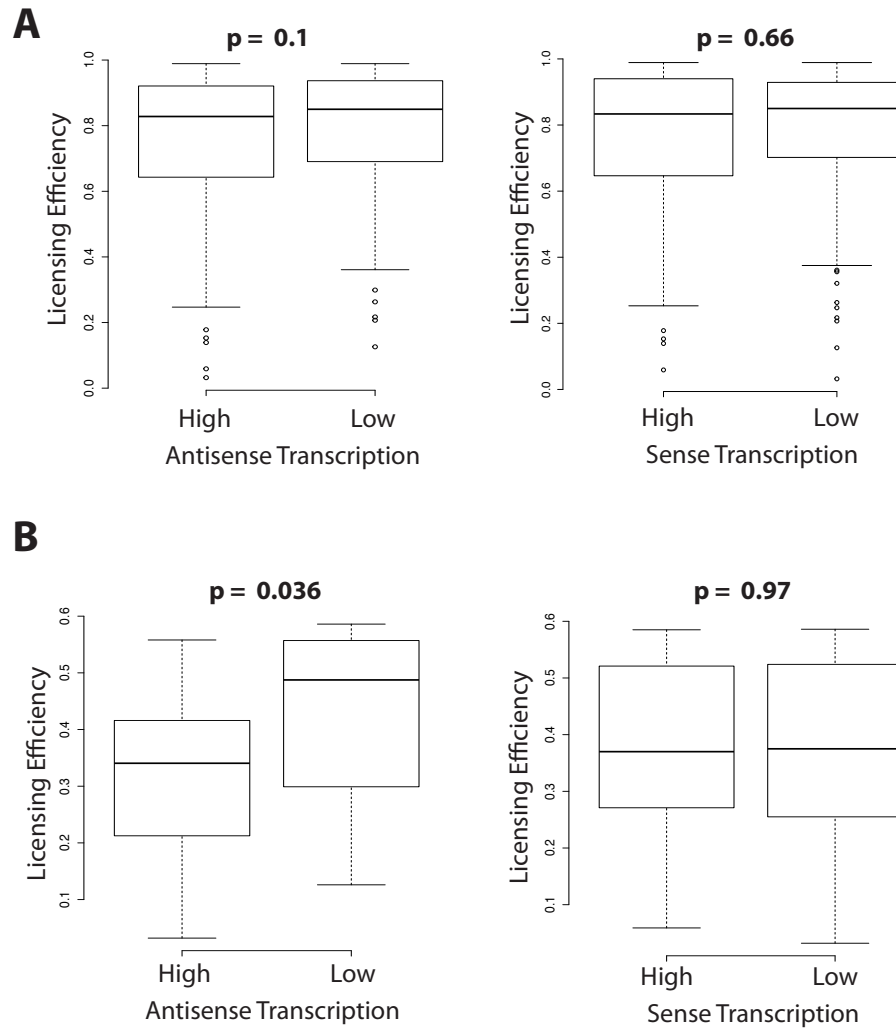
elements within origins, however, ORC would only be able to block transcription coming along the T-rich strand of the ACS before the B elements are invaded. To test this hypothesis, we therefore correlated several measures of origin efficiency with levels of transcription upstream of the ACS on both its T-rich and A-rich strands.

In order to obtain these values, we calculated the average polymerase occupancy in a 100 nucleotide window upstream of the ACS. Transcription levels calculated along the T-rich strand of the ACS were dubbed “sense”, while transcription levels calculated over the A-rich strand of the ACS were dubbed “antisense” (figure 8.4).

As replication occurs in discrete steps, we wanted to know if either sense or antisense transcription were affecting any particular stage of replication initiation. We therefore correlated per-origin estimates of licensing efficiency, firing efficiency, and timing of firing [68] with our estimates of sense and antisense transcription. Our approach was two-fold: for every measure of origin efficiency, we calculated the Pearson’s correlation between it and the levels of sense and antisense transcription. In parallel, we obtained two subpopulations of origins, according to either transcription or efficiency, and compared them using boxplots and t-tests for statistical significance.

### 8.3.1 Licensing Efficiency

Licensing is the first step in DNA replication, however, not all origins are licensed during the cell cycle. Every origin we considered is associated with a value between zero and one, representing the likelihood that the origin will be licensed during the cell cycle [68]. We split the population of origins into two sub-populations according to high and low transcription for both sense and antisense. We then compared the distribution of licensing efficiencies in these two sub-populations. Analysis of the overall populations showed no difference in licensing efficiency whether origins are surrounded by high or low sense or antisense transcription (Fig. 8.5A). A statistically significant anti-correlation between antisense transcription levels and licensing efficiency (pearson’s  $r = -0.15$  with  $p = 0.03$ ) could be observed, but no significant correlation between sense transcription levels and licensing efficiency (pearson’s  $r = -0.04$  with  $p = 0.53$ ). We reasoned that the low levels of endogenous transcription might not be sufficient to affect highly efficient origins. We therefore considered only origins with relatively low licensing efficiency ( $< 0.6$ ) and repeated the experiment (fig 8.5B). A significant difference in the distribution of licensing efficiencies could be



**Figure 8.5:** These boxplots compare the distribution of licensing efficiencies between high- and low-transcription populations. **A:** Boxplots generated using the totality of the origins available to us. high- and low-transcription populations show similar levels of efficiency both according to sense and antisense transcription. **B:** In this experiment, we restricted our boxplots to poorly licensed origins. Higher levels of antisense transcription are now significantly associated with lower efficiency, while high or low sense transcription displays no difference.

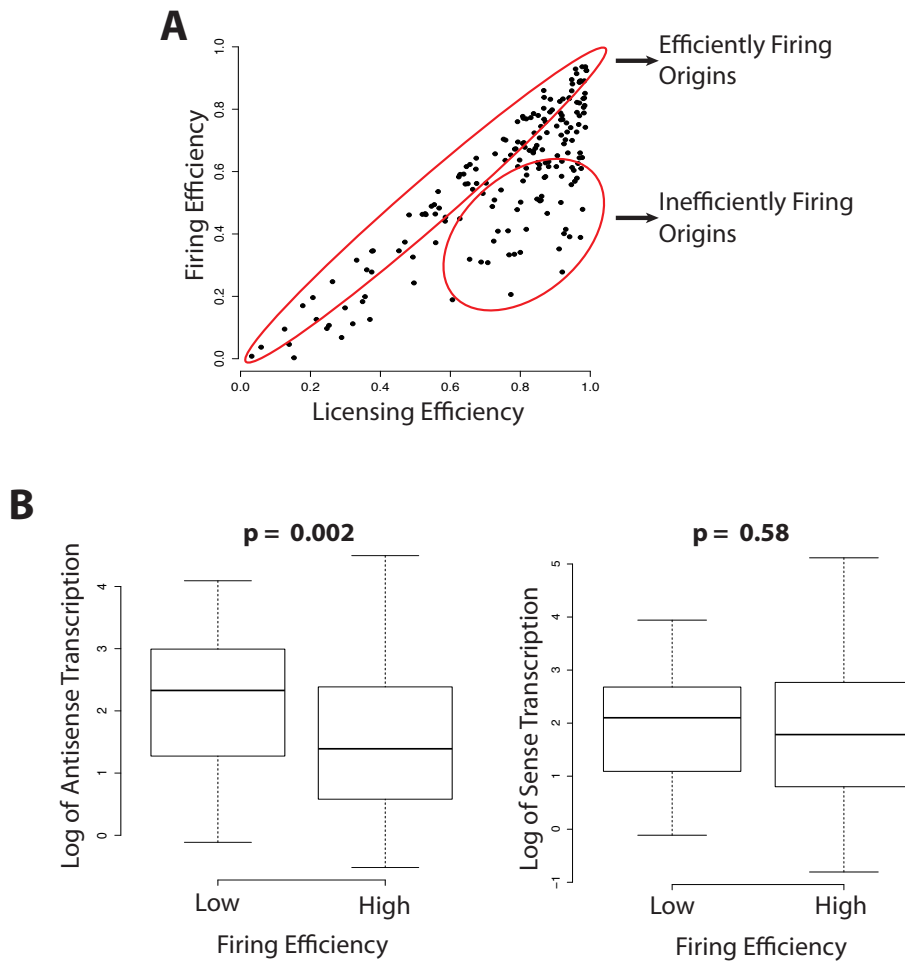


detected between populations with high and low antisense transcription. However, no such a difference could be detected when the two populations were chosen according to sense transcription. These results are supported by the Pearson's correlation coefficients: the anti-correlation between antisense transcription and licensing efficiency is higher (pearson's  $r = -0.32$  with  $p = 0.04$ ), while that between sense transcription and licensing efficiency remains low (pearson's  $r = 0.03$  with  $p = 0.83$ ). Taken together, these results support the notion that physiological levels of transcription antisense to the ACS can negatively affect licensing origins, while sense transcription—regardless of its intensity—does not affect significantly licensing efficiency.

### 8.3.2 Firing Efficiency

Firing is the process that allows licensed origins to activate and begin the replicative process. This step is conditional on the presence of the pre-replication complex, and therefore cannot occur unless the origin has been previously licensed. To define classes of origins with different firing efficiencies, we compared licensing and firing for every origin using published data. As for licensing, the probability of firing has been defined by a value between zero and one [68].

A scatterplot of firing and licensing efficiencies is represented in fig 8.6A. We divided origins into two populations according to their position in this plot. Origins residing around the diagonal are able to fire efficiently, as licensing and firing have similar likelihoods and firing requires licensing. Origins residing below the diagonal, however, fire inefficiently, as their firing efficiency is lower than their licensing efficiency. In figure 8.6B we compare the distribution of antisense and sense transcription levels between these two populations. Inefficiently firing origins display higher levels of antisense transcription relative to efficiently firing origins, however, this relationship is lost when considering sense transcription. These results are supported by Pearson's correlations: antisense transcription is anti-correlated with normalized firing efficiencies (pearson's  $r = -0.18$  with  $p = 0.01$ ) while sense transcription and

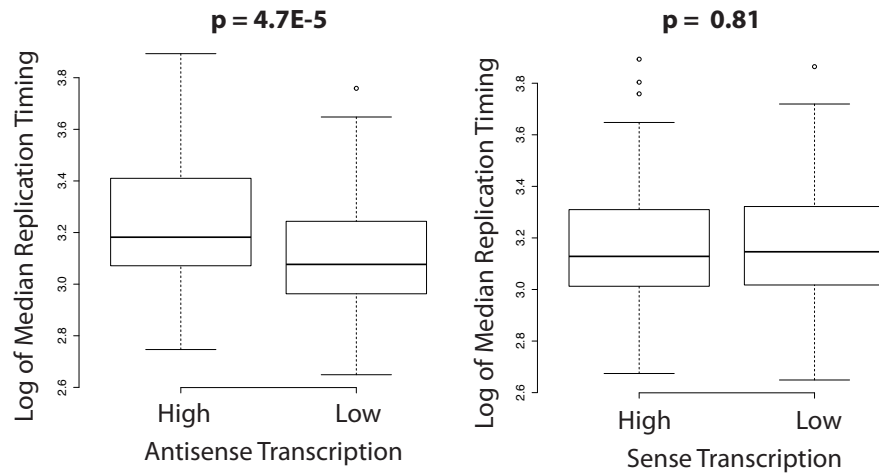


**Figure 8.6:** These boxplots compare the distribution of sense and antisense transcription levels between populations with high and low firing efficiency. **A:** Plot of licensing efficiency vs firing efficiency. Because firing requires licensing, similar efficiency values in these two metrics denote highly efficiently firing origins. We therefore consider those origins close to the diagonal as efficient origins and those below as inefficient. **B:** Boxplots comparing the distribution of sense and antisense transcription levels between origins with low and high firing efficiency. The population with low firing efficiency are significantly associated with higher antisense transcription levels, but this relationship does not hold in the case of sense transcription.

normalized firing efficiencies do not seem to be correlated (pearson's  $r = 0.06$  with  $p = 0.39$ ). Taken together, these results suggest that not only licensing, but also firing is affected by antisense transcription levels, while sense transcription levels have no effect on this step.

### 8.3.3 Timing of Firing

While firing efficiency is a measure of how often a particular origin is able to initiate DNA replication, it does not give information about the elapsed time between the entry in S-phase and activation of the replisome. We wanted to assess whether transcription levels influence the timing of origin firing. In figure 8.7 we compare the distribution of median firing times for high and low, sense and antisense transcription. High antisense transcription is significantly associated with higher median replication times, while no difference in median replication times can be detected between high or low sense transcription levels. These results are also supported by correlations



**Figure 8.7:** These boxplots represent the distribution of median replication times in population with high, low, sense, and antisense transcription. High antisense transcription correlates with higher replication times relative to low antisense transcription. However, sense high and low sense transcription seems to have no effect on replication timing.

that do not rely on subpopulations: antisense transcription levels positively correlate with median replication times (pearson's  $r = 0.19$  with  $p = 0.008$ ) while sense transcription levels show no correlation (pearson's  $r = -0.05$  with  $p = 0.42$ ). These results suggest that even when firing occurs, antisense transcription levels can delay firing, possibly by interfering with the assembly of the replisome, while sense

transcription levels have no impact.

## 8.4 Discussion

In this work, we investigated the relationship between the process of origin specification and that of RNA transcription. We analyzed transcription around replication origins separately on both strands and detected localized increases in polymerase occupancy that coincided with hotspots of transcription termination. We noticed that pausing and termination were arranged asymmetrically relative to the ACS, with a major peak immediately upstream of the ACS in the T-rich orientation (position -25) of and several peaks indicating accumulation of polymerases at different distances (-100 to -125) upstream of the ACS in the A-rich orientation.

We explored the possibility that ORC binding to the ACS might induce roadblock termination at these sites through northern blot analysis of ARS305. This experiment revealed that transcription upstream of the ACS in the T-rich strand orientation is terminated in an ACS-dependent manner. Experiments were also performed to assess the occurrence of termination for transcription entering the ACS from the opposite direction (upstream of the A-rich strand) but the results were not conclusive because a site of NNS-dependent termination was present that masked the possible ACS-dependent termination. At this stage we do not know whether the additional peaks of polymerase pausing at position -100 to -125 upstream of the ACS in the A-rich strand orientation are due to roadblocked polymerases or polymerases paused for other reasons.

However, prompted by the asymmetry revealed by these experiments, we tested the hypothesis that transcription levels could asymmetrically impact origin efficiency depending on origin orientation. We correlated per-origin estimates of licensing efficiency, firing efficiency, and timing of firing with surrounding transcription levels both sense and antisense relative to origins oriented by the T-rich strand of the ACS. High levels of transcription on the antisense strand proved to negatively impact every measure of replication efficiency, while sense transcription had no significant effect.

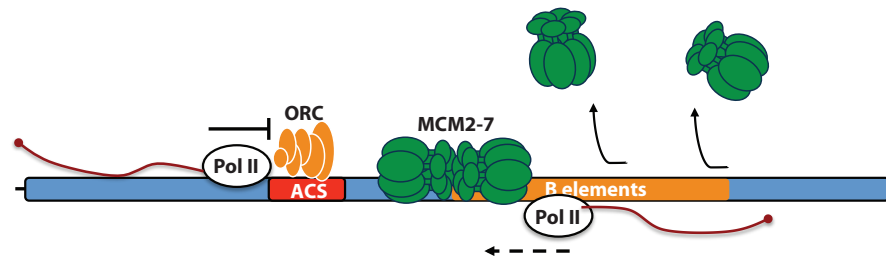
### **8.4.1 Transcription Termination Is a Feature of Replication Origins**

Our meta-site analyses provided insights on the global state of polymerase occupancy and transcription termination around replication origins genome-wide. According to this global view, many origins are associated with distinct peaks of polymerase pausing and transcription termination on both sense and antisense strands. Because of the complexity of the DNA replication process, as well as previous evidence emerging from the literature, we speculate that these termination events protect the origin by preventing transcriptional interference. In accordance with this model, we have evidence that transcription on the sense strand of the ACS is terminated by ORC through road-block, a mechanism already known to protect promoter regions from invading polymerases. This model is supported by preliminary analyses of polymerase occupancy datasets generated in strains defective for either CPF or NNS termination. Both datasets displayed a marked increase in polymerase pausing in the vicinity of ORC, a phenotype consistent with the increased readthrough transcription that is stalled at the site of road-block. While the meta-site analyses provided many elements that suggested road-block by ORC, we could not formally prove that its presence is responsible for the termination. In our case study, ARS305, we show that termination is ACS dependent, but cannot exclude that sequence elements within the ACS could be the determinant for termination.

### **8.4.2 ACS Orientation Determines the Impact of Transcription on Replication Efficiency**

In order to explore the impact of endogenous transcription on DNA replication, we decided to correlate strand specific transcription levels with measures of origin efficiency. Through these analyses we showed that high levels of transcription generally correlate with poor replication performance, however, only transcription entering the origin from upstream of the A-rich strand of the ACS displays such correlations. We propose a model whereby a road-block enacted by ORC is sufficient

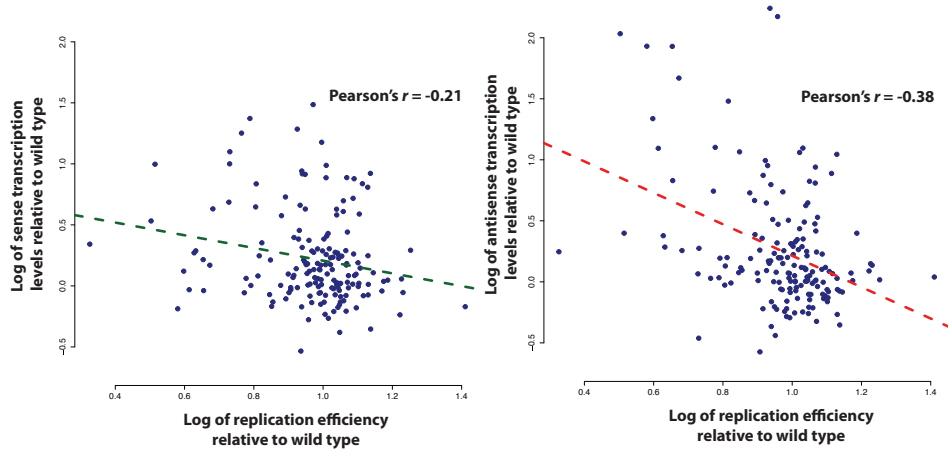
to prevent endogenous levels of RNAPII from elongating into the B elements, thus preventing transcription from interfering with the replicative process (Fig. 8.8). Transcription on the other strand, however, might be terminated less efficiently which might not be sufficient to prevent all incoming polymerases from invading the B elements and affecting one or more DNA replication steps.



**Figure 8.8:** Model of how transcription can asymmetrically affect replication efficiency. While sense transcription can be efficiently blocked by the ACS before reaching the B elements, antisense transcription can invade the B elements more efficiently.

Julien Soudet, one of our collaborators, generated some preliminary data measuring replication efficiency in a strain defective for NNS termination. We calculated Pearson's correlation between antisense transcription levels and replicative efficiency relative to wild type. A strong anticorrelation between the two quantities was observed (Fig. 8.9), implying that stronger antisense transcription relative to wild type is associated with reduced replication activity. Surprisingly, correlation between relative sense transcription levels and relative replicative efficiency also displays a negative trend (Fig. 8.9), albeit lower than in the antisense case. These results suggest that the increased transcription resulting from defects in NNS termination are enough to overcome the road-block and generate defects in replication efficiency independent of the orientation of the origin. However, they also suggest that transcription termination on the sense strand is likely stronger than that on the antisense strand, as it is less associated with poor replicative performance.

Overall, downregulation of replicative activity seems to be a function of the quantity of polymerases that transcribes through the core sequence elements of the origin. However, it is difficult to assess the mechanistic reasons for this phenotype. Tran-



**Figure 8.9:** Scatterplots of relative sense and antisense transcription levels versus relative replication efficiencies. Each axis displays the  $\log_2$  ratio between levels of transcription or replication efficiency in a NNS-defective strain relative to wild type. In this non-physiological condition, both sense and antisense transcription levels anticorrelate with replication efficiency. However, antisense transcription levels remain more strongly associated with poor replicative efficiency.

scription might directly displace or otherwise interfere with elements of the pre-RC. Alternatively, it is tempting to speculate that transcription-dependent nucleosome deposition might interfere with assembly or firing of the replication complex.





# 9

## Dynamics of Nrd1-Nab3 RNA Binding *in vitro* and *in vivo*

In chapter 3 I introduced the various termination pathways known in *S.cerevisiae*. Among them, the NNS pathway is primarily responsible for termination of pervasive transcripts and a limited number of functional non-coding RNAs. Mechanistically, the NNS complex recognizes specific sequence elements on the nascent RNA and, once recruited, the subunit Sen1 is thought to translocate along the RNA and disassemble the elongation complex upon reaching it.

Despite numerous studies, considerable doubt remains on what qualifies NNS terminators *in vivo*. Indeed, while the sequence elements bound by the members of the complex are known, no consistent patterns in number, disposition, or quality emerges from analysis of *in vivo* NNS terminators.

Here we report preliminary analyses of an *in vitro* SELEX experiment performed with a purified Nrd1-Nab3 heterodimer in order to identify the main determinants of heterodimer binding. Additionally, we cross-reference these results with those of an *in vivo* SELEX experiment (Artificial CUT Selection) whose selection criteria is the efficiency transcription termination. In the context of these two experiments we find that particular arrangements and spacings between Nrd1 and Nab3 sites

enhance binding efficiency *in vitro*, but not *in vivo*. Moreover, we identify several clusters of similar sequences that are differentially selected in the two experiments. Finally, we provide evidence that supports two distinct binding modes for nrd1 in binding its cognate sites GUAG and GUAA.

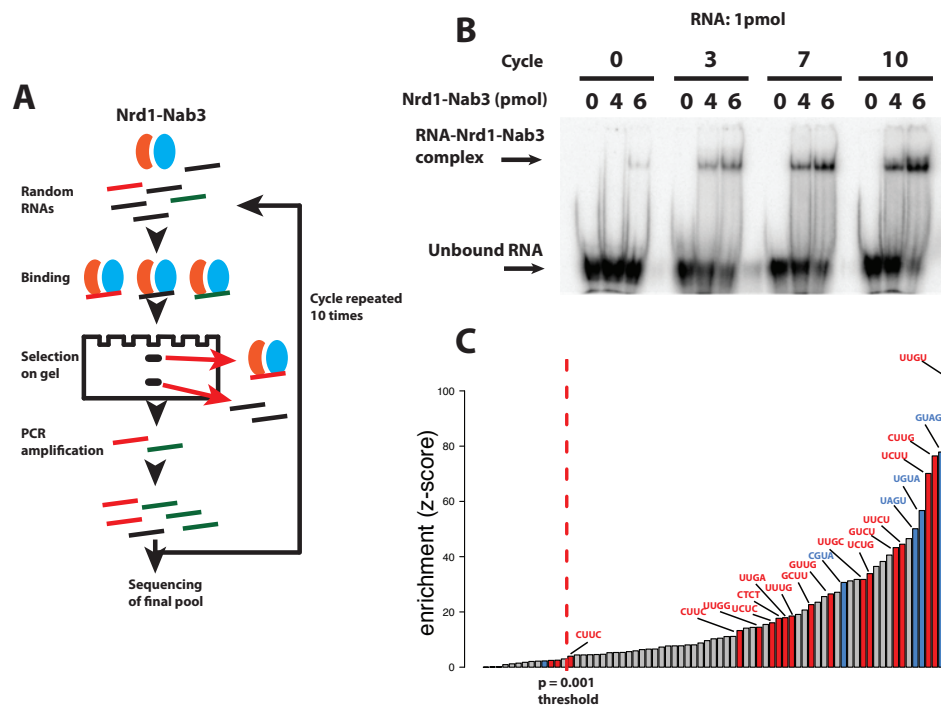
Purification of the Nrd1-Nab3 heterodimer for use in the SELEX experiment was performed by Odil Porrua. The *in vitro* SELEX procedure was performed by Jean-Baptiste Briand. Artificial CUT selection was also performed by Odil Porrua [142]. I performed all the computational analyses and comparisons between the two datasets.

## 9.1 *In vitro* Selection of RNA Sequences With High Affinity for the Nrd1-Nab3 Heterodimer

In order to determine the sequence elements with highest affinity for the Nrd1-Nab3 heterodimer, a SELEX experiment was performed in the laboratory by Jean-Baptiste Briand (Fig. 9.1A).

Full length, HIS-tagged Nrd1 and Nab3 were co-expressed and co-purified from E.coli, obtaining stable heterodimers. The recombinant heterodimer was then incubated with a naïve pool of chemically synthesized RNAs and retained only those that were bound to the Nrd1-Nab3 complex (fig. 9.1B). The selected RNAs underwent reverse transcription, PCR amplification, and *in vitro* transcription, yielding a new pool of sequences. The procedure was iterated for a total of 10 cycles and the final pool of high-affinity binders was submitted to deep sequencing, together with the original naïve pool. The 2000 most represented sequences in the final pool were retained for subsequent analyses.

In order to evaluate the enrichment of specific motifs in the pool of selected sequences, we decided to adapt the Rsat algorithm for oligo-analysis [184, see methods]. This procedure takes into account the nucleotide bias of the naïve pool, comparing the frequency of each motif in the background pool with that encountered in the selected



**Figure 9.1:** **A:** Schematic cartoon of the SELEX procedure. **B:** Electro Mobility Shift Assay (EMSA) performed with different protein concentrations and with sequence pools obtained after different number of SELEX cycles. As the cycles increase, the sequences are more likely to efficiently bind the heterodimer. **C:** Barplot displaying the most enriched motifs obtained through the SELEX procedure. Motifs with at least 3 nucleotides in common with the canonical Nab3 binding sites UCUUG are represented in red. Motifs with at least 3 nucleotides in common with the canonical Nrd1 binding sites GUA[A/G] are represented in blue.

pool and providing an enrichment score. As Nrd1 and Nab3 canonical binding sites have this length, we analysed all 4 nucleotide motifs.

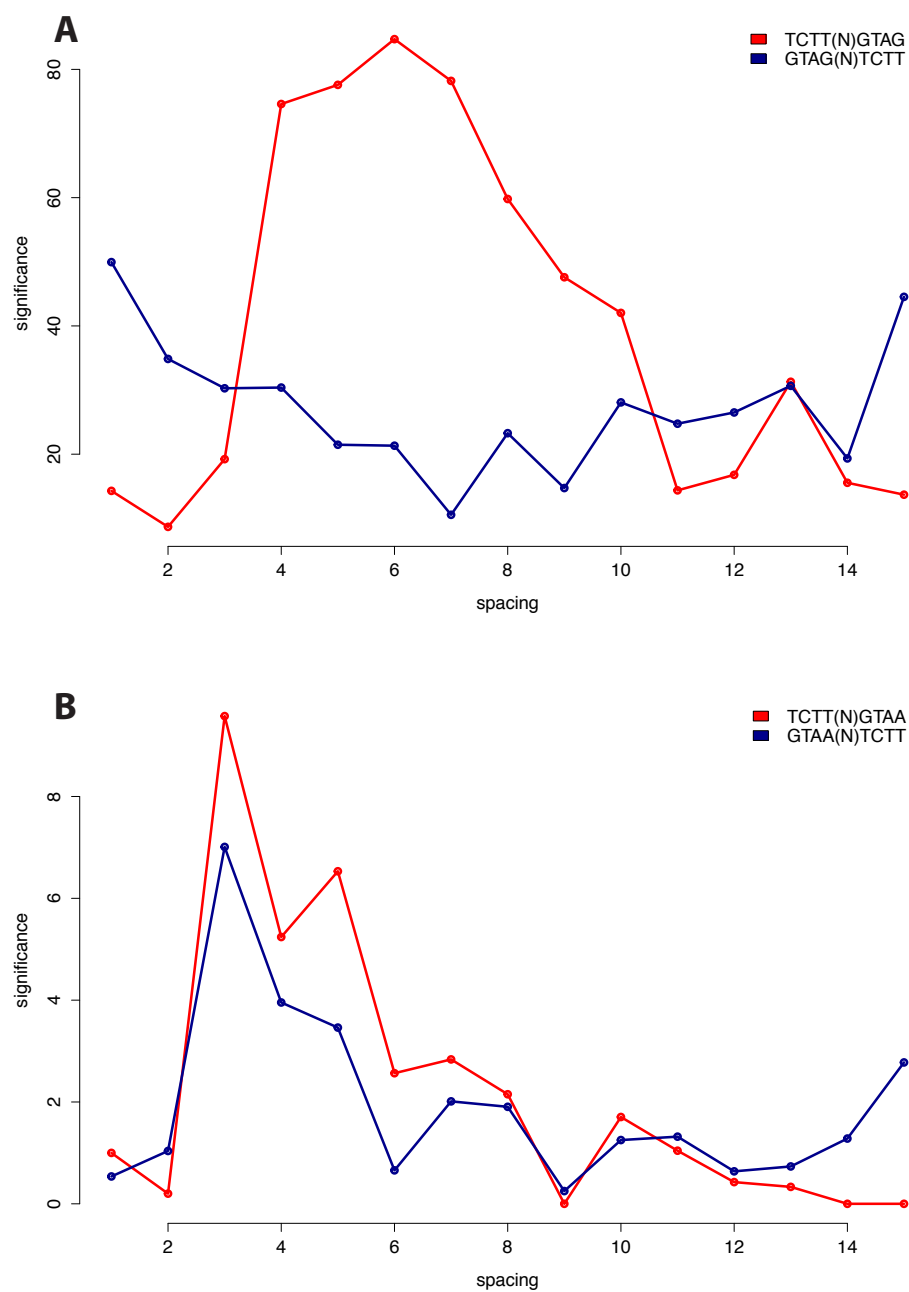
As expected, we managed to identify a large number of known Nab3 and Nrd1 binding sites among the selected sequences (Fig. 9.1C). The canonical sites UCUU and GUAG were among the most enriched motifs, followed by close variants such as CUUG and UGUA. These results suggest that the pool of selected sequences contains high affinity bona fide binding sites for the Nrd1 Nab3 heterodimer.

## 9.2 Arrangement of Binding Sites Influences Heterodimer Affinity *in vitro*

In order to determine whether efficiency of binding could depend on particular arrangements of Nrd1 and Nab3 sites, we analysed the enrichment of motifs containing different arrangements of Nrd1 and Nab3 sites separated by a variable number of random nucleotides. We analysed three canonical binding sites: Nrd1 binding sites GUAG and GUAA, as well as the Nab3 binding site UCUU.

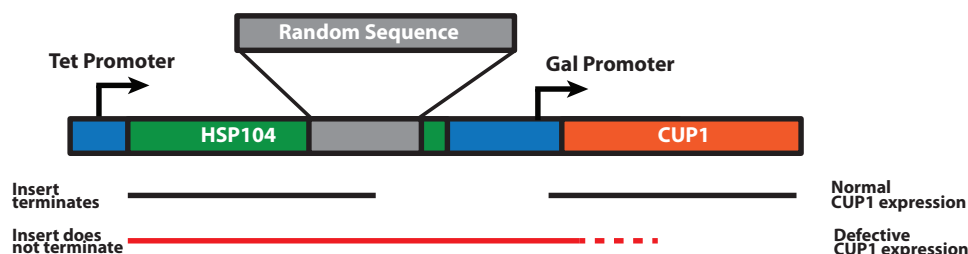
The plot in figure 9.2A shows motif enrichment as a function of spacing between Nab3 binding site UCUU and Nrd1 binding site GUAG. We detect a marked increase in enrichment when UCUU and GUAG are separated by four to ten nucleotides, however, shorter or longer separators cause the significance of the enrichment to drop substantially. When the order of the sites was inverted (GUAG-N-UCUU), no such relationship could be observed. We repeated the same experiment using the Nrd1 binding site GUAA in place of GUAG (Fig 9.2B). Surprisingly, no pattern akin to the one we observed in figure 9.2A could be detected and both site arrangements show very similar enrichment patterns.

In order to assess the importance of these site arrangements within *in vivo* terminators, we decided to repeat these analyses on data coming from a previously published *in vivo* artificial CUT selection [142]. This strategy adopts the same principles of SELEX and allows screening a naïve pool of sequences for efficient termination *in vivo* (Fig. 9.3). The technique relies on a construct containing two strong promoters, pTET and pGAL, arranged in tandem and separated by a test sequence. While the first promoter drives transcription through the randomly selected sequence, the second controls the expression of the CUP1 gene, which allows yeast to grow on copper-containing plates. Transcription from the TET promoter will interfere with the expression of CUP1 unless the test sequence contains an efficient terminator, resulting in copper-sensitive yeast that will not grow on selective medium. After a naïve pool of chemically synthesised sequences was introduced in the construct,



**Figure 9.2:** Significance of motifs with specific site arrangement and spacer length obtained from analysis of the SELEX experiment. Measures of significance were obtained by applying an adapted version of the Rsat algorithm [184, see methods].

yeast underwent two rounds of selection on copper-containing plates. This led to sequencing of inserts containing NNS terminators.



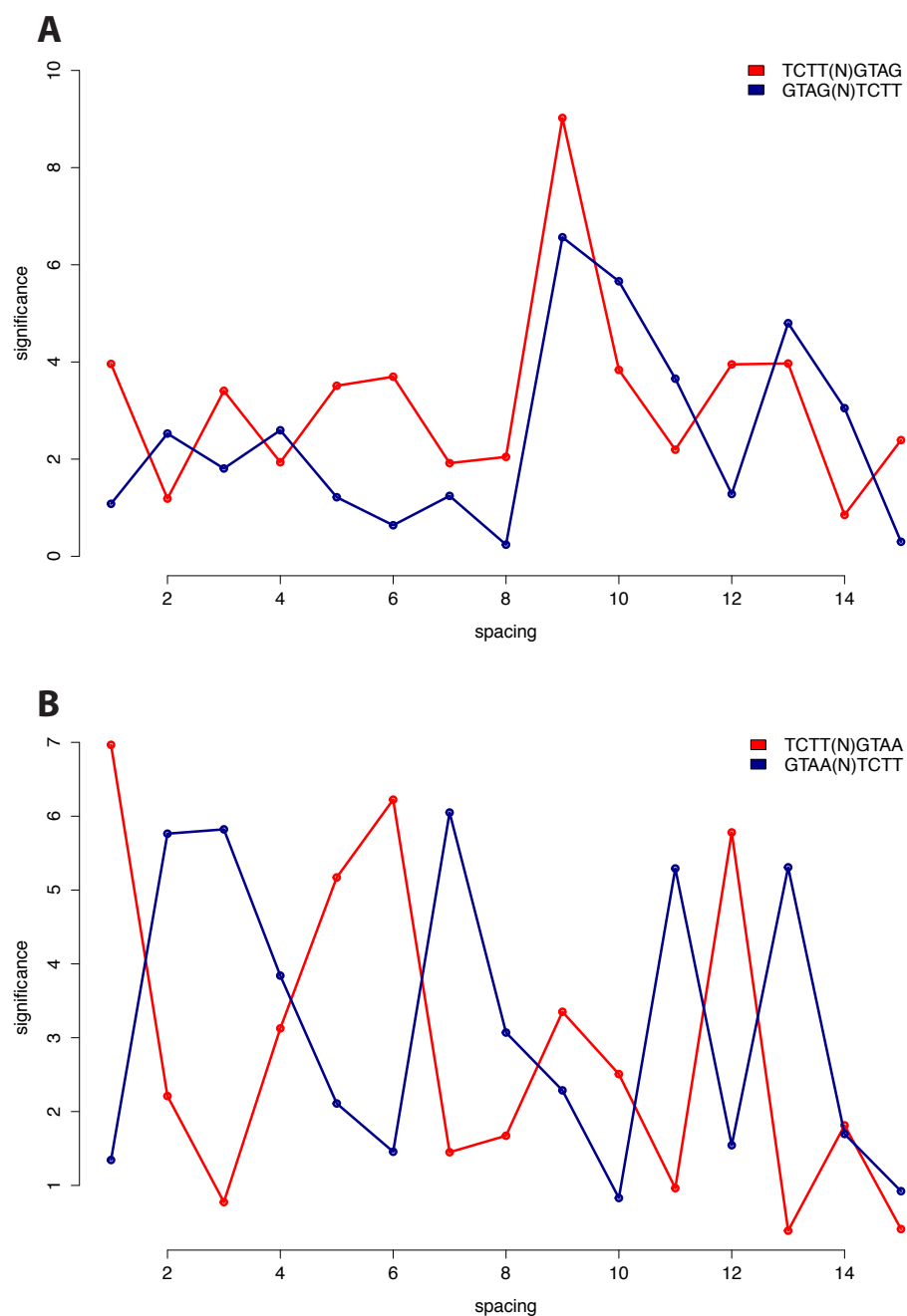
**Figure 9.3:** Cartoon showing the construct used to select sequences according to the efficiency of transcription termination in the artificial CUT selection.

Because of the very similar nature of artificial CUT selection and classical SELEX experiments, results from both techniques could be analysed with the same statistical methods. The two assays, however, significantly differ in environment and selection criteria. While our SELEX experiment relies exclusively on *in vitro* binding of the isolated heterodimer to separate between selected and non-selected pools; artificial CUT selection requires the sequence to be an efficient *in vivo* terminator.

The same analyses conducted on the *in vitro* SELEX winning pool were replicated on the pool of sequences obtained through the *in vivo* artificial CUT selection and are shown in figure 9.4. The enrichment patterns previously observed in fig 9.2A are not replicated in the artificial CUT selection. Spacing and arrangement analysis of UCUU and GUAG in the selected pool of *in vivo* CUT selection shows that no clear spacing-dependent pattern exists and that inverting the order of the sites has little effect. Analysis of UCUU and GUAA, however, reveals a striking alternating enrichment pattern that depends on both spacing and site arrangement (fig 9.4B). The length of the random spacers for which the UCUU-N-GUAA site arrangement is strongly enriched are also the ones for which GUAA-N-UCUU is poorly enriched. Vice versa, the spacer lengths for which GUAA-N-UCUU is strongly enriched, result in poor enrichment when sites are inverted.

Taken together, these results suggest that particular dispositions and spacing of GUAG, GUAA, and UCUU binding sites can affect the binding affinity of the

Nrd1-Nab3 heterodimer and possibly its efficiency in eliciting termination.



**Figure 9.4:** Significance of motifs with specific site arrangement and spacer length obtained from analysis of the artificial CUT selection. Measures of significance were obtained by applying an adapted version of the Rsat algorithm [184, see methods].



### 9.3 Comparison of SELEX and *in vivo* Artificial CUT Selection Unveils Unexpected Dynamics of Nrd1 Binding

In order to obtain a clearer view of the differences between *in vitro* SELEX (RNA binding as selection criterion) and artificial CUT selection (terminator efficiency as selection criterion) we performed motif enrichment analysis and plotted enrichment significance for both experiments (Fig. 9.5A), setting a confidence threshold at  $p = 0.001$ . As expected, the two experiments partially correlate (Pearson's  $r=0.44$ ). This reflects the relationship between binding of the NNS complex to the RNA and subsequent termination. Interestingly, however, several clusters of similar sequences were differentially enriched (Fig. 9.5B).

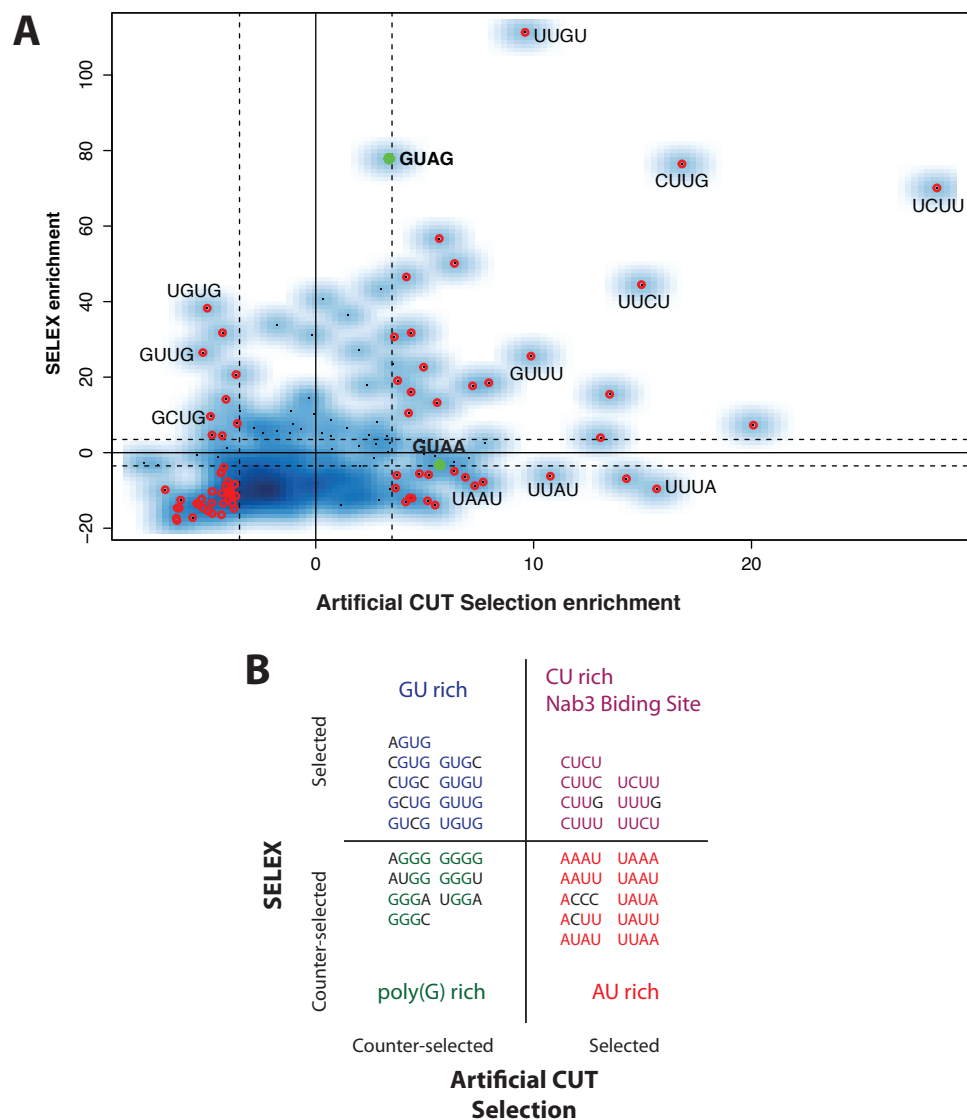
Nab3 binding sites and their variants, along with UC-rich sequences, were strongly selected in both experiments. This is consistent with previous reports that frame Nab3 binding as the most important contributor to overall heterodimer binding affinity.

Heavily G-rich sequences, conversely, were strongly counter-selected in both experiments. This provides evidence that both in the context of the Nab3-Nrd1 heterodimer and *in vivo*, G-rich sequences do not favour binding or termination by the NNS complex.

AU-rich sequences were found to be prevalent both in natural cases and in the artificial CUT selection experiment and were shown to enhance Nrd1 binding [142]. Surprisingly, AU-rich sequences were not enriched in our SELEX experiment and their frequency is significantly lower than expected by chance. A similar but opposite trend was detected for GU-rich sequences. These were enriched in the SELEX experiment, but were counter-selected in the artificial CUT selection.

Interestingly, the enrichment of canonical Nrd1 binding sites in the two experiments mirrors that of GU-rich and AU-rich sequences. In our SELEX experiment, the

more G-rich site, GUAG, was by far the most enriched Nrd1 binding site, but was only moderately enriched in the artificial CUT selection (SELEX enrichment  $p < 1\text{E-}100$ , CUT selection enrichment  $p = 1.9\text{E-}3$ ). Conversely, GUAA is the



**Figure 9.5:** **A:** Scatterplot of SELEX vs artificial cut selection enrichment for all 4 nucleotide motifs. Dashed lines represent  $p = 0.001$  confidence intervals. Significant sequences are highlighted in red. **B:** A schematic view of the enriched and depleted sequences between the two experiments.

most prominent Nrd1 binding site identified in the artificial CUT selection, while it is counter-selected in the SELEX (SELEX enrichment  $p = 0.99$ , CUT selection

enrichment  $p = 1.1\text{E-}7$ ).

Taken together these results confirm the important role of Nab3 as the main contributor to NNS specificity both *in vivo* and *in vitro* and unveil an unexpected dynamic of Nrd1 binding sites selection depending on both context and type of selective pressure.

## 9.4 Nrd1 Binding Sites GUAG and GUAA Possess Different Extended Consensuses

Nrd1's differential affinity for its two major binding sites poses questions regarding its mode of binding. Recent studies proposed that Nrd1's RRM is bipartite, one part has high affinity for AU-rich sequences, while the other prefers GU-rich sequences [7]. According to this study, the two binding surfaces are semi-independent and mutations on one part has only minor effects on the other part's ability to contact its cognate sites.

In order to test the hypothesis that GUAA and GUAG might require two different binding modes, we decided to explore whether these two core motifs have the same preference for flanking nucleotides. Different preferences would indicate that [A] and [G] at position four in the consensus are not simply interchangeable, but impact the conformation of Nrd1 binding to the site.

To achieve this, we measured the nucleotide frequency before and after each of the sites in the SELEX and compared it with the overall nucleotide frequency in the pool of selected sequences. This allowed us to quantify the nucleotide distribution and to determine their over- or under-representation relative to their abundance in the rest of the pool (see methods).  $\text{Log}_2$  of the ratio between the nucleotide frequency flanking Nrd1 sites and the overall frequency in the datasets are shown in figure 9.6A, positive values imply an enrichment around the Nrd1 site, while negative values indicate a depletion, stars indicate a statistically significant depletion or enrichment based on the binomial test (see methods).

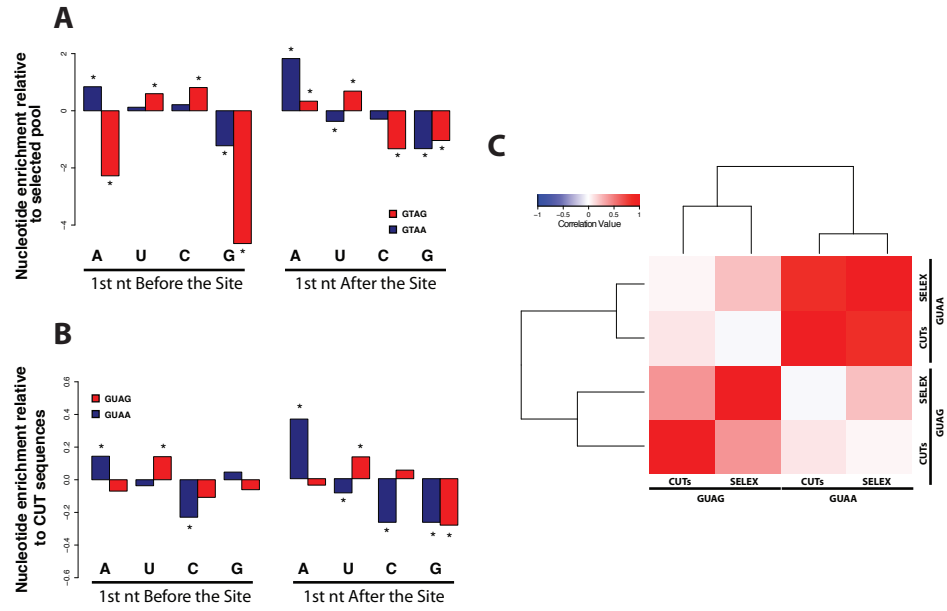
Direct comparison of the extended consensus for these two core Nrd1 binding sites reveals substantial differences. While presence of a G before or after both sites is universally counter-selected, its presence before GUAG is about 10 fold less likely than in the case of GUAA. The two binding sites also substantially differ in their preferences for both preceding and following As. As are heavily enriched following GUAA, while in GUAG only a minor enrichment is detected in this position, giving preference to Us, which are depleted in GUAA.

#### 9.4.1 Confirming the Differences *in vivo*

The stark differences we detected between the two major Nrd1 binding sites in our SELEX experiment prompted us to verify this relationship *in vivo*. To achieve this, we decided to apply the same technique to a pool of CUT sequences extracted from the genome (Fig. 9.6B). CUTs represent the major fraction of NNS terminated non-coding RNAs and are therefore good candidates to test our hypothesis.

As described previously, we analysed the nucleotide frequencies of the position preceding and following the two major Nrd1 binding sites: GUAG and GUAA. The two sites show distinct patterns that globally mirror what was observed in the analysis of the SELEX sequences, although the differences are less stark. GUAA is still significantly associated with As, both in the preceding and following position while GUAG prefers Us. In addition to the similarities, we observed some differences between nucleotide frequencies in CUTs and in the SELEX sequences. Presence of a G before both Nrd1 sites does not seem to be as depleted in CUTs as it is in the SELEX sequences; also, presence of A preceding GUAG seem to be better tolerated in CUTs.

In order to assess the similarity between patterns of nucleotide enrichment across different binding sites and datasets, we decided to perform pairwise correlations. Each combination of site and dataset (e.g. GUAA in CUT sequences) was associated with 8 numerical values corresponding to the relative frequencies of flanking



**Figure 9.6: A and B:** Representation of the comparative enrichment of every nucleotide before and after Nrd1 binding sites GUAG and GUAA. **C:** Heatmap of pairwise Pearson's correlation values between every site-dataset combination. Binding sites prove to correlate well with themselves irrespective of datasets, while different binding sites poorly correlate.

nucleotides (A, T, G, or C, before and after the site). We then calculated the Pearson's correlation between each site-dataset pair. Figure 9.6C shows that pairwise correlation between GUAG and GUAA is very low irrespective of which dataset is used to carry out the analysis. Conversely, SELEX sequences and CUTs agree well on the trend in nucleotide enrichment for the two sites.

Taken together, these results indicate that the extended consensus of the two main Nrd1 binding sites GUAG and GUAA differ substantially both *in vitro* and *in vivo*. This might suggest that Nrd1 has different binding modes depending on the site that is contacted, supporting the hypothesis of a bipartite RRM.

## 9.5 Discussion

In this work, we investigated the binding of the Nrd1-Nab3 heterodimer to its cognate sites in the context of two techniques: an *in vitro* SELEX experiment whose selection criteria is binding; and an *in vivo* CUT selection experiment selecting for sequences able to terminate transcription. We found that a particular arrangement and spacing of UCUU and GUAG sites is strongly enriched in the SELEX experiment, but this enrichment is not mirrored in the CUT selection experiment. Conversely, we found that the sites UCUU and GUAA display a peculiar alternating enrichment pattern in the *in vivo* CUT selection, while no such pattern could be detected in the SELEX experiment.

At the same time, we compared enrichment of 4 nucleotide motifs in both experiments and analyzed clusters of similar sequences across the two experiments. We found that Nab3 sites (and CT-rich sequences in general) are strongly enriched in both experiments, while G-rich sequences are counter-selected for both *in vitro* binding and *in vivo* termination. Additionally, we found that AU-rich sequences—already known to enhance Nrd1 binding and termination *in vivo* [142]—were enriched in the CUT selection, but strongly depleted in our SELEX experiment. Conversely, GU-rich sequences were strongly selected in the selex experiment, but counter-selected in the *in vivo* CUT selection. Curiously, we observed that two versions of the Nrd1 binding site consensus were also following this pattern. Nrd1 site GUAG is strongly enriched in the SELEX and significantly less so in the CUT selection, while GUAA was depleted in the SELEX and strongly enriched in the CUT selection.

We speculate that, depending on the environment and selective pressure, Nrd1 would have higher affinity either for GU-rich sequences or AU-rich sequences and that this altered affinity would be reflected in the choice of binding site: GUAG in the former case, GUAA in the latter. This notion is supported by a recent study where Nrd1 is shown to possess a bipartite RRM that can shift to accommodate either AU-rich sequences or G-rich sequences, potentially resulting in two different

binding modes [7]. To test the hypothesis that GUAG and GUAA might trigger two different binding modalities in Nrd1, we analyzed the frequency of nucleotides flanking these two motifs both in our SELEX experiments and in CUT sequences extracted from the genome. We found that GUAG and GUAA have different preferences for flanking nucleotides and possibly different extended consensus. Pairwise correlation analyses between different datasets and different binding sites reveal that the frequency of flanking nucleotides relative to the background remains similar across datasets, but not across binding sites.

### 9.5.1 Site Arrangement and Spacing Influence Binding Affinity and Termination Efficiency

Our SELEX experiment coupled with the Rsat algorithm for statistical analysis [184] allowed us to determine an enrichment score over the background pool for any sequence. We therefore decided to explore the effect of random nucleotides separating Nrd1 and Nab3 sites on their overall binding affinity. Strikingly, Nab3 site UCUU followed by a spacer of 4-10 nucleotides and then followed by the Nrd1 site GUAG proved to be highly enriched, substantially more than the same sequences spaced by 1-3 or 11+ nucleotides.

We speculate that 4-10 nucleotides represents the most accommodating spacing for the cooperative binding of the heterodimer: a shorter spacing could lead the two proteins to sterically or otherwise interfere with one another, while a longer one would lead the two sites to be bound independently and not cooperatively because of the excessive distance. In our analyses, we also noticed that swapping the sites of Nrd1 and Nab3 did not result in a similar enrichment. This suggests that binding is directional, and that the heterodimer can effectively bind both a Nrd1 and Nab3 site at the same time only if they are present in a particular order. However, it is possible that these enrichment patterns only hold *in vitro* or under these particular conditions, as we found no *in vivo* evidence of this arrangement being preferred.

Another perplexing finding regards the results of this experiment when performed on

the alternative Nrd1 binding site GUAA. While we could not detect any noteworthy pattern when UCUU-N-GUAA is analysed in the context of the SELEX experiment, the same analysis performed on the artificial CUT selection resulted in a striking alternating pattern that depends on both site arrangement and spacer length. At this stage it is difficult to speculate on a reason that would cause this pattern would emerge. However, we think that a fundamental difference in the binding mode of Nrd1 to its cognate sites GUAA and GUAG could affect the overall binding patterns of the heterodimer. Alternatively, we must consider the caveat that SELEX and artificial CUT selections are widely different experiments and that technical or indirect effects could play a substantial role in influencing sequence enrichment dynamics.

### 9.5.2 Comparison of SELEX and *in vivo* CUT Selection Reveals Clusters of Differentially Enriched Sequences

We compared the enrichment of all 4 nucleotide motifs in the SELEX and *in vivo* CUT selection and analyzed which clusters of sequences were enriched or depleted in both experiments, as well as determining the differentially enriched sequences. Because of the different environment and selection criteria for these two experiments, we could speculate on the role that each of these sequence clusters.

We found that UC-rich sequences—among which the Nab3 binding consensus features prominently—were universally enriched in both experiments. This result is consistent with the role of Nab3 in NNS termination, as well as previous reports that put Nab3 binding as the main contributor to heterodimer binding [22]. Second, we found that heavily G-rich sequences were counter-selected in both experiments. Because G-rich sequences are not conducive to Nab3 binding and, to a lesser extent, to Nrd1 binding, we propose that these sequences are not competent for heterodimer binding, and therefore cannot elicit termination, resulting in a depletion in both experiments. Finally, we found that AU-rich sequences were selected in the CUT selection, but counter-selected in the SELEX, while GU-rich sequences followed



the opposite trend, being enriched in the SELEX but counter-selected in the CUT selection. It is tempting to speculate that AU-rich sequences could be required only for termination and not for binding, however, this leaves open the question of why GU-rich sequences are selected for binding *in vitro*, but not for termination *in vivo*. Another possibility is that Nrd1 could differentially interact with these sequences in the two experiments. Previous *in vitro* studies reported that Nrd1 possesses two different binding domains, one with high affinity for G-rich sequences, and with high affinity for AU-rich sequences. We speculate that these two binding domains have different binding modes and binding strengths. The evidence we found is consistent with a model where Nrd1 GU-binding domain is stronger than the AU-binding domain, leading to counter-selection of AU-rich sequences in the SELEX experiment, where heterodimer binding is the only selection criteria. Conversely, we speculate that the AU-rich binding mode, although weaker, is somehow more conducive to transcription termination, possibly due to structural rearrangements that facilitate the process.

In order to further support this model, we tested whether GUAG (enriched in SELEX) and GUAA (enriched in CUT selection and in CUT sequences *in vivo*) had different preferences for flanking nucleotides. This would suggest that GUAG and GUAA are not merely two alternative versions of the same consensus, but make a different set of contacts with Nrd1's RRM. Our results support this view, as the two sites display different flanking nucleotide preferences both in the SELEX experiment and in CUT sequences extracted from the genome.

# Methods

## Cell growth and UV Cross-Linking

For all datasets, 2L of yeast cells expressing Rpb1-HTP tag were grown at 30°C to OD600=0.6 in DO-Trp (drop out without tryptophan). For Anchor Away strains, rapamycin was added at OD600=0.4 for two hours to a final concentration of 1 $\mu$ g/ml. For rna15 experiment, cells were grown at 25°C to OD600=0.6. 1 volume of media preheated at 30 or 37°C was added and cultures were incubated at 30 or 37°C for 1 hour and half. Cells were submitted to UV crosslink in a UV lamp system (Megatron) for 50 seconds. Cells were harvested by centrifugation, washed in cold PBS and resuspended in 2.4 volume / g of cells of TN150 buffer (50 mM Tris pH 7.8, 150 mM NaCl, 0.1% NP-40 and 5 mM beta mercaptoethanol) supplemented with protease inhibitors (cOmplete<sup>TM</sup>, Mini, EDTA-free Protease Inhibitor Cocktail). This suspension was flash frozen in droplets and cells were mechanically broken using the Mixer Mill MM 400 by doing 5 cycles of 3 minutes at 20 Hz. Resulting powder was kept at -80°C.

## RNA-Protein Complexes Purification

Powders were defrozen and the resulting extracts were treated for one hour at 25°C with DNase I (165U/g of cells) to solubilize the chromatin and then clarified by centrifugating 20' at 20000g at 4°C. IgG and nickel purification steps were performed as previously described in the original protocol [61] with minor changes: i) IgG purification steps has been realized using M-280 tosylactivated dynabeads (15mg per

samples) coupled with rabbit IgG. ii) To increase the stringency, washes following both IgG and nickel purification steps were done at high salt concentration (1 M NaCl). After over-night binding on nickel columns, sequencing adaptors were added on the RNA as described in the original procedure. Because we work on nascent transcripts, some changes were required to fit with our purpose: For instance, since nascent transcripts contain a 3' hydroxyl group that is protected inside the polymerase, we decided not to do the dephosphorylation step usually required because of the RNase treatment that leads to 3' phosphate transcripts. Also, we modified both 5' and 3' adaptors in order to sequence RNA molecules from the 3' end so that we can determine the exact position of the polymerase. The 3' ligation has been realized with T4 rnl 2 truncated K227Q enzyme (NEB) instead of classical T4 RNA ligase. RNA-protein complexes were eluted in 400  $\mu$ l of elution buffer (50mM Tris pH 7.8, 50mM NaCl, 150mM Imidazole, 0.1% NP-40, 5mM beta mercaptoethanol). Eluates were concentrated in Vivacon® ultrafiltration spin columns 30 KDa MWCO to a final volume of 120  $\mu$ l. A protein fractionation step was performed using the Gel Elution Liquid Fraction Entrapment Electrophoresis (Gelfree) system from Expedeon. RNA pol II containing fractions were treated with 100  $\mu$ g of proteinase K in a buffer containing 0.5% SDS. RNA was purified by two phenol chloroform isoamyl and one chloroform steps and precipitated in ethanol. RNA were reverse transcribed using the Reverse transcriptase IV (Invitrogen)

## Library Preparation

Between half to  $\frac{3}{4}$  of the cDNA solution was used to performed multiple PCR reactions in a final volume of 25  $\mu$ l using the following conditions: 0.4  $\mu$ M of each primers 0.2 mM dNTP, 2.5 U LA Taq DNA polymerase from Takara, 1X LA PCR Buffer II and 2  $\mu$ l of cDNA per reaction with the programme: 95°C 2', 95°C 30", 58°C 45", 72°C 1', 72°C 5'. PCR were pooled and treated with 200 U of Exonuclease I (NEB) per milliliter of PCR reaction for 1 hour at 37°C. After Exonuclease I inactivation for 20' at 80°C, DNA was purified on PCR clean

up columns (NucleoSpin® Gel and PCR Clean-up, Macherey-Nagel) and sent to sequence using Illumina technology (NextSeq sequencing).

## Dataset Processing

### CRAC

Samples were demultiplexed using the pyBarcodeFilter script from the pyCRAC utility suite [188]. Subsequently, the 3' adaptor is clipped with Cutadapt [117] (-a TGG AATTCTCGGGTGCCAAGGATCTCGTATGCCGTCTTCTGCTTG -m 10) and the resulting insert is quality trimmed from the 3' end using Trimmomatic rolling mean clipping [14] (window size = 5, minimum quality = 25). At this stage the pyCRAC script pyFastqDuplicateRemover is used to collapse PCR duplicates and ensure each insert is represented only once. Each unique insert in our library is associated with a 6 nucleotides random tag within the 5' adaptor. During Demultiplexing, pyBarcodeFilter retains this information in the header of each sequence. This information is used at this stage to better discern between identical inserts and PCR duplicates of the same insert. The resulting sequences are reverse complemented with Fastx\_reverse\_complement (part of the fastx toolkit, [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) and mapped to the R64 genome [27] with bowtie2 (-N 1 -f) [101].

### RNA-seq

Samples are quality trimmed with trimmomatic (see above) and mapped to the R64 genome with bowtie2 with default options.

## Metagene Analyses

For each site included in the analysis we extracted the polymerase occupancy values corresponding to a 5kb window centered on the site. Each position (and its associated RNAPII occupancy value) in this window is then referenced by its relative distance to the ACS, ranging from -2500 to 2500. I calculated the median over all the values associated with the same relative position, and this median represents the value for that position in the final aggregate plot.

In certain cases, the median could not be used because of issues of depth. In this case we used the mean to summarize the data. When using the mean we eliminated extreme outliers to limit noise. For each relative position, we excluded from the analysis every value that was above the mean + 5 standard deviations calculated over all the values at that position.

## Boxplots in fig 7.5C

In order to represent the loss of polymerases due to termination, I calculated the average polymerase occupancy 100 nucleotide across the poly(A) site of all genes upstream of rap1 sites. This value was then divided by the average polymerase occupancy signal across the whole body of the gene. The overall distribution of these values for several datasets was compared with boxplots and t-tests.

## Metagene Analysis of RNAPII Occupancy Around Origins

For each origin included in the analysis I identify the beginning of the ACS as the anchor point (the fixed reference around which all origins are aligned). I then extract the polymerase occupancy values corresponding to a 5kb window centered

Figure	N sites	Sites Origin	Dataset Origin	Notes
7.4A	272	[93]	This Study	Only sites with a gene within 300bp downstream of the site were used
7.4B	294	[203]	This study	we selected the top 500 sites of the "fast" category and analyzed in both orientation. only sites that had an annotated transcript within 300 bp upstream of the site were used.
7.4C	117	[203]	This study	Only sites with a gene within 300 bp upstream of the site were analyzed
7.4D	161	[93]	This study	Only sites with a gene within 300 bp upstream of the site were analyzed
7.5	161	[139]	This Study	I considered the P(A) site of a gene the position with the highest TIF-seq signal within the first 500 bp downstream of the stop codon of that gene. P(A) sites with a Rap1 site within 300 bp downstream (rap1 sites defined according to [93]) were considered for this analysis.
7.9	variable	[33]	[164]	sites for multiple DNA binding factors were considered. Metagene analyses were performed (see above) on all factors using WT data.
7.10	16	centromere annotation	[164]	NA
7.11	259	tRNA annotation	[164]	NA
7.5	117	[203]	[164]	NA
7.6	30	[203]	This Study	we selected the top 500 sites of the "fast" category and analyzed in both orientation. only sites that had an annotated transcript within 300 bp upstream of the site were used.
7.8Top	12	[93]	[164]	We considered sites with CUTs within 300 nucleotodes upstream and with a minimum average coverage of 5 in a 100 bp window upstream of the site.
7.8Bottom	152	[203]	[164]	We considered sites with NNS and antisense transcripts within 300bp upstream of the Reb1 site

**Table 9.1:** Table of metagene analyses. N sites represents the number of sites used in the analysis. "Sites Origin" lists the origin of the sites used for the analysis. Dataset Origin lists the origins of the datasets used, "This study" is used when the datasets were generated in the lab. "Notes" adds details on the selection of the sites or usage of the datasets when needed.

on the ACS for each of the assayed origins. Each position (and its associated RNAPII occupancy value) in this window is then referenced by its relative distance to the ACS, ranging from -2500 to 2500. I calculated the median over all the values associated with the same relative position, and this median represents the value for that position in the final aggregate plot.

To produce this plot, I used a wild type RNAPII parclip dataset [164]. I used a subset of 135 origins [133] that were surrounded by either convergent or tandem features [197]. the final result was smoothed using the `supsmu` R function [145] with a bandwidth of 0.01.

## Metagene Analysis of Termination Events

This metagene analysis was carried out as above, with minor differences. Instead of considering the full value associated with every position, I only considered whether there was a value or not. Presence of a value was considered as a 1, while absence was considered as a 0. Additionally, instead of calculating the median over all values present at the same relative position, I summed them and then divided this number by the total number of origins considered.

For this graph I used a polyadenylated 3' end dataset [191] and used 227 origins [133].

## Per-Origin Estimate of Sense and Antisense Transcription

In order to calculate the amount of average transcription incoming towards the ACS in both the T-rich strand and the A-rich strand, I considered each strand for every

origin and calculated the average RNAPII occupancy signal in a 100 bp window upstream of the ACS.

I used a wild type RNAPII parclip dataset for this purpose [164].

## Statistical Analysis of the Effect of Transcription on Replication Efficiency

I obtained published per-origin estimates of licensing efficiency, timing efficiency and timing of firing [68]. We considered only origins for which 1) ACS annotation was present, 2) estimates of replication efficiency were available, and 3) transcription levels could be calculated. A total of 190 origins were used for these experiments.

I use t-tests to compare the distribution and calculate p-values of different populations (boxplots). Correlations between populations were calculated with Pearson's product moment correlation coefficient. The respective p-values were calculated with the appropriate correlation test (`cor.test()` in R).

### Licensing Efficiency

I divided the total of 190 origins in two equally populated sets according to their sense and antisense transcription levels. I then compared the licensing efficiencies of these two populations. I decided to redo the experiment using only poorly licensed origins. I therefore eliminated all origins with licensing efficiency above 0.6. this left 43 origins.



## **Firing Efficiency**

In this experiment I selected efficiently and inefficiently firing origins according to firing efficiency relative to licensing efficiency. We normalized every firing efficiency by its own licensing efficiency and defined as efficient those origins that had a resulting score higher than 0.66, inefficient if lower.

## **Timing of Firing**

For this experiment I split the total of 190 origins into two halves according to transcription and compared the two population according to the distribution of their timing efficiencies .

## **SELEX and Artificial CUT Selection**

**SELEX** The SELEX experiment was performed in the lab by Jean-Baptiste Briand. The 2000 most represented unique sequences of the final selected pool were kept for further analysis. 200,000 sequences were kept from the naïve pool in order to calculate background distributions.

**Artificial CUT Selection** Artificial CUT selection data was performed by Odil Porrua in the lab [142]. a total of 1000 sequences from the sequencing of the final selected pool was kept for analysis. 200,000 sequences were kept from the naïve pool in order to calculate background distributions.

## Algorithm for Motif Analysis

In order to calculate the enrichment of a specific motif in the selected pool relative to the background nucleotide distribution of the naïve pool of sequences, we decided to employ an algorithm proposed by J. van Helden [184].

Let  $M$  be an RNA motif of length  $l$ . The frequency of this motif in the naïve pool then is:

$$F_{naive}(M) = \frac{occ(M)}{\sum_{i=1}^S L_i - l + 1} = \frac{occ(M)}{T} \quad (9.1)$$

Where  $F_{naive}(M)$  is the frequency of  $M$  in the naïve pool,  $occ(M)$  are the occurrences of  $M$  over all naïve sequences,  $S$  is the total number of sequences, and  $L_i$  represents the length of the  $i$ th sequence.  $T$ , therefore, represents the total number of possible positions that can accomodate motif  $M$  across all sequences in the pool. We will use the frequency  $F_{naive}(M)$  as the probability of observing  $M$  in the selected pool under the assumption that no selection has taken place.

The probability of observing exactly  $n$  occurrences of  $M$  in the selected pool is estimated by the binomial formula:

$$P(occ(M) = n) = \frac{T!}{(T-n)! \times n!} \times (F_{naive}(M))^n \times (1 - F_{naive}(M))^{(T-n)} \quad (9.2)$$

Consequently, the probability to observe  $n$  or more occurrences of motif  $M$  within the selected pool is:

$$\begin{aligned}
P(occ(M) \geq n) &= \sum_{j=n}^T P(occ(M) = j) \\
&= 1 - \sum_{j=0}^{n-1} P(occ(M) = j)
\end{aligned} \tag{9.3}$$

Substituting the number of detected occurrences of  $M$  in the selected pool within 9.3 results in the probability of that number of occurrences emerging by chance given the nucleotide bias of the naive pool.

On this basis, we can define a significance coefficient:

$$Sig = -Log_{10}[P(occ(M) \geq n)] \tag{9.4}$$

This coefficient was used to assess the enrichment of Nrd1 and Nab3 sites with spacers of different length.

## Comparison of Motifs Between SELEX and Artificial CUT Selection

Both the SELEX experiment and the artificial CUT selection follow the same selection principle. A pool of random sequences is subjected to cycles of selection according to variable criteria. The final pool can then be compared to the starting pool of sequences in order to determine enrichment or depletion of specific motifs.

To compare enrichment for all motifs in the two experiments, i analyzed their starting and the final pool with Rsat [1]. I then plotted the z-scores for each motif in figure 9.5. In order to determine significantly enriched motifs, i calculated p-values based on z-scores using the R environment:

```
pvalue = 2 * pnorm(-abs(zscore))
```

I then corrected the p-values for multiple hypothesis testing using the Benjamini-Hochberg correction [70]. After the correction, only motifs with a p-value lower than 0.001 were considered enriched or depleted.

## **Analysis of Nucleotides Flanking GUAG and GUAA**

In this experiment i wanted to assess the over- or under-representation of specific nucleotides flanking GUAG or GUAA in two datasets: the final pool of the SELEX experiment, and a pool of CUT sequences extracted from the genome.

In order to accomplish this, i compared the frequency of specific nucleotides surrounding GUAG and GUAA with the overall frequency of these nucleotides within the pool. The  $\text{Log}_2$  of these ratios is represented in figure 9.6. In order to calculate the statistical significance of the enrichment/depletion, i calculated p-values based on the binomial distribution (binomial test).

## **Correlating the Results Obtained in the Two Datasets**

Because of the environment, the amount of selective pressure, and the higher number of sequences, enrichment scores for Nrd1 and Nab3 binding motifs are much higher in the SELEX experiment than in the artificial CUT selection. This constitute a problem when trying to apply classical clustering approaches in order to determine the similarity of flanking nucleotide enrichment patterns. These techniques rely on euclidian distance between patterns to determine similarity, which is heavily biased by the substantial difference in enrichment values.

I wanted to show that despite overall differences in the magnitude of enrichment scores, the patterns of flanking nucleotides enrichment and depletion hold well even across different datasets. I therefore decided to use Pearson's correlation as a measure of similarity. This ensures that scale is a non-factor in the assessment of

overall similarity between the patterns.

## Glossary

ACS	ARS Consensus Sequence
ARS	Autonomously Replicating Sequences
CDE	centromere DNA element
CF1A	Cleavage Factor 1A
CF1B	Cleavage Factor 1B
CID	CTD Interaction Domain
CMG	Cdc45-MCM-GINS
CPF	Cleavage and Polyadenylation Factor
CPF-CF	Cleavage and Polyadenylation Factor/Cleavage Factor I
CTD	C-Terminal Domain
CUTs	Cryptic Unstable Transcripts
FACT	<u>F</u> acilitates <u>c</u> hromatin <u>t</u> ranscription
GG-NER	global genome nucleotide excision repair mecha- nism
GRF	General Regulatory Factors
HATs	Histone Acetyl-Transferases
HDACs	Histone De-Acetylases

MCM2-7	minichromosome maintenance proteins
NFR	Nucleosome Free Regions
NGD	No-Go Decay
NIM	Nrd1 Interaction Motif
NMD	Non-sense Mediated mRNA Decay
NNS	Nrd1-Nab3-Sen1
NSD	No-Stop Decay
NUTs	Nrd1-dependent Unterminated Transcripts
ORC	Origin Recognition Complex
PIC	Pre-Initiation Complex
pre-RC	pre-replication complex
rDNA	ribosomal DNA
RNAPI	RNA Polymerase I
RNAPII	RNA Polymerase II
RNAPIII	RNA Polymerase III
RP	ribosomal protein
RRM	RNA Recognition Motif
rRNA	Ribosomal RNA
RUTs	Reb1-dependent Unstable Transcripts
SAGA	Spt-Ada-Gcn5-Acetyl transferase
SELEX	<u>S</u> ystematic <u>E</u> volution of <u>L</u> igands by <u>E</u> xponential enrichment

SGD	Saccharomyces Genome Database
snoRNAs	Small Nucleolar RNAs
snRNAs	Small Nuclear RNAs
SUTs	Stable Untranslated Transcripts
TBP	TATA Binding Protein
TEC	Transcription Elongation Complex
TRAMP	<u>Tr</u> f4/ <u>A</u> ir2/ <u>M</u> tr4p <u>P</u> olyadenylation
TSS	Transcription Start Site
XUTs	Xrn1-dependent Unstable Transcripts



## Bibliography

- [1] ALEJANDRA, M., DEFRANCE, M., SAND, O., HERRMANN, C., A, C. J., DELERCE, J., JAEGER, S., BLANCHET, C., VINCENS, P., CARON, C., STAINES, D. M., BRUNO, C., ARTUFEL, M., LUCIE, C., HERNANDEZ, C., THIEFFRY, D., MORGANE, T., AND VAN HELDEN, J. RSAT 2015: Regulatory sequence analysis tools. *Nucleic Acids Res.* *43*, W1 (2015), W50–6.
- [2] ALICIA, S., CARLOS, L., KRIMER, D. B., SCHVARTZMAN, J. B., AND HERNÁNDEZ, P. Transcription termination factor reb1p causes two replication fork barriers at its cognate sites in fission yeast ribosomal DNA in vivo. *Mol. Cell. Biol.* *24*, 1 (2004), 398–406.
- [3] APARICIO, T., GUILLOU, E., COLOMA, J., MONTTOYA, G., AND MÉNDEZ, J. The human GINS complex associates with cdc45 and MCM and is essential for DNA replication. *Nucleic Acids Res.* *37*, 7 (2009), 2087–95.
- [4] APOSTOL, B., AND GREER, C. Copy number and stability of yeast 2  $\mu$ -based plasmids carrying a transcription-conditional centromere. *Gene* *67*, 1 (1988), 59–68.
- [5] ARIGO, J. T., CARROLL, K. L., AMES, J. M., AND CORDEN, J. L. Regulation of yeast NRD1 expression by premature transcription termination. *Mol. Cell* *21*, 5 (2006), 641–51.
- [6] ARIGO, J. T., EYLER, D. E., CARROLL, K. L., AND CORDEN, J. L. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins nrd1 and nab3. *Mol. Cell* *23*, 6 (2006), 841–51.
- [7] BACIKOVA, V., PASULKA, J., KUBICEK, K., AND STEFL, R. Structure and semi-sequence-specific RNA binding of nrd1. *Nucleic Acids Res.* *42*, 12 (2014), 8024–38.
- [8] BADIS, G., CHAN, E. T., VAN BAKEL, H., LOURDES, P., TILLO, D., TSUI, K., CARLSON, C. D., GOSSETT, A. J., HASINOFF, M. J., WARREN, C. L., GEBBIA, M., TALUKDER, S., YANG, A., MNAIMNEH, S., TERTEROV, D., COBURN, D., LI YEO, A., YEO, Z. X., CLARKE, N. D., LIEB, J. D., ANSARI, A. Z., NISLOW, C., AND HUGHES, T. R. A library of yeast transcription factor motifs reveals a widespread function for rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell* *32*, 6 (2008), 878–87.
- [9] BASEHOAR, A. D., ZANTON, S. J., AND PUGH, B. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* *116*, 5 (2004), 699–709.
- [10] BEAUDENON, S., HUACANI, M., WANG, G., DP, M., AND HUIBREGTSE, J. Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in *saccharomyces cerevisiae*. *Mol. Cell. Biol.* *19*, 10 (1999), 6972–9.
- [11] BERBENETZ, N. M., NISLOW, C., AND BROWN, G. W. Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. *PLoS Genet.* *6*, 9 (2010), e1001092.

- [12] BOGENSCHUTZ, N. L., RODRIGUEZ, J., AND TSUKIYAMA, T. Initiation of DNA replication from non-canonical sites on an origin-depleted chromosome. *PLoS ONE* 9, 12 (2014), e114545.
- [13] BOHNSACK, M. T., TOLLERVEY, D., AND GRANNEMAN, S. Identification of RNA helicase target sites by UV cross-linking and analysis of cDNA. *Meth. Enzymol.* 511 (2012), 275–88.
- [14] BOLGER, A. M., LOHSE, M., AND USADEL, B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 15 (2014), 2114–20.
- [15] BURATOWSKI, S., HAHN, S., GUARENTE, L., AND SHARP, P. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 56, 4 (1989), 549–61.
- [16] BURUGULA, B. B., JERONIMO, C., PATHAK, R., JONES, J. W., ROBERT, F., AND GOVIND, C. K. Histone deacetylases and phosphorylated polymerase II c-terminal domain recruit spt6 for cotranscriptional histone reassembly. *Mol. Cell. Biol.* 34, 22 (2014), 4115–29.
- [17] BUSHNELL, D. A., WESTOVER, K. D., DAVIS, R. E., AND KORNBERG, R. D. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 angstroms. *Science* 303, 5660 (2004), 983–8.
- [18] BUSSEMAKER, H., LI, H., AND SIGGIA, E. Regulatory element detection using correlation with expression. *Nat. Genet.* 27, 2 (2001), 167–71.
- [19] BUTLER, J. E., AND KADONAGA, J. T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* 16, 20 (2002), 2583–92.
- [20] CAI, M., AND DAVIS, R. Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. *Cell* 61, 3 (1990), 437–46.
- [21] CARLO, R., AND BRICKNER, J. H. Transcription factors dynamically control the spatial organization of the yeast genome. *Nucleus* 7, 4 (2016), 369–74.
- [22] CARROLL, K. L., GHIRLANDO, R., AMES, J. M., AND CORDEN, J. L. Interaction of yeast RNA-binding proteins nrd1 and nab3 with RNA polymerase II terminator elements. *RNA* 13, 3 (2007), 361–73.
- [23] CARROLL, K. L., PRADHAN, D. A., GRANEK, J. A., CLARKE, N. D., AND CORDEN, J. L. Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol. Cell. Biol.* 24, 14 (2004), 6241–52.
- [24] CASTELNUOVO, M., RAHMAN, S., GUFFANTI, E., INFANTINO, V., STUTZ, F., AND ZENKLUSEN, D. Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat. Struct. Mol. Biol.* 20, 7 (2013), 851–8.
- [25] CELNIKER, S., SWEDER, K., SRIENC, F., BAILEY, J., AND CAMPBELL, J. Deletion mutations affecting autonomously replicating sequence ARS1 of *saccharomyces cerevisiae*. *Mol. Cell. Biol.* 4, 11 (1984), 2455–66.
- [26] CHASMAN, D., LUE, N., BUCHMAN, A., JW, L., LORCH, Y., AND KORNBERG, R. A yeast protein that influences the chromatin structure of UASG and functions as a powerful auxiliary gene activator. *Genes Dev.* 4, 4 (1990), 503–14.

- [27] CHERRY, J., HONG, E. L., AMUNDSEN, C., BALAKRISHNAN, R., BINKLEY, G., CHAN, E. T., CHRISTIE, K. R., COSTANZO, M. C., DWIGHT, S. S., ENGEL, S. R., FISK, D. G., HIRSCHMAN, J. E., HITZ, B. C., KARRA, K., KRIEGER, C. J., MIYASATO, S. R., NASH, R. S., PARK, J., SKRZYPEK, M. S., SIMISON, M., WENG, S., AND WONG, E. D. Saccharomyces genome database: the genomics resource of budding yeast. *Nucleic Acids Res.* *40*, Database issue (2012), D700–5.
- [28] CHEUNG, A. C., AND CRAMER, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* *471*, 7337 (2011), 249–53.
- [29] COLIN, J., CANDELLI, T., PORRUA, O., BOULAY, J., ZHU, C., LACROUTE, F., STEINMETZ, L. M., AND LIBRI, D. Roadblock termination by reb1p restricts cryptic and readthrough transcription. *Mol. Cell* *56*, 5 (2014), 667–80.
- [30] CONRAD, N., WILSON, S., STEINMETZ, E., PATTURAJAN, M., BROW, D., SWANSON, M., AND CORDEN, J. A yeast heterogeneous nuclear ribonucleo-protein complex associated with RNA polymerase II. *Genetics* *154*, 2 (2000), 557–71.
- [31] CUI, M., ALLEN, M. A., LARSEN, A., MACMORRIS, M., HAN, M., AND BLUMENTHAL, T. Genes involved in pre-mRNA 3'-end formation and transcription termination revealed by a lin-15 operon muv suppressor screen. *Proc. Natl. Acad. Sci. U.S.A.* *105*, 43 (2008), 16665–70.
- [32] CZAJKOWSKY, D. M., LIU, J., HAMLIN, J. L., AND SHAO, Z. DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. *J. Mol. Biol.* *375*, 1 (2008), 12–9.
- [33] D, M. K., WANG, T., GORDON, D., GIFFORD, D. K., STORMO, G. D., AND FRAENKEL, E. An improved map of conserved regulatory sites for saccharomyces cerevisiae. *BMC Bioinformatics* *7* (2006), 113.
- [34] DAVID, L., HUBER, W., GRANOVSKAIA, M., TOEDLING, J., PALM, C. J., BOFKIN, L., JONES, T., DAVIS, R. W., AND STEINMETZ, L. M. A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U.S.A.* *103*, 14 (2006), 5320–5.
- [35] DAVIS, J. A., TAKAGI, Y., KORNBERG, R. D., AND ASTURIAS, F. A. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol. Cell* *10*, 2 (2002), 409–15.
- [36] DELLA SETA, F., CIAFRÉ, S., MARCK, C., SANTORO, B., PRESUTTI, C., SENTENAC, A., AND BOZZONI, I. The ABF1 factor is the transcriptional activator of the l2 ribosomal protein genes in saccharomyces cerevisiae. *Mol. Cell. Biol.* *10*, 5 (1990), 2437–41.
- [37] DENGL, S., AND CRAMER, P. Torpedo nuclease rat1 is insufficient to terminate RNA polymerase II in vitro. *J. Biol. Chem.* *284*, 32 (2009), 21270–9.
- [38] DERSHOWITZ, A., SNYDER, M., SBIA, M., SKURNICK, J. H., ONG, L. Y., AND NEWLON, C. S. Linear derivatives of saccharomyces cerevisiae chromosome III can be maintained in the absence of autonomously replicating sequence elements. *Mol. Cell. Biol.* *27*, 13 (2007), 4652–63.
- [39] DIFFLEY, J. Global regulators of chromosome function in yeast. *Antonie Van Leeuwenhoek* *62*, 1-2 (1992), 25–33.

- [40] DIFFLEY, J., AND COCKER, J. Protein-DNA interactions at a yeast replication origin. *Nature* 357, 6374 (1992), 169–72.
- [41] DIFFLEY, J., COCKER, J., DOWELL, S., AND ROWLEY, A. Two steps in the assembly of complexes at yeast replication origins in vivo. *Cell* 78, 2 (1994), 303–16.
- [42] DONOVAN, S., HARWOOD, J., DRURY, L., AND DIFFLEY, J. Cdc6p-dependent loading of mcm proteins onto pre-replicative chromatin in budding yeast. *Proc. Natl. Acad. Sci. U.S.A.* 94, 11 (1997), 5611–6.
- [43] EATON, M. L., GALANI, K., KANG, S., BELL, S. P., AND M, M. D. Conserved nucleosome positioning defines replication origins. *Genes Dev.* 24, 8 (2010), 748–53.
- [44] EGLOFF, S. Role of ser7 phosphorylation of the CTD during transcription of snRNA genes. *RNA Biol* 9, 8 (2012), 1033–8.
- [45] EGLOFF, S., DIENSTBIER, M., AND MURPHY, S. Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet.* 28, 7 (2012), 333–41.
- [46] ESNAULT, C., YAD, G., BRUN, S., SOUTOURINA, J., VAN BERKUM, N., BOSCHIERO, C., HOLSTEGE, F., AND WERNER, M. Mediator-dependent recruitment of TFIIF modules in preinitiation complex. *Mol. Cell* 31, 3 (2008), 337–46.
- [47] FAZAL, F. M., MENG, C. A., MURAKAMI, K., KORNBERG, R. D., AND BLOCK, S. M. Real-time observation of the initiation of RNA polymerase II transcription. *Nature* 525, 7568 (2015), 274–7.
- [48] FELDMANN, E. A., AND GALLETO, R. The DNA-binding domain of yeast rap1 interacts with double-stranded DNA in multiple binding modes. *Biochemistry* 53, 48 (2014), 7471–83.
- [49] FENG, L., WANG, B., AND JONG, A. *Saccharomyces cerevisiae* cdc6 stimulates abf1 DNA binding activity. *J. Biol. Chem.* 273, 3 (1998), 1298–302.
- [50] FERMI, B., BOSIO, M. C., AND DIECI, G. Multiple roles of the general regulatory factor abf1 in yeast ribosome biogenesis. *Curr. Genet.* (2016).
- [51] FERMI, B., BOSIO, M. C., AND DIECI, G. Promoter architecture and transcriptional regulation of abf1-dependent ribosomal protein genes in *saccharomyces cerevisiae*. *Nucleic Acids Res.* 44, 13 (2016), 6113–26.
- [52] FIELD, Y., KAPLAN, N., YVONNE, F., MOORE, I. K., SHARON, E., LUBLING, Y., WIDOM, J., AND SEGAL, E. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* 4, 11 (2008), e1000216.
- [53] FLOER, M., WANG, X., PRABHU, V., BERROZPE, G., AND NARAYAN, S. A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. *Cell* (2010).
- [54] FONG, N., BRANNAN, K., ERICKSON, B., KIM, H., CORTAZAR, M. A., SHERIDAN, R. M., NGUYEN, T., KARP, S., AND BENTLEY, D. L. Effects of transcription elongation rate and xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. *Mol. Cell* 60, 2 (2015), 256–67.

- [55] FOUREL, G., MIYAKE, T., DEFOSSEZ, P. A., LI, R., AND GILSON, E. General regulatory factors (GRFs) as genome partitioners. *J. Biol. Chem.* 277, 44 (2002), 41736–43.
- [56] GAMBUS, A., KHOUDOLI, G. A., JONES, R. C., AND BLOW, J. MCM2-7 form double hexamers at licensed origins in xenopus egg extract. *J. Biol. Chem.* 286, 13 (2011), 11855–64.
- [57] GANAPATHI, M., PALUMBO, M. J., ANSARI, S. A., HE, Q., TSUI, K., NISLOW, C., AND MORSE, R. H. Extensive role of the general regulatory factors, abf1 and rap1, in determining genome-wide chromatin structure in budding yeast. *Nucleic Acids Res.* 39, 6 (2011), 2032–44.
- [58] GAUTHIER, L., DZIAK, R., KRAMER, D. J., LEISHMAN, D., SONG, X., HO, J., RADOVIC, M., BENTLEY, D., AND YANKULOV, K. The role of the carboxyterminal domain of RNA polymerase II in regulating origins of DNA replication in *saccharomyces cerevisiae*. *Genetics* 162, 3 (2002), 1117–29.
- [59] GHAZAL, G., GAGNON, J., JACQUES, P. E., LANDRY, J. R., ROBERT, F., AND ELELA, S. A. Yeast RNase III triggers polyadenylation-independent transcription termination. *Mol. Cell* 36, 1 (2009), 99–109.
- [60] GHAZAL, G., GE, D., JULIEN, G., GAGNON, J., AND ABOU ELELA, S. Genome-wide prediction and analysis of yeast RNase III-dependent snoRNA processing signals. *Mol. Cell. Biol.* 25, 8 (2005), 2981–94.
- [61] GRANNEMAN, S., KUDLA, G., PETFALSKI, E., AND TOLLERVEY, D. Identification of protein binding sites on u3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U.S.A.* 106, 24 (2009), 9613–8.
- [62] GU, B., EICK, D., AND BENSAUDE, O. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res.* 41, 3 (2013), 1591–603.
- [63] GUDIPATI, R. K., VILLA, T., BOULAY, J., AND LIBRI, D. Phosphorylation of the RNA polymerase II c-terminal domain dictates transcription termination choice. *Nat. Struct. Mol. Biol.* 15, 8 (2008), 786–94.
- [64] HARACSKA, L., JOHNSON, R. E., PRAKASH, L., AND PRAKASH, S. Trf4 and trf5 proteins of *saccharomyces cerevisiae* exhibit poly(A) RNA polymerase activity but no DNA polymerase activity. *Mol. Cell. Biol.* 25, 22 (2005), 10183–9.
- [65] HARBISON, C. T., GORDON, D., LEE, T. I., RINALDI, N. J., MACISAAC, K. D., DANFORD, T. W., HANNETT, N. M., TAGNE, J. B., REYNOLDS, D. B., YOO, J., JENNINGS, E. G., ZEITLINGER, J., POKHOLOK, D. K., KELLIS, M., ROLFE, P., TAKUSAGAWA, K. T., LANDER, E. S., GIFFORD, D. K., FRAENKEL, E., AND YOUNG, R. A. Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 7004 (2004), 99–104.
- [66] HARTLEY, P. D., AND MADHANI, H. D. Mechanisms that specify promoter nucleosome location and identity. *Cell* 137, 3 (2009), 445–58.
- [67] HARUKI, H., NISHIKAWA, J., AND LAEMMLI, U. The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Molecular cell* (2008).

- [68] HAWKINS, M., RETKUTE, R., MÜLLER, C. A., SANER, N., TANAKA, T. U., DE MOURA, A. P., AND NIEDUSZYNSKI, C. A. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep* 5, 4 (2013), 1132–41.
- [69] HAZELBAKER, D. Z., MARQUARDT, S., WLOTZKA, W., AND BURATOWSKI, S. Kinetic competition between RNA polymerase II and sen1-dependent transcription termination. *Mol. Cell* 49, 1 (2013), 55–66.
- [70] HOCHBERG, Y., AND BENJAMINI, Y. More powerful procedures for multiple significance testing. *Stat Med* 9, 7 (1990), 811–8.
- [71] HOLSTEGE, F., FIEDLER, U., AND TIMMERS, H. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.* 16, 24 (1997), 7468–80.
- [72] HOLSTEGE, F., VAN DER VLIET, P., AND TIMMERS, H. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* 15, 7 (1996), 1666–77.
- [73] HOUSELEY, J., AND TOLLERVEY, D. The many pathways of RNA degradation. *Cell* 136, 4 (2009), 763–76.
- [74] HSIN, J., SHETH, A., AND MANLEY, J. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* (2011).
- [75] HUANG, R., AND KOWALSKI, D. A DNA unwinding element and an ARS consensus comprise a replication origin within a yeast chromosome. *EMBO J.* 12, 12 (1993), 4521–31.
- [76] HUISINGA, K. L., AND PUGH, B. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *saccharomyces cerevisiae*. *Mol. Cell* 13, 4 (2004), 573–85.
- [77] HUNTER, G. O., FOX, M. J., R, S. W., GOGOL, M., FLEHARTY, B., AND MOSLEY, A. L. Phosphatase rtr1 regulates global levels of serine 5 RNA polymerase II C-Terminal domain phosphorylation and cotranscriptional histone methylation. *Mol. Cell. Biol.* 36, 17 (2016), 2236–45.
- [78] IMBALZANO, A., ZARET, K., AND KINGSTON, R. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *J. Biol. Chem.* 269, 11 (1994), 8280–6.
- [79] JENKS, M., W, O. T., AND REINES, D. Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. *Mol. Cell. Biol.* 28, 12 (2008), 3883–93.
- [80] JIA, H., WANG, X., ANDERSON, J. T., AND JANKOWSKY, E. RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19 (2012), 7292–7.
- [81] JIANG, C., AND PUGH, B. A compiled and systematic reference map of nucleosome positions across the *saccharomyces cerevisiae* genome. *Genome Biol.* 10, 10 (2009), R109.
- [82] JOHN, L., HOUSELEY, J., SAVEANU, C., PETFALSKI, E., THOMPSON, E., JACQUIER, A., AND TOLLERVEY, D. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* 121, 5 (2005), 713–24.

- [83] KAMENOVA, I., WARFIELD, L., AND HAHN, S. Mutations on the DNA binding surface of TBP discriminate between yeast TATA and TATA-less gene transcription. *Mol. Cell. Biol.* *34*, 15 (2014), 2929–43.
- [84] KF, O., Y, S., AND BAKER, R. Role of the *saccharomyces cerevisiae* general regulatory factor CP1 in methionine biosynthetic gene transcription. *Mol. Cell. Biol.* *15*, 4 (1995), 1879–88.
- [85] KILCHERT, C., WITTMANN, S., AND VASILJEVA, L. The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* *17*, 4 (2016), 227–39.
- [86] KIM, J., WEI, S., LEE, J., YUE, H., AND LEE, T. H. Single-Molecule observation reveals spontaneous protein dynamics in the nucleosome. *J Phys Chem B* (2016).
- [87] KIM, M., AHN, S. H., KROGAN, N. J., GREENBLATT, J. F., AND BURATOWSKI, S. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J.* *23*, 2 (2004), 354–64.
- [88] KIM, M., KROGAN, N. J., VASILJEVA, L., RANDO, O. J., NEDEA, E., GREENBLATT, J. F., AND BURATOWSKI, S. The yeast *rat1* exonuclease promotes transcription termination by RNA polymerase II. *Nature* *432*, 7016 (2004), 517–22.
- [89] KIM, M., VASILJEVA, L., RANDO, O. J., ZHELKOVSKY, A., MOORE, C., AND BURATOWSKI, S. Distinct pathways for snoRNA and mRNA termination. *Mol. Cell* *24*, 5 (2006), 723–34.
- [90] KNIGHT, B., KUBIK, S., GHOSH, B., BRUZZONE, M. J., GEERTZ, M., MARTIN, V., DÉNERVAUD, N., JACQUET, P., OZKAN, B., ROUGEMONT, J., MAERKL, S. J., NAEF, F., AND SHORE, D. Two distinct promoter architectures centered on dynamic nucleosomes control ribosomal protein gene transcription. *Genes Dev.* *28*, 15 (2014), 1695–709.
- [91] KORDE, A., ROSSELOT, J. M., AND DONZE, D. Intergenic transcriptional interference is blocked by RNA polymerase III transcription factor TFIIIB in *saccharomyces cerevisiae*. *Genetics* *196*, 2 (2014), 427–38.
- [92] KUBICEK, K., CERNA, H., HOLUB, P., PASULKA, J., HROSSOVA, D., LOEHR, F., HOFR, C., VANACOVA, S., AND STEFL, R. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of *nrd1*. *Genes Dev.* *26*, 17 (2012), 1891–6.
- [93] KUBIK, S., BRUZZONE, M. J., JACQUET, P., FALCONE, J. L., ROUGEMONT, J., AND SHORE, D. Nucleosome stability distinguishes two different promoter types at all Protein-Coding genes in yeast. *Mol. Cell* *60*, 3 (2015), 422–34.
- [94] KULAEVA, O. I., HSIEH, F. K., CHANG, H. W., LUSE, D. S., AND STU-DITSKY, V. M. Mechanism of transcription through a nucleosome by RNA polymerase II. *Biochim. Biophys. Acta* *1829*, 1 (2013), 76–83.
- [95] KURTZ, S., AND SHORE, D. RAP1 protein activates and silences transcription of mating-type genes in yeast. *Genes Dev.* *5*, 4 (1991), 616–28.
- [96] L, M. A., AND GREENLEAF, A. L. Cotranscriptional association of mRNA export factor *yra1* with c-terminal domain of RNA polymerase II. *J. Biol. Chem.* *286*, 42 (2011), 36385–95.

- [97] L, M. K., AND CHEESEMAN, I. M. The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* (2015).
- [98] LANG, W., MORROW, B., JU, Q., WARNER, J., AND REEDER, R. A model for transcription termination by RNA polymerase i. *Cell* (1994).
- [99] LANG, W., AND REEDER, R. The REB1 site is an essential component of a terminator for RNA polymerase i in *saccharomyces cerevisiae*. *Mol. Cell. Biol.* 13, 1 (1993), 649–58.
- [100] LANG, W., AND REEDER, R. Transcription termination of RNA polymerase i due to a t-rich element interacting with reb1p. *Proc. Natl. Acad. Sci. U.S.A.* 92, 21 (1995), 9781–5.
- [101] LANGMEAD, B., AND SALZBERG, S. L. Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 4 (2012), 357–9.
- [102] LARDENOIS, A., LIU, Y., WALTHER, T., CHALMEL, F., EVRARD, B., GRANOVSKAIA, M., CHU, A., DAVIS, R. W., STEINMETZ, L. M., AND PRIMIG, M. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit rrp6. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3 (2011), 1058–63.
- [103] LIPFORD, J., AND BELL, S. Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol. Cell* 7, 1 (2001), 21–30.
- [104] LORCH, Y., JW, L., AND KORNBERG, R. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* 49, 2 (1987), 203–10.
- [105] LUCAS, I. A., AND RAGHURAMAN, M. The dynamics of chromosome replication in yeast. *Curr. Top. Dev. Biol.* 55 (2003), 1–73.
- [106] LUNDE, B. M., HÖRNER, M., AND MEINHART, A. Structural insights into cis element recognition of non-polyadenylated RNAs by the Nab3-RRM. *Nucleic Acids Res.* 39, 1 (2011), 337–46.
- [107] LUO, W., JOHNSON, A. W., AND BENTLEY, D. L. The role of rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev.* 20, 8 (2006), 954–65.
- [108] LUSTIG, A., KURTZ, S., AND SHORE, D. Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. *Science* 250, 4980 (1990), 549–53.
- [109] M, B., EDLER, K., AND BARBERIS, A. RNA polymerase II and III transcription factors can stimulate DNA replication by modifying origin chromatin structures. *Nucleic Acids Res.* 29, 22 (2001), 4570–80.
- [110] MAK, H., PILLUS, L., AND IDEKER, T. Dynamic reprogramming of transcription factors to and from the subtelomere. *Genome Res.* 19, 6 (2009), 1014–25.
- [111] MAKINO, D. L., SCHUCH, B., STEGMANN, E., BAUMGÄRTNER, M., BASQUIN, C., AND CONTI, E. RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature* (2015).
- [112] MALABAT, C., FEUERBACH, F., MA, L., SAVEANU, C., AND JACQUIER, A. Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *Elife* 4 (2015).



- [113] MARAHRENS, Y., AND STILLMAN, B. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* 255, 5046 (1992), 817–23.
- [114] MARQUARDT, S., HAZELBAKER, D. Z., AND BURATOWSKI, S. Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. *Transcription* 2, 3 (2011), 145–154.
- [115] MARQUARDT, S., RENAN, E., PHO, N., WANG, J., CHURCHMAN, L., SPRINGER, M., AND BURATOWSKI, S. A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* 157, 7 (2014), 1712–23.
- [116] MARTENS, J. A., LAPRADE, L., AND WINSTON, F. Intergenic transcription is required to repress the *saccharomyces cerevisiae* SER3 gene. *Nature* 429, 6991 (2004), 571–4.
- [117] MARTIN, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 1 (May 2011), 10–12.
- [118] MAYAN, M. D. RNAP-II molecules participate in the anchoring of the ORC to rDNA replication origins. *PLoS ONE* 8, 1 (2013), e53405.
- [119] MAYER, A., HEIDEMANN, M., LIDSCHREIBER, M., SCHREIECK, A., SUN, M., HINTERMAIR, C., KREMMER, E., EICK, D., AND CRAMER, P. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336, 6089 (2012), 1723–5.
- [120] MAYER, A., LIDSCHREIBER, M., SIEBERT, M., LEIKE, K., SÖDING, J., AND CRAMER, P. Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* 17, 10 (2010), 1272–8.
- [121] MELLOR, J., JIANG, W., FUNK, M., RATHJEN, J., BARNES, C., HINZ, T., HEGEMANN, J., AND PHILIPPSEN, P. CPF1, a yeast protein which functions in centromeres and promoters. *EMBO J.* 9, 12 (1990), 4017–26.
- [122] MILLIGAN, L., A, H. V. A., CLÉMENTINE, D., TUCK, A., PETFALSKI, E., LOMBRAÑA, R., SANGUINETTI, G., KUDLA, G., AND TOLLERVEY, D. Strand-specific, high-resolution mapping of modified RNA polymerase II. *Mol. Syst. Biol.* 12, 6 (2016), 874.
- [123] MITCHELL, P., PETFALSKI, E., SHEVCHENKO, A., MANN, M., AND TOLLERVEY, D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'–{>}5' exoribonucleases. *Cell* 91, 4 (1997), 457–66.
- [124] MOEHLE, C., AND HINNEBUSCH, A. Association of RAP1 binding sites with stringent control of ribosomal protein gene transcription in *saccharomyces cerevisiae*. *Mol. Cell. Biol.* 11, 5 (1991), 2723–35.
- [125] MORALES, J. C., RICHARD, P., ROMMEL, A., FATTAH, F. J., MOTEA, E. A., PATIDAR, P. L., XIAO, L., LESKOV, K., WU, S. Y., HITTELMAN, W. N., CHIANG, C. M., MANLEY, J. L., AND BOOTHMAN, D. A. Kub5-Hera, the human rtt103 homolog, plays dual functional roles in transcription termination and DNA repair. *Nucleic Acids Res.* 42, 8 (2014), 4996–5006.
- [126] MOSLEY, A. L., PATTENDEN, S. G., CAREY, M., VENKATESH, S., GILMORE, J. M., FLORENS, L., WORKMAN, J. L., AND WASHBURN, M. P. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol. Cell* 34, 2 (2009), 168–78.

- [127] MOYER, S. E., LEWIS, P. W., AND BOTCHAN, M. R. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl. Acad. Sci. U.S.A.* *103*, 27 (2006), 10236–41.
- [128] NEIL, H., MALABAT, C., YVES, D., XU, Z., STEINMETZ, L. M., AND JACQUIER, A. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* *457*, 7232 (2009), 1038–42.
- [129] NEWLON, C., COLLINS, I., DERSHOWITZ, A., DESHPANDE, A., GREENFEDER, S., ONG, L., AND THEIS, J. Analysis of replication origin function on chromosome III of *saccharomyces cerevisiae*. *Cold Spring Harb. Symp. Quant. Biol.* *58* (1993), 415–23.
- [130] NGUYEN, V., CO, C., AND LI, J. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* *411*, 6841 (2001), 1068–73.
- [131] NIEDENTHAL, R., M, S., WILMEN, A., AND HEGEMANN, J. Cpf1 protein induced bending of yeast centromere DNA element i. *Nucleic Acids Res.* *21*, 20 (1993), 4726–33.
- [132] NIEDUSZYNSKI, C. A., BLOW, J., AND DONALDSON, A. D. The requirement of yeast replication origins for pre-replication complex proteins is modulated by transcription. *Nucleic Acids Res.* *33*, 8 (2005), 2410–20.
- [133] NIEDUSZYNSKI, C. A., KNOX, Y., AND DONALDSON, A. D. Genome-wide identification of replication origins in yeast by comparative genomics. 1874–1879.
- [134] PAL, M., PONTICELLI, A. S., AND LUSE, D. S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol Cell* *19*, 1 (2004), 101–110.
- [135] PAPAI, G., TRIPATHI, M. K., RUHLMANN, C., LAYER, J. H., WEIL, P., AND SCHULTZ, P. TFIIA and the transactivator rap1 cooperate to commit TFIID for transcription initiation. *Nature* *465*, 7300 (2010), 956–60.
- [136] PARK, J., KANG, M., AND KIM, M. Unraveling the mechanistic features of RNA polymerase II termination by the 5'-3' exoribonuclease rat1. *Nucleic Acids Res.* *43*, 5 (2015), 2625–37.
- [137] PATEL, P. K., ARCANGIOLI, B., BAKER, S. P., BENSIMON, A., AND RHIND, N. DNA replication origins fire stochastically in fission yeast. *Mol. Biol. Cell* *17*, 1 (2006), 308–16.
- [138] PEARSON, E. L., AND MOORE, C. L. Dismantling promoter-driven RNA polymerase II transcription complexes in vitro by the termination factor rat1. *J. Biol. Chem.* *288*, 27 (2013), 19750–9.
- [139] PELECHANO, V., WEI, W., AND STEINMETZ, L. M. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* *497*, 7447 (2013), 127–31.
- [140] PILPEL, Y., SUDARSANAM, P., AND CHURCH, G. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.* *29*, 2 (2001), 153–9.
- [141] PLANTA, R., GONÇALVES, P., AND MAGER, W. Global regulators of ribosome biosynthesis in yeast. *Biochem. Cell Biol.* *73*, 11-12 (1995), 825–34.

- [142] PORRUA, O., HOBOR, F., BOULAY, J., KUBICEK, K., YVES, D., GUDIPATI, R. K., STEFL, R., AND LIBRI, D. In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J.* **31**, 19 (2012), 3935–48.
- [143] PORRUA, O., AND LIBRI, D. A bacterial-like mechanism for transcription termination by the sen1p helicase in budding yeast. *Nat. Struct. Mol. Biol.* **20**, 7 (2013), 884–91.
- [144] QIU, H., HU, C., AND HINNEBUSCH, A. G. Phosphorylation of the pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and ser2 CTD phosphorylation near promoters. *Mol. Cell* **33**, 6 (2009), 752–62.
- [145] R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [146] RAGHURAMAN, M. K., AND LIACHKO, I. Sequence determinants of yeast replication origins. In *The Initiation of DNA Replication in Eukaryotes*, D. L. Kaplan, Ed. Springer, Tallahassee , FL , USA, 2016, ch. 7, pp. 123–143.
- [147] RANI, P., RANISH, J. A., AND HAHN, S. RNA polymerase II (Pol II)-TFIIF and pol II-mediator complexes: the major stable pol II complexes and their activity in transcription initiation and reinitiation. *Mol. Cell. Biol.* **24**, 4 (2004), 1709–20.
- [148] REED, S., AKIYAMA, M., STILLMAN, B., AND FRIEDBERG, E. Yeast autonomously replicating sequence binding factor is involved in nucleotide excision repair. *Genes Dev.* **13**, 23 (1999), 3052–8.
- [149] REEDER, R., GUEVARA, P., AND ROAN, J. *Saccharomyces cerevisiae* RNA polymerase i terminates transcription at the reb1 terminator in vivo. *Mol. Cell. Biol.* **19**, 11 (1999), 7369–76.
- [150] REINBERG, D., AND SIMS, R. J. de FACTo nucleosome dynamics. *J. Biol. Chem.* **281**, 33 (2006), 23297–301.
- [151] REITER, A., HAMPERL, S., SEITZ, H., MERKL, P., JORGE, P., WILLIAMS, L., GERBER, J., NÉMETH, A., LÉGER, I., GADAL, O., MILKEREIT, P., GRIESENBECK, J., AND TSCHOCHNER, H. The reb1-homologue Ydr026c/Nsi1 is required for efficient RNA polymerase i termination in yeast. *EMBO J.* **31**, 16 (2012), 3480–93.
- [152] REJA, R., VINAYACHANDRAN, V., GHOSH, S., AND PUGH, B. Molecular mechanisms of ribosomal protein gene coregulation. *Genes Dev.* **29**, 18 (2015), 1942–54.
- [153] REMUS, D., BEURON, F., TOLUN, G., GRIFFITH, J. D., MORRIS, E. P., AND DIFFLEY, J. F. Concerted loading of mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139**, 4 (2009), 719–30.
- [154] RHEE, H. S., AND PUGH, B. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 6 (2011), 1408–19.
- [155] RHEE, H. S., AND PUGH, B. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 7389 (2012), 295–301.

- [156] RHODE, P., SWEDER, K., OEGEMA, K., AND CAMPBELL, J. The gene encoding ARS-binding factor i is essential for the viability of yeast. *Genes Dev.* 3, 12A (1989), 1926–39.
- [157] RONDÓN, A. G., MISCHO, H. E., KAWAUCHI, J., AND PROUDFOOT, N. J. Fail-safe transcriptional termination for protein-coding genes in *s. cerevisiae*. *Mol. Cell* 36, 1 (2009), 88–98.
- [158] ROWLES, A., TADA, S., AND BLOW, J. Changes in association of the xenopus origin recognition complex with chromatin on licensing of replication origins. *J. Cell. Sci.* 112 ( Pt 12) (1999), 2011–8.
- [159] ROY, K., GABUNILAS, J., GILLESPIE, A., NGO, D., AND CHANFREAU, G. F. Common genomic elements promote transcriptional and DNA replication roadblocks. *Genome Res.* 26, 10 (2016), 1363–1375.
- [160] S, M., FONG, N., ROSONINA, E., YANKULOV, K., BROTHERS, G., SIDEROVSKI, D., HESSEL, A., FOSTER, S., SHUMAN, S., AND BENTLEY, D. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.* 11, 24 (1997), 3306–18.
- [161] S, M., FONG, N., YANKULOV, K., BALLANTYNE, S., PAN, G., GREENBLATT, J., PATTERSON, S., WICKENS, M., AND BENTLEY, D. The c-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 6614 (1997), 357–61.
- [162] SAINSBURY, S., BERNECKY, C., AND CRAMER, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 16, 3 (2015), 129–43.
- [163] SANTOCANALE, C., SHARMA, K., AND DIFFLEY, J. Activation of dormant origins of DNA replication in budding yeast. *Genes Dev.* 13, 18 (1999), 2360–4.
- [164] SCHAUGHENCY, P., MERRAN, J., AND CORDEN, J. L. Genome-wide mapping of yeast RNA polymerase II termination. *PLoS Genet.* 10, 10 (2014), e1004632.
- [165] SCHREIECK, A., EASTER, A. D., ETZOLD, S., WIEDERHOLD, K., LIDSCHREIBER, M., CRAMER, P., AND PASSMORE, L. A. RNA polymerase II termination involves c-terminal-domain tyrosine dephosphorylation by CPF subunit glc7. *Nat. Struct. Mol. Biol.* 21, 2 (2014), 175–9.
- [166] SCHULZ, D., SCHWALB, B., KIESEL, A., BAEJEN, C., TORKLER, P., GAGNEUR, J., SOEDING, J., AND CRAMER, P. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* 155, 5 (2013), 1075–87.
- [167] SEKI, T., AND DIFFLEY, J. Stepwise assembly of initiation proteins at budding yeast replication origins in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 97, 26 (2000), 14115–20.
- [168] SELTH, L. A., SIGURDSSON, S., AND SVEJSTRUP, J. Q. Transcript elongation by RNA polymerase II. *Annu. Rev. Biochem.* 79 (2010), 271–93.
- [169] SIOW, C. C., NIEDUSZYNSKA, S. R., MÜLLER, C. A., AND NIEDUSZYNSKI, C. A. OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.* 40, Database issue (2012), D682–6.

- [170] SØGAARD, T. M. M., AND SVEJSTRUP, J. Q. Hyperphosphorylation of the c-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem.* *282*, 19 (2007), 14113–20.
- [171] SOUTOURINA, J., WYDAU, S., AMBROISE, Y., BOSCHIERO, C., AND WERNER, M. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* *331*, 6023 (2011), 1451–4.
- [172] SPAIN, M. M., ANSARI, S. A., PATHAK, R., PALUMBO, M. J., MORSE, R. H., AND GOVIND, C. K. The RSC complex localizes to coding sequences to regulate pol II and histone occupancy. *Mol. Cell* *56*, 5 (2014), 653–66.
- [173] STEINMETZ, E., AND BROW, D. 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 (1996), 6993–7003.
- [174] STEINMETZ, E., AND BROW, D. Control of pre-mRNA accumulation by the essential yeast protein nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. *Proc. Natl. Acad. Sci. U.S.A.* *95*, 12 (1998), 6699–704.
- [175] STEPHEN, M., AND BROW, D. A. *Saccharomyces cerevisiae* sen1 helicase domain exhibits 5′- to 3′-Helicase activity with a preference for translocation on DNA rather than RNA. *J. Biol. Chem.* *290*, 38 (2015), 22880–9.
- [176] SUH, H., FICARRO, S. B., KANG, U. B., CHUN, Y., MARTO, J. A., AND BURATOWSKI, S. Direct analysis of phosphorylation sites on the rpb1 C-Terminal domain of RNA polymerase II. *Mol. Cell* *61*, 2 (2016), 297–304.
- [177] TANAKA, S., HALTER, D., M, L., RESZEL, B., AND THOMA, F. Transcription through the yeast origin of replication ARS1 ends at the ABFI binding site and affects extrachromosomal maintenance of minichromosomes. *Nucleic Acids Res.* *22*, 19 (1994), 3904–10.
- [178] THIEBAUT, M., COLIN, J., NEIL, H., JACQUIER, A., SÉRAPHIN, B., LACROUTE, F., AND LIBRI, D. Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in *s. cerevisiae*. *Mol. Cell* *31*, 5 (2008), 671–82.
- [179] THIEBAUT, M., ELENA, K., ROUGEMAILLE, M., BOULAY, J., AND LIBRI, D. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol. Cell* *23*, 6 (2006), 853–64.
- [180] THOMPSON, D. M., AND PARKER, R. Cytoplasmic decay of intergenic transcripts in *saccharomyces cerevisiae*. *Mol. Cell Biol.* *27*, 1 (2007), 92–101.
- [181] TUDEK, A., CANDELLI, T., AND LIBRI, D. Non-coding transcription by RNA polymerase II in yeast: Hasard or nécessité? *Biochimie* *117* (2015), 28–36.
- [182] TUDEK, A., PORRUA, O., KABZINSKI, T., LIDSCHREIBER, M., KUBICEK, K., FORTOVA, A., LACROUTE, F., VANACOVA, S., CRAMER, P., STEFL, R., AND LIBRI, D. Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell* *55*, 3 (2014), 467–81.

- [183] VAN DIJK, E., CHEN, C., Y, D., GOURVENNEC, S., KWAPISZ, M., ROCHE, V., BERTRAND, C., SILVAIN, M., P, L., LOEILLET, S., NICOLAS, A., THERMES, C., AND MORILLON, A. XUTs are a class of xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475, 7354 (2011), 114–7.
- [184] VAN HELDEN, J., ANDRÉ, B., AND J, C. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.* 281, 5 (1998), 827–42.
- [185] VASILJEVA, L., AND BURATOWSKI, S. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol. Cell* 21, 2 (2006), 239–48.
- [186] VASILJEVA, L., KIM, M., MUTSCHLER, H., BURATOWSKI, S., AND MEINHART, A. The Nrd1-Nab3-Sen1 termination complex interacts with the ser5-phosphorylated RNA polymerase II c-terminal domain. *Nat. Struct. Mol. Biol.* 15, 8 (2008), 795–804.
- [187] W, O. T., LOYA, T. J., HEAD, P. E., HORTON, J. R., AND REINES, D. Amyloid-like assembly of the low complexity domain of yeast nab3. *Prion* 9, 1 (2015), 34–47.
- [188] WEBB, S., HECTOR, R. D., KUDLA, G., AND GRANNEMAN, S. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* 15, 1 (2014), R8.
- [189] WELLINGER, R. J., AND ZAKIAN, V. A. Everything you ever wanted to know about saccharomyces cerevisiae telomeres: beginning to end. *Genetics* 191, 4 (2012), 1073–105.
- [190] WEST, S., GROMAK, N., AND PROUDFOOT, N. J. Human 5'  $\{-\}$  3' exonuclease xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432, 7016 (2004), 522–5.
- [191] WILKENING, S., PELECHANO, V., JÄRVELIN, A. I., TEKKEDIL, M. M., ANDERS, S., BENES, V., AND STEINMETZ, L. M. An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res.* 41, 5 (2013), e65.
- [192] WILMES, G. M., AND BELL, S. P. The b2 element of the saccharomyces cerevisiae ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1 (2002), 101–6.
- [193] WILSON, M. D., HARREMAN, M., AND SVEJSTRUP, J. Q. Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. 151–7.
- [194] WILTSHIRE, S., RAYCHAUDHURI, S., AND EISENBERG, S. An abf1p c-terminal region lacking transcriptional activation potential stimulates a yeast origin of replication. *Nucleic Acids Res.* 25, 21 (1997), 4250–6.
- [195] WLOTZKA, W., KUDLA, G., GRANNEMAN, S., AND TOLLERVEY, D. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J.* 30, 9 (2011), 1790–803.

- [196] WYERS, F., ROUGEMAILLE, M., BADIS, G., ROUSSELLE, J. C., DUFOUR, M. E., BOULAY, J., RÉGNAULT, B., DEVAUX, F., NAMANE, A., SÉRAPHIN, B., LIBRI, D., AND JACQUIER, A. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121, 5 (2005), 725–37.
- [197] XU, Z., WEI, W., GAGNEUR, J., PEROCCHI, F., SANDRA, C., CAMBLONG, J., GUFFANTI, E., STUTZ, F., HUBER, W., AND STEINMETZ, L. M. Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 7232 (2009), 1033–7.
- [198] YANKULOV, K., TODOROV, I., ROMANOWSKI, P., LICATALOSI, D., CILLI, K., S, M., LASKEY, R., AND BENTLEY, D. MCM proteins are associated with RNA polymerase II holoenzyme. *Mol. Cell. Biol.* 19, 9 (1999), 6154–63.
- [199] YARRINGTON, R. M., RICHARDSON, S. M., LISA HUANG, C. R., AND BOEKE, J. D. Novel transcript truncating function of rap1p revealed by synthetic codon-optimized ty1 retrotransposon. *Genetics* 190, 2 (2012), 523–35.
- [200] YU, S., SMIRNOVA, J. B., FRIEDBERG, E. C., STILLMAN, B., AKIYAMA, M., TOM, O., WATERS, R., AND REED, S. H. ABF1-binding sites promote efficient global genome nucleotide excision repair. *J. Biol. Chem.* 284, 2 (2009), 966–73.
- [201] YU, S., TOM, O., FRIEDBERG, E. C., WATERS, R., AND REED, S. H. The yeast Rad7/Rad16/Abf1 complex generates superhelical torsion in DNA that is required for nucleotide excision repair. *DNA Repair (Amst.)* 3, 3 (2004), 277–87.
- [202] ZENKLUSEN, D., LARSON, D. R., AND SINGER, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* 15, 12 (2008), 1263–71.
- [203] ZENTNER, G. E., KASINATHAN, S., XIN, B., ROHS, R., AND HENIKOFF, S. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat Commun* 6 (2015), 8733.
- [204] ZHANG, H., RIGO, F., AND MARTINSON, H. G. Poly(A) Signal-Dependent transcription termination occurs through a conformational change mechanism that does not require cleavage at the Poly(A) site. *Mol. Cell* 59, 3 (2015), 437–48.
- [205] ZHANG, Z., FU, J., AND GILMOUR, D. S. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, pcf11. *Genes Dev.* 19, 13 (2005), 1572–80.
- [206] ZHANG, Z., AND GILMOUR, D. S. Pcf11 is a termination factor in drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol. Cell* 21, 1 (2006), 65–74.

## **Résumé de thèse en français**

### **La terminaison de la transcription par roadblock**

Au cours de ma thèse j'ai participé à mettre en évidence une nouvelle voie de terminaison de la transcription par l'ARN polymérase II (RNAPII) qui participe au contrôle de la transcription cachée (voir travaux antérieurs et Colin *et al*, 2014). Cette terminaison fait intervenir l'activateur transcriptionnel Reb1p et repose sur un mécanisme de roadblock : Reb1p lié à l'ADN constitue un obstacle pour les polymérases en cours d'élongation. La collision de la polymérase sur Reb1p induit un arrêt transcriptionnel qui est résolu par l'ubiquitylation et peut-être la dégradation de la polymérase. Les ARN ainsi libérés sont, comme les CUT, en partie polyadénylés par la poly(A)-polymérase Trf4p (probablement au sein du complexe TRAMP) et dégradés par l'exosome nucléaire. L'existence d'une pause transcriptionnelle fait que cette terminaison est associée à la détection d'espèces non polyadénylées relativement stables. En effet, l'extrémité 3' des ARN en cours de synthèse reste à l'intérieur de la polymérase pendant la pause transcriptionnelle, de ce fait, elle est inaccessible aux exonucléases. Les caractéristiques distinctives de la terminaison par roadblock sont donc l'occurrence d'une pause transcriptionnelle marquée et la génération de transcrits instables présentant une extrémité 3' bien définie qui sont en partie non adénylés. Cette voie est utilisée pour terminer des ARN non codants que nous avons nommés RUTs (Reb1p-dependent Unstable Transcripts) du fait de leur similarité avec les CUT



(polyadénylation par Trf4p et dégradation par l'exosome nucléaire). Elle est également employée pour neutraliser la transcription readthrough c'est-à-dire les événements de transcription d'un ARNm qui se prolongent au-delà du terminateur. Ceci permet de protéger les régions régulatrices d'autres gènes contre l'invasion par des événements transcriptionnels dérivant des gènes voisins (voir travaux).

Dans un génome compact et hautement transcrit tel que celui de *S. cerevisiae*, il est indispensable de gérer la cohabitation entre transcription fonctionnelle (produisant par exemple des ARNm) et transcription « envahissante » (due par exemple à des défauts de terminaison ou des initiations inappropriées). La terminaison de la transcription par roadblock, par ses caractéristiques, joue un rôle important dans ce mécanisme de contrôle. En effet, elle requiert peu d'information pour arrêter une polymérase (un site de liaison à l'ADN pour une protéine) et elle est induite par d'autres facteurs que Reb1p, chose que j'ai aussi participé à démontrer. En revanche, elle conduit à la dégradation du transcrit ainsi terminé et très probablement de la polymérase, ce qui représente un certain coût énergétique. Cependant, ce coût est très probablement justifié par sa fonction dans le contrôle de la qualité transcriptionnelle et par le fait que le mécanisme utilisé soit simple et économique d'un point de vue évolutif. De plus, la terminaison de la transcription par roadblock pourrait être un mécanisme mis en jeu non seulement pour protéger les promoteurs des ARNm mais pourrait également protéger d'autres structures qui doivent être mises à l'abri d'une invasion par des polymérases.

### Mécanisme de la terminaison de la transcription par roadblock

La première étape d'une tentative de résolution d'un arrêt transcriptionnel fait intervenir le facteur général d'élongation TFIIIS. En effet, les polymérases commencent par reculer, et l'extrémité 3' de l'ARN en cours de synthèse se retrouve ainsi déplacée hors du site catalytique. TFIIIS stimule le clivage de cet ARN par RNAPII afin de repositionner l'extrémité 3' au niveau du site catalytique. Si cela ne suffit pas à résoudre l'arrêt, alors la grande sous-unité de la polymérase, Rpb1p, est ubiquitinylée. Cela nécessite l'action séquentielle des ubiquitine ligases Rsp5p et Elc1p-Cul3p. Rsp5p réalise la mono-ubiquitylation de Rpb1p puis un complexe comprenant Elc1p et Cul3p catalyse l'allongement de la chaîne de polyubiquitine. A cette étape, cette polyubiquitylation est réversible par l'action de la déubiquitinase Ubp3p. Il est admis que toutes ces étapes sont réalisées au site d'arrêt de la polymérase, sur le complexe ternaire RNAPII-ADN-ARN. Rpb1p est ensuite dégradée par le protéasome qui pourrait être recruté au site d'arrêt. Cette dégradation nécessite l'intervention de l'ATPase Cdc48p. En effet, la sous-unité ubiquitinylée doit être désassemblée de l'holoenzyme pour pouvoir être suffisamment dénaturée afin d'être dégradée par le protéasome. Cdc48p prend en charge cette étape.

Cette dissection de la résolution de l'arrêt transcriptionnel a principalement été réalisée dans le cadre de lésions au niveau de l'ADN (pour revue voir Wilson *et al*, 2013). Dans le cas d'un arrêt de la transcription dû à un événement de roadblock, nous avons déterminé que les ubiquitine ligases Rsp5p et Cul3p intervenaient. Cependant, le mécanisme précis de libération de la polymérase et les facteurs impliqués ne sont pas connus. Par exemple, nous n'avons pas déterminé si la polymérase est toujours dégradée ou

si son inactivation par ubiquitinylation est réversible. Ces questions relatives au mécanisme de résolution de l'arrêt transcriptionnel dû à un roadblock sont importantes et seront adressées dans la laboratoire.

### Identification d'autres facteurs capables de terminer la transcription par un mécanisme de roadblock

Il est clair que tous les facteurs de transcription ne sont pas des roadblockers. Reb1p lie son site avec une bonne affinité (de l'ordre de 70nM). Compte-tenu du mécanisme mis en jeu, on pourrait imaginer que tout facteur ayant une affinité suffisante pour son site soit capable de réaliser de la terminaison par roadblock. Cependant, ce n'est pas le cas. En effet, Gal4p lie également son site avec une bonne affinité mais ne semble pas capable de terminer la transcription par roadblock (Greger & Proudfoot, 1998; Porrua *et al*, 2012).

J'ai contribué à Identifier d'autres facteurs capables d'induire la terminaison de la transcription par roadblock ce qui a permis également d'avoir une vision plus claire de l'étendue du processus de son rôle protecteur.

### Rap1p et Abf1p sont des facteurs de roadblock

Nous avons identifié deux candidats. Le premier, Rap1p, est, comme Reb1p, un facteur de transcription essentiel et abondant de la famille Myb-like. L'éventualité d'un rôle de Rap1p comme facteur de terminaison selon un mécanisme de roadblock avait déjà été évoquée, cependant ces travaux ne permettait pas d'affirmer que le mécanisme était bien du roadblock

et surtout, aucune piste n'était avancée concernant le rôle physiologique de cette terminaison (Yarrington *et al*, 2012). Les travaux réalisés dans notre laboratoire permettent de confirmer un rôle de Rap1p dans la terminaison et soutiennent un mécanisme de terminaison par roadblock. Un cas similaire est le facteur Abf1p/Reb2p. Ce facteur de transcription possède également un rôle dans la réplication et il lie de nombreuses origines de réplication. Comme dans le cas de Rap1p, nous avons montré l'existence d'une pause transcriptionnelle en amont du site de liaison d'Abf1p. D'autres arguments issus de la littérature indiquent qu' Abf1p pourrait fonctionner comme un facteur de roadblock (Valerius *et al*, 2002). L'étude de l'impact à l'échelle du génome de la terminaison dépendante de Rap1p et Abf1p est décrite dans le manuscrit.

Puisque Rap1p et Abf1p sont tous les deux des facteurs de transcription, leurs sites de liaison sont souvent localisés dans des promoteurs. Ils pourraient donc, comme Reb1p, participer à la protection de ces promoteurs d'une extinction par interférence transcriptionnelle.

#### Autres cas potentiels de « roadblockers »

Les promoteurs utilisés par l'ARN polymérase III (RNAPIII) présentent certaines similarités avec les promoteurs contenant des sites Reb1p, Rap1p ou Abf1p. En effet, il avait été montré que le facteur d'initiation de la transcription par RNAPIII TFIIC joue, comme les facteurs Reb1p, Rap1p et Abf1p, un rôle de séparateur vis-à-vis de l'hétérochromatine (Fourel *et al*, 2002; Simms *et al*, 2008).

Il a été récemment montré que le promoteur d'un ARNt (transcrit par RNAPIII) peut se comporter comme un terminateur vis-à-vis de la transcription par RNAPII (Korde *et al*, 2014). Les auteurs ont montré que l'extinction du gène situé en aval (*CIS1*) par interférence transcriptionnelle n'était pas sensible aux facteurs de remodelage de la chromatine, comme c'est le cas pour la terminaison roadblock par Reb1p. En revanche, les facteurs d'initiation de la transcription par RNAPIII, en particulier TFIIIB, sont requis. Cette étude ne porte que sur un locus et ne fait pas la preuve d'une terminaison par roadblock. Cependant le locus étudié présente une accumulation précise d'extrémités 3' de transcrits instables (Neil *et al*, 2009) ainsi que de la pause transcriptionnelle (Churchman & Weissman, 2011) ce qui est tout à fait compatible avec une terminaison par roadblock. J'ai étendu mes analyses génomiques aux tRNAs et démontré que le roadblock a bien lieu en 5' et en 3' des tRNAs.

Enfin, les origines de réplication (ARS) sont réputées silencieuses et cette absence de transcription est nécessaire à leur fonction. Nos études montrent que les ARS sont capables de se comporter comme des terminateurs vis-à-vis de la transcription par RNAPII. Cette terminaison pourrait être provoquée par un mécanisme de roadblock puisqu'elle semble précise, bidirectionnelle et que les espèces terminées présentent des extrémités polyA(+) et polyA(-) caractéristique que l'on observe au niveau des transcrits terminés par un mécanisme de roadblock. Le facteur à l'origine de cette terminaison reste à identifier. En effet, il ne s'agit *a priori* d'Abf1p (qui est présent dans un certain nombre d'ARS dont *ARS1*) et pourrait être le complexe de liaison de l'origine de réplication (ORC). La terminaison de la transcription par roadblock au niveau des origines de réplication pourrait avoir un rôle dans la protection de

l'activité de ces ARS. Mes analyses suggèrent que cette « protection » a un impact important sur la fonction des origines de réplication.

Pour générer les données d'analyse transcriptionnelle à l'échelle génomique, nous avons utilisé la technique du CRAC (crosslink and cDNA analysis). Le CRAC consiste à réaliser un pontage covalent aux UV entre protéines et ARN et à purifier en deux étapes, dont une dénaturante, la polymérase (Bohnsack *et al*, 2012; Granneman *et al*, 2009). Cette technique a été utilisée pour générer les données que j'ai analysé. Le CRAC a été utilisé pour analyser les sites de pause de la polymérase (l'un des indicateurs caractéristiques de la terminaison par roadblock) et sa distribution dans des mutants de terminaison.

L'étude de données de CRAC obtenues en absence de nos roadblockers a permis de détecter le défaut de terminaison à l'échelle du génome avec la disparition de la pause transcriptionnelle due à ces facteurs et l'apparition de « reads » en aval du site de terminaison.

Ces défauts de terminaison sont susceptibles d'altérer l'expression de gènes voisins. Lors de nos études, l'utilisation d'une forme tronquée de Rap1p permettant de terminer mais pas d'activer la transcription nous a permis de discriminer quels gènes étaient sous-exprimés du fait d'un événement d'interférence transcriptionnelle (voir Figure 5).

## Analyse de la terminaison secondaire

L'étude de la terminaison par Reb1p et Rap1 nous a permis de mettre à jour un rôle majeur de cette voie de terminaison comme mécanisme de secours pour palier aux défauts d'efficacité de terminateurs d'un autre type, en l'occurrence des terminateurs d'ARNm, donc dépendants du complexe de terminaison CPF-CF. Lorsque le roadblock induit par Reb1p est utilisé pour neutraliser les fuites d'un terminateur primaire, ces fuites sont relativement faibles. Il est donc délicat de visualiser clairement un pic de pause transcriptionnelle. Pour visualiser par CRAC les événements de terminaison secondaire, il a été donc nécessaire d'accroître ces fuites. Pour cela, nous avons inactivé indépendamment différentes voies de terminaison primaires (CPF-CF, NNS, roadblock par Reb1p/Rap1p...) à l'aide du système anchor away ou des mutants (Haruki *et al*, 2008). Le système anchor away consiste à délocaliser le facteur d'intérêt du compartiment cellulaire dans lequel il agit (dans notre cas le noyau) vers un autre (le cytoplasme). Le système anchor away fonctionne très rapidement ce qui permet d'éliminer du noyau le facteur d'intérêt dans des délais suffisamment courts pour limiter les effets indirects. De plus, l'utilisation de ce mode d'inactivation a déjà fait ses preuves quant à l'étude de la terminaison de la transcription (Schaughency *et al*, 2014; Schulz *et al*, 2013).

L'inactivation des voies de terminaison primaires permet la poursuite de l'élongation jusqu'à l'éventuel site de roadblock localisé en aval. Cet afflux de polymérases, en augmentant localement la transcription en amont du site de roadblock utilisé comme terminateur secondaire, a permis de faciliter la détection des pics de pause associés (nous définissons les pics de pause par rapport à la moyenne locale de la transcription). Cette étude

nous a permis d'avoir une vision plus claire de l'importance de la terminaison par roadblock comme mécanisme de secours.

### Régulation de et par la terminaison de la transcription par roadblock

Comme pour la terminaison dépendant du complexe NNS, il pourrait exister des cas de régulation par roadblock conditionnel. Cela pourrait être le cas de facteurs capables de réaliser de la terminaison par roadblock mais qui ne sont exprimés (ou suffisamment exprimés) que dans certaines conditions physiologiques (comme par exemple Hsf1p en cas de choc thermique). De même, des sites de liaison de facteurs de roadblock pourraient dans certaines conditions être inaccessibles du fait de l'encombrement stérique d'autres facteurs ou plus simplement du fait de la présence d'un nucléosome. Il sera intéressant dans le futur de détecter ces cas de roadblock conditionnels en réalisant des expériences de CRAC sur des cellules cultivées dans différentes conditions de milieu, de température.

### **Bibliographie**

- Bohnsack MT, Tollervey D & Granneman S (2012) Identification of RNA helicase target sites by UV cross-linking and analysis of cDNA. *Meth. Enzymol.* **511**: 275–288
- Churchman LS & Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**: 368–373
- Colin J, Candelli T, Porrua O, Boulay J, Zhu C, Lacroute F, Steinmetz LM & Libri D (2014) Roadblock termination by reb1p restricts cryptic and readthrough transcription. *Mol. Cell* **56**: 667–680



- Fourel G, Miyake T, Defossez P-A, Li R & Gilson E (2002) General regulatory factors (GRFs) as genome partitioners. *J. Biol. Chem.* **277**: 41736–41743
- Granneman S, Kudla G, Petfalski E & Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U.S.A.* **106**: 9613–9618
- Greger IH & Proudfoot NJ (1998) Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of *Saccharomyces cerevisiae*. *EMBO J.* **17**: 4771–4779
- Haruki H, Nishikawa J & Laemmli UK (2008) The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol. Cell* **31**: 925–932
- Korde A, Rosselot JM & Donze D (2014) Intergenic transcriptional interference is blocked by RNA polymerase III transcription factor TFIIIB in *Saccharomyces cerevisiae*. *Genetics* **196**: 427–438
- Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM & Jacquier A (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**: 1038–1042
- Porrua O, Hobor F, Boulay J, Kubicek K, D'Aubenton-Carafa Y, Gudipati RK, Stefl R & Libri D (2012) In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J.* **31**: 3935–3948
- Schaughency P, Merran J & Corden JL (2014) Genome-wide mapping of yeast RNA polymerase II termination. *PLoS Genet.* **10**: e1004632
- Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J & Cramer P (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* **155**: 1075–1087
- Simms TA, Dugas SL, Gremillion JC, Ibos ME, Dandurand MN, Toliver TT, Edwards DJ & Donze D (2008) TFIIIC binding sites function as both heterochromatin barriers and chromatin insulators in *Saccharomyces cerevisiae*. *Eukaryotic Cell* **7**: 2078–2086
- Valerius O, Brendel C, Düvel K & Braus GH (2002) Multiple factors prevent transcriptional interference at the yeast ARO4-HIS7 locus. *J. Biol. Chem.* **277**: 21440–21445

Wilson MD, Harreman M & Svejstrup JQ (2013) Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. *Biochim. Biophys. Acta* **1829**: 151–157

Yarrington RM, Richardson SM, Lisa Huang CR & Boeke JD (2012) Novel transcript truncating function of Rap1p revealed by synthetic codon-optimized Ty1 retrotransposon. *Genetics* **190**: 523–535

**Title :** Genomewide analysis of road-block termination

**Keywords :** road-block, transcription, termination

**Abstract :** Transcription of DNA into RNA intermediates constitutes the first step in gene expression. During the last decade, several studies showed that about 80-90% of the genome is transcribed, and that transcription can initiate almost anywhere. This process—known as pervasive transcription—represents a serious threat to proper gene expression as it has the potential to interfere with not only other transcription events, but any DNA-based process. Selective transcription termination is therefore a mechanism of paramount importance for genome transcriptome stability

and correct regulation of gene expression. Here we describe road-block termination, a novel termination mechanism for RNA polymerase II that functions to limit pervasive transcription and buffer the consequences of readthrough transcription at canonical terminators in *S.cerevisiae*. We show that several transcription factors can elicit this termination and that a number of unexpected genomic loci are associated with it. Additionally, we explore the possibility that road-block termination might contribute to specification of replication origins.

**Titre :** Analyse de la voie de terminaison “road-block”

**Mots clefs :** Roadblock, terminaison, transcription

**Résumé :** La transcription de l'ADN en ARN constitue la première étape de l'expression d'un gène. Durant les dix dernières années, plusieurs études ont montré qu'environ 80-90% du génome est transcrit et que la transcription peut démarrer presque partout. Ce phénomène, connu sous le nom de transcription envahissante, représente une menace sérieuse contre l'expression correcte du génome car il peut interférer non seulement avec d'autres événements de transcription mais également avec n'importe quel procédé impliquant l'ADN. Une terminaison sélective est donc un mécanisme de la plus haute importance pour la stabilité du génome

et la correcte régulation de l'expression des gènes. Ici nous décrivons la terminaison road-block, un nouveau mécanisme de la terminaison par l'ARN polymérase II, qui a pour fonction de limiter la transcription envahissante et de limiter les conséquences d'une translecture au niveau des sites de terminaison canoniques de *S.cerevisiae*. Nous démontrons également que plusieurs facteurs de transcription peuvent entraîner cette terminaison et que certains sites génomiques y sont associés. De plus, nous explorons également la possibilité que ces terminaisons road-block puissent contribuer à rendre spécifiques les origines de réplication.