Role of general regulatory factors in the control of gene expression and transcription fidelity

Drice Challal

NNT : 2019SACLS176

Thèse de doctorat

Role of General Regulatory Factors in the Control of Gene Expression and Transcription Fidelity

Thèse de doctorat de l'Université Paris-Saclay
préparée à Université Paris-Sud

École doctorale n°577 Structure et Dynamique des Systèmes Vivants
Spécialité de doctorat: Sciences de la vie et de la santé

Thèse présentée et soutenue à Paris, le 2 juillet 2019, par

Drice Challal

Composition du Jury :

M. Olivier Namy Directeur de recherche, CNRS (I2BC)	Président
M. Torben Heick Jensen Professeur, Aarhus University (Department of Molecular Biology and Genetics)	Rapporteur
M. Cosmin Saveanu Directeur de recherche, Institut Pasteur (Génétique des Interactions Macromoléculaire)	Rapporteur
Mme. Claire Rougeulle Directeur de recherche, CNRS (Epigénétique et Destin Cellulaire)	Examinateur
M. Stéphane Marcand Directeur de recherche, CEA (SGCSR)	Examinateur
M. Domenico Libri Directeur de recherche, CNRS (Institut Jacques Monod)	Directeur de thèse

ACKNOWLEDGMENTS

First, I would like to thank the members of the jury, Torben Heick Jensen, Cosmin Saveanu, Stephan Marcand, Claire Rougeulle and Olivier Namy for accepting to read, comment and evaluate my work.

À Domenico Libri, mon directeur de thèse, je tiens à te remercier de m'avoir accueilli au sein de ton laboratoire. Je suis sincèrement reconnaissant de la confiance que tu m'as accordée durant ces années. Merci d'avoir été à l'écoute autant sur le plan scientifique qu'humain et de m'avoir donné la possibilité de m'émanciper. Merci d'avoir cru en moi et d'avoir fait preuve de soutien. Mon seul regret aura été de ne jamais avoir eu de grands désaccords avec toi sur le plan scientifique.

Merci à Jessie Colin de m'avoir encadré durant mes premières années au sein du laboratoire. Je n'oublierai pas ton obsession pour le lab book bien organisé et en Arial Narrow, tes conseils précieux sur le temps d'incubation minimum pour avoir le temps d'aller fumer et tes conseils sur l'importance de fédérer une équipe autour d'une bière. Merci d'avoir partagé avec moi ces longs mois de CRACage à optimiser. Tu m'auras également enseigné que la persévérance est un élément clef du chercheur. Merci de ta confiance, de m'avoir poussé à être plus autonome et d'avoir pris le temps de répondre à mes innombrables questions.

I would like to thank all the past and present members of the Libri team for their kindness, their precious help and the good moments spent together. Merci plus particulièrement à Julien Gros, Tommaso Villa, Odil Porrua et Mathieu Rougemaille d'avoir partagé votre expérience. Merci de votre bienveillance, vos conseils et votre aide au quotidien. Tommaso, si je t'avais donné un euro par question posée, tu serais probablement déjà à Rome dans une immense maison. Pour cela je te suis sincèrement reconnaissant. À toutes les personnes ayant partagé avec moi les différents projets de thèse (Tito Candelli, Mara Barucco, Jean-Baptiste Briand…), merci de votre aide, vos idées et du temps passé à parler de science.

Merci à Slawomir Kubik et Frank Feuerbach sans qui mon projet n'aurait pu si bien terminer. Vous avez été d'une aide incroyable. Frank, le wikipedia de la levure, ton impressionnant savoir et ta rigueur scientifique ont été des éléments plus qu'appréciables durant ces journées de manip.

A David Shore, Anna Babour, Vincent Geli et Sebastian Chavez, merci de votre confiance. Vous m'avez donné la chance de participer à vos projets et ce fut un honneur et un plaisir de travailler avec vous.

Gwenaël Badis-Bréard, Simon Lebaron, Benoit Palancade et Jessie Colin je vous remercie d'avoir été présents à mes comités de thèse. Merci pour vos conseils et pour le temps que vous avez su m'accorder. Gwenaël, je suis vraiment ravi d'avoir discuté avec toi de la suite de ma carrière. Cela m'aura notamment permis de trouver mon labo de post-doc.

Merci aux équipes du 3ème étage de l'institut Jacques Monod d'avoir facilité mon travail. Je remercie plus particulièrement Benoit Palancade et Sébastien Léon d'avoir pris le temps de discuter avec moi quand j'en avais besoin. Vos conseils ont été très enrichissants. Merci également à la plateforme NGS de l'i2BC et à Yan Jaszczyszyn pour votre efficacité et votre travail de qualité.

A l'équipe d'enseignants de Paris-Sud, merci d'avoir fait de ces heures d'enseignement un moment si agréable. Je remercie notamment l'équipe de TP Bio Mol et d'Anglais scientifique de leur confiance. Votre aide a été très appréciable dans ces moments où il fallait prétendre être un enseignant expérimenté. Merci d'avoir transmis votre passion.

À tous mes formidables amis, je tiens à vous transmettre ma gratitude. À la JMTeam et aux +1, +2… (vous êtes trop pour tous vous citer), à Marion, Manon, Laurine, Kevin, Vicky… un grand merci pour votre folie et votre soutien à toute heure du jour et de la nuit. Vous avez su rendre ces années beaucoup plus agréables et entretenir ma vie sociale hors du laboratoire. Ann-Sé, merci d'avoir fait semblant de t'extasier devant les photos de ma thèse à mesure que le nombre de pages augmentait. À Amandine, compagnon de paillasse au départ et aujourd'hui amie, un immense merci d'avoir su me faire rire au quotidien, merci pour ta joie de vivre, ton aide sans limite et tes heures/jours passés à lire et relire ma thèse, mon projet de post-doc, mes mails … pour corriger les innombrables fautes. Merci d'avoir rendu ces meetings en Espagne, en Italie et à New York si instructifs socialement. À Steffie et Maeva, je vous remercie sincèrement d'avoir été présentes depuis tant d'années, d'avoir toujours fait preuve d'un soutien incroyable et de votre amitié sans faille, et je sais que vous avez, à votre façon, contribué à ma réussite.

Merci à ma famille, à mes frères, à Soukaina et Louis et à la famille Guilbert/Treuil d'avoir participé à mon équilibre de vie pendant ces 4 ans (et plus). Je remercie Céline, Clément, Isabelle et le reste de ma belle-famille d'aussi avoir été très présents et à l'écoute. Merci à ma mère de m'avoir toujours montré qu'elle croyait en moi et de n'avoir cessé de me rappeler comment j'en étais arrivé là. C'est à travers tes yeux que j'ai toujours pu ressentir le plus de fierté et d'admiration. Tu représentes un exemple de combativité et une source de courage incroyable.

Enfin, à Jeremy, c'est à toi que je dois probablement le plus. Tu as fait preuve d'une patience inimaginable, et d'un soutien presque infini. Tu as su m'aider au quotidien et faciliter chaque étape de ma vie de thésard (et je ne parle pas uniquement de faire le ménage, la vaisselle, les courses, t'occuper des taches administratives, et des petits travaux de l'appartement). Tu as toujours cru en moi et en ma réussite, tu as pris le temps d'écouter mes répétitions de talk en anglais même quand tu ne comprenais rien, tu m'as aidé dans les moments de faiblesse et de doute et m'as redonné le sourire quand j'en avais le plus besoin. Aujourd'hui, tu es même prêt à tout sacrifier pour me suivre dans mon projet de l'autre côté de l'Atlantique sans te poser de questions. Pour tout ça et pour bien plus encore, je te serai éternellement reconnaissant.

ABSTRACT

The last decades have been marked by the discovery of pervasive transcription. Indeed, many studies have shown that transcription by RNA polymerase II is not restricted to annotated regions but is widespread in eukaryotic genomes, leading to the production of a plethora of non-coding RNAs. Precise delimitation of transcriptional units appears to be essential to ensure robust fidelity of gene expression and to maintain the integrity of DNA-associated events by preventing the occurrence of conflicts with transcription. In this respect, accurate transcription initiation and termination represent crucial mechanisms to partition the genome and define the correct processing of RNA molecules. Here, we show that yeast general regulatory factors (GRFs), a class of highly expressed transcription regulators, control pervasive transcription at the level of initiation and termination and are also involved in the fidelity of initiation of mRNA-coding genes. We demonstrate that GRFs bound at promoter regions can elicit transcription termination by physically impeding the progression of polymerases mainly deriving from readthrough transcription at upstream canonical termination sites. We provide evidence that this termination pathway named roadblock is widespread throughout the yeast genome and protects promoter regions from transcriptional interference. Furthermore, we establish that the presence of general regulatory factors limits pervasive transcription at the level of initiation, notably by occluding spurious transcription start sites present in the vicinity of their binding sites. We also unveil the importance of these factors in promoting correct transcription start site selection at mRNA-coding genes thus favouring the synthesis of transcripts with an appropriate coding potential. Finally, we determine that the role of GRFs in controlling proper initiation is intimately linked to their ability to correctly position nucleosomes in promoters, a role that occurs independently from but in cooperation with chromatin remodelers.

RESUME

Ces dernières décennies ont été marquées par la découverte de la transcription dite « cachée » ou « pervasive ». Il a été en effet montré que la majeure partie du génome des eucaryotes est transcrite, donnant naissance à la formation de nombreux ARNs non-codants. La délimitation des unités de transcription apparait essentielle dans le contrôle de l'expression des gènes mais également dans le maintien de l'intégrité des processus associés à l'ADN en limitant notamment l'apparition de conflits avec la transcription. Dans ce contexte, l'initiation et la terminaison de la transcription représentent des étapes clés dans le partitionnement du génome et le métabolisme des ARNs. Nous avons montré que certains facteurs de transcription, appelés GRFs (General Regulatory Factors) chez la levure *S. cerevisiae*, jouent un rôle important dans le contrôle de la transcription pervasive à la fois au niveau de l'initiation mais également de la terminaison de la transcription et sont également requis pour assurer la fidélité de la transcription des gènes codant les ARN messagers. Nous avons prouvé que les GRFs liés au niveau des régions promotrices sont capables d'induire la terminaison de la transcription en bloquant physiquement la progression d'ARN polymérases issues de la translecture des terminateurs situés en amont. D'après nos études, cette voie de terminaison appelée « roadblock » est très répandue à l'échelle du génome et joue un rôle important dans la protection des promoteurs contre l'interférence transcriptionnelle. Nous avons également découvert que les GRFs limitent la transcription pervasive en obstruant les sites d'initiations ectopiques situés à proximité de leur site de fixation sur l'ADN. Ces facteurs sont aussi impliqués dans le contrôle de l'expression des gènes codants en favorisant l'utilisation de sites d'initiations les plus appropriés, c'est-à-dire, permettant la synthèse d'ARNs ayant un fort potentiel codant. Le rôle des GRFs dans le contrôle de l'initiation apparait intimement lié à leur capacité à correctement positionner les nucléosomes au niveau des promoteurs en collaboration avec les facteurs de remodelage de la chromatine.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

INTRODUCTION

I - Mechanism of Class II Genes Transcription in Eukaryotes

Transcription is a fundamental process present in all living organisms. It allows the transfer of the genetic information carried on the deoxy-ribonucleic acid (DNA) to the ribonucleic acid (RNA). Synthesis of RNA molecules is carried out by enzymatic complexes named RNA Polymerases (RNAP). Although RNAPs are found in all three-domain of life (i.e. bacteria, eukaryotes and archaea), their composition and number vary across evolution. While a single RNAP is found in bacteria and archaea, eukaryotic cells contain three main RNAPs: RNAPI, responsible for the transcription of Ribosomal RNAs (rRNA), RNAPII, dedicated to the transcription of protein-coding genes and a few classes of Non-Coding RNAs (ncRNA) and RNAPIII that mainly synthesizes Transfer RNAs (tRNA) and the 5S rRNA molecule. In plants, specialized forms of RNAPII (RNAPIV and V) have been described. These polymerases play a major role in RNA-directed DNA methylation and the production of small interfering transcripts (McKinlay et al., 2018).

Every transcription cycle event can be divided into three main steps: transcription initiation, elongation and termination. The first chapter of this manuscript aims at describing the mechanisms of RNAPII-dependent transcription events and their layers of regulation. A particular focus will be devoted to transcription occurring in the yeast *Saccharomyces cerevisiae*, the model organism used in the host laboratory. When necessary, parallels will be made with transcription occurring in bacteria, metazoan or other higher eukaryotes.

1. Transcription Initiation

1.1 Core promoter elements

Transcription of RNAPII-dependent genes (class II genes) initiates with the assembly of General Transcription Factors (GTF) and RNAPII on a DNA region referred to as the core promoter, which together, define the Pre-Initiation Complex (PIC).

Core promoters are organised into modular DNA elements that usually extend between -40 to +40 Base Pairs (bp) from the Transcription Start Site (TSS) (Figure 1). Although many sequence motifs have been characterized, core promoters do not contain universal sequence elements but rather vary in structure and function. Metazoan core promoters are composed of the Initiator (Inr), the TATA box, the TFIIB Recognition Element (BRE) motif, the Downstream Promoter Element (DPE) and Motif Ten Element (MTE) (Juven-Gershon and Kadonaga, 2010). However, these elements are generally scarce in *Saccharomyces cerevisiae*.

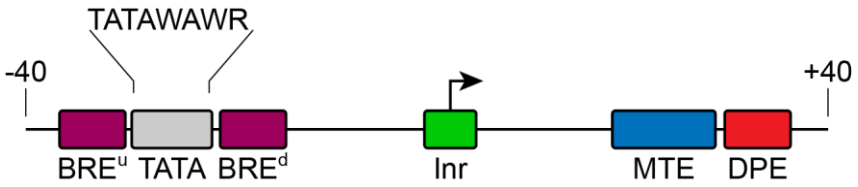


Figure 1. The core promoter elements. -40 and +40 represent the distance in base pair from the transcription start site (black arrow). Inr, BRE, MTE and DPE correspond to the initiator, recognition element, motif ten element and downstream promoter element respectively. Inspired from Juven-Gershon and Kadonaga, 2010.

The TATA box is the first core promoter element that was discovered and is the most conserved from yeast to metazoan. The consensus sequence (TATAWAWR, with W = A/T and R = A/G) serves as a docking site for the TATA-Binding Protein (TBP, a component of the general transcription factor TFIID, see below) (Dikstein, 2011). The TATA box has long been thought to be present at every promoter of genes. In reality, in mammals, the TATA box is present in only 10 to 15% of genes and is located from -35 to -25 bp from the TSS. In *S. cerevisiae*, this element is present at ~20% of all genes and is located farther from the TSS compared to most eukaryotes (-120 to -40 bp upstream from the TSS) (Basehoar et al., 2004; Yang et al., 2007). Interestingly however, Rhee and Pugh reported that most TATA-less promoters actually contain deviant TATA-box consensus that vary from the canonical sequence at one or two positions (Rhee and Pugh, 2012). Presence of these mismatches is

not neutral and has consequences on the organisation and assembly of the PIC (discussed later).

The initiator is a metazoan conserved element that encompasses the TSS. In yeast, Inr-like sequences have been shown to be present at 40% of all core promoters including both TATA and TATA-like associated regions (Yang et al., 2007).

The BRE, MTE and DPE sequences are only found in metazoan. These sequences are all bound by GTFs and together, influence the level of basal transcription and the assembly of the PIC (Juven-Gershon and Kadonaga, 2010).

1.2 PIC assembly on DNA

Formation of the PIC on DNA is the result of the sequential assembly of basal transcription factors together with the RNAPII (Figure 2). For thirty years, many efforts have been devoted to understanding the successive steps that lead to the loading of RNAPII on the template DNA and to solving the crystal structure of components of the PIC. While no universal DNA element controlling gene transcription can be defined in gene promoters, the RNAPII and initiation factors are rather strongly conserved in evolution.

General transcription factors are divided into 5 main complexes: TFIIB, TFIID, TFIIE, TFIIF and TFIIH. TFIID is the first complex loaded on the DNA. Its recruitment occurs via the recognition of the TATA box by TBP. TFIID comprises 14 other subunits named TAFs (1 to 14) for TBP-Associated Factors. The binding of TFIID is subsequently stabilized by the recruitment of TFIIA and TFIIB, two smaller complexes composed of 2 and 1 subunits respectively. In addition to its role in stabilization of the PIC, TFIIB is also involved in TSS selection (see I.2.1) and favours the recruitment of the RNA polymerase. Recruitment of TFIIF (2 to 3 subunits) together with the RNAPII (12 subunits) then leads to the formation of a stable complex called “core initiation complex”. The interaction of TFIIE with the core initiation complex enables the recruitment of TFIIH and favours the melting of the DNA double helix. TFIIH, composed of 10 subunits contains an ATPase, a helicase and a kinase activity. The helicase activity is necessary for promoter opening and transcription start while the kinase activity carried by Kin28 (Cdk7 in human) is responsible for the phosphorylation of the serine 5 and 7 of the Carboxy-Terminal Domain (CTD) of the largest RNAPII subunit (discussed later) (Buratowski et al., 1989; Cheung and Cramer, 2012; Sainsbury et al., 2015; Sikorski and Buratowski, 2009).

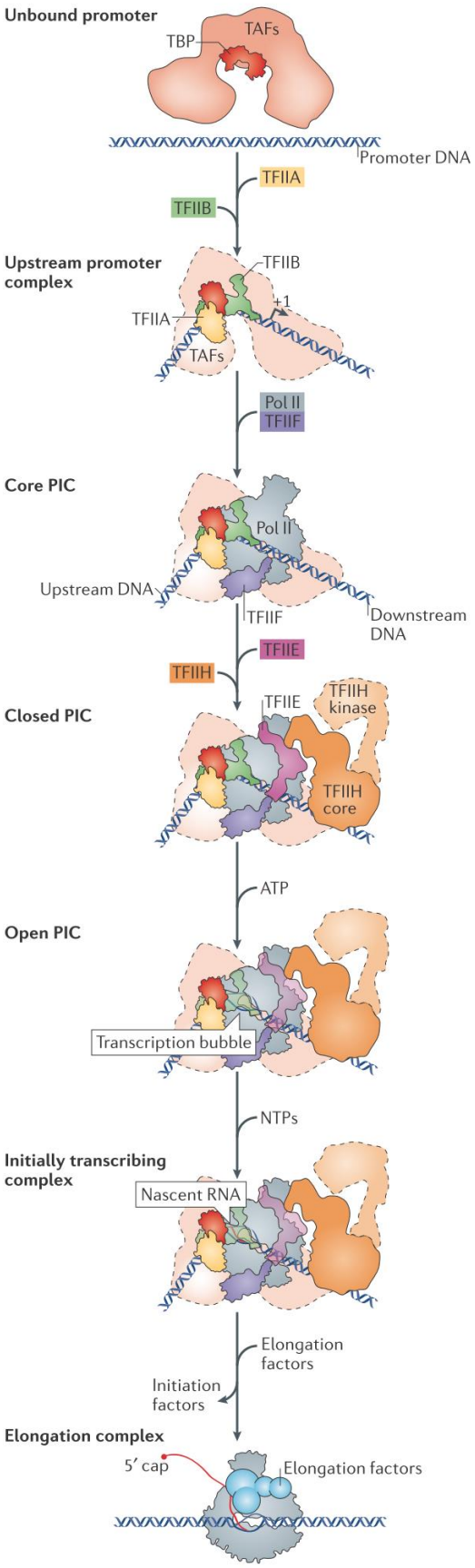


Figure 2. Schematic representation of PIC assembly. The binding of TBP and TAFs (TFIID) at promoters leads to the bending of the DNA fiber. TFIIB and TFIIA stabilize TFIID prior to the recruitment of RNAPII and TFIIF which constitutes the core initiation complex. The formation of the PIC is achieved after the recruitment of TFIIE and TFIIH. The subsequent formation of the transcription bubble requires the hydrolysis of Adenosine TriphosPhate (ATP). The polymerase then initiates the synthesis of the RNA molecule and enters a productive elongation mode. From Sainsbury et al., 2015.

As previously mentioned, a vast majority of core promoters contains a less conserved TATA-box consensus (TATA-like). The presence of this sequence has been shown to be more prevalent among constitutively expressed genes (housekeeping genes) and correlates with the presence at these genes of TBP associated with TAFs to form the TFIID complex. In contrast, highly regulated genes such as stress-induced genes are generally associated with canonical TATA sequences and with the presence of the SAGA (Spt-Ada-Gcn5 Acetyltransferase) coactivator (Huisinga and Pugh, 2004; Kubik et al., 2017). These two complexes share common subunits including the TATA-binding protein and other TAFs. Interestingly, two recent studies published in 2017 challenge this notion of SAGA *versus* TFIID-dominated genes. Using Chromatin Endogenous Cleavage and sequencing (ChEC-Seq), a formaldehyde-independent method for mapping DNA-binding proteins, the authors revealed that both TFIID and SAGA are present at TATA and TATA-less promoters. Consistent with this notion, depletion of either complex leads to the perturbation of nearly all yeast Messenger RNA (mRNA) coding genes suggesting a synergistic role for SAGA and TFIID in regulating RNAPII transcription (Baptista et al., 2017; Warfield et al., 2017).

An additional important coactivator of transcription is the mediator. In budding yeast, the mediator is composed of 25 subunits organised into four distinct modules (head, tail, kinase and middle). This well conserved complex plays a role in the formation and stabilisation of the PIC through multiple and direct interactions with GTFs. A well-established function of the mediator is also to constitute a bridge between sequence specific transcription regulators (activators or repressors) and the polymerase (Hahn and Young, 2011; Poss et al., 2013).

1.3 Chromatin organisation at promoters

In eukaryotes, nuclear DNA is wrapped around histone proteins that arrange DNA into a more compact structure called chromatin. The single unit of chromatin, called nucleosome, consists of two copies each of the four core histones H2A, H2B, H3 and H4, surrounded by approximatively 147 bp of DNA (Kornberg, 1974; Luger, 1997). The presence of histones interferes with the binding of proteins to the DNA thus influencing the occurrence of many DNA associated events including transcription initiation, replication or DNA repair (Field et al., 2008; Han and Grunstein, 1988; Lee et al., 2004). Importantly, nucleosomes are highly dynamic structures, which represents an important source of regulation. For instance, nucleosomes can slide along the DNA, can be evicted from the DNA and are also subject to a myriad of Post-Translational Modifications (PTM) (review: Lai and Pugh, 2017).

The position of nucleosomes can be assessed by the deep sequencing of DNA that is resistant to digestion by Micrococcal DNase (MNase-Seq). This has allowed the genome-wide cartography of nucleosomes. Within genes, nucleosomes are usually regularly spaced and separated by a short DNA linker. In yeast, the average length of the linker sequence is 15 bp but varies between species (Jansen and Verstrepen, 2011; Jiang and Pugh, 2009). Unlike gene body, promoters of actively transcribed genes have been shown to be depleted in nucleosomes (Nucleosome Free or Depleted Regions, NFR / NDR) (Figure 3). NDRs in promoters is a conserved feature of all eukaryotes from yeast to human. These regions are usually located immediately upstream of transcription start sites (Lee et al., 2004; Mavrich et al., 2008a; Schones et al., 2008; Yuan, 2005). NDRs are bordered by two well-positioned nucleosomes referred to as +1 and -1 nucleosomes. The "+1" is the first of the array of nucleosomes present along the gene body. Its exact position and histone composition influences transcription, notably by affecting the binding of transcription factor (Lai and Pugh, 2017; Lee et al., 2007b; Shivaswamy et al., 2008).

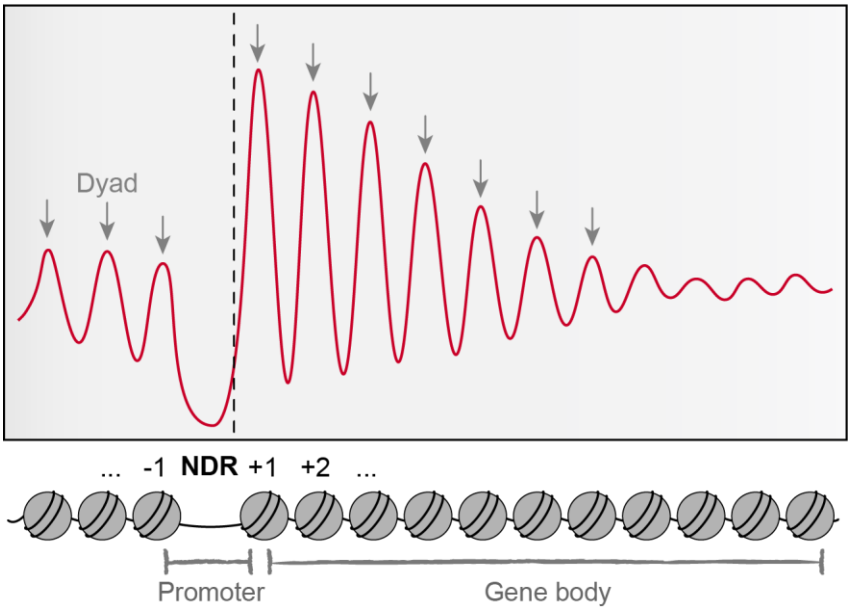


Figure 3. Nucleosome organisation at class II genes. The red line corresponds to the average nucleosome occupancy along the gene. Peaks indicated by a grey arrow represent the position of the nucleosome center and valleys indicate regions with low occupancy. Promoters are characterized by the presence of nucleosome depleted regions (NDR) flanked by two well positioned nucleosomes termed +1 and -1. The nucleosome occupancy becomes less defined (or "fuzzy") at the 3' end of the gene body. Adapted from Lai et Pugh 2017.

The +1 nucleosome has the particularity to contain the H2A.Z variant of the H2A histone. Such feature has been proposed to favour the eviction of the latter by destabilizing the promoter-proximal side of the nucleosome, thus facilitating the passage of the polymerase during transcription elongation (Albert et al., 2007; Lai and Pugh, 2017). In yeast, the -1 nucleosome

is also characterized by this H2AZ variant. In *Drosophila* and human, H2AZ-containing nucleosomes are absent in the -1 position, but usually extend beyond the +1 position across the gene body (nucleosomes +2 and +3) (Mavrich et al., 2008a; Schones et al., 2008).

Importantly, the position of the +1 nucleosome correlates with the position of the TSS. In *S.cerevisiae*, the TSS is buried in the +1 nucleosome and located 12 to 15 NucleoTides (nt) downstream of its upstream border (Hughes et al., 2012; Lee et al., 2007b; Mavrich et al., 2008b; Tsankov et al., 2010). Unlike NDR formation or H2A.Z presence, the distance of the +1 nucleosome relative to the TSS is not a conserved feature among eukaryotes. The 5' border of the +1 nucleosome is located ~60 bp downstream of the TSS in *Drosophila*, (Mavrich et al., 2008a) and at ~40 bp downstream of the TSS of actively transcribed genes in human (Schones et al., 2008). Such differences may have important consequences regarding the mechanisms of TSS selection in different species.

In addition to be associated with a canonical TATA element (see I.1.2), promoters of regulated genes are also characterized by a dynamic and generally higher nucleosome occupancy close to TSSs. Presence of nucleosomes is believed to outcompete the binding of transcription factors by burying *cis*-regulatory sequences. Upon transcriptional activation, nucleosomes have been shown to be evicted from promoters, as for heat-shock genes upon stress response (Shivaswamy et al., 2008) or genes involved in phosphate metabolism (Ertel et al., 2010). This is distinct from what is observed for promoters of housekeeping genes that have a static and well-defined architecture of nucleosomes, which enables the constant accessibility of the transcription machinery to core elements (Lee et al., 2004; Tirosh and Barkai, 2008). Taken together, these studies confirm the clear relationship that exists between the nucleosome architecture at promoters and gene activation.

In 2015, a study conducted by Kubik and colleagues unveiled the presence of an additional and non-canonical nucleosome within yeast NDRs, hence challenging the above-mentioned idea that active promoters are simply flanked by well-positioned +1 and -1 nucleosomes (Kubik et al., 2015). By using a decreased concentration of MNase for mapping nucleosome position, the authors found that 40% of mRNA-coding genes display an unstable nucleosome named "Fragile Nucleosome" (FN) that is surrounded by only 100 bp of DNA. FNs are found at broad NDRs (i.e. for which the distance between two stables -1 and +1 nucleosomes exceeds 300 bp) and are characteristic of highly expressed genes. The occupancy of a FN at NDRs would be a consequence of the presence of a long naked DNA fragment and would serve to restrict the access of RNAPII to specific regions. The existence of FNs remains, however, a matter of debate. While Chereji and colleagues propose that this particle actually correspond to non-

histone protein complexes (Chereji et al., 2017), Brahma and Henikoff could confirm the presence of FNs by immunoprecipitating chromatin following MNase-digestion (Brahma and Henikoff, 2019).

In summary: Transcription initiation is a complex and conserved process that involves many factors that assemble in a sequential manner at promoter regions to form the pre-initiation complex. Assembly of the PIC is dependent both on the presence of cis-regulatory sequences recognized by general transcription factors and on the formation of a nucleosome free region that exposes core promoter elements to GTFs.

2. Transcription Elongation

2.1 Promoter clearance and TSS selection

During transcription elongation, the RNA polymerase travels along DNA, catalysing the addition of nucleotides on the nascent transcript. Elongation starts with the release of the polymerase from promoters, i.e. promoter clearance (also called promoter escape). The transition from the initiation complex to a fully committed transcription elongation state represents a real challenge for the polymerase. Within the PIC, TFIIE and TFIIF complexes respectively bind to the Rpb1 clamp domain and Rpb2 protrusion domain of the polymerase (Chen et al., 2007; Luse, 2013). Moreover, the TFIIB complex occupies the RNA exit channel of RNAPII where the nascent transcript will next reside. These interactions between components of the PIC need to be lost in order to allow the release of the polymerase. One important requirement for polymerase initiation is the opening of the DNA template that forms the transcription bubble upstream of the TSS. This step is carried out by Ssl2 (XPB in human), a component of the TFIIH complex that possesses an ATP-dependent DNA-translocase activity and that acts by pulling the downstream DNA sequence into the polymerase (Fishburn et al., 2015; Sainsbury et al., 2015).

During the early stages of elongation, the polymerase still in contact with TFIIB undergoes many events of abortive transcription (Luse, 2013). As the neo-synthetized RNA lengthens, the elongation complex gets stabilized by the formation of the DNA-RNA hybrid within the RNAPII holoenzyme. In parallel, the activity of TFIIH leads to the progressive unwinding of the transcription bubble until it reaches ~18 bp, a size that is typical of the early stages of elongation. At one point, the upstream part of the bubble suddenly closes (bubble collapse transition) to trap and stabilize the polymerase into a smaller (~10 bp long) bubble, more characteristic of a productively elongating complex. About 13 nt downstream of the TSS, the interaction between TFIIB and RNAPII is lost (Čabart et al., 2011). Once the transcript reaches ~17 nt, the 5' end is released from the template DNA and enters the exit channel of the polymerase. Promoter clearance is achieved when the transcript is about 30 nt long which corresponds to the distance that is required for the polymerase to adopt all the characteristics of a proficient elongation complex (Liu et al., 2011; Luse, 2013). This transition state is characterized by the hyperphosphorylation of the polymerase CTD at serine 5 (see section I.2.2). The phosphorylation of this residue by Kin28 (TFIIH component) has been shown to impede the interaction between Rpb1 and the mediator, thus favouring the release of the polymerase from the promoter (Max et al., 2007; Wong et al., 2014). Finally, apart from TFIIB

that gets destabilized and TFIIH that travels with the polymerase during the first part of transcription elongation across the gene, most of the other GTFs remain bound to the core promoter, thus favouring the reassembly of the PIC for further rounds of initiation (Dvir et al., 1997; Luse, 2013; Sainsbury et al., 2015).

Despite the wealth of studies on the topic, it remains unclear how the exact position of the transcription start site is defined. In higher eukaryotes, transcription initiation usually occurs at a single site and in a narrow region located about 30 bp downstream of the TATA-box (see 1.1). In this context, the architecture of the PIC is thought to be the major determinant of the TSS selection by directly placing the active center of the RNAPII holo-enzyme on top of the initiation site (Bushnell et al., 2004; Leuther et al., 1996). Although this pattern is to some extent conserved in *S. pombe* (+30 to +70 from the TATA-box), it diverges completely in *S.cerevisiae* (+40 to +120 from the TATA-box) (Chen and Struhl, 1985; Nagawa and Fink, 1985; Zhang and Dietrich, 2005) where TSSs are selected after a scanning step that does not require RNA synthesis and is driven by the Ssl2 translocase motor (Fishburn et al., 2016; Kuehner and Brow, 2006).

Prior studies performed in the 90s defined the polymerase and TFIIB as being the major determinants of the TSS utilization in budding yeast. For instance, deletion of the non-essential *RPB9* gene (RNAPII subunit) (Hull et al., 1995), mutations of the largest subunit of the polymerase (*RPB1*) (Berroteran et al., 1994) and mutation of *SUA7* (TFIIB) (Pinto et al., 1992, 1994) have been shown to be associated with upstream (Rpb9) or downstream (Rpb1 and Sua7) shifts of the TSS (Kwapisz et al., 2008; Sun et al., 1996; Thiebaut et al., 2008). Consistent with the role of these factors in TSS selection, *in vitro* assays using the *S. pombe* polymerase and TFIIB in a *S. cerevisiae* transcription context (i.e. *S. cerevisiae* promoter and other GTFs) confer a species-specific transcription profile that resembles the one of *S. pombe* (Li et al., 1994; Yang and Ponticelli, 2012). TFIIF has also been involved in TSS selection since mutation of *TFG1* or *TFG2* (components of TFIIF) is accompanied by an overall upstream shift of the TSS that is exacerbated in a *rpb9*Δ (Δ = deletion) strain and partially suppressed in *SUA7* mutants (Ghazy et al., 2004). Finally, in two later studies that aimed at mapping the 5' ends of transcripts in yeast, the authors observed a strong bias for transcription to start at the $A(A_{rich})_5NYR$ motif, with Y = C/T and R = A/G and where <u>R</u> correspond to the mapped TSS and is mostly an A (Malabat et al., 2015; Zhang and Dietrich, 2005). Taken together, these data suggest that the selection of the TSS during early elongation in yeast may result from the synergic action of both components of the PIC, and *cis*-regulatory elements present at or close to the TSS.

2.2 Mechanism and function of CTD phosphorylation

The unstructured carboxy-terminal domain of Rpb1, the largest subunit of the yeast RNAPII, is composed of tandem repeats of the heptapeptide Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7. While the consensus sequence of the CTD is conserved, it varies in number of repeats across the eukaryotic kingdom (26 in *S. cerevisiae*, 29 in *S. pombe*, 37 in *D. melanogaster* and *A. thaliana* and 52 in *H. sapiens*). The CTD is a substrate of various kinases and phosphatases that act subsequently to modify the phosphorylation state of the polymerase during the transcription cycle. These post-translational modifications are highly regulated in space and time and play a crucial role in the establishment of specific interactions between the polymerase and various cofactors. Among the 7 residues, 5 (Tyr1, Ser2, 5 and 7 and Thr4) are subject to phosphorylation (Harlen and Churchman, 2017). In this section, I will describe the pattern of the CTD code during transcription and define its interplay with co-transcriptional events.

2.2.1 Dynamic of CTD phosphorylation

The polymerase is recruited to the PIC under its unphosphorylated form, which favours the tight interaction with the mediator (Lu et al., 1991). The Kin28 cyclin-dependent kinase triggers phosphorylation of the Ser5 and Ser7 residues of the CTD leading to the release of the polymerase from promoters (see 2.1) (Kim et al., 2009; Komarnitsky et al., 2000; Max et al., 2007; Wong et al., 2014). The Ser5/7-P (Phosphorylated) state is characteristic of polymerases close to the TSS and undergoing early elongation (Figure 4). As the elongation complex moves away from the TSS, the polymerase undergoes the Ser5-Ser2 transition under the effect of the Ser5-phosphatase Rtr1 (Hunter et al., 2016; Mosley et al., 2009) and the Ser2-kinase Bur1 (Liu et al., 2009; Qiu et al., 2009) even though the extent to which Bur1 participate to the deposition of Ser2-P remains a matter of debate (Bataille et al., 2012; Keogh et al., 2003). The Ser5-Ser2 transition starts occurring about 150 bp downstream of the TSS (Milligan et al., 2016) and is not dependent of gene length but rather on the absolute distance from the initiation site (Bataille et al., 2012; Mayer et al., 2010). Later steps in transcription are characterized by the increase in the Ser2-P mark due to the action of the Ser2-kinase Ctk1 (Bataille et al., 2012; Qiu et al., 2009) and a massive drop in Ser5 phosphorylation caused by Ssu72 (Bataille et al., 2012; Krishnamurthy et al., 2004), a component of the Cleavage and Polyadenylation Factor (CPF) (see I.3.1). Prior to termination, Ssu72 also abrogates the Ser7 phosphorylated mark. Finally, the Ser2 phosphatase Fcp1 is required to bring back the polymerase to its unphosphorylated form thereby allowing its recycling for new rounds of transcription (Bataille et al., 2012; Egloff et al., 2012a).

The phosphorylation events of Ser2, 5 and 7 are the best documented modifications occurring on the CTD. More recently, two other residues have been shown to be reversibly phosphorylated: Tyr1 and Thr4. In yeast, the distribution of Tyr1-P resembles the one of Ser2-P except that it drops before the termination site (Mayer et al., 2012). Glc7, which belongs to the CPF complex, has been proposed to be the major phosphatase of Tyr1 *in vivo* (Schreieck et al., 2014). Thr4 is also increasing in the 5' to 3' direction, however, the phosphatases and kinases responsible for this PTM are still unknown.

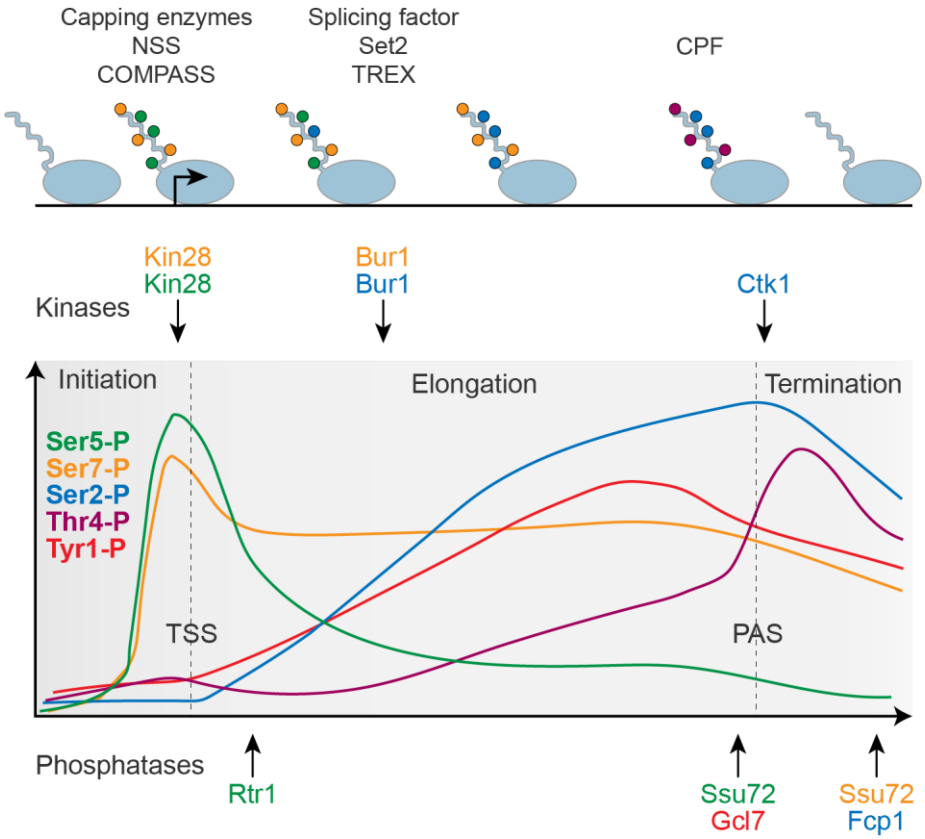


Figure 4. The CTD pattern across yeast protein-coding genes. The average level of Ser5, Ser7, Ser2, Thr4 and Tyr1 phosphorylation at protein-coding genes during the transcription cycle is represented. TSS and PAS indicate the position of the Transcription Start Site and the PolyAdenylation Site respectively. Phosphatases (bottom) and kinases (top) responsible for the establishment of the pattern are indicated with the corresponding colour. The top panel illustrates the average occupancy of the different phosphorylated form of the RNAPII CTD. Co-transcriptional events and factors associated with specific CTD forms are indicated. Adapted from Egloff et al., 2012a; Harlen and Churchman, 2017.

2.2.2 Co-transcriptional events and the CTD code

Several studies have reported the importance of the CTD code in coupling transcription with co-transcriptional events by favouring the recruitment of specific factors such as elongation or termination factors, the capping enzymes or the splicing machinery. Yet, the role of some CTD marks remains unclear and is still under debate.

In addition to its role in promoter clearance, the Ser5-P state serves as a platform for the binding of the 5'-capping enzymes (Komarnitsky et al., 2000; Mayer et al., 2010; McCracken et al., 1997a) whose action close to the promoter enables the rapid protection of the nascent transcript from degradation. Interestingly, cells defective for the TFIIH kinase activity display a strong decrease in mRNA steady state level and a limited effect on RNAPII occupancy along the gene. Moreover, the reduction of mRNA level can be partially counteracted by deletion of *XRN1*, a gene coding for the major cytoplasmic 5'-3' exonuclease, suggesting that the major role of Ser5-P mark might actually be related to its interaction with the capping enzyme and not to transcription (Hong et al., 2009). Supporting the notion that Ser5-P is important for capping, the fusion of the CTD of Rpb1 with the capping enzyme prevent the lethality observed in *S. pombe* cells expressing a S5A CTD form (i.e. mutation of every Ser5 to alanine) (Schwer and Shuman, 2011). The Ser5-P also favours the interaction with the COMPASS (COMplex Protein ASsociated with Set1, described in I.2.3.1) at 5' regions (Ng et al., 2003). Finally, the combined presence of specific RNA sequences and the Ser5-P mark also promotes the recruitment of the Nrd1-Nab3-Sen1 (NNS) complex to trigger early transcription termination (see I.3.2) (Gudipati et al., 2008; Tudek et al., 2014; Vasiljeva et al., 2008). The persistence of the NNS complex within the coding region is limited by the phosphorylation of Tyr1 that prevents its binding (Mayer et al., 2012).

The phosphorylation of Ser2 and Ser5 residues mediates the recruitment of the elongation factor Set2 (see 2.3.1) and the conserved TREX complex which plays a role in the export of mRNA molecules from the nucleus to the cytoplasm (MacKellar and Greenleaf, 2011; Meinel et al., 2013). The splicing machinery is also recruited co-transcriptionally and interacts with both Ser2-P and Ser5-P (Gu et al., 2013; Harlen et al., 2016; McCracken et al., 1997b).

At 3' ends of genes, the increase in RNAPII Ser2-P mark is associated with an enrichment of RNA processing and termination factors. Pcf11 and Rtt103, two components of the CPF-CF (Cleavage and Polyadenylation Factor and Cleavage Factor I) complex in yeast interact with the Ser2-P CTD to trigger termination and catalyse the addition of the polyA tail on the nascent RNA (Harlen and Churchman, 2017; Harlen et al., 2016). Moreover, although Thr4-P has been originally shown to only affect the processing of histone mRNAs (Hsin et al., 2011), a recent study from the Churchman group unveiled the role of this residue in transcription termination. Indeed, in budding yeast, Rtt103 also recognizes Thr4-P after the polyA site to ensure efficient termination (Harlen et al., 2016). Finally, throughout the coding region, the presence of the Tyr1-P mark impairs the early binding of termination factors thus preventing premature termination from the CPF-CF complex (Mayer et al., 2012; Schreieck et al., 2014). Altogether,

these three marks contribute to the accurate localisation of the CPF-CF complex at the 3' end of genes.

In 2019, Collin and colleagues reinvestigated the role of the CTD code in transcription termination using a dual tag system (Collin et al., 2019). A *rpb1* allele carrying different CTD mutations and a flag tag were co-expressed with the wild-type (WT) version of *RPB1*. The distribution of RNAPII was monitored by ChIP and NET-seq. In this context, the authors could confirm the genome-wide role of Ser2 and Thr4 in transcription termination by the CPF-CF complex. Interestingly however, they revealed that the implication of Tyr1 in termination is only restricted to ~100 mRNA-coding genes and does not seem to be dependent on its phosphorylation state. Unlike mRNAs, ncRNAs are more sensitive to mutations of Tyr1 and display a major termination defect upon expression of a Tyr1 phosphomimic form of the polymerase. This difference between mRNAs and ncRNAs is explained by the different termination pathways associated with these two types of transcripts (see I.3). Finally, although Ser7 had been proposed to participate in termination of Small-Nuclear RNAs (snRNA) in mammals (Egloff et al., 2012b), no termination defect was reported in this study (Collin et al., 2019). This dual system represents a powerful tool in the study of the CTD code as it limits secondary and indirect effects that can be observed upon depletion of Rpb1-targeted phosphatases and kinases.

2.3 RNAPII transcription through chromatin

Akin to transcription initiation, the progression of the transcription machinery is also impeded by the presence of nucleosomes. Thus, numerous elongation factors are recruited in order to facilitate the elongation process.

In 1991, Izban and Luze reported that RNAPII transcription through a nucleosomal template is associated with a severe decrease in the elongation rate as compared to a naked DNA template (Izban and Luse, 1991). This work was the first clear evidence exposing the repressive role of chromatin organisation on elongation. Since then, a lot of efforts have been devoted to the characterization of factors and mechanisms that support transcription elongation *in vivo*. Transcription elongation factors the operate on chromatin can be divided into two main groups: (i) elongation factors that participate to PTMs of histones which alter the interaction between DNA and histones and favour the recruitment of other factors such as chromatin remodelers, (ii) elongation factors that are able to slide, evict and reassemble nucleosomes as the polymerase progresses along the gene. Among these factors are the well characterized histone modifiers, the ATP-dependent chromatin remodelers and the histone

chaperones (Selth et al., 2010). In this section, I will briefly describe these different factors and their implication in transcription elongation.

2.3.1 Histone modifiers

Covalent histone modifications are often concomitant with transcription elongation of RNA polymerase throughout the gene. The three main characterized modifications are acetylation, methylation and ubiquitylation (Figure 5).

Acetylation

The addition and removal of acetylation groups are carried out by histone acetyltransferases (HAT) and histone deacetylases (HDAC) respectively. Acetylation occurs mainly on lysines located at the N-terminal region of histones. This modification neutralizes the positive charges of histones, which in turn reduces electrostatic interactions with the surrounding DNA molecule (Hong et al., 1993). The decreased affinity between the DNA and acetylated histones destabilises nucleosome architecture thus favouring the binding of other factors to the DNA at promoters and facilitating transcription through chromatin (Lee et al., 1993; Selth et al., 2010). The acetylation mark is predominantly found at promoters of active genes and less abundantly in coding regions (Pokholok et al., 2005). Yet, many HAT and HDAC complexes have been shown to be enriched at coding regions (Gilbert et al., 2004; Govind et al., 2007; Keogh et al., 2005; Wang et al., 2002) and mutations in HATs associated with human neuropathies cause a gradual decrease in RNAPII density along the coding regions of targeted genes (Close et al., 2006). Together, these data suggest a role for HDAC and HAT in both transcription initiation and elongation. A possible model is that HAT and HDAC would both travel with the polymerase to ensure the decompaction (acetylation) of the chromatin downstream the elongation complex and the direct re-compaction (deacetylation) of the transcribed region upstream (Selth et al., 2010).

Methylation

Unlike acetylation, methylation does not directly impact chromatin structure but is rather important for the subsequent recruitment of many other factors. Methylation essentially occurs on lysines and arginines of H3 and H4 histones and are associated with either active (H3K4, H3K36) or silent (H3K9, not present in *S. cerevisiae*) transcription regions (Freitag, 2017). In *S. cerevisiae*, two factors have been identified as carrying a methyltranferase activity: the Set1 protein encompasses into the COMPASS and Set2.

COMPASS is composed of eight proteins including the catalytic subunit Set1 that is responsible for the mono- di- or tri-methylation of lysine 4 in histone H3 (H3K4me1, 2, 3) (Briggs, 2001; Roguev, 2001). Active genes are characterized by a decreasing gradient of the H3K4 methylation profile. While 5' regions are enriched in the H3K4me3 mark, downstream nucleosomes are characterized by the presence of H3K4me2 and H3K4me1 modifications (Liu et al., 2005; Pokholok et al., 2005). The mechanisms and effects underlying the establishment of this gradient still need to be fully determined. Tri-methylation of histones is recognized by the HAT proteins Nua3, Hbo1 and SAGA and is thought to influence the acetylation profile of histone H4 and to destabilize nucleosomes around promoters (Woo et al., 2017). Di-methylation instead has been shown to favour the recruitment of the HDAC complex Set3C (Set3 Hos1 Hst1) to promote histone deacetylation in the body of genes. This deacetylation triggers the compaction of chromatin and somehow promotes efficient transcription elongation (Kim and Buratowski, 2009). The exact mechanism by which the action of Set3C would positively act on transcription is not clear. The Set1–Set3C pathway has also been investigated for its role in gene silencing during transcriptional interference (discussed in section III).

Another well characterized methyltransferase associated with transcription is Set2. This factor methylates lysine 36 in histone H3. H3K36 is more prevalent inside the gene body and is associated with highly expressed genes (Pokholok et al., 2005). The main function of this modification is the recruitment of factors, among which the Rpd3S histone deacetylase complex that recognizes both H3K36me2 and me3 marks and maintains an hypoacetylated state at coding regions (Keogh et al., 2005; Li et al., 2009). This pathway is particularly important to counteract intragenic transcription initiation and is also involved in the mechanism of transcriptional interference (see III), similarly to Set3C.

The recruitment of both Set1 and Set2 is linked with the phosphorylation-state of the polymerase. At the 5' end of active genes, phosphorylation of the serine-5 of the CTD is associated with the recruitment of COMPASS histone methyltransferase complex. The recruitment of COMPASS by Ser5-P CTD is not direct but is mediated by the Paf1 elongation complex (Krogan et al., 2003a; Ng et al., 2003). In 2003, four independent studies have reported the ability of Set2 to interact with both the Ser-2 and Ser-5 phosphorylated forms of the polymerase (Krogan et al., 2003b; Li et al., 2003; Schaft et al., 2003; Xiao et al., 2003). Supporting this notion is the fact that deletion of the *CTK1* kinase (Ser2 phosphorylation) or partial deletion of the CTD affect both H3K36me3 profile and the recruitment of Set2.

Histone methylation has long been thought to be a permanent mark that could be depleted from chromatin only by replacing the corresponding nucleosome. In 2004, Shi and colleagues

were the first to provide evidence for the existence of an enzyme (Lysine Specific histone Demethylase, LSD1) endowed with a demethylase activity. LSD1 is a conserved factor from *S. pombe* to human but is absent in *S. cerevisiae*. Knock-down of *LSD1* causes an increase in both H3K4 methylation and transcription level of targeted genes (Shi et al., 2004). Since then, more than 20 additional demethylases belonging to two main families have been identified from bacteria to human (Shi and Tsukada, 2013).

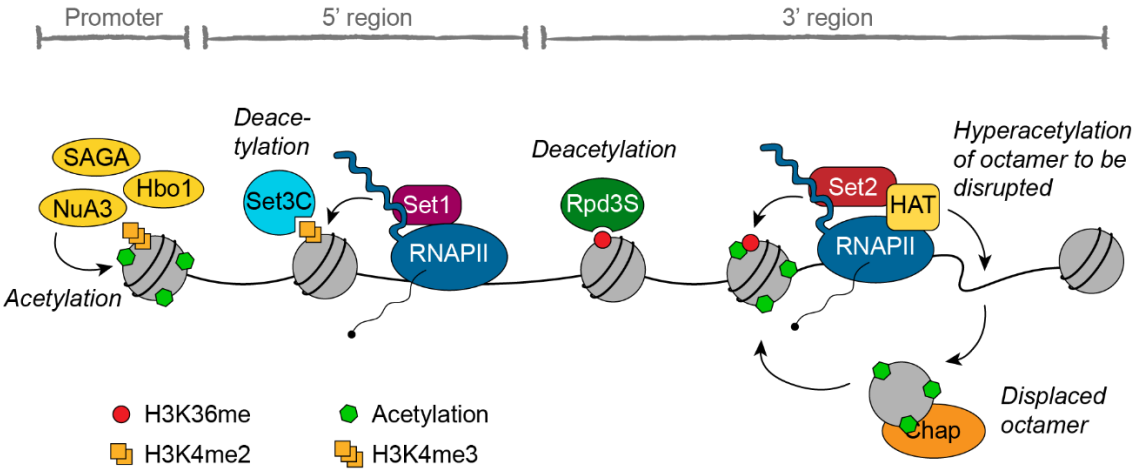


Figure 5. Model of transcription elongation through nucleosomes. The main factors implicated in the progression of RNAPII though chromatin is represented. The scheme essentially recapitulates the central role of acetylation and methylation marks along the gene. Histone methylation mediates the recruitment of various complexes including histone deacetylases. The removal of the acetylation group enables the restoration of the chromatin structure behind the polymerase. Adapted from Selth et al., 2010; Kim and Buratowski 2009.

Ubiquitylation

Ubiquitin is a mall polypeptide of 76 amino acids that can be added to protein under different forms. Its presence is often considered as a signal that triggers degradation by the proteasome. Nonetheless, mono-ubiquitylation is also a source of regulation of protein activity. The addition of ubiquitin is divided into three main steps: (i) an ubiquitin activating enzyme (E1) activates the polypeptide (ii) that is then conjugated via a thioester bond to a cysteine residue (ubiquitin-conjugating enzyme E2) and (iii) finally swaps from the cysteine of E2 onto the lysine residue of the targeted protein by an ubiquitin-protein isopeptide ligase E3 (Weake and Workman, 2008).

The mono-ubiquitylation of histones occurs on H2B protein and is carried out, in *S. cerevisiae*, by a complex composed of Rad6, Bre1 and Lge1 (Weake and Workman, 2008). This mark is present all along the genes and increases gradually into the transcribed regions (Minsky et al., 2008; Xiao et al., 2005). The mono-ubiquitylation mark is associated with highly transcribed genes (Henry et al., 2003; Kao et al., 2004; Xiao et al., 2005) and works in cooperation with

the histone chaperone Spt16 (see below) to reassemble histones during transcription elongation (Fleming et al., 2008; Pavri et al., 2006). Perturbation of the deposition of this mark is associated with a decrease in RNAPII occupancy at the 3' ends of genes but does not prevent its recruitment, thus supporting a specific role of mono-ubiquitylation in transcription elongation (Tanny et al., 2007). Histone ubiquitylation has also been shown in many instances to promote histone methylation by favouring the recruitment of COMPASS and other methyltransferase complexes (Briggs et al., 2002; Dover et al., 2002; Henry et al., 2003; Lee et al., 2007a; Ng et al., 2002; Sun and Allis, 2002). Nonetheless, the positive effect of ubiquitylation on the elongation rate of RNAPII is not mediated by methylation (Shukla and Bhaumik, 2007; Tanny et al., 2007). Finally, histone-ubiquitylation is reversible and de-ubiquitylation is executed by two distinct enzymes in yeast: Ubp8, component of the SAGA complex and Ubp10 (Henry et al., 2003; Weake and Workman, 2008).

2.3.2 Histone chaperones

Up to date, two main histone chaperone complexes have been shown to play an important role in transcription elongation: the FACT (FAcilitating Chromatin Transcription) complex and Spt6. FACT has first been identified *in vitro* by addition of a Hela cell extract on a chromatinized template (Orphanides et al., 1998). It comprises two proteins, Spt16 and Pob3 in yeast (SPT16 and SSRP1 in human), that are able to destabilize nucleosomes by displacing the histone dimer H2A-H2B (Orphanides et al., 1999). FACT also has the ability to promote the re-deposition of the H2A/H2B dimer onto the DNA after the passage of the elongation complex (Belotserkovskaya et al., 2003; Kwak and Lis, 2013).

Spt6 is well-conserved throughout eukaryotes and has the ability to control the structure of chromatin *via* its interactions with histones H3 and H4, which is thought to be particularly important for the reestablishment of nucleosomes in the wake of transcribing RNAPII (Bortvin and Winston, 1996; Kaplan et al., 2003; Selth et al., 2010). Spt6 is recruited early during the 5' transition and has been shown in many instances to interact with the polymerase (Endoh et al., 2004; Kaplan et al., 2000; Mayer et al., 2010). However, unlike most of the transcription associated factors, its recruitment is not mediated by the CTD of the polymerase although its distribution along the gene coincides with the Serine-2 phosphorylated form of RNAPII (Sdano et al., 2017). Last but not least, the elongation rate of RNAPII on naked DNA *in vitro* can be enhanced by the addition of Spt6, suggesting an additional, more direct role for Stp6 in promoting transcription elongation (Endoh et al., 2004).

2.3.3 Chromatin remodelers

Chromatin remodelers are able to slide, evict or deposit histones in an ATP-dependent manner. Their role has been extensively studied in the context of transcription initiation and NDR formation (see II.2). However, the extent to which these complexes affect the transcription elongation rate is not clear. *In vitro* studies revealed that the presence of RSC and to a lesser extent SWI/SNF, ISW1 and Chd1 (detailed in II.2) facilitates the passage of RNAPII through chromatin in an ATP-dependent manner (Carey et al., 2006). *In* vivo, depletion of RSC causes a decrease in RNAPII occupancy within the Open Reading Frame (ORF) of a subset of RSC-targeted genes but not at promoters (Mas et al., 2009; Spain et al., 2014). It is however unclear how RSC would specifically affect RNAPII progression at some, and not all, of its targets.

The role of Chd1 in transcription elongation is clearer. Chd1 interacts with many different elongation factors such as the FACT chaperone complex and is found at coding regions of actively transcribed genes. During elongation, Chd1 together with Isw1 acts to maintain the structure of chromatin by preventing hyperacetylation and helps the compaction of the chromatin upstream of the polymerase (Simic et al., 2003; Smolle et al., 2012).

2.4 Transcriptional pausing within genes

Transcription elongation is not a linear process. During its journey along the DNA, RNA polymerase encounters various obstacles that can slow down or stall its progression. As already mentioned, this includes nucleosomes, but also other DNA-binding factors, DNA damages, nucleotides misincorporation, DNA sequences or NTP depletion. Different mechanisms have evolved to help the polymerase to reach the 3' end of genes and to complete the synthesis of the full-length RNA molecule.

When pausing, RNA polymerase can backtrack, i.e. move backwards along the DNA from 2 to 14 bp (Wilson et al., 2013). During backtracking, the neo-synthetized RNA slides forward relative to the enzyme and enters the front channel of the polymerase, thus exposing the 3' end of the RNA. An elegant work from the Cramer laboratory that aimed at characterizing the structure of the backtracked complex revealed that up to 9 nucleotides of the nascent RNA can be extruded from the polymerase. The DNA-RNA hybrid into the RNA polymerase is maintained and may contribute to the preservation of the stability of the elongation complex (Cheung and Cramer, 2011). Such configuration of the RNAPII-RNA-DNA complex favours the recruitment of an elongation complex called TFIIS (Dst1 in yeast) that stimulates the intrinsic endonucleolytic activity of the polymerase, resulting in the cleavage of the backtracked

RNA fragment. This step is required for re-aligning the 3'-OH of the nascent RNA with the catalytic center of the polymerase, thus allowing the resumption of elongation (Izban and Luse, 1993; Reines, 1992).

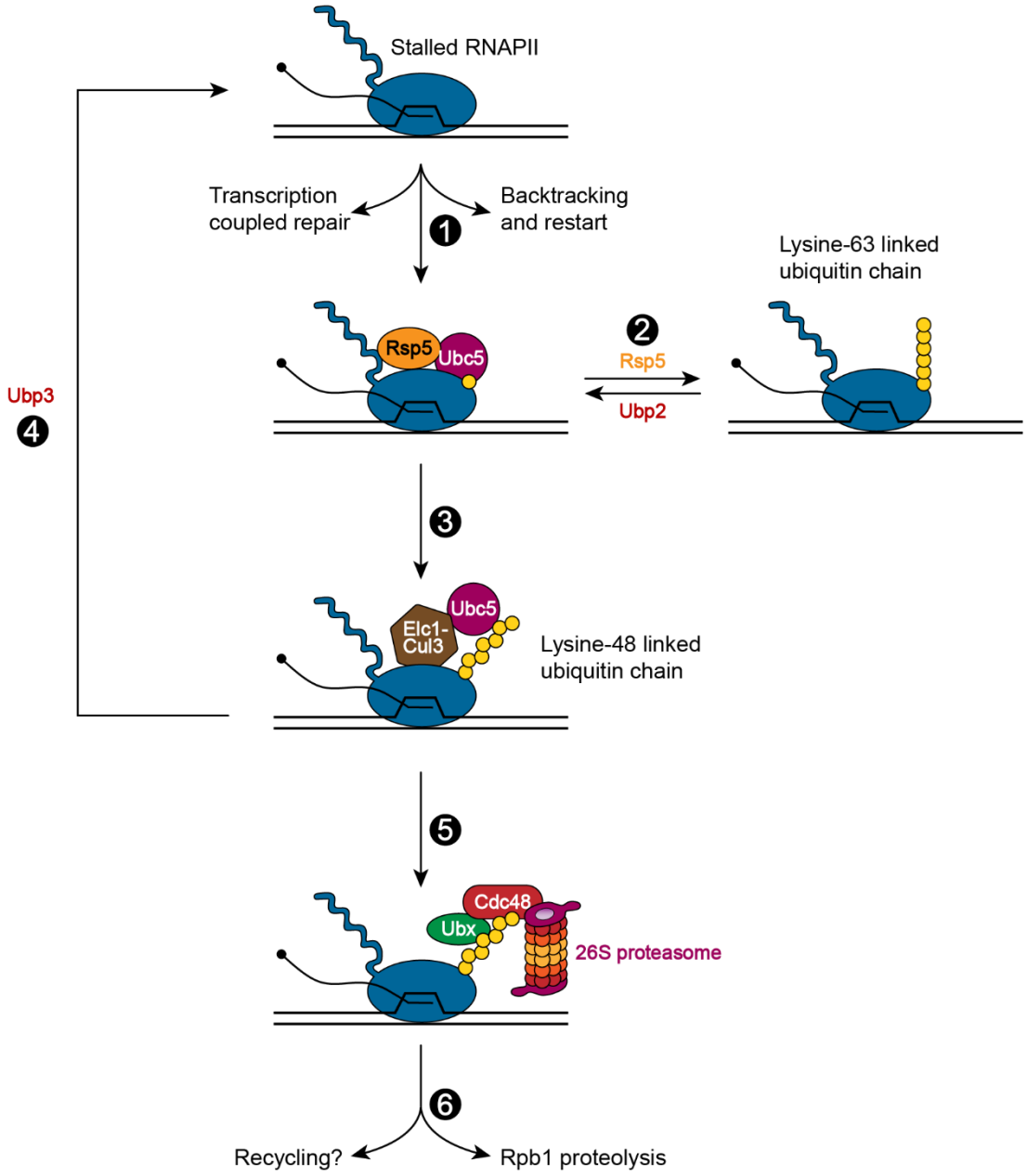


Figure 6. Mechanism of RNAPII ubiquitylation and degradation during transcriptional arrest. Paused polymerases can be removed *via* an ubiquitylation mechanism leading to the degradation of Rpb1. First, the Rsp5 and Ubc5 trigger the mono-ubiquitylation of Rpb1 (1). The mono-ubiquitin chain can be extended on lysine-63 (and reversed by the ubiquitin protease Ubp2) which is however not associated with RNAPII degradation (2). The Elc1-Cul3 complex together with Ubc5 recognize the mono-ubiquitin form of Rpb1 and promote the formation of the lysine-48 poly-ubiquitin chain (3) that can be removed by Ubp3 to prevent undesirable proteolysis (4). The lysine-48 poly-ubiquitin chain serves as a signal for the recruitment of the ATPase Cdc48-Ubx and the 26 proteasome (5) which promotes the dismantling of the polymerase and the degradation of Rpb1 (6). Adapted from (Wilson et al., 2013).

The past decades have been accompanied by the emergence of techniques allowing the genome-wide cartography of the polymerase at nucleotide resolution. In 2011, Churchman and Weissman described a technique called NET-seq (Native Elongating Transcript sequencing), a method based on the purification of elongated polymerases and the sequencing of the 3' ends of the nascent RNAs (Churchman and Weissman, 2011). Analyses of RNAPII distribution in *S. cerevisiae* have revealed the accumulation of polymerases upstream of the first four nucleosomes of the array. More interestingly, in cells deleted for *DST1*, most of the observed peaks are increased and slightly shifted downstream, which indicate the occurrence of backtracking in wild type cells. Finally, using a dominant-negative TFIIS mutant, Sheridan and colleagues confirmed the role of this elongation factor in favouring rapid and efficient elongation in human cells (Sheridan et al., 2019).

When the backtracked state is not resolved and becomes more prolonged, RNAPII enters transcriptional arrest. In this particular situation the paused polymerase needs to be removed from the DNA, possibly to prevent additional accumulation of elongation complexes upstream. To do so, ubiquitin ligases are recruited at the pausing site to ubiquitylate RNAPII, which will favour its degradation *via* the proteasome (Ribar et al., 2006, 2007; Somesh et al., 2005, 2007). Among the different actors, the Rsp5 E3 ubiquitin ligase has been particularly studied, especially in the context of DNA damages (Beaudenon et al., 1999; Wilson et al., 2013). The mechanism through which RNAPIIs undergo degradation is presented in Figure 6.

In summary: One of the main phenomena associated with transcription elongation is the reversible and dynamic phosphorylation of the carboxy-terminal domain of the polymerase. PTMs of Rpb1 play a role in every step of the elongation process from the promoter clearance to the processing of pre-mRNA molecules. It also contributes to the progression of the polymerase along the gene through the recruitment of elongation factors.

3. Transcription Termination

Once the synthesis of the pre-mRNA is complete, the elongation complex must be disassembled in order to release the transcript and recycle the RNAPII holo-enzyme for new rounds of transcription. In *S. cerevisiae*, two main transcription termination mechanisms have been described: The Cleavage and Polyadenylation Factor-Cleavage Factor I (CPF-CF) and the Nrd1-Nab3-Sen1 (NNS) pathway. Accurate termination by these two pathways requires the recognition of specific sequences on the nascent RNA and the interaction with Rpb1. The CPF-CF and NNS pathways are mainly dedicated to the termination of mRNAs and ncRNAs respectively and are associated with specific fates of the neo-synthetized transcript.

In this section, I will describe the mechanisms of these termination pathways and explain how they determine the fate of the resulting RNA molecule. Other, minor pathways of termination are also present in *S. cerevisiae*, which I will also discuss below. Finally, I will present a quick overview of termination pathways employed by different RNA polymerases or existing in other organisms.

3.1 The CPF-CF termination pathway

The CPF-CF pathway is the first termination mechanism described in eukaryotes. It employs generally conserved multiprotein sub-complexes that act collectively to promote termination of mRNA-coding genes as well as some non-coding RNAs. The action of the CPF-CF can be divided into three successive steps: i) the recruitment of the complex via its interaction with the polymerase and the recognition of *cis*-regulatory elements on the neo-synthetized transcript, ii) the cleavage of the nascent RNA and iii) the dismantling of the elongation complexes from the DNA.

3.1.1 Organisation and function of the CPF-CF complex

The CPF-CF is composed of about 20 proteins organised into four main modules: CFI and CFII that constitute the CPF complex and carry the catalytic activity and CFIB and IA that are important for the recognition of the RNA (see I.3.1.2). The role of the different factors is presented in table 1 (adapted from (Kuehner et al., 2011; Lidschreiber et al., 2018; Mandel et al., 2008)). Many proteins of the complex are able to interact with specific sequences on the nascent RNA (Rna15, Hrp1, Yhh1/Cfp1, Ydh1/Cfp2 and Yth1, see I.3.1.2). Another key factor of the CPF-CF is Pcf11 (CFIA). This protein contains a CTD-interacting domain (CID) and has

been shown to specifically contact the Ser2 phosphorylated form of Rpb1, thus favouring its recruitment at the 3' region of genes (Barilla et al., 2001; Kim et al., 2004a; Lunde et al., 2010; Meinhart and Cramer, 2004; Sadowski et al., 2003). Once the polyA sequence is transcribed and the complex assembled on the RNA, the endoribonuclease Ysh1 (CPF) catalyses the cleavage of the RNA. A stretch of adenosines is then added to the newly formed 3' end by the polyA polymerase Pap1 (PAP in human) (Figure 7). The activity of Pap1 is regulated by the Nab2 protein that interacts with the CPF complex. Nab2 contains a zinc finger domain important for its binding to the polyA tail. The presence of this factor limits the number of A residues added by Pap1 and is also important to protect the newly synthetized RNA molecules and promote their rapid and efficient export (Green et al., 2002; Hector et al., 2002; Tudek et al., 2018; Viphakone et al., 2008).

Table 1. Components of the yeast CPF-CF complex.

Factor	Protein	Function
CFIA	Rna14	*Scaffolding protein*
CFIA	Rna15	*RNA recognition*
CFIA	Pcf11	*Interaction with Ser2-P CTD, scaffolding protein*
CFIA	Clp1	*Scaffolding, interaction with CPF and Pcf11*
CFIB	Hrp1	*RNA recognition*
CPF (PFI)	Pfs2	*Interaction with Fip1 (PFI), Rna14 (CFI) and Ysh1 (CFII)*
CPF (PFI)	Fip1	*Pap1 recruitment and interaction with Yth1 (Scaffolding)*
CPF (PFI)	Yth1	*Interaction with Fip1 and Ysh1, RNA recognition*
CPF (CFII) / APT	Pta1	*Interaction with Ysh1, Ydh1, Pti1, Syc1 and Ssu72*
CPF (CFII)	Ysh1/Brr5	*Endoribonuclease, cleavage of the polyA site*
CPF (CFII)	Yhh1/Cft1	*RNA and CTD-binding*
CPF (CFII)	Ydh1/Cft2	*RNA recognition and Scaffolding (Ysh1, Yhh1/Cft1, Pta1, Pfs2, Ssu72, Pcf11)*
CPF	Mpe1	*RNA interaction, scaffolding, mediates ubiquitylation*
CPF	Pap1	*Synthetizes the polyA tail after pre-mRNA cleavage*
CPF / APT	Ssu72	*Phosphatase (Ser5 and 7)*
CPF / APT	Glc7	*Phosphatase (Tyr1)*
CPF / APT	Pti1	*Scaffolding*
CPF / APT	Swd2	*Scaffolding, also member of the COMPASS*
CPF / APT	Ref2	*RNA-binding, regulation of Glc7,*
APT	Syc1	*Only belongs to the APT complex*

Purification of the core-CPF has revealed the presence of 6 additional subunits present in a substoichiometric manner: Glc7, Ssu72, Ref2, Pti1, Swd2 and Syc1. These proteins form a subcomplex called APT (Associated with PTa1) that promotes efficient transcription termination (Nedea et al., 2003) of Small-NucleOlar RNAs (snoRNA). Among these proteins,

two (Ssu72 and Glc7) carry a phosphatase activity involved in the modification of the CTD (see I.2.2). In 2018, the Passmore laboratory reinvestigated the structural and functional connection between the APT and the CPF complexes (Lidschreiber et al., 2018). By performing purification of TAP-tagged (Tandem-affinity purification) factors followed by LC-MS/MS (Liquid chromatography - mass spectrometry), the authors discovered that the APT could form a distinct complex from the CPF. While Glc7, Ssu72, Ref2, Pti1 and Swd2 can be found in both the CPF and APT, the Syc1 protein is present only in the APT complex. Pta1, the protein that mediates the interaction between Ssu72 and Pti1 (i.e. the APT) and the CPF complex can also be present independently within the APT. In terms of function, cells defective for the APT complex (*syc1*Δ) show a decrease in the transcription level at sn/snoRNA but not at mRNA-coding genes. Consistent with this observation, Syc1/APT is more abundant at sn/snoRNA-coding genes and binds more efficiently than CPF-CF to sn/snoRNA. Together with previously published data, these results unveil a specific and independent role of the APT in the processing of some ncRNAs in *S. cerevisiae* (Dheur et al., 2003; Lidschreiber et al., 2018). Because of its role in processing RNAs that do not possess canonical polyA tail, Syc1/APT has been proposed to be connected to another termination pathway with a similar role: the NNS pathway (discussed below).

3.1.2 The core cleavage and polyadenylation signal sequence

In budding yeast, the polyA site can contain five distinct RNA sequences: an AU-rich efficiency element (EE), an A-rich positioning element (PE), an upstream uridine-rich element (UUE), the cleavage site and a downstream uridine-rich element (DUE) (Mischo and Proudfoot, 2013) (Figure 8).

The EE element consists of the hexanucleotide TAYRTA (Y = C/T and R = A/G) and is located at a variable distance from the cleavage site (Mischo and Proudfoot, 2013). This sequence is recognized by the Hrp1 component of the CFIB sub-complex. Neither the binding sequence nor Hrp1 are strictly required for the cleavage of the nascent transcript. Instead, they play a role in the efficiency of the 3’ end processing and the selection of the position of the polyA cleavage site (Dichtl and Keller, 2001; Guo et al., 1995; Minvielle-Sebastia et al., 1998). The TATATA sequence is important for efficient 3’ end formation and is present in more than half of the 3’ UnTRanslated regions (UTR) of genes in yeast (Guo et al., 1995).

The PE element is located 10 to 30 nt upstream of the cleavage site. This A-rich sequence is conserved in eukaryotes and mutation of the consensus motif is associated with human diseases (Mandel et al., 2008). As for the EE sequence, the PE module is dispensable for

polyA site recognition (Dichtl and Keller, 2001). Once the PE is transcribed, it is recognized by Rna15, a protein that belongs to the CFIA complex. Interestingly however, Rna15 alone does not show any RNA sequence specificity *in vitro*. The presence of Rna14 (CFIA) and Hrp1 (CFIB) instead promotes the specific binding of Rna15 to the A-rich element. Consistent with its crucial role in mRNA recognition, mutation in the RNA recognition motif (RRM) of Rna15 is lethal (Gross and Moore, 2001).

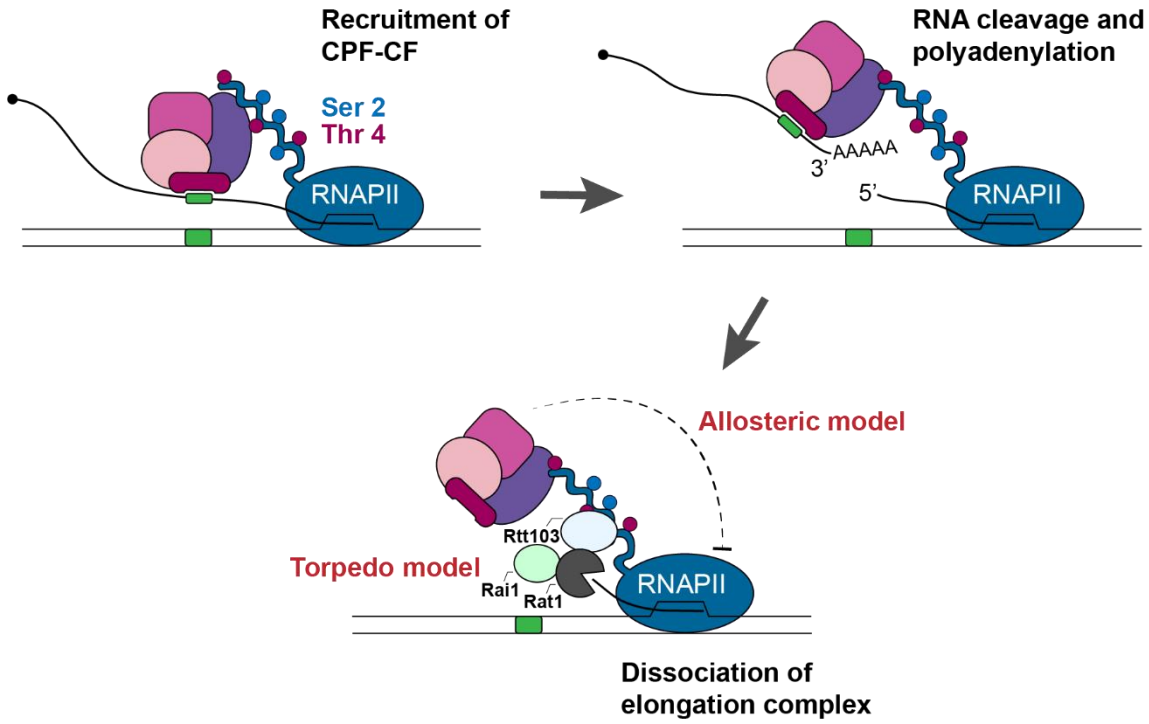


Figure 7. Mechanism of transcription termination by the CPF-CF pathway. The recruitment of the CPF-CF at the 3' end of mRNA-coding genes depends on its interaction with the CTD of the polymerase and the presence of specific sequences on the nascent transcript (green box). Endonucleolytic cleavage of the RNA by Ysh1 leads to the formation of an uncapped 5' end that serves as an entry point for the 5'-3' exonuclease Rat1. The dismantling of the elongation complex occurs *via* a torpedo mechanism (Rat1 dependent), an allosteric mechanism (conformational changes of the complex) or a combination of both. Adapted from (Porrua and Libri, 2015)

In *S. cerevisiae*, the cleavage site is surrounded by two U-rich sequences that act in concert to enhance and position the cleavage of the polyA site (Dichtl and Keller, 2001; Graber et al., 1999). Components of the CPF complex (Yhh1/Cft1, Ydh1/Cft2, Mpe1 and Yth1) bind to the U-rich element and the cleavage site to promote the association of the mature CPF complex (Barabino et al., 2000; Dichtl and Keller, 2001; Dichtl et al., 2002; Kyburz et al., 2003; Lee and Moore, 2014). The cleavage of the nascent RNA requires the lose sequence $Y(A)_n$ (Y = C/T) and occurs at the 3' end of an adenosine (Mandel et al., 2008).

3.1.3 Disassembly of the polymerase: The torpedo vs allosteric model

In budding yeast, the elongation complex is released from the template about ~200 bp downstream of the polyA site (Baejen et al., 2017; Schaughency et al., 2014). An important and still debated question in the field of transcription termination is to determine whether the cleavage of the nascent transcript is required for the disassembly of the polymerase. In this respect, two main and not mutually exclusive models have been proposed to explain the dismantling of the elongation complex: the allosteric model and the torpedo model, both supported by independent findings (Figure 7).

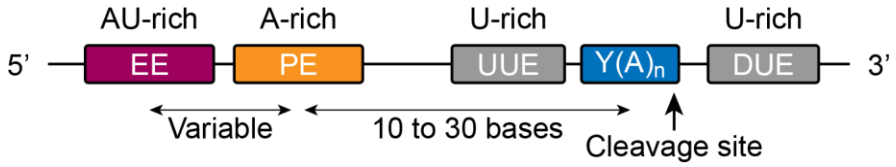


Figure 8. The CPF-CF transcription termination signal in yeast. 3' UnTranslated Regions (UTRs) encompass the polyadenylation signal composed of up to five distinct modules involved in the recognition of the CPF-CF and that signal the position of the cleavage site. EE, PE, UUE and DUE correspond to the AU-rich efficiency element, A-rich positioning element, upstream uridine-rich element and the downstream uridine-rich element respectively. Y = pyrimidine (C or T). Adapted from (Mandel et al., 2008; Mischo and Proudfoot, 2013)

The allosteric model proposes that the cleavage of the nascent transcript is not a strict requirement for RNAPII dismantling, although the polyA site itself is important (Orozco et al., 2002). Many different studies and approaches support this notion. For instance, visualization of transcription by electron microscopy in drosophila suggests that RNA cleavage occurs in a majority of cases post-transcriptionally and therefore does not participate to termination (Osheim et al., 2002). In addition, Pcf11 has been proposed to dissociate the elongation complex independently of cleavage *in vitro* in both *S. cerevisiae* and *D. melanogaster* by interacting with the nascent RNA and the CTD (Zhang and Gilmour, 2006; Zhang et al., 2005). More recently, Zhang and colleagues (Zhang et al., 2015) showed that the cleavage of the RNA is not required for termination in a mammalian *in vitro* reconstituted system, proposing that a conformational change of the elongating RNAPII holo-enzyme induces its release from the DNA template.

The co-transcriptional cleavage of the nascent RNA leads to the formation of an uncapped 5' end attached to the elongation complex. The torpedo model posits that the 5'-P end would serve as an entry point for exonucleases that would progressively degrade the RNA to finally displace the polymerase from the DNA. The first evidence supporting this model derives from studies published in 2004 by the Buratowski and Proudfoot groups (Kim et al., 2004b; West et

al., 2004). In the yeast model, the authors described the role of Rat1-Rtt103 complex in eliciting transcription termination (Kim et al., 2004b). Rtt103 is a scaffold protein able to interact with the Ser2-P and Thr4-P of the CTD at the 3' end of genes (Harlen et al., 2016; Jasnovidova et al., 2017a, 2017b; Kim et al., 2004b; Lunde et al., 2010). Rat1 is a 5' to 3' exonuclease originally described for its role in 5.8 S and snoRNA processing. The non-essential protein Rai1 interacts with Rat1 and enhances its activity. ChIP experiments reveal that Rat1, Rai1 and Rtt103 localized together at the 3' end of protein-coding genes *in vivo*. Furthermore, deletion of *RAI1* or expression of a thermosensitive mutant for *RAT1* are both associated with accumulation of RNAPII downstream or the normal region of termination (Kim et al., 2004b). Similar results were also obtained in a parallel study on the human homologue of Rat1, Xrn2 (West et al., 2004). In the past years, the effect of both Rat1 and Xrn2 on termination was validated genome-wide using ChIP-Seq, sequencing of newly synthesized transcripts or NET-seq experiments (Baejen et al., 2017; Eaton et al., 2018; Fong et al., 2015). In addition, mutations causing a decrease or an increase in the elongation rate are respectively linked with early or late termination events (Fong et al., 2015). This effect is in agreement with the idea that termination is a dynamic process occurring when Rat1/Xrn2 degrades the uncapped molecule and catches up with the polymerase. Despite all the *in vivo* evidence however, whether Rat1 can elicit termination *in vitro* remains a controversial question (Dengl and Cramer, 2009; Park et al., 2015; Pearson and Moore, 2013). Lastly, a combined model arguing that *in vivo,* the torpedo and allosteric model may probably act in concert to promote the disassembly of the elongation complex has also been proposed (Luo et al., 2006).

3.2 The NNS termination pathway

The NNS termination pathway is the second canonical termination mechanism known is *S. cerevisiae*. It operates on termination of ncRNAs such as snoRNAs (Steinmetz et al., 2001), and various additional non-functional and unstable RNAs (describe later) (Arigo et al., 2006a; Thiebaut et al., 2006). It relies on the action of three essential proteins, Nrd1, Nab3 and Sen1, that are all required for efficient termination of ncRNAs. In *S. pombe*, homologues of these proteins are found but only Seb1 (the Nrd1 homologue) appears to be involved in termination (Larochelle et al., 2018) although only for the production of mRNAs. The human homologue of Sen1 (senataxin) has also been proposed to elicit termination in cooperation with Xrn2 (Rat1 in yeast) but the evidence remains controversial (Proudfoot, Natoli, personal communication).

In addition to be essentially dedicated to the termination of distinct categories of RNAs, the NNS also differs from the CPF-CF for the intrinsic mechanism that leads to termination and the processing of the resulting RNA. Unlike CPF-CF, termination by the NNS is not linked to

the endonucleolytic cleavage of the nascent RNA. Instead, the release of the transcript and the dismantling of the elongation complex rely on the action of a helicase of the complex (see below). Another important difference with the CPF-CF pathway is that termination by the NNS is mechanistically associated with the action of nuclear degradation pathways, which function in the processing or the complete degradation of the released transcript (Porrua and Libri, 2015).

3.2.1 The NNS components

Nrd1

Nrd1 (nuclear pre-mRNA down-regulation) is an RNA-binding protein that is also able to interact with many different factors. In this regard, it plays an essential role in the recruitment of the NNS complex at target genes and contributes to coordinating transcription termination with RNA degradation by the nuclear exosome.

Nrd1 is a 68KDa protein that contains a central RRM motif necessary for the recognition of a GUAA/G motif present on the nascent RNA (Carroll et al., 2004; Porrua et al., 2012; Schaughency et al., 2014; Steinmetz and Brow, 1998; Wlotzka et al., 2011). It also contacts the RNAPII CTD through its N-terminal CID domain and interacts directly with Nab3 to form a heterodimer (Conrad et al., 2000). The CID domain of Nrd1 is required for accurate termination (Arigo et al., 2006a; Tudek et al., 2014) and is also important for the interaction with Tfr4, a component of the TRAMP (Trf4-Air2-Mtr4) complex to mediate RNA degradation (see later) (Tudek et al., 2014).

Nab3

The 90KDa Nab3 (nuclear polyadenylated RNA-binding) factor recognizes the RNA sequence UCUUG through its RRM motif (Carroll et al., 2004; Hobor et al., 2011; Porrua et al., 2012; Schaughency et al., 2014; Wlotzka et al., 2011). It encompasses a Nrd1-interaction domain and complexes with Sen1 *in vivo* (Chinchilla et al., 2012; Conrad et al., 2000). Nab3 is capable of forming homotetramers *in vitro*. The multimerization of Nab3 is carried out by its low complexity carboxy-terminal domain that is distinct from the RRM and Nrd1-interaction. This domain is important for transcription termination and has been proposed to favour the formation of multiple Nrd1-Nab3 heterodimer *in vivo* (Loya et al., 2012, 2013).

Sen1

Among the three factors that constitute the NNS, Sen1 (splicing endonuclease) is the only one to bear enzymatic activities. It is also the biggest (252KDa) and less abundant protein of the

complex (30 to 100 times less than Nab3 and Nrd1) (Ghaemmaghami et al., 2003), suggesting that the three proteins are not always present stoichiometrically in the complex.

Sen1 is a helicase that belongs to the superfamily 1 (SF1) helicases and resembles to the Upf1 helicase, a protein involved in the Nonsense-Mediated mRNA Decay (NMD) pathway. Sen1 binds and can translocate on both RNA and single stranded DNA in a 5' to 3' manner (Han et al., 2017; Martin-Tumasz and Brow, 2015). Sen1 alone is sufficient to trigger termination *in vitro*. In this system, is has been shown that the dissociation of the elongation complex requires the presence of the nascent RNA and hydrolyses of ATP but not the CTD of the polymerase (Porrua and Libri, 2013). Later studies have revealed that the sole helicase domain is sufficient to ensure robust termination *in vitro*, thus placing Sen1 as the main terminator factor of the NNS pathway (Han et al., 2017). *In vivo,* however, the N-terminal domain of Sen1 is also required for termination, possibly because it is required for the interaction with the CTD of Rpb1 (Ursic et al., 2004).

3.2.2 NNS-dependent termination mechanism

The current model posits that the recruitment of the NNS complex depends both on the binding of the two RNA-binding proteins on short RNA sequences and the interaction of Nrd1 with the CTD of the polymerase. NNS terminators often contain clusters of binding sites that are believed to be bound by Nrd1-Nab3 heterodimers. Consistent with the essential role of the RNA recognition, mutation of these sites drastically impairs transcription termination *in vivo.* In addition to the previously mentioned motifs, *in vivo* selection experiments have highlighted the presence of AU-rich sequences located close to the Nab3 site that are important for efficient termination (Porrua et al., 2012).

Unlike Pcf11 or Rtt103 (see I.3.1), the interaction between Nrd1 CID and the CTD is not mediated by the Ser2-P mark. Instead, structural and biochemical analyses of the NNS complex have revealed an important role of Ser-5 phosphorylated CTD in the recruitment of Nrd1 (Gudipati et al., 2008; Heo et al., 2013; Kubicek et al., 2012; Mayer et al., 2012; Vasiljeva et al., 2008). Ser5-P is characteristic of the early stages of elongation (See I.2.2). This finding is consistent with a model according to which NNS-dependent termination occurs early in the transcription cycle (i.e. close to the TSS), which explains why NNS-targeted genes are usually shorter (less than ~600 bp) than mRNA-coding genes. Consistently, it has been shown that displacing NNS terminators far from the TSS (~1000 bp) decreases the Nrd1-dependency of termination (Gudipati et al., 2008). Interestingly however, the extent to which the Ser5-P contributes to termination remains unclear. It has been shown that deletion of the CID of Nrd1

decreases the efficiency of transcription termination at many loci (Tudek et al., 2014), which also depends on the integrity of the Kin28 kinase (Gudipati et al., 2008). Yet, a more recent study did not observe such readthrough events when using Rbp1 variants that are mutated at the essential Ser5 residues of the CTD (Collin et al., 2019).

The crystal structure of Nrd1 shows an interaction between an aspartate present in the CID and the Tyr1 residue of the polymerase (Kubicek et al., 2012). Tyr1 has a significant effect on termination as shown by the mutation to phenylalanine of this residues in all of the CTD repeats (Collin et al., 2019). However, this residue has been proposed to play a distinct role in the termination than recruiting Nrd1 (Collin et al., 2019), notably by inducing RNAPII pausing around the site of termination.

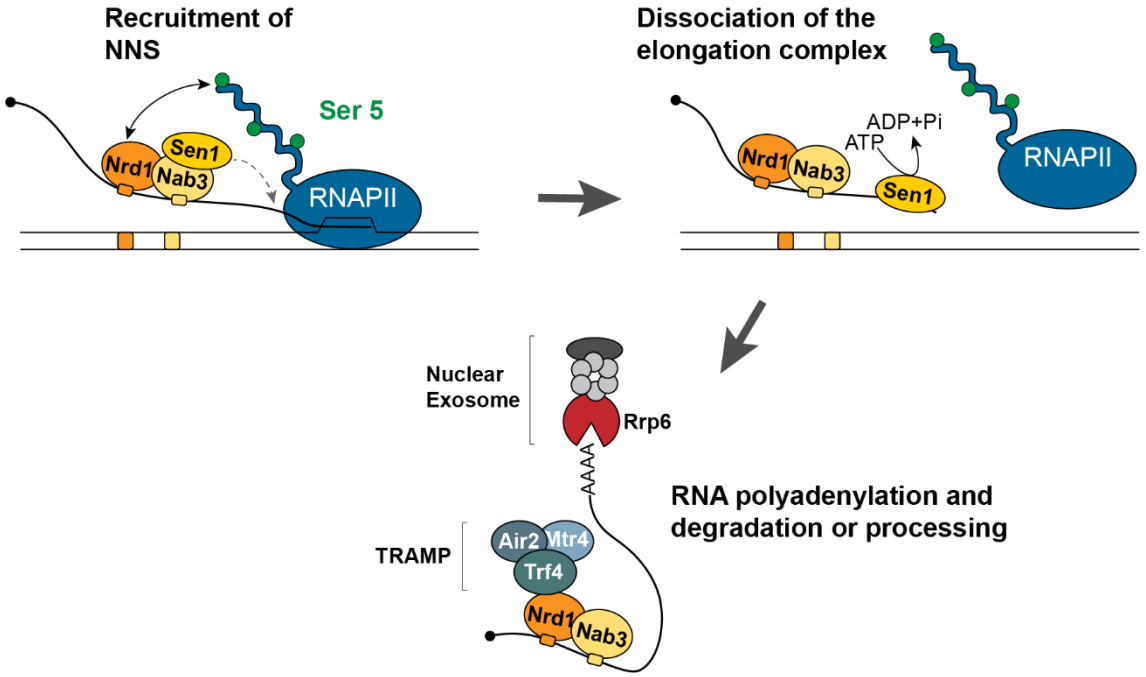


Figure 9. Overview of the NNS transcription termination pathway. Nrd1 and Nab3 are recruited to non-coding transcripts by the recognition of specific motifs present on the nascent RNA (yellow and orange boxes). The interaction between Nrd1 and the Ser5-phosphorylated form of the CTD contributes to the specificity of the NNS complex for termination of short genes. By hydrolysing ATP, the Sen1 helicase is able to translocate along the RNA and disassemble the polymerase from the DNA template. Subsequently, the interactions between Nrd1 and Nab3, the TRAMP complex and the nuclear exosome promote the polyadenylation of the RNA (TRAMP) and its degradation or processing (exosome). Adapted from Porrua and Libri, 2015.

The assembly of the Nrd1-Nab3 heterodimer on the nascent RNA is believed to facilitate the recruitment of the Sen1 helicase. The present model posits that once Sen1 is recruited, it translocates along the nascent RNA to reach the polymerase and dismantles the elongation complex (Porrua and Libri, 2015) (Figure 9). It has been shown that the NNS termination defects can be partially or completely suppressed by slowing down the elongation rate of the

polymerase across the gene (Collin et al., 2019; Hazelbaker et al., 2013; Lenstra et al., 2013). This has led to the suggestion that robust termination of the NNS pathway relies on a kinetic competition between the translocation of Sen1 on the nascent RNA and RNAPII elongation (Hazelbaker et al., 2013; Porrua and Libri, 2013).

3.2.3 RNA metabolism and NNS termination pathway

Following termination, NNS-dependent transcripts are targeted to the nuclear exosome that degrades some RNAs (Cryptic Unstable Transcripts (CUT) described later) and trims snoRNAs into the mature and functional molecule. An important actor that participates to the processing of the transcript is the TRAMP complex, a cofactor of the nuclear exosome that recruits and stimulates its activity (Jensen et al., 2013; Porrua and Libri, 2015) (Figure 9).

The nuclear exosome is composed of 11 subunits and is endowed with 3 ribonucleic activities. The enzymatic activities are organised around the core exosome arranged in 2 distinct layers: 6 factors forming a bottom ring-like structure and a top layer of 3 “cap” proteins with an RNA-binding capacity. Two 3’ to 5’ exonucleases, Dis3 and Rrp6, are located at each extremity of the core complex (Chlebowski et al., 2013; Makino et al., 2015). Dis3 also carries an endonucleolytic activity mediated by a different domain (Lebreton et al., 2008). Rrp6 is only present in the nuclear form of the exosome and functions partially redundantly with Dis3 (Gudipati et al., 2012). In the nucleus, the exosome machinery degrades CUTs (Wyers et al., 2005) and trims snoRNAs up to the size of the mature form, most likely because the snoRNP core ribonucleoprotein complex prevents further progression of the exonuclease. As expected for a role of Rrp6 in RNA turnover and processing, deletion of the gene leads to accumulation of polyadenylated sn/snoRNA precursors and normally unstable transcripts (Porrua and Libri, 2015). Another striking phenotype associated with the deletion of *RRP6* is the occurrence of transcription termination defects at NNS gene targets (Castelnuovo et al., 2013; Fox and Mosley, 2016; Vasiljeva and Buratowski, 2006). Therefore, it has been proposed that the exosome may directly impact termination by a mechanism that remains unclear. An ongoing study in the Libri laboratory, however, supports a model whereby the excess of RNAs that accumulate in the absence of Rrp6 leads to the titration of the NNS complex, therefore preventing its efficient recruitment at nascent RNAs (Villa et al., in preparation).

The TRAMP complex is composed of three factors: the polyA polymerase Trf4, the RNA-binding factor Air2 and the RNA helicase Mtr4. Together, they are responsible for the polyadenylation of a plethora of non-coding transcript including sn/snoRNAs, some rRNAs, hypomodified $tRNA^{Met}$ and CUTs, which are a major product of pervasive transcription (Kadaba

et al., 2004; LaCava et al., 2005; Vanácová et al., 2005; Wyers et al., 2005). Structurally, the Trf4 polyA polymerase shares similarity with the CPF-CF Pap1 polymerase. Deletion of *TRF4* results in an important accumulation of NNS-dependent transcripts *in vivo*. *In vitro*, Trf4 alone is not sufficient to catalyse the addition of a polyA tail and requires the Air2 factor. Consistently, cells defective for *AIR2* show a similar defect in RNA degradation as compared to a *trf4*Δ strain (Wyers et al., 2005). The Mtr4 protein limits the number of adenosines added on the transcript (Jia et al., 2011). PolyA tails of steady states RNA molecules that are terminated by the NNS pathway are usually shorter (~5 nt) than the CPF-CF terminated transcripts. This difference probably explains why these RNAs are not stable but rather directed to the nuclear exosome (Wlotzka et al., 2011).

The accurate coupling between transcription and RNA processing is mediated by successive interactions involving the polymerase, the NNS complex, the TRAMP and the nuclear exosome machinery. The Trf4 component of the TRAMP interacts *via* its CTD mimic domain (also called NIM, for Nrd1-interacting motif) with the CID domain of Nrd1. Interestingly, the Nrd1-Trf4 and Nrd1-Rpb1 interactions both rely on the same CID domain of Nrd1. Hence, these interactions are mutually exclusive which may participate to the sequential coordination of transcription and degradation (Tudek et al., 2014). The recruitment of the exosome at NNS targets is promoted by the interaction between Rrp6 and Trf4, Rrp6 and Nrd1 (Tudek et al., 2014) and Rrp6 and Nab3 (Fasken et al., 2015).

3.3 Alternative transcription termination pathways

Together, the CPF-CF and NNS pathways contribute to the termination and 3' end processing of most of the RNAPII-dependent transcripts expressed in *S. cerevisiae*. Yet, at least two other mechanisms have been shown to impact termination: a Rnt1-dependent and a roadblock-associated pathway.

3.3.1 The Rnt1-dependent termination

Rnt1 is the eukaryotic homologue of bacterial dsRNA-specific endonuclease Rnase III. This factor has been extensively studied for its role in the maturation of precursor RNA molecules such as pre-rRNA, sn/snoRNAs as well as intron-encoded snoRNAs (Ghazal et al., 2005). Rnt1 cleavage sites are mostly characterized by the presence of a stem-loop structure containing an AGNN tetraloop (Chanfreau et al., 2000) although other type of substrates have also been identified (Gagnon et al., 2015; Ghazal et al., 2005). It has been shown that Rnt1-

dependent cleavages can trigger transcription termination (Ghazal et al., 2009; Rondón et al., 2009).

The *NPL3* gene has been used as a model to study the role of Rnt1 in termination. The 3' region of *NPL3* is characterized by the presence of a weak polyA site (inefficient for transcription termination) followed by an AGNN motif predicted to form a tetraloop hairpin. Mutation of the Rnt1 motif or deletion of *RNT1* leads to readthrough transcription downstream of the Rnt1 cleavage site. Importantly, no termination defect was detected at the primary site (CPF-CF) when Rnt1-dependent termination was impaired demonstrating that Rnt1 acts independently of the CPF-CF pathway. Nonetheless, similarly to CPF-CF termination (3.1.1), Rnt1-dependent termination requires the exonuclease Rat1 (Ghazal et al., 2009; Rondón et al., 2009). The Rnt1 termination pathway also shares similarity with the NNS pathway since transcripts that are produced are stabilized in cells defective for either Rrp6 (nuclease exosome) or Trf4 (TRAMP complex) (Egecioglu et al., 2006; Rondón et al., 2009).

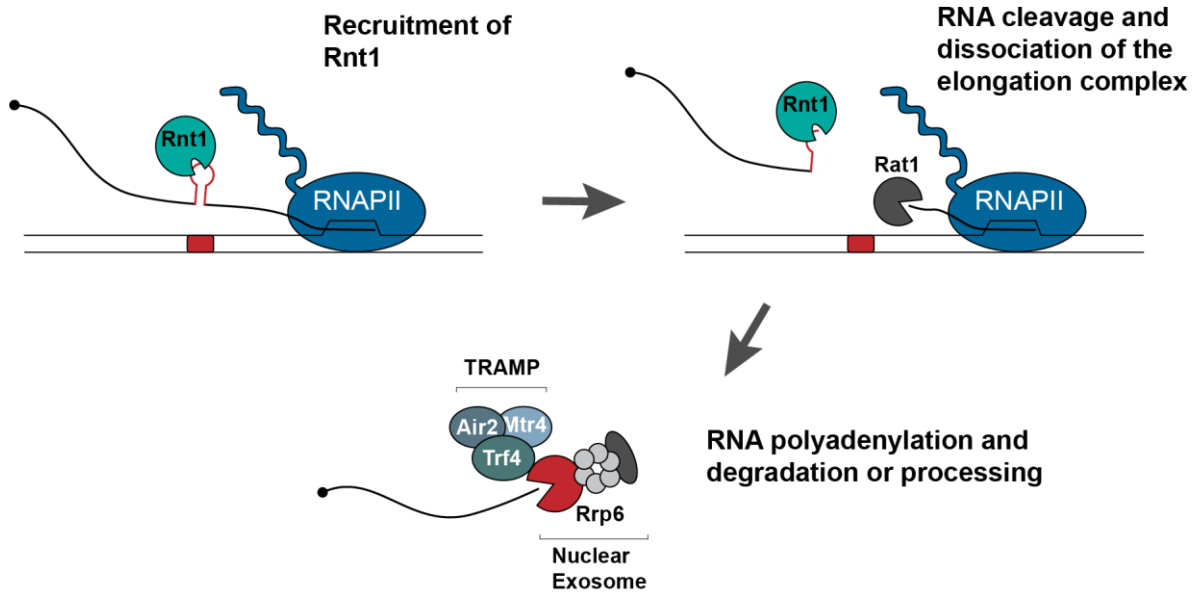


Figure 10. Rnt1-dependent transcription termination pathway. Recruitment of the endonuclease Rnt1 requires the presence of specific stem-loop structures on the nascent transcript (in red). The 5' end of the downstream portion of the RNA is targeted by Rat1 to trigger termination and the released transcript is degraded by the nuclear exosome.

The current model therefore posits that Rnt1 is recruited at specific sites through the recognition of a stem-loops structure, cleaves the RNA and generates a free 5'-OH extremity that serves as an entry point for Rat1 (Ghazal et al., 2009; Rondón et al., 2009). The released transcript is subsequently subjected to degradation (or maturation) by the TRAMP / Nuclear exosome (Figure 10).

Finally, in addition to *NPL3*, the Rnt1 pathway is involved in termination of about 30 other genes in yeast, including snRNA and mRNA-coding genes (Ghazal et al., 2009). Importantly, it can serve as the main (snRNA) or back-up (mRNA) termination pathway. In the latter cases, Rnt1 has been proposed to act as a fail-safe mechanism to prevent the progression of polymerases that fail to terminate at primary sites.

3.3.2 The roadblock termination pathway

The roadblock termination mechanism relies on the ability of the Reb1 DNA-binding factor to physically impede the progression of transcribing polymerases. Unlike the previously described pathways, roadblock events do not involve the recognition of the nascent transcript or the transcription of specific signals present on DNA (Figure 11). The first extensive characterisation of this mechanism has been made by our laboratory, in a study published in 2014 (Colin et al., 2014).

In order to discover new termination mechanisms, Colin and colleagues performed an *in vivo* selection from a pool of naïve DNA sequences, and assessed their ability to terminate transcription. From this experiment, the authors found that a short motif of ~ 10 bp was particularly enriched, and corresponds to the binding site of the essential Reb1 DNA-binding factor. Reb1 is a general regulatory factor known to regulate hundreds of genes in yeast by binding at specific sequences present within promoter regions (discussed later). Using a reporter system, they confirmed the implication of the Reb1 motif and protein in termination. In addition, the expression of a truncated form of Reb1 lacking the activator domain is sufficient to induce termination, thus revealing that the transcription activation activity and the termination activity of Reb1 are independent.

Importantly, Reb1-dependent termination is characterized by a marked pausing peak of RNAPII upstream of Reb1-binding sites. This accumulation of polymerases is abolished upon depletion of Reb1 or mutation of the binding site and enhanced in the absence of TFIIS/Dst1 or in cells defective for the Rsp5 ubiquitin ligase (see I.2.4). A plausible model is that polymerases encountering the Reb1 factor are stalled and backtracked, leading to the recruitment of TFIIS. However, the persistence of Reb1-binding on DNA prevents RNAPII to restart transcription and triggers its ubiquitylation and most probably its degradation by the proteasome.

Akin to the Rnt1 mechanism, the roadblock pathway can account for the primary termination mechanism of a few non-coding transcripts or can function as a back-up mechanism to restrict

the disruptive effects of leaky CPF-CF termination. In both cases however, the RNAs derived from Reb1-dependent termination are unstable and degraded in the nucleus by the TRAMP and exosome pathway. This class of RNA has been called RUTs, for Reb1 unstable transcripts.

Prior study had already suggested a role for various DNA-binding proteins in inducing RNAPII roadblock termination. This includes for instance other general regulatory factors such as Rap1 (Yarrington et al., 2012) or Abf1 (Valerius et al., 2002), or the RNAPIII transcription factor TFIIIB (Korde et al., 2014). These studies were however restricted to a limited amount of natural case study and did not address the molecular mechanism behind such termination events.

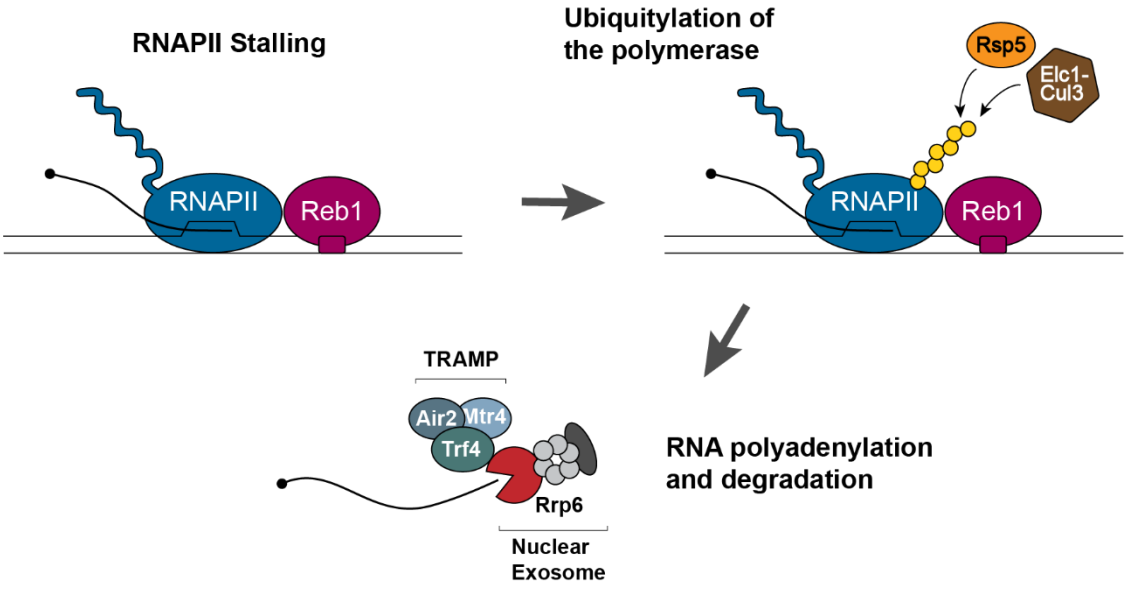


Figure 11. Reb1-mediated roadblock termination pathway. RNA polymerases encountering the Reb1 factor are stalled upstream of the Reb1-binding site. The paused polymerases are targeted to degradation *via* the Rsp5-dependent ubiquitylation pathway and the released RNA is degraded by the nuclear exosome.

My first thesis project aimed at evaluating the genome-wide occurrence and the functional role of roadblock termination in *S. cerevisiae*. To do so, I have analysed the distribution of the polymerase genome-wide using a modified version of the CRAC (UV crosslinking and analysis of cDNA) technique (Granneman et al., 2009). By sequencing nascent RNAs after purification of RNAPII elongation complexes, this method allows the detection of elongated complexes in a strand specific manner and with nucleotide resolution. This work is described in the first result section.

3.4 Diversity of transcription termination pathways

Mechanisms that govern transcription termination are diverse and generally specific of the polymerase that is considered. In this section, I will give a rapid overview of the mechanisms associated with termination of RNAPI and III in yeast, or present in different organisms such as bacteria or mammals.

3.4.1 Transcription termination of RNA polymerase I and III

RNA polymerase I

RNAPI is responsible for the transcription of the 35S ribosomal RNA whose post-transcriptional maturation leads to the formation of the 5.8S, 18S and 25S. The yeast genome contains 150 to 200 copies of rDNA organised in tandem repeat on chromosome XII. The termination region is encompassed within the InterGenic Sequence (IGS) containing specific sequences and recognition motifs for DNA-binding factors. In both yeast and higher eukaryotes, RNAPI-dependent termination occurs through a similar mechanism to the previously mentioned roadblock mechanism (Porrua et al., 2016). In *S. cerevisiae* the binding of Nsi1 at IGS induces an arrest of the polymerase and is sufficient to induce termination *in vitro* and in an artificial termination system *in vivo* (Merkl et al., 2014). Nsi1 and its paralogue Reb1 contain similar Myb-like DNA-binding domains and bind an identical site. For this reason, Reb1 was long thought to be the main terminator factor of RNAPI (Lang and Reeder, 1993, 1995; Lang et al., 1994; Reeder et al., 1999).

The termination region also contains an Rnt1 site upstream the Nsi1-binding site. During the termination process, the endonuclease Rnt1 is recruited at the site of transcription and cleaves the nascent transcript after recognition of the stem-loop structure. The presence of the newly formed RNA 5’ end attached to the paused polymerase has been proposed to be a substrate for the Rat1 exonuclease and the Sen1 helicase. These two factors may act in concert to dislodge the polymerase and elicit termination (El Hage et al., 2008; Kawauchi et al., 2008). Finally, in yeast, a T-rich motif (absent in higher eukaryotes) has also been proposed to favour RNAPI termination efficiency in cooperation with Nsi1 (Lang and Reeder, 1995; Lang et al., 1994; Reiter et al., 2012).

RNA polymerase III

RNAPIII transcribes relatively short and abundant RNA molecules (tRNAs, U6 snRNA, 5.S rRNA). The RNAPIII termination mechanism differs remarkably from the other mechanisms in

the sense that it does not involve any specific factors in addition to the DNA sequence (Porrua et al., 2016).

Termination of RNAPIII requires the presence of a short T-stretch at the 3' end of genes. The number of Ts varies in a species-specific manner. In budding yeast, the stretch contains 7 Ts on average but can also be longer (up to 10 nt long) at some genes. In *S. pombe* and *H. sapiens*, the motif is usually shorter and rarely exceed 8 and 5 nt respectively (Braglia et al., 2005). Despite the simplicity of the termination signal, the structural modifications leading to the release of the polymerase are still poorly understood. Three RNAPIII subunits, Rpc11, Rpc37 and Rpc53, have been proposed to be particularly important for termination. The current model posits that during transcription of the first four As of the template strand (T-stretch on the non-template strand), the polymerase enters a metastable pre-termination complex (PTC). This transition requires the three above-mentioned factors but is not sufficient to induce termination. Once the polymerase reaches the 5th A, the interaction between this nucleotide and the C-terminal region of Rpc7 may destabilise the complex and switch the PTC to a transcript release mode, which leads to termination (Arimbasseri and Maraia, 2015).

3.4.2 Transcription termination among the living world

Transcription termination in bacteria

Two main distinct termination pathways are known in bacteria. The first one is called intrinsic termination and is characterized by the strict requirement for a specific sequence. Intrinsic terminators are composed of GC-rich motif followed by a stretch of Us. Transcription of the GC-rich sequence leads to the formation of a hairpin structure on the nascent RNA. The presence of the hairpin and the formation of an unstable RNA:DNA hybrid at the U-rich site destabilises the elongation complex and induces termination. The precise function of the hairpin in promoting termination is still under debate.

The second known pathway, the Rho-dependent termination, depends on the action of the Rho helicase factor. Rho is a molecular motor structured into a homo-hexametric ring complex that acts in an ATP dependent manner. It binds the nascent RNA at clustered Rho utilization elements composed of C-rich G-poor motifs and translocates towards the polymerase for dismantling the elongation complex. Because of their analogous mechanism and despite the difference in structure, the Rho factor is often compared to the yeast Sen1 helicase of the NNS complex (Porrua et al., 2016; Ray-Soni et al., 2016).

RNAPII transcription termination in mammals

Akin to yeast, different mechanisms of RNAPII termination exist in mammals and are often linked with the 3' end formation and processing of distinct classes of transcript. As previously mentioned, the components and function of the CPF-CF are conserved in mammals (CPSF-CF) while the NNS pathway is instead mainly present in the budding yeast. In mammals, termination of short transcription units relies on a different set of factors that are mainly dependent on the essential 5' cap-binding complex (CBP). The CBP is a complex of two proteins (CBP20 and CBP80) involved in various steps of RNAs metabolism including RNA stability, transport and translation.

During transcription, CBP has been shown to interact with the ARS2 protein to promote 3' end termination and processing of short transcripts. As a proof of concept, knock-down of either factors leads to readthrough transcription at various RNAPII-dependent coding genes (Hallais et al., 2013). This pathway is mainly dedicated to the termination of snRNAs, as well as some classes of highly unstable transcripts that are relatively similar to CUTs (PROMoter-Proximal Transcripts, PROMPT) or Enhancer RNAs (eRNA)) (Iasillo et al., 2017). The CBP-ARS2 together with the zinc-finger ZC3H18 protein interact with the Nuclear EXosome Targeting (NEXT) complex which in turn favours the recruitment of the exosome and the degradation of these RNAs (Andersen et al., 2013; Preker et al., 2008). In this respect, the CBP-ARS2 shares a common feature with the NNS pathway as it also connects transcription termination with RNA decay. Finally, it has been proposed that two components of the CPSF-CF pathway (PCF11 and CLP1) may also play a role in the CBP-ARS2 pathway, perhaps by participating to the release of the elongation complex (Hallais et al., 2013).

In summary: Across the living world, a plethora of different mechanisms have evolved to ensure robust transcription termination of RNA polymerases. Many of them rely on the recognition of specific signals present on the nascent RNA and the combined action of endonucleases and exonucleases or the recruitment of a 5' to 3' helicase. Some other pathways on the contrary are characterized by the strict requirement of a DNA motif and may in this case be marked by a structural switch of the polymerase to favour its release from the template strand. Importantly, all these termination mechanisms share common functions: on one hand, they often couple transcription with the 3' end processing of the nascent transcript, which is important to determine its metabolic pathway (i.e. export, translation, degradation…); on the other hand, they are crucial to delimit the 3' border of transcription units and avoid extensive readthrough of elongation complexes. In this respect, some pathways have been selected to act as a fail-safe mechanism in order to counteract the leakage of RNAPs at canonical terminators.

II - Role of General Regulatory Factors and Chromatin Remodelers in Gene Regulation

General regulatory factors (GRFs) and chromatin remodelers (CRs) are two classes of factors whose function is intimately correlated with the formation and the maintenance of the nucleosomal structure at inter- and intragenic regions. Because of their impact on NDR formation, GRFs and CRs have been extensively studied in the context of transcription regulation. During my thesis, I have studied the role of GRFs in delimiting transcription units and investigated, in collaboration with the Shore laboratory, the function of CRs on transcription initiation.

This second chapter will present the yeast GRFs and CR complexes. A particular attention will be devoted to the effect of their depletion on NDR formation and gene expression.

1. General Regulatory Factors

GRFs are essential and abundant sequence-specific DNA-binding proteins involved in the regulation (activation or repression) of several classes of genes in *S. cerevisiae*. The three most documented GRFs are Rap1, Reb1 and Abf1. As many transcription factors, these master regulators are structurally organised into at least two distinct domains: a DNA-binding domain and an activator (or regulatory) domain. These domains are important for the function of the protein and are implicated in the interaction with various additional factors.

1.1 Rap1

Rap1, for Repressor Activator Protein 1, is one of the best characterized GRFs. It is a 827 amino acid protein composed of central DNA-Binding Domain (DBD) flanking by a C-terminal domain harbouring most of the regulatory function, and a N-terminal domain important for the maintenance of cell wall homeostasis (for a recent review see Azad and Tomar, 2016). Its DBD is similar to the human oncogene Myb (Myb-like domain) and binds to a G-rich DNA motif whose consensus is TGTAC/TGGGTG (Badis et al., 2008; Rhee and Pugh, 2011).

Rap1 is involved in the regulation of about 5% (~300) of *S. cerevisiae* genes, which are mainly highly expressed (Lieb et al., 2001). It is responsible for the activation of more than 90% of the ribosomal protein-coding genes (RPG: 129 out of the 138 present in yeast) (Knight et al., 2014). It also controls the expression of genes implicated in the glycolytic pathway (Brindle et al., 1990; Lieb et al., 2001; Mizuno et al., 2004) as well as some low-glucose dependent targets (Buck and Lieb, 2006) and genes related to the production of rRNAs (some RNAPI-associated subunits and transcription factors or rRNA processing factors) (Lieb et al., 2001). Lastly, Rap1 has also been described as a repressor of the HML and HMR mating-type loci during vegetative growth (Kimmerly et al., 1988; Kurtz and Shore, 1991).

At ribosomal protein genes, Rap1 acts in concert with additional factors. The FIS complex (Fhl1-Ifh1-Sfp1) has been shown to be particularly enriched at all Rap1 RPG-targets. In addition, about half of the Rap1-dependent RPGs are regulated by the Hmo1 DNA-binding factor (Knight et al., 2014; Reja et al., 2015; Wade et al., 2004). Rap1 is often positioned in NDRs at the most upstream position with respect to the TSS and is followed by the FIS and Hmo1 (Knight et al., 2014; Reja et al., 2015).

Downregulation of RPGs is known to occur as a result of various stress conditions including heat shock or inhibition of the TOR (Target Of Rapamycin) pathway (Cardenas et al., 1999; Powers and Walter, 1999). Interestingly, in this particular context, Ifh1, Sfp1 and Hmo1, have been shown to dissociate from the template DNA while Rap1 and Fhl1 remain strongly bound to their promoter sequence (Reja et al., 2015; Wade et al., 2004). This might suggest that the binding of Rap1 is not submitted to regulation but rather is ubiquitous, at least in the tested conditions.

The regulation of HMR and HML is also under the dependence of various factors. Sir1, 3, 4, the histone deacetylase Sir2, (Silent Information Regulators) and Abf1 have been all shown to participate to the transcriptional silencing of these two loci (Azad and Tomar, 2016; Kimmerly et al., 1988). In the case of the Sir proteins, their recruitment at mating-type loci is mediated by the direct interaction of Sir3 and 4 with the C-terminal domain of Rap1 (Liu and Lustig, 1996; Luo et al., 2002).

Rap1 has been shown to interact with components of the PIC as well as some chromatin remodeler complexes. These interactions have been proposed to underlie the gene activation capacity of Rap1. Notably, Rap1 interacts with the TFIID and TFIIA general transcription factors (Bendjennat and Weil, 2008; Garbett et al., 2007; Johnson and Weil, 2017; Papai et al., 2010; Tomar et al., 2008). *in vitro* pull-down experiments using recombinant versions of Rap1 or TFIID have revealed that their association is mainly mediated by the C-terminal domain of Rap1 and the Taf12, 4 and 5 subunits of TFIID. Nonetheless, a truncated version of Rap1 containing only the DBD is also able to interact with TFIID albeit to a lesser extent (Garbett et al., 2007; Tomar et al., 2008). These associations could be confirmed by cryo-electron microscopy experiments (Papai et al., 2010) and are also supported by *in vivo* data showing that Rap1 is sufficient to recruit Tafs factors to promoters in a TBP and RNAPII independent manner (Mencía et al., 2002). By using a variant of Rap1 with altered DNA-binding specificity, Johnson and Weil recently showed that the vast majority of the activation function of Rap1 is carried by a 41 amino acid long region located at the C-terminal part of the protein, whose mutation partially impairs the interaction with Taf5 (Johnson and Weil, 2017). Affinity chromatography of yeast extract using Rap1-bound columns show that, as for TFIID, the SWI/SNF remodeler complex associates with the C-terminal domain of Rap1 and to a lesser extent with the DBD (Tomar et al., 2008). Collectively, these data indicate that Rap1 has the ability to interact with factors involved in gene activation, suggesting a possible mode of action as a transcriptional activator.

Another well characterized and extensively studied function of Rap1 is related to telomere maintenance. Telomeres are DNA regions located at the extremity of linear eukaryotic chromosomes. *S. cerevisiae* telomeres consist in tandem repeats of the DNA sequence TG_{1-3} (T followed by 1 to 3 Gs). In all eukaryotes, telomeres serve as docking sites for various factors that work together to control their size during aging and protect them from fusion and action of DNA repair mechanisms (Wellinger and Zakian, 2012). Rap1 is one of these factors. Cells expressing various *rap1^{ts}* (thermosensitive) alleles are known to be associated with variations in telomere lengths (Kyrion et al., 1992; Lustig et al., 1990). Similarly to HML and HMR loci, the interaction between Rap1, and SIR factors (Sir2, 3 and 4) is crucial for establishing a silent heterochromatin-like structure near telomeric regions (Luo et al., 2002; Mattarocci et al., 2016). Finally, Rap1 also associates with the Rap1-Interacting Factor 1 and 2 (Rif1, Rif2) via its C-terminal domain. Deletion of either RIF factors leads to an increase in telomere length, suggesting that Rap1 and Rifs act together to maintain telomere length homeostasis (Hardy et al., 1992; Wotton and Shore, 1997).

1.2 Abf1

The structure of the essential Abf1 (ARS-Binding Factor 1) protein is close to that of Rap1 and other GRFs. Abf1 is composed of 731 amino acids organised into a C-terminal activation domain and an N-terminal myb-like DNA binding domain. It binds a consensus sequence CGTNN(T_4)TGAT where the underlines positions correspond to nucleotides that are most frequently found (Badis et al., 2008; Beinoravičiūtė-Kellner et al., 2005; Yarragudi et al., 2007). Abf1 plays a role in DNA replication (Wiltshire et al., 1997), DNA repair (Nucleotide excision repair) (Reed et al., 1999) and repression or activation of many genes in *S. cerevisiae* (Bosio et al., 2017a; Fermi et al., 2017).

Abf1 governs the expression of about 10% of RPGs (Fermi et al., 2016; Knight et al., 2014) and also associates with hundreds of genes involved in the biogenesis of ribosomes (Ribi genes) (Bosio et al., 2017b, 2017a). It plays a role in silencing of the HMR locus although to a lesser extent compared to Rap1 (Kimmerly et al., 1988; Shore et al., 1987).

A recent analysis of Abf1-associated RPG promoters has revealed the presence of Fhl1 (FIS component) binding sites located at the 3' end of Abf1 motifs (Fermi et al., 2016). Unlike Rap1, the occupancy of Abf1 at promoters can be modulated by external signals. Notably, inhibition of the TOR pathway is associated with a concomitant decrease in the expression of Abf1-dependent RPG and Ribi genes and an increase in Abf1 association (Fermi et al., 2016). The reason for this is however unclear since Abf1 promotes the expression of these genes in a WT

context. The link between Abf1 and the TOR pathway had already been suggested by intendent studies showing that Abf1 is one of the downstream effectors of the TORC1 complex (Oliveira et al., 2015) that is phosphorylated depending on growth conditions (e.g. nitrogen starvation, carbon sources…) (Silve et al., 1992). Hence, unlike other GRFs whose binding is thought to be ubiquitous, the occupancy of Abf1 on DNA can be regulated by external signals through post-translational modifications (Bosio et al., 2017a).

1.3 Reb1

Akin to Rap1 and Abf1, Reb1 (RNA polymerase I Enhancer-Binding protein 1) binds to specific DNA sequences through its Myb-like DNA-binding domain that recognizes the TTACCCGG motif (Badis et al., 2008; Morrow et al., 1990, 1993).

Reb1 has been first characterized as an RNAPI transcription factor. By binding at spacer regions between rRNA-coding genes, Reb1 promotes the synthesis of the 35S precursor (Kulkens et al., 1992; Morrow et al., 1989). Parallel studies also demonstrated a role for Reb1 in the regulation of class II genes (Chasman et al., 1990; Ju et al., 1990; Wang et al., 1990). Notably, Reb1 governs the expression of hundreds of Ribi genes where it can be found either in association with Abf1, or alone (Bosio et al., 2017b).

Reb1 has also been implicated in the termination of RNAPII through a roadblock mechanism (see I.3.3.2). Importantly, the termination function of Reb1 is distinct from its regulation function as expression of the Reb1-DBD alone is sufficient to induce termination, but not to ensure robust gene expression at its targets (Colin et al., 2014).

1.4 Cbf1 and Tbf1

Cbf1 (Centromere-Binding Factor 1) and Tbf1 (TTAGGG repeat-Binding Factor 1) are two additional GRFs whose function is however less spread as compared with Rap1, Abf1 and Reb1.

The non-essential factor Cbf1 has been particularly studied for its role in centromere binding. Centromeres are ~120 bp long DNA regions organised into three different modules called CDEI, CDEII and CDEIII (Centromere-Determining Elements) to which multiple DNA factors are bound (Biggins, 2013). The dispensable CDEI element is characterized by the presence of the RTCACRTG (with R = A/G) sequence that serves as a binding site for the helix-loop-helix factor Cbf1 (Niedenthal et al., 1991). Cbf1 is also involved in the transcriptional regulation

of a few genes including the MET genes, i.e. genes implicated in the S-adenosyl-methionine metabolism (Mellor et al., 1990).

Tbf1 is an essential factor containing a SANT/Myb type DNA-binding domain able to interact with the TTAGGG DNA motif (Bilaud et al., 1996; Brigati et al., 1993). Tbf1 binds sub-telomeric regions where it acts to prevent the recognition of DNA ends as DNA damage sites (Koering et al., 2000; Ribaud et al., 2012). Tbf1 has also been shown to bind to most snoRNAs promoters and to be required for robust expression of these genes (Preti et al., 2010).

2. Chromatin Remodeler Complexes

Chromatin remodelers are ATP-dependent complexes that are able to modify the chromatin structure by sliding, evicting or depositing nucleosomes along the DNA. Their action not only impacts the formation of the NDR at promoters but also influences the occupancy of nucleosomes at intragenic regions. In all eukaryotes, chromatin remodelers can be divided into four main subfamilies, depending on their ATPase domain and their respective partners: the SWItch/Sucrose Non-Fermentable (SWI/SNF), INOsitol requiring INO80, Imitation SWItch (ISWI) and Chromodomain Helicase DNA-binding (CHD) subfamily (Clapier et al., 2017; Lai and Pugh, 2017)

2.1 The SWI/SNF subfamily

In *S. cerevisiae* the SWI/SNF subfamily comprises two main subtype multi-protein complexes, the SWI/SNF and the essential RSC (Remodel the Structure of Chromatin) remodeler, whose catalytic activity is carried out by the Snf2 and Sth1 factors respectively. *In vitro* studied have revealed that RSC and SWI/SNF are able to eject (Clapier et al., 2017; Lorch et al., 2006, 2011) and slide nucleosomes along and beyond the end of a linear DNA template (Flaus and Owen-Hughes, 2003; Kassabov et al., 2003). *In vivo*, the SWI/SNF subfamily mediates the establishment of the NDR by sliding the +1 and -1 nucleosomes away from the promoter (Badis et al., 2008; Ganguli et al., 2014; Hartley and Madhani, 2009; Kubik et al., 2015; Parnell et al., 2008, 2015; Yen et al., 2012) and/or by destabilizing or ejecting the fragile nucleosome present within promoter regions (Floer et al., 2010; Kubik et al., 2015).

Overall, the RSC complex has been shown to target more promoters than the SWI/SNF complex (Ganguli et al., 2014). Cells defective for the RSC complex show an increase in nucleosome density at hundreds of gene promoters (Badis et al., 2008; Ganguli et al., 2014;

Hartley and Madhani, 2009; Kubik et al., 2015; Parnell et al., 2015). The action of RSC is intimately linked to the presence of two DNA sequences, a GC-rich motif and a poly(A) tract sequence (Badis et al., 2008; Krietenstein et al., 2016; Kubik et al., 2015). Notably, two components of the RSC, Rsc3 and Rsc30 are able to directly bind the GC-rich motif present within NDRs (Angus-Hill et al., 2001; Badis et al., 2008). More recently, Kubik and colleagues confirmed the role of the DNA sequence in the recruitment of RSC and demonstrated that the arrangement of the two motifs (i.e. orientation and spacing) influences the activity and binding of RSC (Kubik et al., 2018). The ability of RSC to promote the formation of larger NDRs by evicting nucleosomes has been shown to positively affect the recruitment of the TATA-binding protein by exposing its binding site (Kubik et al., 2018).

2.2 The ISWI and CHD subfamily

The yeast ISWI subfamily comprises the ISW1 (ISW1a and ISW1b) and ISW2 subtypes. They are composed of 2 to 3 proteins organised around the Isw1 and Isw2 ATPase-translocase factor. The Chd1 protein is the only representative of the CHD subfamily in budding yeast (Clapier et al., 2017). ISW1a, ISW1b, ISW2 and CHD1 are all able to slide nucleosomes along the DNA *in vitro*, although to a different extent (Krajewski, 2013; Stockdale et al., 2006). By using an *in vitro* nucleosome assembly set up with purified components, the Korber lab demonstrated that ISW1 and CHD1 promote the regular spacing of nucleosomes independently of nucleosome concentration (Lieleg et al., 2015). This important result suggests that ISW1 and CHD1 are capable to sense the distance between two adjacent particles *in vitro*.

In vivo, chromatin remodelers of the ISWI and CHD subfamily act mainly at intragenic regions where they participate to the phasing of the array of nucleosomes (Gkikopoulos et al., 2011; Ocampo et al., 2016; Parnell et al., 2015; Tirosh et al., 2010; Whitehouse et al., 2007). Unlike ISW2, deletion of *CHD1* and *ISW1* leads to a dramatic redistribution of nucleosomes at most genes. Moreover, cells defective for both Chd1 and Isw1 proteins display a much stronger defect in intragenic nucleosome phasing as compared with either single mutants, suggesting a cooperative effect of the two remodelers in determining the nucleosome architecture (Gkikopoulos et al., 2011; Ocampo et al., 2016). Interestingly, the size of the linker between two adjacent nucleosomes was found to depend on the remodeler: CHD1 target genes have on average a shorter distance between consecutive dyads (~160 bp) as opposed to ISW1 (~175 bp) and ISW2 (~200 bp) target genes. This difference between CHD1 and ISW1 has led the authors to speculate that they may not act in concert but rather compete at most yeast expressed genes to set the correct spacing (Ocampo et al., 2016).

Supporting the role of ISW1 and CHD1 in setting up the organisation of intragenic nucleosomes, and unlike the SWI/SNF subfamily, their depletion has only a minor effect on the position of the +1 nucleosome and on transcription initiation (Gkikopoulos et al., 2011; Lenstra et al., 2011; Ocampo et al., 2016; Parnell et al., 2015; Vary et al., 2003). Instead, they are important to promote efficient elongation of the polymerase across the genes (Simic et al., 2003; Smolle et al., 2012). However, *isw2*Δ strains also show a downstream shift of the +1 nucleosome indicating that it may function differently with respect to the other subtypes of the ISWI subfamily (Whitehouse et al., 2007; Yadon et al., 2010; Yen et al., 2012).

2.3 The INO80 subfamily

Ino80 and Swr1 compose the catalytic subunit of the INO80 and SWR-C subtypes. They both belong to the INO80 subfamily and share many common factors (Clapier et al., 2017). *In vitro* INO80 (but not SWR-C) moves nucleosomes towards the center of a linear DNA fragment (Udugama et al., 2011).

In the early 2000's, three independent studies have demonstrated that the Swr1 ATP-ase is responsible for the deposition of the H2A.Z histone variant (Kobor et al., 2004; Krogan et al., 2003b; Mizuguchi et al., 2004). The SWR-C complex catalyses the exchange of the H2A-H2B with the H2A.Z-H2B dimer (Kobor et al., 2004; Mizuguchi et al., 2004) at the +1 nucleosome (see I.1.3). The recruitment of SWR-C requires a long nucleosome-free region adjacent to a nucleosome (which usually correspond to the +1 nucleosome core particle) and is also dependent on the acetylated pattern of histones (Ranjan et al., 2013; Watanabe et al., 2013). INO80 has been originally proposed to be implicated in the reverse process, i.e. replacing the H2A.Z-H2B by the H2A-H2B dimer at the +1 position (Papamichos-Chronakis et al., 2011; Yen et al., 2013) even though its role in evicting H2A.Z remains controversial (Wang et al., 2016; Watanabe and Peterson, 2016; Watanabe et al., 2013).

3. Function of GRFs and Remodelers in Chromatin Organisation

3.1 The central role of general regulatory factors on NDR formation

It is now well established that GRFs not only bind at NDRs but also actively take part in their formation and maintenance. This notion has emerged about 20 years ago, notably with studies from the Morse lab showing that the insertion of a GRF DNA-binding motif into a chromatinized

plasmid was sufficient to alter its nucleosomal architecture (Morse, 2000; Yarragudi et al., 2004; Yu and Morse, 1999; Yu et al., 2003).

Later on, independent studies could confirm the genome-wide implication of GRFs in NDRs formation by showing that their depletion is associated with an increase in nucleosome occupancy at promoters of targeted genes (Badis et al., 2008; Ganapathi et al., 2011; Hartley and Madhani, 2009; Tsankov et al., 2011). In addition, Hartley and Madhani have demonstrated a connection between GRFs and the RSC remodeler complex. Indeed, NDRs affected by Reb1 depletion show a similar degree of sensitivity to the depletion of the catalytic subunit of RSC, Sth1. This also appears to be true for Abf1 albeit to a lesser extent. Supporting the connection between Reb1 and RSC, an artificial NDR resulting from the intragenic insertion of a Reb1 site could be similarly disrupted by depletion of either factor. Taken together, this results allowed the authors to propose a model whereby the binding of GRF at NDR could promote the recruitment of the RSC remodeler in order to position the +1 and -1 nucleosome (Hartley and Madhani, 2009). The connection between GRFs and chromatin remodelers in NDR formation is also supported by later studies showing that DNA-binding proteins with a nucleosome displacing activity (including Reb1 and Abf1) are all known to interact directly with remodeler complexes (Ozonov and van Nimwegen, 2013).

The function of GRFs in NDRs formation has been confirmed more recently in a study published by the Shore lab (Kubik et al., 2015). In this report, the authors proposed that the binding of Reb1, Rap1 and Abf1 could not only promote the sliding of the +1 and -1 nucleosomes away from the promoter region, but also destabilize an MNase sensitive nucleosome particle present at large NDRs (see I.1.3).

Collectively, these studies have demonstrated the role of GRFs in modulating the chromatin organisation in the vicinity of their binding sequences. This function is crucial to favour the formation of the NDRs which in turn ensures the assembly of the pre-initiation complex.

3.2 Cooperative and antagonist action of chromatin remodelers

In vivo, different chromatin remodelers have been shown to bind the 5’ end of the same sets of genes (Yen et al., 2012), suggesting that they function together at these targets. Chromatin remodelers with similar activity such as SWI/SNF and RSC (Rawal et al., 2018) but also CHD1 and ISW1 (Gkikopoulos et al., 2011) can cooperate to promote gene expression and nucleosome positioning at common genes. Antagonistic effects have also been reported for RSC and ISW1 (Parnell et al., 2008) or SWI/SNF and ISW2 (Tomar et al., 2009) at some

genes. Collectively, these studies and others have established that accurate organisation of the nucleosome landscape in eukaryotes is the result of the synergistic and antagonistic action of different chromatin remodelers (Gkikopoulos et al., 2011; Morris et al., 2014; Parnell et al., 2015; Rawal et al., 2018; Tomar et al., 2009; Yen et al., 2012).

In 2016, the Korber lab set up an *in vitro* chromatin assembly system using purified histones and genomic DNA to assesse the ability of nucleosomes to correctly organise depending on the presence of chromatin remodelers and GRFs (Krietenstein et al., 2016). This *in vitro* reconstituted system constitutes a remarkable tool as it allows to precisely define the major determinants of nucleosome positioning and the connection between different complexes and factors. In this system, it has been shown that INO80 by its own is sufficient to correctly position the +1 nucleosome in presence of DNA and histones only. In the absence of INO80, the position of the +1 could also be obtained by the combined action of RSC and ISW2 (and to a lesser extent ISW1a) although in this context, the presence of a GRF (Abf1 or Reb1) is required. This result demonstrates that chromatin remodelers with opposite action can collaborate *in vitro* to establish nucleosome organisation. In addition, the authors could confirm the crucial role of the ISWI subfamily, and most particularly ISW1a, and INO80 in the establishment of the nucleosome spacing. More importantly, the ability of GRFs and chromatin remodelers to properly reconstitute the pattern of nucleosomes demonstrates that these factors are the main actors of the chromatin organisation.

In summary: In eukaryotes, the establishment of the chromatin architecture is crucial for proper gene expression and other DNA associated events (DNA repair, replication, telomere maintenance…). GRFs and chromatin remodelers have been particularly studied for their ability to shape the nucleosomal landscape. GRFs (also called pioneer factors in mammals) are DNA-binding factors that, despite the absence of any enzymatic activity, play a role in the formation of NDRs at hundreds of promoters. Chromatin remodelers act in an ATP-dependent manner to slide, evict or deposit nucleosomes along the DNA. These two categories of factors and complexes have been shown to act in concert and at common targets to define nucleosome positioning. Importantly, the activity of GRFs and remodelers is an evolutionary conserved feature of eukaryotes and perturbation of the latter can be associated with diseases in human.

III - Pervasive Transcription in *S. cerevisiae*

The development of high-throughput techniques has revealed an unexpected complexity of the transcriptome in virtually all organisms analysed. At the beginning of the 21st century, many different groups identified a plethora of mainly non-functional RNAPII-dependent transcripts derived from intergenic regions of eukaryotic genomes. This discovery gave rise to the notion of “pervasive” or “hidden” transcription, i.e. transcription occurring beyond regions annotated for the production of functional molecules such as tRNAs, snRNAs, snoRNA, rRNA, and mRNAs.

In this section, I will summarize the major studies that have led to the discovery of such transcriptional events and will briefly describe the landscape of pervasive transcription in yeast. I will describe the mechanisms that are devoted to the control of pervasive transcription and will finally show how pervasive transcription can represent a source of regulation for canonical RNAPII transcripts.

1. The Landscape of Pervasive Transcripts

1.1 Discovery of pervasive transcripts

In 2005, the Libri, Jacquier and Seraphin laboratories reported the existence of cryptic unstable transcripts in yeast (Wyers et al., 2005). CUTs are a class of RNAs that are very unstable and rapidly degraded in the nucleus by the Rrp6 component of the nuclear exosome (see I.3.2.3). The instability of the largest share of pervasive transcripts partially explains their late observation. This discovery was soon after confirmed by two independents studies (Davis and Ares, 2006; Houalla et al., 2006). Despite the fact that CUTs are probably the most abundant pervasive transcripts in yeast, other kinds of RNAs were later characterized.

Stable Unannotated Transcripts (SUTs), as opposed to CUTs, are stable transcripts that can be observed in a WT context (Xu et al., 2009). SUTs are primarily degraded in the cytoplasm by the Xrn1 5'-3' exonuclease (Marquardt et al., 2011) and are also sensitive to the nuclear exosome, albeit to a lesser extent (Gudipati et al., 2012; Marquardt et al., 2011). CUTs and SUTs are respectively terminated by the NNS (Arigo et al., 2006b; Thiebaut et al., 2006) and CPF-CF (Marquardt et al., 2011) pathway although the two classes often overlap with one another.

Table 2. Yeast transcripts and their associated termination and processing/degradation pathways.

Transcript	Termination pathway	Stability	Degradation factors
mRNA	CPF-CF	Stable	
sn/snoRNA	NNS; Pcf11	Stable (3' end processed)	Nuclear exosome
CUT	NNS	Unstable	Nuclear exosome
SUT	CPF-CF and possibly NNS	Partially unstable	Nuclear exosome, Xrn1 (NMD)
XUT	CPF-CF	Unstable	Xrn1 (NMD)
RUT	Reb1 roadblock	Unstable	Nuclear exosome & NMD

In 2011, the Morillon group characterized the XUT RNAs (Xrn1-sensitive Unstable Transcripts) using cells defective for Xrn1 (van Dijk et al., 2011). XUTs are terminated by the CPF-CF pathway and exported in the cytoplasm where they are rapidly degraded. Unlike SUTs, XUTs are more difficult to observe in WT cells, but the difference between these two categories is largely relates to the stability of the molecule.

As previously mentioned, RUTs are specifically terminated by the Reb1-dependent roadblock mechanism. They are stabilized in *rrp6*Δ cells but can also be targeted to the nonsense-

mediated mRNA decay pathway (see III.2.3) for instance when they overlap mRNA-coding regions (Colin et al., 2014) (see the general discussion and perspectives section).

CUTs, SUTs, XUTs, and RUTs are defined according to the termination and degradation pathway to which they are associated (Table 2 adapted from Porrua and Libri, 2015). Importantly, however, the distinction between these different classes of pervasive transcripts is sometimes blurry. Indeed, a given pervasive transcript can be primarily targeted by the NNS pathway and degraded in the nucleus by the TRAMP and nuclear exosome pathway (CUTs). Nonetheless, as biological processes are not fully efficient, a fraction of the transcripts can escape the main termination pathway and be subsequently terminated by another mechanism (CPF-CF or roadblock) and thus, be considered as a XUT, SUT or RUT. Similarly, some RNAs may escape nuclear degradation and be exported and targeted by cytoplasmic quality control pathways. Consistent with this idea is the fact that most of the considered pervasive transcripts are sensitive to both the cytoplasmic and nuclear degradation pathways (Malabat et al., 2015; Smith et al., 2014). Common characteristics of these RNAs are the absence of any clear function and their poor coding potential. Even though they can sometimes be found associated with polysomes similar to mRNAs, they do not appear to code for functional peptides (Carvunis et al., 2012; Smith et al., 2014; Wilson and Masel, 2011).

1.2 Origins of pervasive transcription

Akin to canonical RNAPII transcripts, cryptic RNAs also emanate from nucleosome depleted regions (NDRs) present at 5' and 3' of genes or from cryptic promoters present in intragenic regions (Malabat et al., 2015; Neil et al., 2009). An essential notion that has emerged about 10 years ago is the fact that most promoters are bidirectional. The ability of promoters to fire in both direction has been proposed to be the major source of pervasive transcription in yeast (Neil et al., 2009; Xu et al., 2009) (Figure 12). Importantly, non-coding transcription depends on the formation of a distinct PIC with respect to the divergent or sense coding gene (Murray et al., 2012; Rhee and Pugh, 2012) and, in some instances, it has been proposed that the two PICs may compete with one another for the same pool of general transcription factors. For instance, mutation of the TATA box of the *TPI1* gene leads to an upregulation of the divergent CUTs (Neil et al., 2009).

Despite the clear bidirectional state of promoters in yeast, the level of transcription is often higher towards the "mRNA-coding" direction with respect to the non-coding direction (Churchman and Weissman, 2011; Jin et al., 2017). In a recent study, Jin and colleagues have introduced in *S.cerevisiae* a Yeast Artificial Chromosomes (YAC) containing DNA from other

yeast species, in order to study the bidirectional nature of promoters. Remarkably, they found that the preference for the sense direction tends to be lost when promoters are transferred in a different species (Jin et al., 2017). This result suggests that promoters are intrinsically bidirectional and that active mechanisms have likely contributed to the progressive selection of the "correct" orientation across evolution. Supporting this concept, newly evolved promoters (i.e. only present in the *Saccharomyces sensu stricto* genus) are more markedly bidirectional as opposed to promoters of genes that are also present in more divergent yeast species.

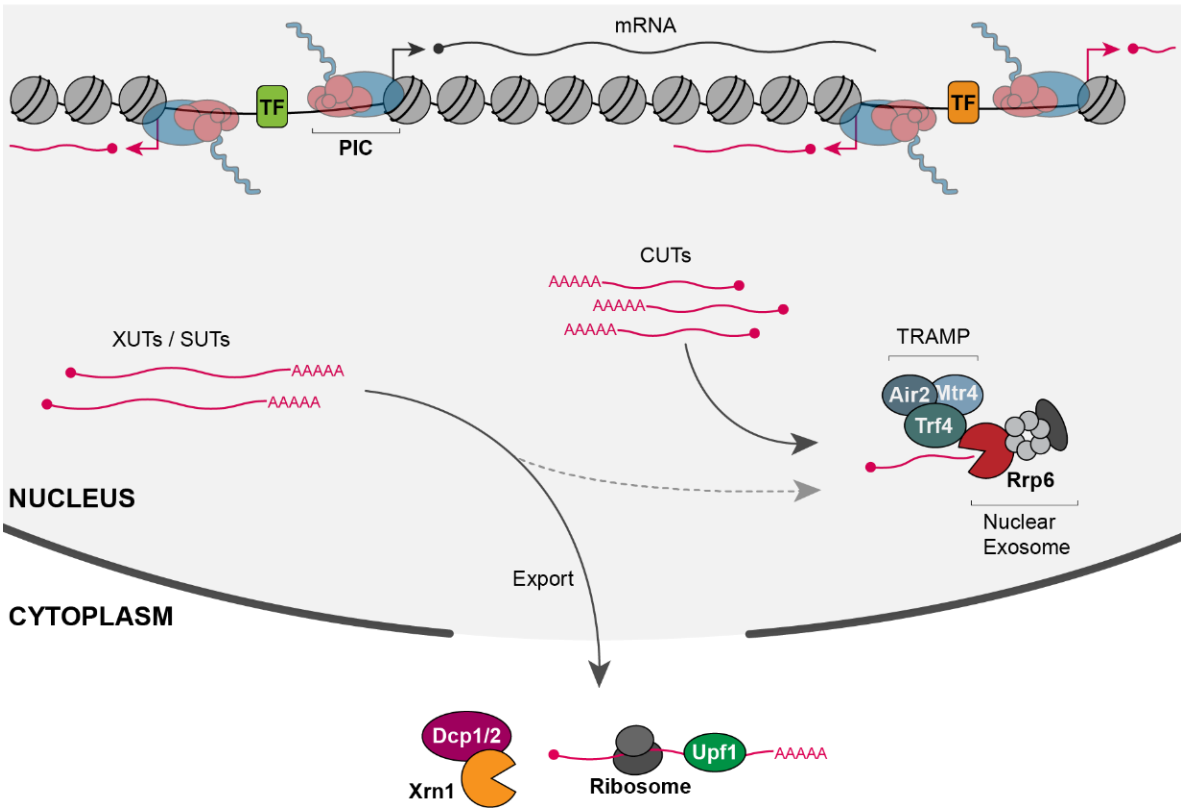


Figure 12. Origin and fate of yeast pervasive transcripts. Pervasive transcripts mainly originate from bidirectional promoters where distinct pre-initiation complex assemble. After their release, non-coding RNAs are immediately degraded in the nucleus or are exported to the cytoplasm and targeted by quality control mechanisms including the nonsense-mediated decay. The action of the decapping enzymes Dcp1 and Dcp2 enables the digestion of the RNA by the Xrn1 5'-3' exonuclease. Adapted from (Tudek et al., 2015).

The intrinsic bidirectionality of promoters is a shared feature of all eukaryotic cells (Wei et al., 2011). In humans, various non-coding RNAs originate from NDRs located at the 5' or 3' of genes. These include stable ncRNAs such as short and long RNA (sRNA and lRNA), transcription initiation RNAs (tiRNA), promoter or terminator associated RNA (PASR, PALR, TASR) as well as unstable RNAs like the promoter proximal transcript (PROMPT) (Kapranov et al., 2007; Preker et al., 2008; Seila et al., 2008; Sigova et al., 2013; Taft et al., 2009).

2. Control of Pervasive Transcription

As demonstrated above, pervasive transcription is widespread in the *S. cerevisiae* genome. The occurrence of spurious transcription initiation events is a potential danger for the cell, since, if uncontrolled, it could notably interfere with other DNA-associated events. Also, production of aberrant transcripts may generate unwanted translational products or titrate processing/export factors. Hence pervasive transcription needs to be tightly controlled. This control is commonly considered to occur *via* three distinct mechanisms: (i) by limiting the number of spurious initiation events, (ii) by promoting termination of non-coding transcription (iii) by degrading non-functional transcripts post-transcriptionally *via* nuclear and cytoplasmic quality control mechanisms (Jensen et al., 2013).

2.1 Control of pervasive transcription initiation

Many studies in the last decade have highlighted the notion that transcription initiation is tightly controlled by many factors, and is not only strictly dependent on *cis*-acting sequence signals. In this section, I will describe some of the factors that have been characterized in the literature and explain how they affect transcription initiation. This section is relevant to one of my main thesis projects (results section).

2.1.1 The Spt6 & Spt16 histone chaperones

The first factor that has been reported to function in restricting spurious transcription initiation is the Spt6 histone chaperone (see I.2.2.1). In 2003, Kaplan, Laprade, and Winston found that mutation of *SPT6* (*spt6-1004* thermosensitive mutant) causes the emergence of aberrant initiation events within coding regions (Kaplan et al., 2003). These newly synthesized RNA molecules are often shorter than the full-length mRNA and are transcribed in the same orientation. Moreover, careful analysis of the *FLO8* model gene has revealed that the internal initiation site is located close to a TATA sequence element bound by the TBP factor and whose mutation abolishes intragenic initiation.

More recently, the effect of the same *spt6-1004* mutant was reinvestigated using various genome-wide approaches (Doris et al., 2018). Mapping of TSSs in wild-type and *spt6-1004* cells at non-permissive temperature have revealed thousands of new or upregulated intragenic (sense ~6000, or antisense ~2000), as well as some intergenic (~400) initiation events. In agreement with the results from Kaplan et al, components of the PIC complex (TFIIB) are enriched at new intragenic initiation sites. Importantly, internal TSSs arise at regions containing

motifs and features that are characteristic of bona fide intergenic promoters: they are located at loci where the nucleosome occupancy decreases in the Spt6 mutant (NDR-like), they take place within AT-rich regions (like promoters), they often contain the $A(A_{rich})_5NY\underline{R}$ motif associated with the transcription initiation site (Malabat et al., 2015) and a consensus TATA motif (Doris et al., 2018; Uwimana et al., 2017).

Note: The study of antisense transcripts emerging within mRNA-coding regions in Spt6 mutant revealed that these RNAs often terminate at the CPF-CF terminator located at the 3' end of the upstream gene. This appears to be true even when the upstream gene is tandemly oriented as compared with the gene from which the antisense Spt6-dependent transcript arises. These set of experiments demonstrate that most CPF-CF terminators are bidirectional which might be required to limit the extension of pervasive transcripts (Uwimana et al., 2017).

Similarly, the Spt16-comprising FACT chaperone complex is also important to ensure transcription fidelity by limiting the occurrence of initiation from internal coding regions in yeast (Cheung et al., 2008; Mason and Struhl, 2003) as well as in more complex eukaryotes such as *A. thaliana* (Nielsen et al., 2019). Mutation of the yeast Spt16 causes the activation of internal and bidirectional cryptic promoters and results in the production of ~1000 "Spt6-suppressed Non-coding Transcripts" (SNTs) (Feng et al., 2016).

The occurrence of spurious intragenic initiation has been shown to depend on the transcription of the gene (Kaplan et al., 2003), suggesting that histone chaperones might act by favouring the redeposition of nucleosomes behind the polymerase, which limits the production of pervasive transcripts.

2.1.2 The Set2-Rpd3S pathway

During transcription elongation, the deposition of the H3K36 methylation mark by Set2 promotes the recruitment of the histone deacetylase Rpd3S. Perturbation of either complex has a disruptive effect on chromatin organisation and leads to an increase in spurious transcription initiation events (Carrozza et al., 2005; Cheung et al., 2008; Churchman and Weissman, 2011; Li et al., 2009; Malabat et al., 2015; Venkatesh et al., 2016). For instance, cells defective for Rco1 (component of Rpd3S) show a higher proportion of antisense transcription at divergent promoters (Churchman and Weissman, 2011), and an increased production of internal transcripts at the *FLO8* and *STE11* loci (Carrozza et al., 2005).

Set2 is also involved in the repression of many internal TSSs (Malabat et al., 2015) and its deletion induces internal initiation in the *FLO8* and *STE11* genes similar to what observed in

Rco1 or Spt6 mutants (Carrozza et al., 2005). In 2016, Venkatesh and colleagues gave the name of SRATs (Set2-repressed antisense transcripts) to antisense RNAs whose expression is governed by Set2. The authors have found more than 800 antisense transcripts that are produced in the absence of Set2. Importantly, many of these RNAs are already pre-existing in WT cells and can be better detected when they are stabilized in the absence of Xrn1 (involved in the main cytoplasmic degradation pathway) (Malabat et al., 2015; Venkatesh et al., 2016). It has been proposed that the production of these transcripts is also related to the redeposition of nucleosomes behind the transcribing polymerase, which is facilitated by the histone deacetylation activity of Rpd3S (see I.2.3.1).

2.1.3 Chromatin remodelers

At coding regions, ISW1b and CHD1 function together to repress intragenic transcription, notably by limiting the exchange between histones incorporated onto the DNA fiber and free particles present in the cell (Smolle et al., 2012). Moreover, the cryptic transcripts observed in a double mutant (*Isw1*Δ and *Chd1*Δ) often overlap with those detected in *Set2*Δ strains. This is explained at least partially by the fact that the recruitment of ISW1 (and more particularly ISW1b) requires the presence of the H3K36 methylation mark (Smolle et al., 2012).

The chromatin remodelers ISW2 and INO80 restrict the size of NDRs (See II.2). Perturbation of their function is associated with an increase in the levels of pervasive transcription emanating from bidirectional promoters (Whitehouse et al., 2007; Xue et al., 2015, 2017; Yadon et al., 2010). Deletion of components of the MINC (Mot1-Ino80-NC2) complex leads to a drastic increase in the steady state level of XUTs, SUTs or CUTs initiating from yeast bidirectional promoters. This role of MINC in limiting transcription at bidirectional promoters is also conserved in humans where it prevents the synthesis of PROMPTs (Xue et al., 2017). The MINC has been shown to co-purify with the PICs at promoters. Based on this observation, the authors have proposed that the MINC may contribute to the silencing of pervasive transcription by binding at PICs thereby preventing the recruitment of the polymerase at promoters (Xue et al., 2017). However, since INO80 impacts chromatin structure, it is also possible that the observed defect in transcription initiation results from the altered intragenic nucleosome positioning (Whitehouse et al., 2007; Yadon et al., 2010).

2.2 Transcription termination limits pervasive transcription

Despite the strong control on initiation, pervasive transcription events occur genome-wide. In this context, transcription termination plays an important role as it allows to partition the

genome and restrict the progression of polymerases thereby limiting the chances to encounter other DNA associated events (e.g. transcription of coding genes or replication forks) and to perturb their function (see III.3 and the general discussion and perspectives section).

Although all the previously mentioned termination pathways participate to this layer of regulation (see I.3 and III.1.1), the NNS pathway appears to be particularly involved in the control of pervasive transcription. Indeed, unlike CPF-CF, NNS termination events occur rapidly after initiation (Gudipati et al., 2008) and is coupled with efficient turn-over of the neo-synthetized non-coding transcript, thus preventing its accumulation (see below).

2.3 Nuclear and cytoplasmic quality control mechanisms

As previously mentioned, a vast majority of pervasive transcripts are degraded in the nucleus by the TRAMP and nuclear exosome pathway. Many of these RNAs contain Nrd1 and Nab3-binding sites. The action of the NNS pathway ensures the efficient and rapid degradation of the transcripts by coupling termination with degradation through direct interactions between components of the NNS, TRAMP, and exosome (detailed in I.3.2). The degradation of cryptic RNAs is considered as being important to avoid the accumulation of non-functional species that could otherwise compete with functional RNAs (mRNA, rRNA…) for the recruitment of various RNA-binding factors.

Despite the efficient degration of pervasive transcripts in the nucleus, a non-negligible amount of non-coding transcripts reaches the cytoplasm. This is the case for instance for the XUTs, SUTs and for transcripts that have escaped NNS termination. The latter molecules generally derived from transcription events that terminate at downstream terminators by the CPF-CF pathway, and are therefore directed to the export pathway.

In the cytoplasm, many different quality control mechanisms coexist to control the presence of aberrant transcripts. Most of these pathways are activated during translation and result in the degradation of the RNA, generally *via* the 5’ to 3’ exoribonuclease Xrn1. The main cytoplasmic quality control mechanism associated with the decay of cryptic transcript is the NMD pathway.

The NMD degrades mRNA molecules containing premature termination codons, generally derived from inefficiently spliced pre-mRNAs, mRNAs containing upstream ORFs or mRNAs undergoing a frameshift of the ribosome during translation. Overall, the NMD signals the presence of abnormally long 3’ UTR regions. One of the central effectors of the NMD pathway is the Upf1 helicase. Upf1 interacts with the Dcp1-Dcp2 decapping enzymes and the 5’ to 3’

exoribonuclease Xrn1. The action of the decapping complex offers an entry point for Xrn1 that degrade the targeted RNA (Kervestin and Jacobson, 2012). The helicase activity of Upf1 is also required for NMD, eventhough, the exact mechanism through which it triggers degradation is still under debate. A plausible model is that Upf1 favours termination and release of the paused ribosome, which somehow enhances degradation efficiency (Serdar et al., 2016)

Remarkably, deletion of *XRN1* in an *upf1*Δ background does not further enhance the stability of most XUTs and SUTs as is the case for mRNA (Malabat et al., 2015). This result indicates that the two factors function in the same pathway and therefore that NMD is the main mechanism limiting the accumulation of XUTs and SUTs in the cytoplasm. Supporting this idea, the analysis of the sequence of these two categories of transcripts has revealed the frequent presence of small ORFs, often located close to the 5' region. Furthermore, the 3' UTR region following these spurious ORFs are longer, on average, than mRNA 3' UTRs thus representing an efficient substrate for the NMD pathway.

3. Role of Pervasive Transcription

What is the role of pervasive transcription? This question has obviously been raised soon after the discovery of pervasive transcription. Several studies have revealed that, even though the non-coding RNA molecules have no clearly established role *per se* in budding yeast, the process of pervasive transcription is a powerful mechanism that participates to gene expression regulation. In addition to its role on regulation, pervasive transcription may represent an important source of novel protein-coding genes across evolution.

3.1 Pervasive transcription and regulation of protein-coding genes

The *SER3* gene is the first reported case of gene regulation by a pervasive transcription event (Martens et al., 2004). *SER3* is implicated in the biosynthetic pathway of serine and glycine. The *SRG1* (*SER3* Regulatory Gene 1) gene is located upstream and transcribed in tandem relative to *SER3*. *SRG1* transcription is under the control of various activators including the Cha4 transcription factor, the SAGA complex and the SWI/SNF chromatin remodeler. Importantly, *SRG1* transcription overlaps the *SER3* promoter and causes its downregulation by a transcriptional interference mechanism when serine is present in the growth medium (Martens et al., 2004, 2005). It has been shown that the repressive effect of *SRG1* on the expression of *SER3* requires the action of Spt6 and the FACT complex (Spt16) that act in a

synergistic manner to increase the nucleosome occupancy at *SER3* promoter after the passage of the polymerase (Hainer et al., 2011).

The expression of *DCI1* and *GLO4* is also associated with the transcription of a non-coding gene producing a CUT initiating in the 5' region (sense) or 3' region (antisense) relative to *DLC1* and *GLO4,* respectively. In both cases, Set3, a component of the Set3C histone deacetylase, is required for the repression of the targeted genes by favouring the compaction of the chromatin upstream of the polymerase (Kim et al., 2012) (see I.2.3.1).

Like *GLO4*, *PHO84* repression is also controlled by antisense transcription that is activated during cell aging (Camblong et al., 2007). Interestingly, single cell FISH analysis reveals that the expression of the sense and antisense transcript are anticorrelated and never occur concomitantly. This suggests that transcriptional interference occurs by epigenetic mechanisms and not by preventing elongation because of collisions between polymerases transcribing in the opposite direction. Consistently, silencing of the gene depends on the action of an histone deacetylases, Hda1 and its cofactors Hda2 and Hda3, as well as the Set1 H3K4 histone methyltransferase (Camblong et al., 2007, 2009; Castelnuovo et al., 2013).

PHO5 and *CDC28* belong to the rare cases for which antisense transcription leads to the increased production of the mRNA (Nadal-Ribelles et al., 2014; Uhler et al., 2007). In the case of *PHO5*, the ncRNA reaches the promoter of the mRNA-coding gene and influences the kinetic of *PHO5* activation by a mechanism that involves chromatin re-organisation (Uhler et al., 2007). Regarding *CDC28*, the expression of the long antisense non-coding transcript is driven by the binding of the Hog1 SAPK (Stress-activated protein kinases) at the 3' region upon stress condition. Then, it has been suggested that Hog1 contacts the +1 nucleosome of *CDC28* through a gene looping phenomenon depending on the Ssu72 phosphatase. The presence of Hog1 in close proximity of the +1 nucleosome promotes the recruitment of the RSC complex and the upregulation of *CDC28*. This mechanism plays a role in the re-entry of cells into the cell cycle following stress-induced response (Nadal-Ribelles et al., 2014).

The process of gametogenesis (sporulation in yeast) requires the transient expression of *IME1* (Inducer of MEiosis 1) in diploid cells. In haploid cells, the transcriptional activator Rme1 promotes the expression of the SUT *IRT1* located upstream and transcribed in the same orientation of *IME1*. The *IRT1*-induced transcriptional interference requires the histone methyltransferase Set2 and the Set3 histone deacetylase that increase the density of nucleosome around *IME1* promoter, thus preventing the binding of its transcriptional activator Pog1. Remarkably, the introduction of an early terminator close to the *IRT1* initiation site

causes the activation of *IME1,* which is sufficient to force the entry in meiosis even in haploid cells. In the same system, it has been shown that antisense transcription giving rise to the production of *RME2* represses the *IME4* coding gene that is also implicated in efficient entry into sporulation. *RME2* and the *IRT1*-activator Rme1 are both under the control of the a1-α2 repressor expressed in diploid cells (van Werven et al., 2012). Thus, the gametogenesis process is under the control of genes producing non-coding RNAs whose transcription negatively controls the expression of mRNA-coding meiotic genes that are silenced during mitotic growth conditions.

Many additional examples of gene regulation by a sense or antisense ncRNA have been described in the literature (e.g. *GAL1*, *GAL10*, *NDC80*, *ADH1*, *FLO11*, *HMS2*…). One of the common features of all these transcriptional interference events is the implication of methyltransferases and deacetylases to promote gene silencing. In this context, the role of the Set1–H3K4me2–Set3C and the Set2–H3K36me3–Rpd3S pathway have been extensively studied genome-wide and under different physiological conditions (Kim et al., 2016, 2012; Nevers et al., 2018). The dependency on either one of the two pathways is intimately linked to the distance between the TSS of the ncRNA and the mRNA. Indeed, while the Set1-Set3C pathway functions when the non-coding RNA TSS is located at ~900bp on average from the TSS of the coding gene, the Set2-Rpb3S is involved in the silencing of promoters located on average 2kb downstream from the initiation site of the non-coding gene (Kim et al., 2016).

Note: Mutants of the NNS pathway are known to induce a high frequency of transcriptional interference events due to readthrough transcription of pervasive transcripts. Remarkably, deletion of both SET2 and SET3 leads to a less severe growth phenotype in some NAB3 mutants (personal communication from the Jacquier lab), suggesting that an important role of the NNS pathway is to protect promoters from inappropriate silencing.

Pervasive transcription is also associated with other types of regulatory mechanisms that are independent of transcription interference. This is, for instance, the case for *URA2* and *IMD2*, two genes whose expression is linked to the concentration of nucleotides (uracil and guanine respectively). Those two genes share a similar organisation with an upstream 5’ CUT transcribed from a distinct TSS with respect to the mRNA but regulated by a unique promoter and TATA element. When the pool of nucleotide is high, transcription starts from the upstream TSS and gives rise to an unstable transcript terminated by the NNS pathway. Upon nucleotide starvation, transcription initiates from the downstream TSS, thereby bypassing the NNS terminator and promoting the synthesis of a full-length and functional mRNA molecule (Kuehner and Brow, 2008; Thiebaut et al., 2008).

3.2 Pervasive transcription: a source of new genes?

Despite the large variety of non-coding RNAs produced in *S. cerevisiae* (CUTs, XUTs, SURs…), they do not appear to play a major role in the cellular fate. However, there is clear evidence that some of these molecules can be translated (Carvunis et al., 2012; Smith et al., 2014; Wilson and Masel, 2011). Thus, strikingly, a lot of energy seems to be devoted to pervasive transcripts. An interesting idea that has emerged during the last decades is that some of the non-coding transcription units may evolve over time to form new genes.

Apart from horizontal transfers, the acquisition of new coding genes can occur *via* two main distinct strategies: they can derive from the re-organisation or duplication of pre-existing coding sequences or emerge from non-coding RNAs that acquire functional ORFs (Carvunis et al., 2012; Thiebaut et al., 2006). In the latter case, evolution would be gradual as it would not affect (or to a lesser extent) the pool of pre-existing mRNA-coding genes.

Interestingly, by comparing genomes of 15 yeast species from two genera (*Lachancea* and *Saccharomyces*), Vakirlis and colleagues could identify 30 non-coding transcriptional units whose mutations have led to the formation of an ORF in *S. cerevisiae* (Vakirlis et al., 2018). Moreover, by analysing a set of *de novo* gene candidates, the authors demonstrated that new genes are more likely to emerge from bi-directional promoters where they may benefit from the transcription of the divergent coding region. Taken together, these data suggest that non-coding units have the potential to evolve towards coding genes.

In summary: Pervasive transcription emanates from nucleosome depleted regions that form at 5' and 3' regions of genes. It gives rise to the production of a plethora of different RNA molecules that are mostly characterized by their poor coding potential and low stability. The role of these ncRNAs in S. cerevisiae, if any, is not clear. However, the act of transcription represents an efficient strategy for the regulation of mRNA-coding genes. In the light of evolution, this is of particular interest in the budding yeast where no RNA interference machinery is present. Finally, because pervasive transcription can interfere with the expression of canonical genes, it is important to limit its occurrence. Different layer of regulations acts in concert to this purpose. These include the restriction of initiation events, the early termination of non-coding RNAs coupled with nuclear degradation and the presence of cytoplasmic quality control mechanisms that promote the turnover of aberrant transcripts.

RESULTS

I - High-resolution Transcription Maps Reveal the Widespread Impact of Roadblock Termination in Yeast

When I initially joined the laboratory, the Reb1-dependent roadblock termination (described in I.3.3.2) was still under investigation. In addition, a Ph.D. student from the laboratory had also started to address the role of another GRF, Rap1, in promoting RNAPII termination by a similar mechanism as Reb1. The ability of these two factors to physically prevent the progression of elongation polymerases raised an important question: *How general and widespread is the roadblock termination across the yeast genome?*

In order to successfully answer this question, we needed a technique that would allow the accurate detection of the RNAPII elongation complexes at a nucleotide resolution and in a strand specific manner. Thus, I started my project by setting up and improving the CRAC (UV crosslinking and analysis of cDNA) technique, a method originally developed by the Tollervey lab that allows the detection of transcripts bound by any RNA-binding protein (Granneman et al., 2009). The goal was to apply the CRAC technique to isolate and sequence nascent transcripts bound to the RNAPII. Under the supervision of Jessie Colin, we could significantly improve the CRAC method through different modifications that are detailed in Candelli et al., 2018 and Challal et al., 2018 (see below).

RNAPII CRAC in various mutant for termination pathways have revealed that the roadblock termination is a widespread mechanism employed in *S. cerevisiae* to limit the progression of natural readthrough arising at canonical termination pathways. Importantly, we have demonstrated that roadblock termination occurs independently from the other known termination mechanisms. Finally, we have provided evidence that roadblock events can be carried out by various DNA-binding proteins including GRFs, centromere-binding proteins and RNAPIII transcription factors.

Detailed contribution: D.L and J.C supervised the project. D.L wrote the paper and was responsible for funding acquisition. J.C, T.C, D.C and O.P reviewed and edited the draft. *In vivo* selection and prior investigations were performed by O.P and J.B. The bioinformatic part (data processing and analysis) was performed by T.C. Northern blot experiments were mainly performed by J-B.B. Yeast strains and plasmids were constructed and designed by J.C, J-B.B and D.C. RNA-Seq data were performed by J.C and D.C. RNAPII CRAC experiments were carried out by D.C.

Published online: January 19, 2018

Article

SOURCE DATA | TRANSPARENT PROCESS

THE EMBO JOURNAL

High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast

Tito Candelli[1,2,†,‡], Drice Challal[1,2,†], Jean-Baptiste Briand[1,2,†], Jocelyne Boulay[3], Odil Porrua[1], Jessie Colin[1,*,§] & Domenico Libri[1,**]

Abstract

Transcription termination delimits transcription units but also plays important roles in limiting pervasive transcription. We have previously shown that transcription termination occurs when elongating RNA polymerase II (RNAPII) collides with the DNA-bound general transcription factor Reb1. We demonstrate here that many different DNA-binding proteins can induce termination by a similar roadblock (RB) mechanism. We generated high-resolution transcription maps by the direct detection of RNAPII upon nuclear depletion of two essential RB factors or when the canonical termination pathways for coding and non-coding RNAs are defective. We show that RB termination occurs genomewide and functions independently of (and redundantly with) the main transcription termination pathways. We provide evidence that transcriptional readthrough at canonical terminators is a significant source of pervasive transcription, which is controlled to a large extent by RB termination. Finally, we demonstrate the occurrence of RB termination around centromeres and tRNA genes, which we suggest shields these regions from RNAPII to preserve their functional integrity.

Keywords pervasive transcription; Rap1; roadblock termination; transcription readthrough; transcription termination mechanism
Subject Categories RNA Biology; Transcription
DOI 10.15252/embj.201797490 | Received 2 June 2017 | Revised 14 December 2017 | Accepted 15 December 2017 | Published online 19 January 2018
The EMBO Journal (2018) 37: e97490

Introduction

The compact genome of *Saccharomyces cerevisiae* is covered by several machineries that need to be temporally and spatially coordinated for allowing the robust reading and perpetuation of the genetic information.

The complexity of the transcriptional landscape is paradigmatic in this regard. Transcription initiation occurs frequently in regions and direction that largely overrun the canonical annotation of genes, a phenomenon known as pervasive transcription. This is due to the inherently loose control imposed on initiation by the structure of chromatin and to the intrinsic bi-directionality of promoters, which is generally conserved in evolution (Porrua & Libri, 2015). This promiscuity of transcription events is a potential threat to the stability of gene expression programs because many transcription events are susceptible to interfere with each other. Pervasive transcription might also affect other DNA-related events, such as replication, chromosome segregation, or the expression of RNA polymerase I- and III-dependent genes. The integration of widespread transcription with other cellular processes is a complex process, requiring tools to limit and coordinate concurrent events.

Transcription termination plays essential roles in the control of pervasive transcription. In yeast, two main termination pathways exist. The first depends on the cleavage and polyadenylation factor–cleavage factor (CPF-CF, referred to as CPF hereafter) and terminates transcription of genes producing mRNAs and some non-coding RNAs. The CPF complex recognizes signals on the nascent RNA and cleaves the latter, producing a 5′ fragment that is polyadenylated by the Pap1 poly(A) polymerase and exported to the cytoplasm. The 3′ fragment, still associated with the transcribing polymerase, is recognized and degraded by a 5′-3′ exonuclease, Rat1, which contributes to dismantling the elongation complex by a still elusive mechanism. The CPF is also believed to be directly involved in the polymerase release step of termination by allosterically modifying the properties of the transcription elongation complex (for a recent review, see Porrua *et al*, 2016).

The second canonical pathway depends on the NNS (Nrd1-Nab3-Sen1) complex and was first associated with the production of sn- and snoRNAs (Steinmetz *et al*, 2001). Nrd1 and Nab3 bind the

1 Institut Jacques Monod, CNRS, UMR 7592, Univ Paris Diderot, Paris, France
2 Ecole doctorale Structure et Dynamique des Systèmes Vivants, Université Paris Saclay, Gif sur Yvette, France
3 Institut de Biologie Intégrative de la Cellule (I2BC), CNRS, UMR 9198, Univ Paris-Saclay, Centre Energie Atomique, Gif sur Yvette, France
*Corresponding author. Tel: +33 170429451; E-mail: jessie.colin@uvsq.fr
**Corresponding author. Tel: +33 157278065; E-mail: domenico.libri@ijm.fr
†These authors contributed equally to this work
‡Present address: Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands
§Present address: Laboratoire de Génétique et Biologie Cellulaire, EA4589, UVSQ/Université Paris-Saclay, EPHE/PSL Research University, Montigny-le-Bretonneux, France

nascent RNA at short motifs containing a well-conserved 4–5 nucleotides core and are thought to recruit the Sen1 helicase that translocates on the nascent RNA to release the polymerase. Peculiar to this pathway is that the released RNA is polyadenylated by a different poly(A) polymerase, Trf4, functioning within the TRAMP4/5 (Trf4/5-Air2/1-Mtr4-polyadenylation) complex, and trimmed to its mature size in the nucleus by the exosome, a large multisubunit complex that is endowed with 3′–5′ exonuclease activities (Porrua & Libri, 2015).

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 with their metabolism. CUTs (cryptic unstable transcripts) have been first described based on their extreme instability (Wyers *et al*, 2005). These RNAs derive from transcription events terminated by the NNS pathway and are degraded to completion by the TRAMP-exosome pathway (Wyers *et al*, 2005; Arigo *et al*, 2006; Thiebaut *et al*, 2006). When NNS termination is defective, elongated forms of CUTs are produced that have been recently named NUTs (Nrd1 unterminated transcripts, Schulz *et al*, 2013). Some of the non-coding RNAs produced by pervasive transcription are sufficiently stable to be detected in wild-type cells (SUTs, stable unannotated transcripts, David *et al*, 2006) or are degraded in the cytoplasm by the nonsense-mediated decay (NMD) and Xrn1 pathways (XUTs, Xrn1-sensitive unstable transcripts, van Dijk *et al*, 2011; Malabat *et al*, 2015). Finally, some are only detected in particular physiological conditions (MUTs, meiotic unannotated transcripts, Lardenois *et al*, 2011).

We have recently described an additional pathway of transcription termination that depends on the general regulatory factor (GRF) Reb1. We have shown that the elongating polymerase pauses upstream of DNA-bound Reb1, which prompts its release by a mechanism that involves its ubiquitylation and presumably degradation (Colin *et al*, 2014). Insertion of a Reb1 binding site in a region of elongation is sufficient for termination, indicating that this "roadblock" (RB) pathway does not require additional sequence elements. 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; Colin *et al*, 2014).

Here we demonstrate that many additional DNA-binding complexes or factors can elicit RB termination and studied the overall impact of RB termination in the yeast genome. Using an improved crosslinking and cDNA analysis protocol (CRAC, Granneman *et al*, 2009), we sequenced nascent transcripts to generate the first high-resolution transcription maps upon depletion of two RB factors, and analyzed the genomewide impact of roadblock termination in wild-type cells or under conditions defective for NNS- or CPF-dependent termination. We directly demonstrate that RNAPII pausing depends on the roadblock factor and not on sequence elements or other events. We show that many RB events are associated with natural readthrough at canonical CPF or NNS terminators and that RB termination plays a general quality control role in limiting such pervasive transcription events. We studied the mutual relationships between RB termination and the other pathways and conclude that they are functionally independent and act redundantly to provide robust demarcation of adjacent transcription units. Finally, we show that roadblock termination also occurs around centromeres and tRNAs, which we suggest to be protected from the potentially negative interference of surrounding pervasive transcription events.

The faculty of DNA-associated factors to alter the processivity of elongation complexes, and the diversity of these factors, reveals a major role of RB termination in shaping the transcription landscape. This also underlies a large potential for regulation that likely extends to many organisms.

Results

In vivo selection reveals Rap1-dependent transcription termination

We have previously described a procedure to select transcription terminators from pools of naïve sequences (Porrua *et al*, 2012; Colin *et al*, 2014). Briefly, test sequences are inserted within a transcription unit driven by the tetracycline-repressible promoter (TET_P), roughly 200 nt downstream of the transcription start site. A second promoter, from the *GAL1* gene ($GAL1_P$), is inserted downstream and drives expression of a selectable marker, *CUP1*, the expression of which is required for yeast growth in copper-containing medium (Fig 1A). In the absence of a terminator in the test sequence, transcription driven from TET_P silences $GAL1_P$ by transcription interference and prevents *CUP1* expression, leading to copper sensitivity. When the test sequence induces termination, the *CUP1* gene is expressed and yeasts grow on copper-containing plates. 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 roadblock pathway (Porrua *et al*, 2012; Colin *et al*, 2014). We also selected sequences that do not belong to either class, some of which contain a motif resembling a Rap1 binding site (Figs 1A and EV1A). 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 (for a review, see Azad & Tomar, 2016). Rap1 is also strongly associated with the positioning and formation of nucleosome-free regions (NFR; Hartley & Madhani, 2009; Kubik *et al*, 2015 and references therein).

RNA species of a size compatible with termination occurring immediately upstream of the Rap1 site were observed when a selected terminator was present in the reporter construct (Figs 1B, lane 1 and EV1A). These transcripts are only detected when the Rap1 binding site is present (Fig 1B, lanes 3–4) and are strongly sensitive to degradation, as indicated by their marked steady-state increase in *rrp6*Δ or *rrp6*Δ*trf4*Δ mutants of the nuclear 3′–5′ RNA degradation pathway (Fig 1B and C). A major fraction of the transcripts detected in *rrp6*Δ cells are non-adenylated (Fig 1C, compare lanes 5 and 6), and a fraction is polyadenylated by Trf4 (compare lanes 5 and 8) and is strongly sensitive to exosomal degradation (Fig 1C, compare lanes 2–3 to 5–6). Non-adenylated RNAs are also subject to degradation by the exosome (compare lanes 3 and 6), but can be detected in the wild-type strain (Fig 1C, lanes 1,3), consistent with the notion that they represent the nascent RNAs associated with stalled polymerases.

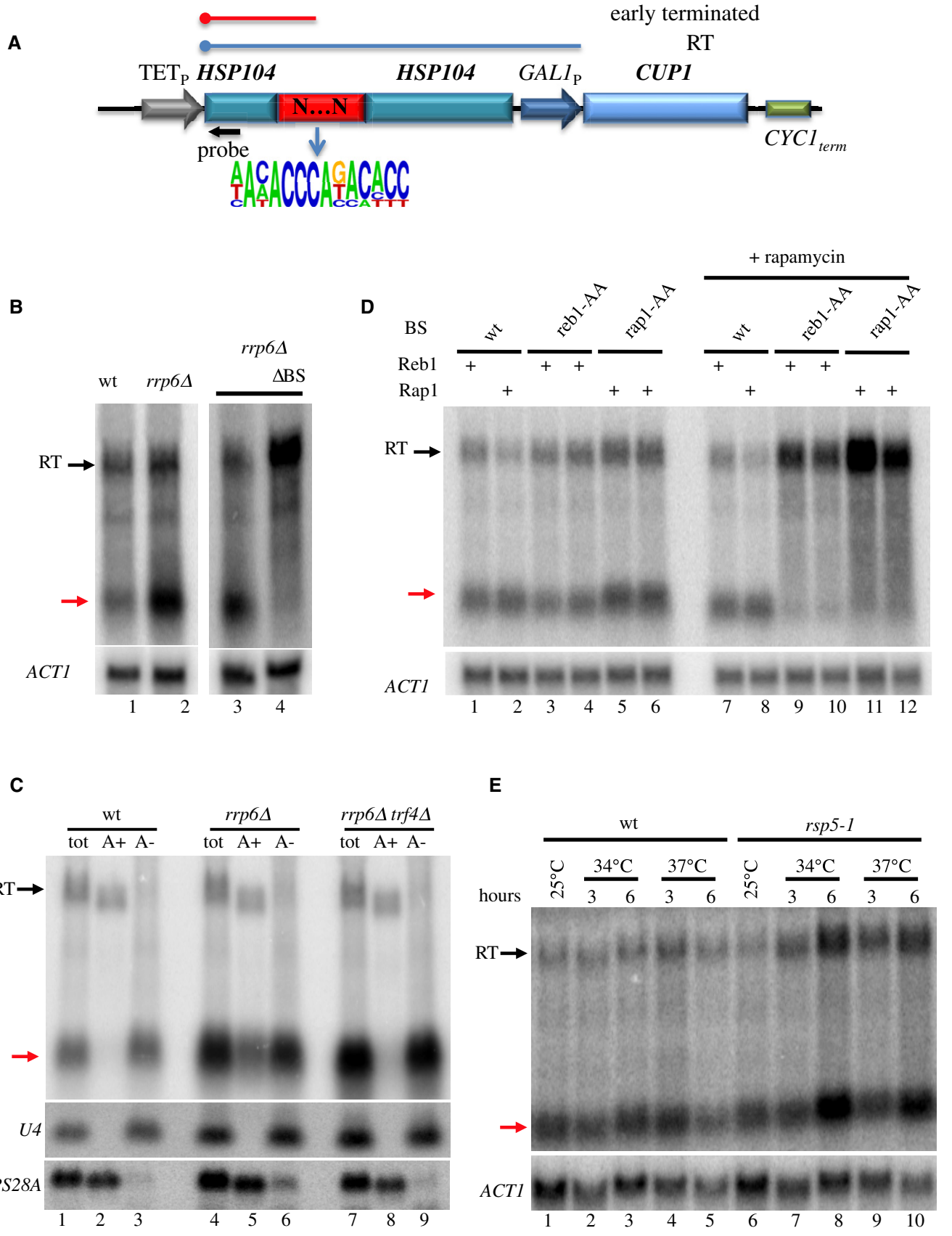

A
early terminated
RT
TETP HSP104
HSP104
GAL1P
CUP1
N...N
probe
CYC1term
B
wt
rrp6Δ
rrp6Δ
ΔBS
RT
ACT1
1 2 3 4
D
+ rapamycin
BS
wt
reb1-AA
rap1-AA
wt
reb1-AA
rap1-AA
Reb1
Rap1
RT
ACT1
1 2 3 4 5 6 7 8 9 10 11 12
C
wt
rrp6Δ
rrp6Δ trf4Δ
tot A+ A-
RT
U4
RPS28A
1 2 3 4 5 6 7 8 9
E
wt
rsp5-1
hours
25°C
34°C
37°C
3 6 3 6
RT
ACT1
1 2 3 4 5 6 7 8 9 10

Figure 1.

Figure 1. Analysis of the transcripts produced upon transcription termination induced by Rap1.

A Scheme of the reporter used for selecting terminators from naïve sequences. TET_P: doxycycline-repressible promoter; $GAL1_P$: *GAL1* promoter. The random sequence (120 nt, red box) was inserted within *HSP104* sequences upstream of $GAL1_P$. The transcripts produced in the presence (red) or absence (blue) of termination signals are indicated. The readthrough (RT) transcript terminates at a cryptic terminator within the *GAL1* promoter. A logo (http://weblogo.berkeley.edu/logo.cgi) derived from the putative Rap1 binding sites found in the selected terminators is shown. The approximate position of the oligonucleotide probe used for Northern blot analysis is indicated by a black arrow.

B Northern blot analysis of transcripts produced in the presence of a Rap1-dependent terminator in wt or *rrp6*Δ cells as indicated. ΔBS: RNAs derived from a construct containing a precise deletion of the Rap1 binding site (BS). A red arrow indicates the position of the short transcript produced at the Rap1 termination site. RT transcripts are indicated by a black arrow.

C Analysis of the polyadenylation status of transcripts derived from Rap1-dependent termination. Total, polyadenylated (A+, oligo dT-selected) and non-adenylated (A−, oligo dT-depleted) fractions are analyzed in the strains indicated. Rap1-terminated and RT transcripts indicated as in (B). U4snRNA and *RPS28A* RNAs are used as controls for non-adenylated and adenylated species, respectively.

D Northern blot analysis of strains containing reporters bearing a Rap1-dependent or a Reb1-dependent terminator (clone X3, Colin *et al*, 2014) as indicated. Reb1 (Reb1-AA) or Rap1 (Rap1-AA) anchor away strains were used to deplete either protein by the addition of rapamycin (lanes 7–12, two biological replicates). Red and black arrows indicate short and readthrough transcripts as in (B).

E Northern blot analysis of RNAs derived from a reporter containing a Rap1-dependent terminator in a wild-type (lanes 1–5) or a thermosensitive *rsp5-1* strain (lanes 6–10) grown at different temperatures as indicated. Note that the short RNA (red arrow) mainly represents nascent RNA associated with the roadblocked polymerase.

Source data are available online for this figure.

To demonstrate that Rap1, and not overlapping termination signals, is responsible for releasing RNAPII, we analyzed the RNAs produced upon nuclear depletion of Rap1 with the anchor away methodology (Haruki *et al*, 2008; Kubik *et al*, 2015). In the absence of Rap1, we observed the disappearance of the short RNA species, to the profit of a longer species earmarking termination at a downstream site (Fig 1D). As a control, we also show the effect of the nuclear depletion of Reb1 at a site of Reb1-dependent termination (Colin *et al*, 2014).

Finally, we have previously shown that release of the roadblocked polymerase from the DNA template occurs following its ubiquitylation that depends on the Rsp5 ubiquitin ligase. The failure to clear the roadblocked RNAPII results in increasing levels of both the nascent transcript (preferentially detected in a wild-type strain) and the RT transcripts, due to increased opportunity to overcome the RB when the polymerase is not released (Colin *et al*, 2014). Northern blot analysis confirmed such expected increase when the Rap1-roadblocked polymerase is less efficiently removed in a thermosensitive *rsp5-1* mutant strain (Fig 1E).

Together, these results demonstrate that Rap1 induces transcription termination by a roadblock mechanism.

Rap1-dependent transcription termination in the *Saccharomyces cerevisiae* genome

To assess the natural extent of Rap1-dependent RB termination, we analyzed the occurrence of RNAPII pausing immediately upstream of Rap1 sites, which is a hallmark of roadblock termination (Colin *et al*, 2014). We profiled RNAPII occupancy in a wild-type and a Rap1 anchor away (Rap1-AA) strain by an improved version of the crosslinking and cDNA analysis protocol (CRAC, Granneman *et al*, 2009). This approach allows assessing the position of the polymerase at the nucleotide resolution level by sequencing the nascent transcript associated with the largest subunit of the enzyme after *in vivo* UV crosslinking (RNAPII CRAC, Milligan *et al*, 2016). The analysis indeed detects nascent transcripts, as demonstrated by the coverage of intronic regions in the RNAPII CRAC dataset but not in the sequencing of mature, total RNAs (Figs 2C and EV1B).

Notable examples of Rap1-dependent roadblock sites are shown in Fig 2. CRAC analysis revealed a marked RNAPII peak upstream of sites of Rap1 binding (Rhee & Pugh, 2011; Knight *et al*, 2014) at the *HYP2, RPL11B,* and *RPS24A* loci.

At the *HYP2* locus (Fig 2A), a Rap1-dependent RB terminates transcription of an upstream, non-annotated transcription unit (dubbed *uHYP2*), leading to the production of a cryptic transcript as revealed by SAGE analysis (Fig 2A, Neil *et al*, 2009). At the *RPL11B* and *RPS24A* loci (Fig 2B and C), roadblocked polymerases most likely derive from transcription events reading through the upstream terminator (see below). Nuclear depletion of Rap1 by the addition of rapamycin led to the significant decrease of the RNAPII peak and to the spreading of a readthrough signal downstream of the RB site (Fig 2, insets). Rap1-dependent termination could be confirmed by inserting a short region only containing the two Rap1 sites present at the *HYP2* locus in the heterologous context of our reporter system (Appendix Fig S1A).

To extend these findings genomewide, we generated aggregate plots by profiling the average distribution of the RNAPII CRAC signal around aligned sites of Rap1 occupancy (Fig 3). A major peak of average RNAPII occupancy is present downstream of the aligned occupancy sites, due to general presence of genes regulated by Rap1 (Fig 3A, left). Importantly, however, a significant roadblock peak was observed upstream of Rap1 binding, which strongly decreased upon Rap1 depletion (compare the red and blue traces).

Similar RNAPII CRAC analyses were also performed upon depletion of Reb1. Peaks of RNAPII pausing were readily observed at individual sites of Reb1 occupancy that significantly decreased upon Reb1 depletion (Appendix Fig S1B and data not shown). Aggregate plots (Fig 3A, right) show, as for Rap1, a major peak of transcription initiation mainly due to Reb1-regulated genes, and a prominent RB peak that is Reb1-dependent.

The detection of a RB at Rap1 sites was not due to crosslinking of Rap1 to the DNA, because it could be observed using techniques that do not rely on crosslinking (NET-seq, Churchman & Weissman, 2011; Appendix Fig S2A and B) or that rely on the sole crosslinking of the RNA to proteins (PAR-CLIP, Schaughency *et al*, 2014; Appendix Fig S2B and data not shown). Finally, the occurrence of transcription termination at Rap1 and Reb1 occupancy sites is consistent with a peak of RNA 3′ ends that coincides with the site of RB and is generally more prominent in a degradation defective *rrp6*Δ strain (Appendix Fig S2C and D).

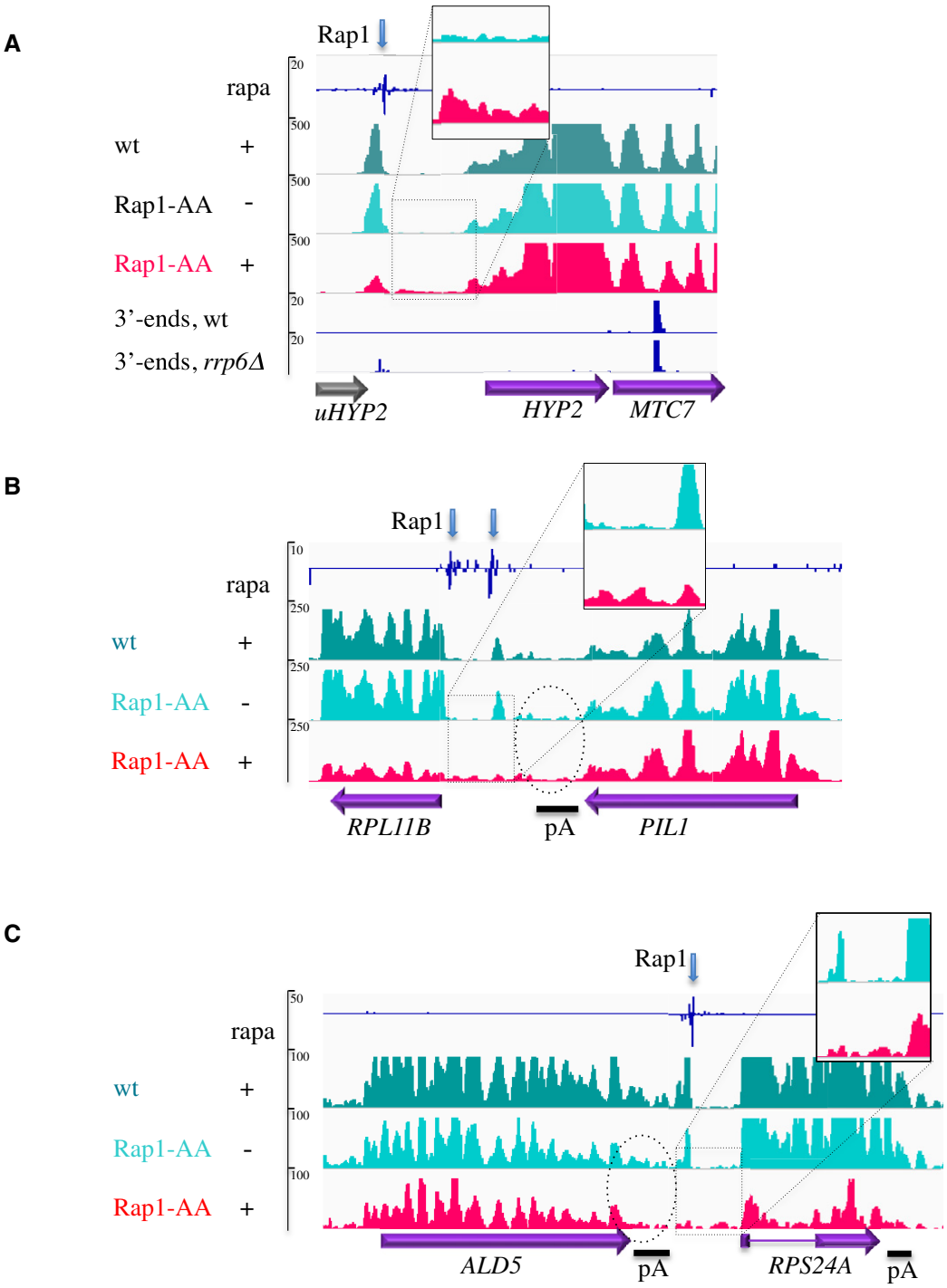


Figure 2. RNAPII occupancy at sites of Rap1 roadblock detected by CRAC analysis.

A RNAPII CRAC profile at a site of roadblock upstream of *HYP2* (only the signals on the strand of the annotated features are shown). A peak of CRAC RNAPII signal is visible upstream the site of Rap1 occupancy (blue arrow, ChIP exo data, Rhee & Pugh, 2011) in a wild-type strain in the presence of rapamycin (dark green track) or Rap1-AA in the absence of the drug (light green track). The roadblock peak is markedly diminished when Rap1 is depleted from the nucleus by the addition of rapamycin to Rap1-AA cells (red track). Transcription termination at the RB site is accompanied by the production of a non-annotated cryptic transcript (*uHYP2*, gray arrow) with a predominant 3′ end located 13 nt upstream of the Rap1 site (data from Roy *et al*, 2016). The maximum value of the RNAPII peak is 26 nt upstream of the sequence of the Rap1 site. The position of multiple polyadenylation (pA) sites for *HYP2* as defined by 3′-T-Fill analysis (Wilkening *et al*, 2013) is indicated. Note the occurrence of transcriptional readthrough after the roadblock when Rap1 is depleted (inset).

B, C Same as in (A), with Rap1 sites located between two genes arranged in tandem. The dotted oval underscores the level of polymerase occupancy between the CPF terminator and the roadblock, which is not affected by depletion of the roadblock factor. The maxima of the RNAPII peak are located 33 nt (*PIL1*) and 15 nt (*ALD5*) upstream of the first Rap1 binding site.

Overall, these results demonstrate the widespread occurrence of Rap1- and Reb1-dependent, roadblock transcription termination in *S. cerevisiae*.

Roadblock termination limits widespread readthrough transcription in the *Saccharomyces cerevisiae* genome

Many Reb1 and Rap1 sites are located in intergenic regions, frequently downstream of genes. Based on a few model cases, we have previously proposed that roadblock termination might function to limit transcription reading through canonical, CPF-dependent terminators (Colin *et al*, 2014), but neither the general validity of this concept, nor the generalized occurrence of readthrough transcription could be demonstrated.

If polymerases fail to terminate at canonical sites with a significant frequency, they are expected to accumulate at sites of Reb1 and Rap1 binding downstream of genes, where they should be easily detected because of the roadblock.

To address this possibility, we restricted our RNAPII CRAC metasite analyses to Reb1 and Rap1 occupancy sites located within 300 nt downstream of mRNA-coding genes. In these conditions, only polymerases escaping CPF-dependent termination, if any, should contribute to the metaprofile upstream of the roadblock. As shown in Fig 3B, polymerases accumulate at Rap1 and Reb1 sites downstream of canonical CPF terminators in the wild-type strain strongly suggesting the existence of a constitutive transcriptional readthrough. To substantiate this notion, we also performed a parallel RNAPII CRAC analysis using a thermosensitive *rna15-2* allele, which impairs CPF termination. In these conditions, we observed a clear increase in the roadblock peak relative to what observed in wt cells, supporting the notion that the flux that aliments roadblocked polymerases originates from upstream transcription units and increases when upstream termination is defective (Fig 3B, compare green and blue traces). As a control, we profiled RNAPII distribution at the same set of CPF-dependent features upon nuclear depletion of Nrd1 (Schaughency *et al*, 2014), 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 roadblock peak (Fig EV2A and data not shown). Manual inspection of a significant number of these locations ruled out the existence of intergenic transcription initiation based on the recent published repertoire of RNAPII transcripts 5′ ends (data not shown; see also Fig EV2B; Malabat *et al*, 2015).

Similarly to Reb1 and Rap1, Abf1 belongs to the class of general regulatory factors and contains a myb-like DNA-binding domain (Fermi *et al*, 2017 and references therein). We profiled the RNAPII CRAC signal around sites of Abf1 occupancy downstream of CPF terminators. Although less prominent, a RB peak was observed, which increased, as for Rap1 and Reb1, when termination was impaired in an *rna15-2* mutant (Fig EV2C). A metaprofile analysis using a larger set of Abf1 occupancy sites is shown below in Fig EV3.

Transcriptional readthrough was not restricted to sites containing a downstream Rap1, Reb1, or Abf1 roadblock but could be consistently revealed by the significant detection of intergenic transcription downstream of genes in the absence of dedicated initiation sites. A few representative snapshots are shown in Fig EV2B, in which the levels of readthrough transcription are comparable to the levels of transcription of the downstream gene.

Overall, these results demonstrate the widespread occurrence of transcription readthrough at CPF terminators in strains that are proficient for transcription termination. Such pervasive readthrough events are restricted, to a significant extent, by downstream roadblock termination.

Roadblock termination and the CPF pathway function independently

RNA polymerase II pausing is generally considered to promote termination by favoring "chasing" of the polymerase by Rat1 at CPF-dependent genes. It could be conceived that RB pausing functions as part of the CPF pathway for the efficient release of the polymerase. In this perspective, removing the roadblock should significantly affect the overall efficiency of termination. To address this possibility, we assessed whether termination failure could be observed at CPF terminators when the downstream Reb1- or Rap1-dependent roadblocks were removed by nuclear depletion of either factor. Two examples of CPF-dependent genes with a downstream roadblock are shown in Fig 2. In both cases, transcription termination occurs efficiently at the CPF sites even in the absence of the roadblock as witnessed by the similar decrease in the RNAPII signal at and downstream of the termination region (Fig 2B and C, dotted oval).

To generalize these observations, we compared the RNAPII metaprofile in regions of CPF termination upstream of a Rap1 binding site in the presence and absence of the roadblock factor (Fig 4). The precise location of transcription termination for each gene is not known, but we reproducibly observed a decrease in the RNAPII CRAC signal in the region around the sites of poly(A) addition (Fig 2 and data not shown). This early decrease in the RNAPII signal was to some extent unexpected, but was also observed using NetSeq (data from Harlen *et al*, 2016; not shown). It might be due to termination at cryptic or earlier sites of poly(A) addition or to the higher speed of the polymerase in this region. Irrespective of its possible components, this signal was clearly different in *rna15-2* cells (see below, Fig 4B), which are termination impaired at the non-permissive temperature, suggesting that it is linked to the occurrence of termination. We therefore anchored the alignment to the strongest site of poly(A) addition for each gene (Pelechano *et al*, 2014) and focused our analysis on the region of early termination, to avoid interference with the RNAPII signals at the roadblock. As shown in Fig 4A, a progressive decrease in the average RNAPII signal was observed in wild-type cells in this region, consistent with the occurrence of termination. As a control, transcription readthrough was clearly observed when termination was impaired in *rna15-2* cells at the non-permissive temperature (Fig 4B, compare green and blue traces). Importantly, however, upon depletion of Rap1, CPF-dependent termination occurred efficiently, as witnessed by the identical decline in the RNAPII CRAC signal in the termination region (Fig 4A, compare red and blue traces). Similar results were obtained for the set of CPF-dependent genes upstream of a Reb1-dependent roadblock (data not shown). To quantitate these results, we calculated the fractional level of readthrough for each CPF-dependent gene upstream of a Rap1-dependent roadblock by dividing the density of reads in the termination region by the density in the gene body (Fig 4C). The distribution of the values obtained is strongly affected by the *rna15-2* mutation, as expected for a *bona fide*

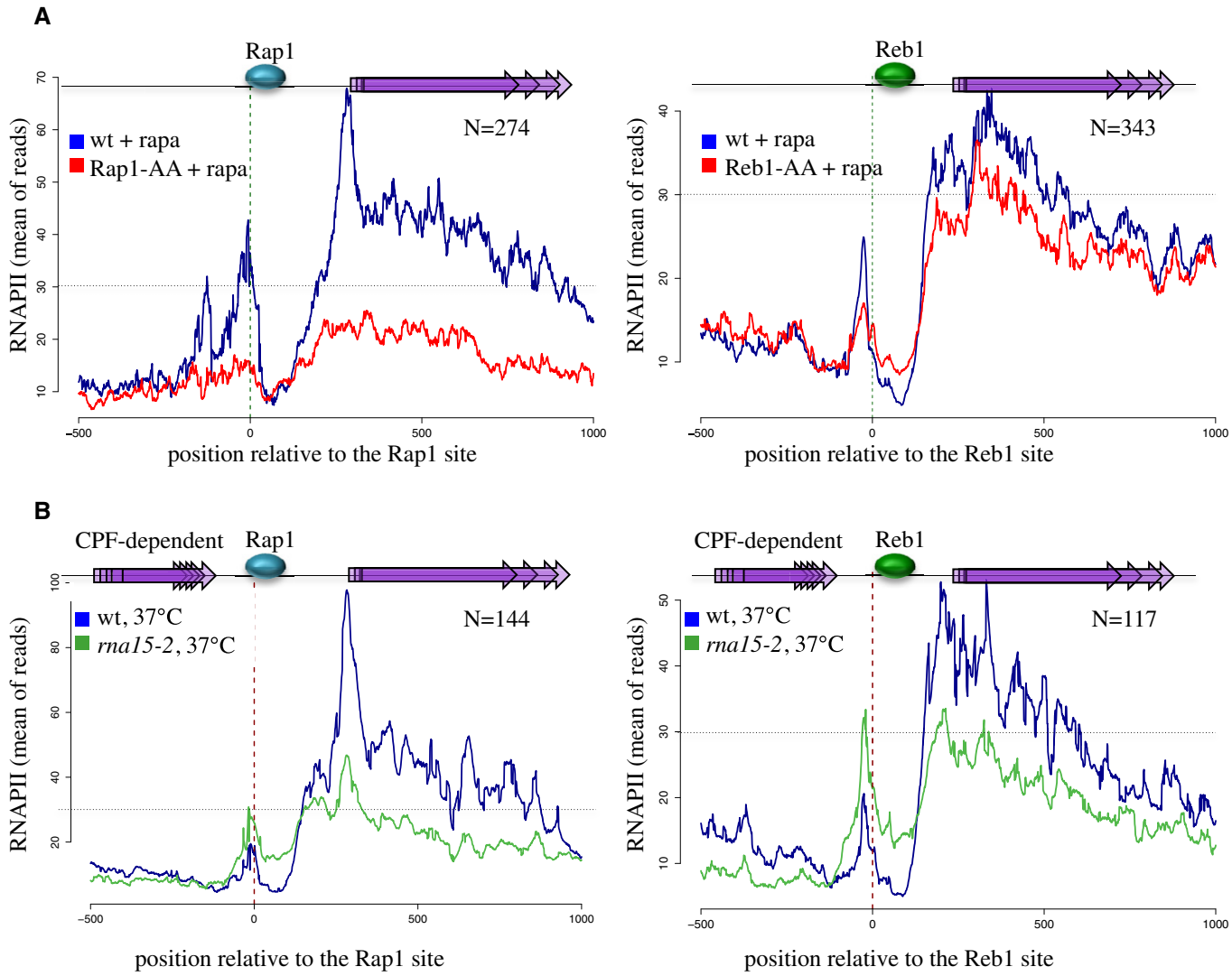


Figure 3. Metasite analysis of roadblock termination at Rap1 and Reb1 sites.

A Average RNAPII CRAC profile at genomic regions aligned on Rap1 (left panel) or Reb1 (right panel) occupancy sites, in the presence (blue) or absence (red) of the roadblock factor. In both cases, the latter was depleted from the nucleus by the addition of rapamycin. Overlapping purple arrows represent features transcribed downstream of Rap1 or Reb1 occupancy sites. Note that the two panels have a different *y*-axis scale due to the average higher expression of Rap1-dependent genes; a dotted horizontal line marks the same average occupancy for comparison.

B Same as in (A), using only Rap1 or Reb1 sites located within 300 nt downstream of genes terminated by the CPF pathway. The RNAPII average profile was determined for the wild-type strain (blue) or an *rna15-2* (green) strain at the non-permissive temperature of 37°C for the mutant; data from the same cells at permissive temperature have not been plotted, but are available.

Data information: For all panels, the number of sites used is indicated. Rap1 and Reb1 sites used in these analyses are listed in Dataset EV1.

termination defect ($P = 10^{-5}$), but not by the absence of Rap1 ($P = 0.4$), demonstrating that removing the roadblock does not significantly impact CPF termination.

Relationships between RB- and NNS-dependent termination

While this work was in progress, another study (Roy *et al*, 2016) proposed that roadblock- and NNS-dependent termination are functionally linked, notably suggesting that (i) roadblocked polymerases are released by the NNS pathway and (ii) the roadblock is part of the mechanism of snoRNA termination. We revisited these important questions using our high-resolution RNAPII CRAC in cells defective for the CPF, NNS, and roadblock pathways.

Roadblock peaks have been shown to increase in strains defective for NNS termination, which was interpreted as evidence that roadblocked polymerases are not efficiently cleared when NNS termination is impaired (Roy *et al*, 2016). Alternatively, it is possible that this increase in RB peaks is due to the accumulation of polymerases failing to terminate at upstream NNS-dependent terminators. Consistent with this notion, we observed increased RB peaks only at sites downstream of NNS terminators when the NNS complex is defective (Fig 5 and data not shown). When the RB site

follows a CPF terminator, depletion of Nrd1, Sen1, or mutation of Nab3 does not affect the levels of roadblocked polymerases. This is illustrated at the *PIL1* and *ALD5* loci (Appendix Fig S3A), and more generally in the aggregate RNAPII CRAC profile at Rap1 RB sites downstream of CPF terminators (Fig EV2A). Here, the depletion of Nrd1 poorly affects the signal at the Rap1 roadblock, which is, on the contrary, strongly increased upon impairment of CPF termination in the *rna15-2* mutant (Fig 3B and Appendix Fig S3A). These data demonstrate that the NNS pathway is not generally required for the clearance of roadblocked polymerases.

We also addressed the converse possibility, that is, that the RB pathway could be required for termination of snoRNAs (Roy *et al*, 2016). We analyzed the polymerase profile around four snoRNAs for which a Reb1 (*SNR161, SNR8, SNR48*)- or Rap1-dependent (*SNR39B)* roadblock peak of variable intensity was detected in the termination region (Fig 5). We compared the distribution of polymerases at these NNS-dependent targets under conditions of defective RB termination by depleting either one of the RB factors. As a control, we generated RNAPII CRAC data upon depletion of Nrd1 with the auxin degron method (Nishimura *et al*, 2009). Depletion of Nrd1 led to transcription readthrough at the NNS-dependent terminator as expected, which fed the flow of polymerases accumulating at the downstream roadblock peak (Fig 5A–D, compare red and blue tracks in the insets, red arrows; see also Fig 5E, left scheme). This was clearly visible at the *SNR8* and *SNR48* loci, where the roadblock is slightly more distal (Fig 5, panels A and B), but also observed at *SNR161* and *SNR39B* where the readthrough signal merges to some extent with the roadblock signal (Fig 5C and D).

Upon depletion of the roadblock factor (Rap1 or Reb1), we did not observe alterations in termination at the primary NNS site, which occurred with similar overall efficiency as in the presence of the RB (Fig 5A–D, note that the RB-less tracks, pink/red, are always beneath the wt tracks, light blue). A small readthrough was only detected downstream of the RB site (Fig 5, blue arrows), due to the release of polymerases that had accumulated at the roadblock (see Fig 5E, right scheme). These results strongly suggest that the RB is not required for NNS-dependent termination at these sites.

The small readthrough at the RB in the absence of Rap1 or Reb1 leads to the production of longer transcripts that might have diverse fates and stability, depending on many factors including their sequence and the nature of downstream termination. We analyzed the levels of these RNAs by RNAseq in the presence or absence of the RB. To visualize the primary product of termination that is trimmed to the mature snoRNA by the nuclear exosome, we performed this analysis in an *rrp6*Δ strain. As shown in Fig 5A–D, variable levels of readthrough transcripts accumulated at three of the four snoRNAs studied in the absence of the RB. The strongest accumulation was observed at the *SNR39B* site and intermediate levels at *SNR161* and *SNR8,* but in all cases the transcript levels hardly mirrored the levels of polymerases reading through the site of RB detected by CRAC. This indicates that the abundance of these RNAs is mainly dictated by their stability and not by the levels of readthrough transcription.

Together with the results shown in the previous section, our data strongly support the notion that the roadblock pathway functions as a fail-safe mechanism to neutralize natural readthrough transcription at both the CPF- and NNS-dependent canonical terminators.

Functional importance of fail-safe transcription termination

As shown in Fig 2, depletion of Rap1 strongly affects transcription of *RPL11B* and *RPS24A*. These genes might be downregulated either because the absence of the roadblock exposes their promoters to transcriptional interference or because they require Rap1 for transcriptional activation. To distinguish between these non-exclusive possibilities, we investigated whether the RB alone could be sufficient to restore, at least partially, their expression. 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, but supports roadblock termination, as verified by RT–qPCR upstream of *HYP2* (Appendix Fig S3B). As a control, we used strains containing the wild-type Rap1 or an empty plasmid and sequenced the RNAs produced in these cells. Because expression of Rap1-DBD alone affects growth in a dominant-negative manner, we could not perform reliable CRAC experiments in these conditions, but sequenced the transcriptome at two different time points after Rap1 depletion. Consistent with the RNAPII CRAC data, expression of *RPL11B* and *RPS24A* RNAs was markedly affected by the depletion of endogenous Rap1 and restored by the concomitant expression of wt Rap1 (Fig 6A and B, compare red and blue tracks). Importantly, expression of the DNA-binding domain alone of Rap1 is sufficient to restore *RPL11B* and *RPS24A* RNAs to wild-type levels (Fig 6A and B, purple tracks). This is not due to Rap1-DBD retaining a general activation function as demonstrated by the failure of the latter to restore expression of genuine Rap1 targets sites such as *RPS0A* (Fig 6C), *RPL29,* or *MF(ALPHA)1* (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 roadblock factors.

Extent of roadblock termination in the *Saccharomyces cerevisiae* genome

In light of the results shown here on Rap1 and Reb1, we assessed more generally the occurrence of roadblock termination at sites of occupancy for DNA-binding proteins or complexes. We first used published data on the genomewide distribution of transcription factors (Harbison *et al*, 2004) and profiled RNAPII occupancy at genomic regions aligned on sites of binding as defined by MacIsaac *et al* (2006). We found evidence for RB termination at many such sites, some of which are shown in Fig EV3. These profiles are indicative of roadblock occurring at a variable distance upstream of the protein-binding site, likely reflecting the topology of the collision between RNAPII and the DNA-bound factor or complex of factors.

We also observed prominent levels of RNAPII roadblocks at centromeres and tRNA genes. In *S. cerevisiae*, centromeres are defined by a set of short, well-conserved sequence elements located in a 125-nt region. These sequences, CDEI, CDEII, and CDEIII (Fig 7), are specifically occupied by DNA-binding factors that are part of the kinetochore (for a review, see Westermann *et al*, 2007; Biggins, 2013). The 8-bp CDEI is directly recognized by Cbf1, a DNA-binding factor that is also a transcriptional activator. CDEIII (26 nt) is instead recognized by CBF3, a complex of four proteins, while the CDEII sequence (78–86 nt) is wrapped around a specific centromeric nucleosome (CENP-A), containing a histone H3 variant, Cse4. We

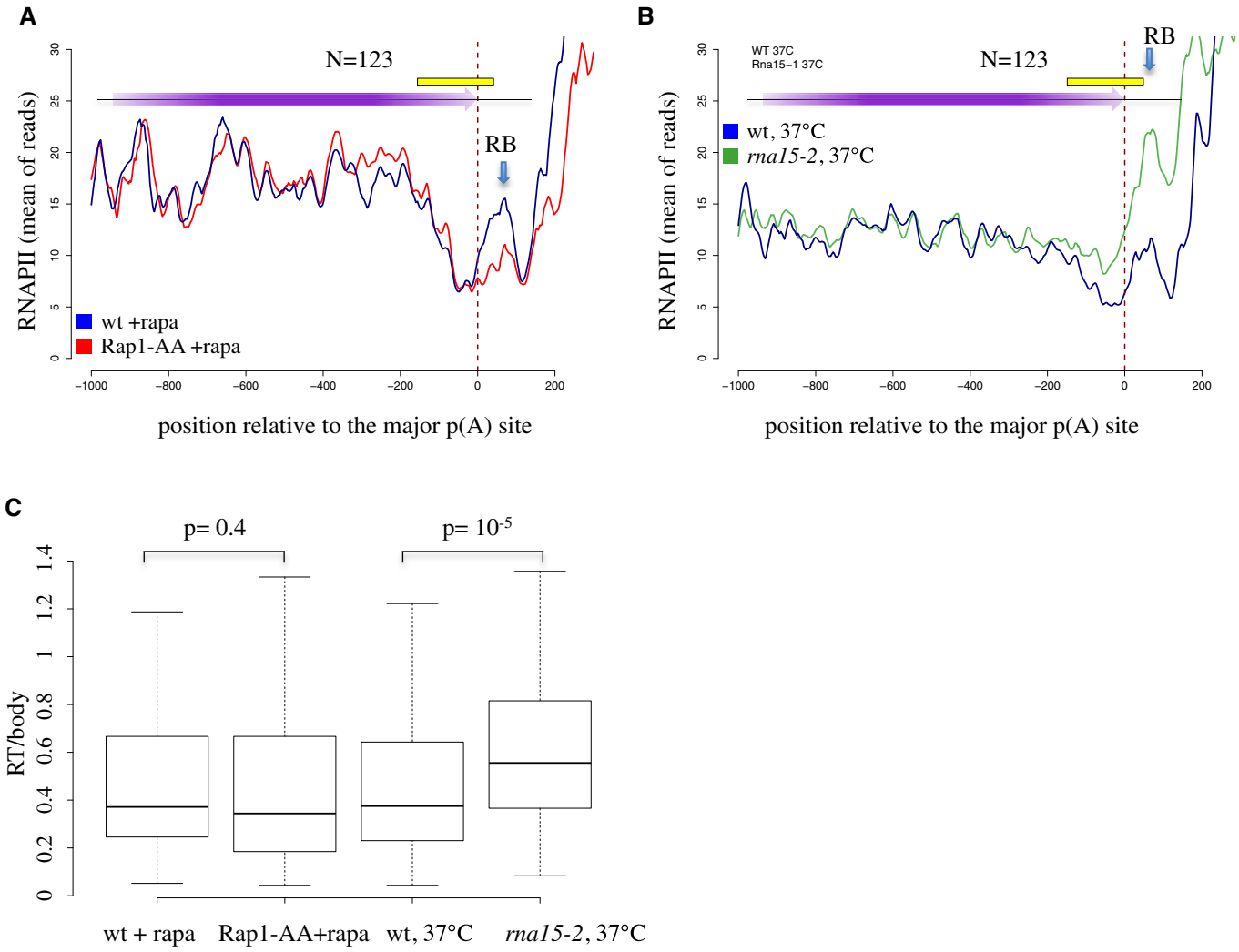


Figure 4. Aggregate RNAPII profile at genes followed by a Rap1 site aligned on the poly(A) site.

A Genes terminated by the CPF pathway and followed by a Rap1 site were aligned on the major poly(A) site and the average RNAPII CRAC signal was plotted for the wild-type (blue) or Rap1-AA (red) strain in the presence of rapamycin to induce the nuclear depletion of Rap1 in Rap1-AA. The termination region, defined by the region displaying an average decrease of the RNAPII signal, is indicated by a yellow rectangle. Note that the signal in this region is not affected by the depletion of the RB. The roadblock peak (RB), indicated by a blue arrow, is broader and smaller in these plots because genomic regions are not precisely aligned on the RB site.

B As in (A), the average RNAPII profile in wild-type (blue) and *rna15-2* (green) cells at the non-permissive temperature was plotted to highlight a *bona fide* termination defect. To visually appreciate the occurrence of readthrough, we normalized the read counts so that the average RNAPII CRAC signals in gene bodies are comparable in wt and *rna15-2* cells.

C Boxplots representing the distribution of the ratios between the density of reads in the 100 nt immediately preceding the major poly(A) site (RT) and the density of reads in each ORF (body) in the indicated strains and conditions. The plots were generated by the standard R boxplot function. The central line represent the 50th percentile. The top and bottom of the box represent the 75th and 25th percentile respectively. The bottom whisker is the lowest value still within 1.5 Interquartile range (IQR); the top whisker is the highest values still within 1.5 IQR. The number of sites used for the analysis corresponds to the rap1 sites for which an experimentally defined polyadenylation site could be found within 300 bp upstream of the site (123). A two-sided *t*-test has been used to assess statistical significance. Sites used in these analyses are listed in Dataset EV1.

analyzed the distribution of polymerase around centromeres using both PAR-CLIP (Schaughency *et al*, 2014) and our RNAPII CRAC data. A marked peak of localized RNAPII pausing was clearly observed at individual centromeres (Figs 7B and EV4A and B), the average position of which was roughly 25 nt upstream of CDEI as shown in the aggregate plot (Fig 7A), strongly suggesting that Cbf1 induces roadblock termination, at least in the context of centromeres.

A RB peak was also observed around CDEIII and upstream of CDEII, as shown in the RNAPII CRAC metaprofile and at individual centromeres (Figs 7B and EV4). In many cases, the RB was more prominently observed when incoming transcription was increased by affecting termination of convergent genes in *rna15-2* cells (Figs 7B and EV4C–E, *rna15-2* tracks). Termination by the RB pathway occurred in these regions, as witnessed by the presence of RNA 3′ ends peaks that overlap the peaks of pausing and that often represent unstable transcripts (compare the wt and the *rrp6*Δ profile, Fig EV5A and B). Interestingly, two peaks of termination can be observed around CDEIII (Fig EV5B), one within and another

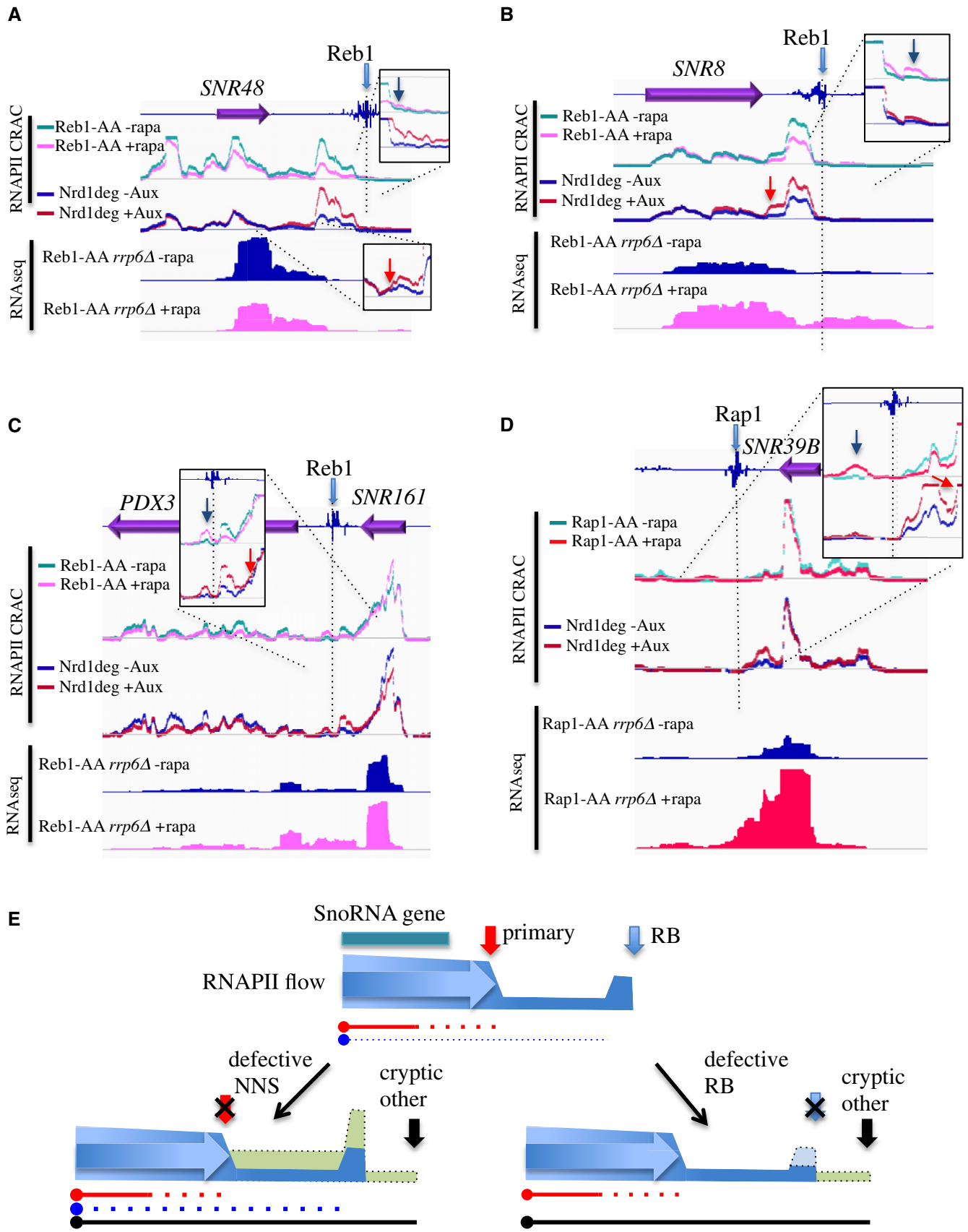


Figure 5.

Figure 5. Analysis of the impact of RNAPII roadblocks in termination of snoRNAs.

A–D RNAPII CRAC and RNAseq profiles at the indicated genomic loci and conditions (only the signals on the strand of the annotated features are shown). The position of the Reb1 or Rap1 occupancy (Rhee & Pugh, 2011) is indicated (light blue filled arrow). The occupancy profile in the presence or absence of the RB factor or Nrd1 has been overlapped for ease of comparison. The regions of the readthrough after the roadblock (dark blue arrow) or after the NNS terminator (red arrow) have been enlarged in the insets.

E Model of primary (NNS) and secondary (RB) termination at snoRNAs that contain a downstream RB site. The flow of RNAPII is indicated in blue, and the internal arrow indicates the direction of transcription. The sites of NNS and RB termination are indicated, respectively, by red and blue arrows; a site of cryptic or alternative (e.g., CPF-dependent) termination downstream of the RB is indicated by a black arrow. A low level of natural readthrough at the primary site is indicated by a low schematic flow of polymerases (blue) between the NNS and RB sites, which feeds the RB peak. Under defective NNS termination, this readthrough flux increases, together with the RB (left scheme, dotted line, light green). When the RB is affected (right scheme), only the readthrough due to unblocked polymerases (dotted line, light green) downstream of the RB site is observed, terminating at downstream sites (black arrows). The transcripts produced in the different conditions are indicated by plain or dotted lines, which roughly represent the stability and steady-state levels of the different species. The colors represent the kind of termination (NNS, RB, or cryptic) that leads to the production of a given species.

upstream of it, indicating heterogeneity in the position of the RB around (Fig 7A).

In apparent contrast with published data (Ohkuni & Kitagawa, 2011, 2012), no evidence of transcription within CDEII was observed, which we find to be virtually free of polymerases. Note that this is unlikely due to failure from mapping A-T-rich reads to this region, because reads with virtually identical sequences (i.e., containing one or two mismatches) could be efficiently mapped to other regions in the genome (data not shown).

Transcription of tRNA genes depends on internal promoters, which harbor sequences (A and B boxes) that are bound by the TFIIIC hexameric complex (Arimbasseri & Maraia, 2016). TFIIIC covers the whole tRNA sequence and interacts with TFIIIB, composed of three subunits, which binds at position −60 relative to the transcription start site (Nagarajavel *et al*, 2013). The aggregate profile of RNAPII distribution around tRNA genes is presented in Fig 7C and one representative example in Fig 7D, together with the mapped "bootprints" of TFIIIB and TFIIIC (Nagarajavel *et al*, 2013). A prominent peak of RNAPII accumulation was observed at position −75 relative to the start site consistent with a roadblock induced by TFIIIB bound at position −60. Interestingly, a major roadblock was also observed for RNAPII transcription running antisense to tRNA genes, peaking roughly 50 nt upstream of the aligned U-rich tract that defines the RNAPIII termination signal. Evidence for the production of unstable transcripts for termination occurring upstream of tRNA start sites and for antisense transcription downstream of tRNA terminators is provided by the distribution of stable and unstable RNA 3′ ends around these features (Fig EV5C and D).

Evidence for RNAPII roadblocks was also observed around the gene coding for the ribosomal 5S RNA subunit, which is an RNAPIII gene with a structure similar to that of tRNA genes (Fig 7E).

TFIIIC was also found to bind at locations distinct from tRNA genes, in the absence of TFIIIB (ETC, extra TFIIIC sites, Roberts *et al*, 2003; Moqtaderi & Struhl, 2004; Nagarajavel *et al*, 2013). We could not find evidence of transcriptional roadblock at ETC sites, suggesting that the sole binding of TFIIIC is not sufficient to prevent RNAPII elongation (data not shown). Together, these data illustrate the genomewide extension of roadblock termination and underscore its large potential for modeling the yeast transcriptional landscape.

Discussion

In a previous study, we have demonstrated that transcription termination occurs when the RNAPII encounters the factor Reb1 bound to the DNA. Here we generated high-resolution genomewide RNAPII transcription maps data under conditions of defective RB, CPF, or NNS termination to study the overall impact of RB termination on the yeast genome, and the functional relationships with the other pathways.

We provide evidence that natural readthrough at canonical CPF and NNS terminators constitutes an additional and functionally significant source of pervasive transcription in *S. cerevisiae*. We demonstrate that the canonical pathways and RB termination function independently from each other but act redundantly at the end of transcription units, limiting pervasive readthrough and favoring insulation of transcription events. Finally, we extend the repertoire of roadblocking factors, which we propose to play major roles in determining the distribution of transcription events.

Rap1 is a roadblock termination factor

We demonstrated that Rap1, a DNA-binding factor that has roles in transcription activation, gene silencing, and telomere homeostasis (Azad & Tomar, 2016), is also a roadblock termination factor. An earlier study showed that the fortuitous introduction of a Rap1 site in a Ty1 retrotransposon leads to RNAPII stalling and repression of gene expression (Yarrington *et al*, 2012). Based on the analysis of the RNA produced, which was reported to be non-adenylated and insensitive to exosome degradation, it was concluded that termination of transcription does not occur in this system. In contrast to this early study, we show that roadblock termination occurs at Rap1 binding sites, leading to the production of RNAs that can be polyadenylated by Trf4 and are degraded for a large part by the nuclear exosome. Importantly, nuclear depletion of Rap1 prevents termination, indicating that the protein—and not the presence of termination signals that might overlap its binding site—is essential for the release of the polymerase. Failure to detect the nuclear degradation of the adenylated and non-adenylated RNAs for technical reasons in the study by Yarrington *et al* (2012) might account for the discrepancies; alternatively, release of the polymerase might not occur in the Ty1 retrotransposon model for unknown reasons.

The mechanism of roadblock termination

Similar to what previously shown for Reb1 (Colin *et al*, 2014), release of the polymerase paused upstream of the roadblock depends on its ubiquitylation by Rsp5 and possibly its degradation. Thus, this pathway is not restricted to Reb1-dependent termination but presumably extends to all cases of roadblock, and possibly of

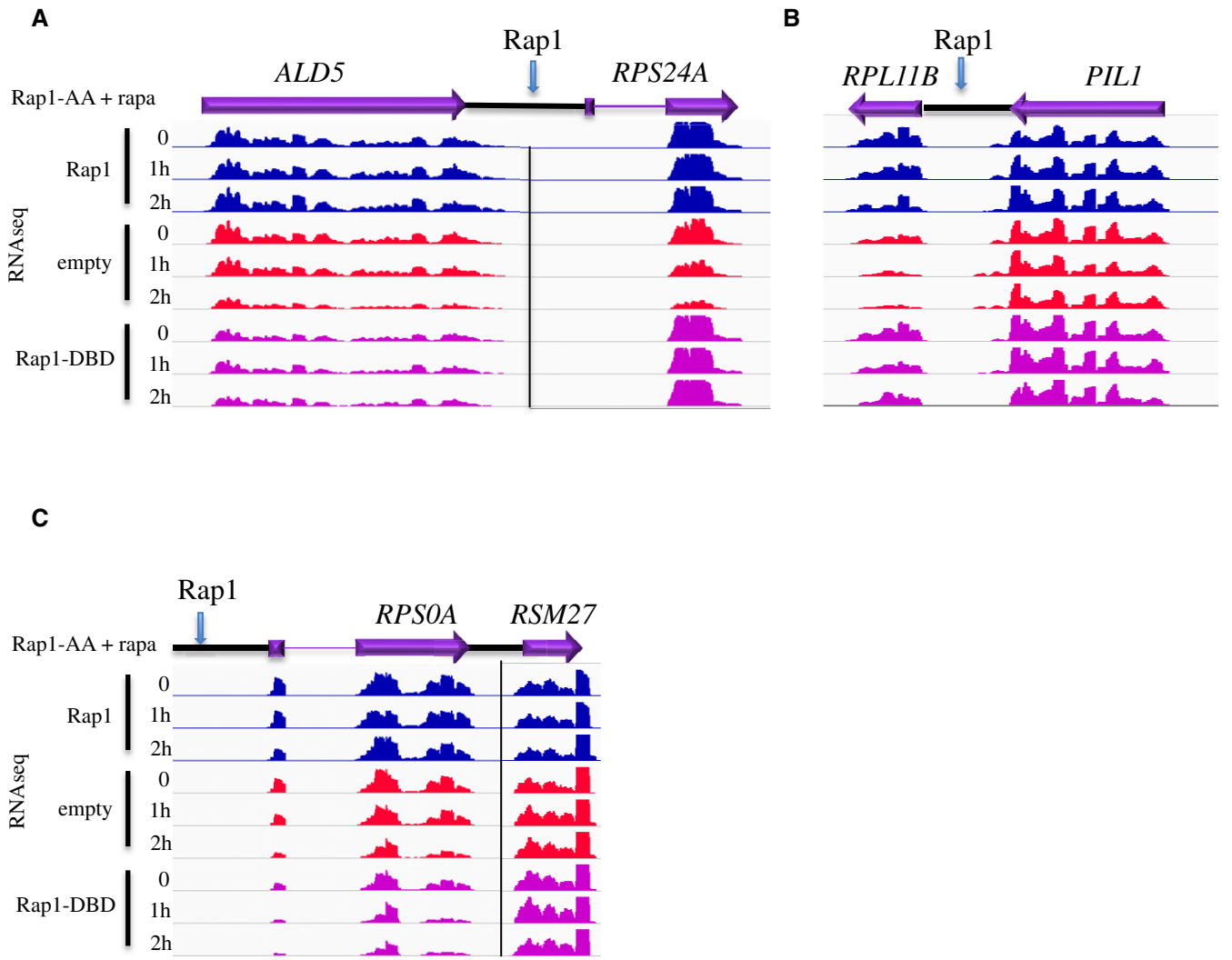


Figure 6. Expression of the Rap1 DNA-binding domain restores expression of genes containing an upstream roadblock.

A–C RNAseq profiles at two tandem features containing an upstream Rap1 roadblock site (*RPS24A* and *RPL11B*, A and B). A feature (RPS0A) that depends on Rap1 for transcription activation but for which no upstream RB can be detected is shown in (C) as a control. Only the signals on the strand of the annotated features are shown. Wild-type Rap1, the Rap1 DNA-binding domain (Rap1-DBD), or an empty plasmid was expressed in the Rap1-AA strain and the endogenous protein was depleted from the nucleus upon addition of rapamycin for the times indicated. RNAseq tracks are shown for the target genes and for neighboring genes as a control. Adjacent loci have been separated by vertical lines when the two features have very different expression levels and different scales have been used.

most RNAPII pausing that cannot be resolved in a more "conservative" manner (Wilson *et al*, 2013). We generally observed very sharp peaks of stalling at the roadblock sites, which is compatible with one or two polymerases on average roadblocked at a time and indicates that, at steady state, the clearance due to the Rsp5 pathway must be as efficient as the feeding of the peak by incoming polymerases.

It has been recently proposed, mostly based on the analysis of RNA 3′ ends in several mutant conditions, that the NNS and the RB termination pathways are functionally interconnected, in that the NNS is required for releasing roadblocked polymerases and, conversely, that the presence of a RB is necessary for NNS termination (Roy *et al*, 2016). The analyses presented here based on the direct and high-resolution detection of RNAPII transcription in conditions defective for RB, CPF, or NNS termination do not generally support this model. We observed that RB termination is largely insensitive to depletion of Nrd1 (*e.g.*, see Fig EV2 and Appendix Fig S3), or mutation of Nab3 (data not shown), which is also consistent with the findings that the insertion of the sole Reb1 (Colin *et al*, 2014) or Rap1 sites (Fig 1 and Appendix Fig S1A) in the heterologous context of the *HSP104* gene is sufficient for efficient, NNS-independent termination. Similarly to what reported by Roy *et al* (2016), we detected increased RNAPII occupancy at some roadblock sites upon impairment of NNS function (*e.g.*, Fig 5), but we show that this is due to the accumulation of polymerases that fail to terminate at NNS terminators upstream of the RB rather than to the general defective clearance of stalled elongation complexes.

We favor a model according to which polymerases are not recycled for further steps of transcription when encountering a RB

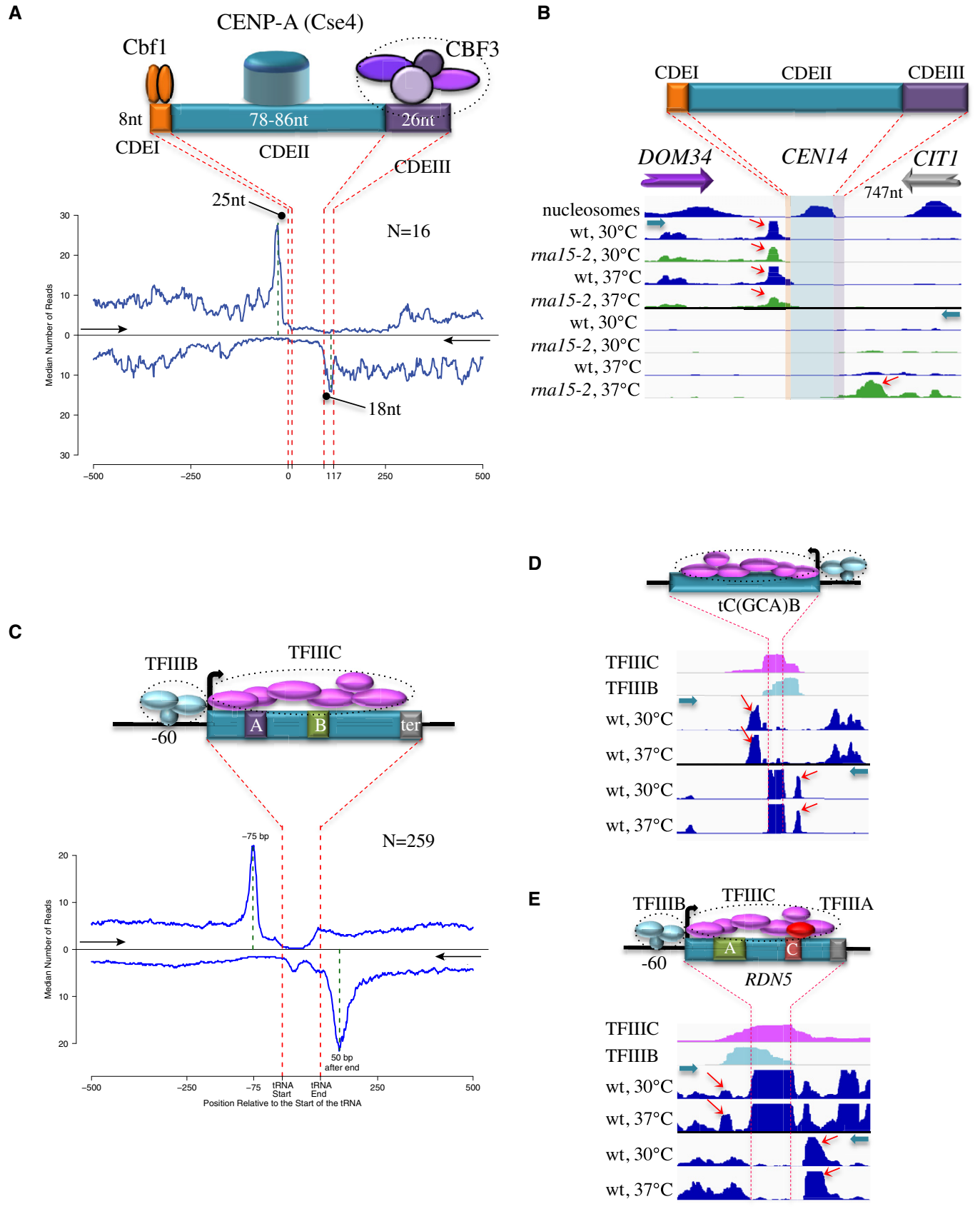

A
CENP-A (Cse4)
Cbf1
CBF3
8nt
78-86nt
26nt
CDEI
CDEII
CDEIII
25nt
N=16
18nt
Median Number of Reads
B
CDEI
CDEII
CDEIII
DOM34
CEN14
CIT1
747nt
nucleosomes
wt, 30°C
rna15-2, 30°C
wt, 37°C
rna15-2, 37°C
wt, 30°C
rna15-2, 30°C
wt, 37°C
rna15-2, 37°C
C
TFIIIB
TFIIIC
-60
A
B
ter
N=259
-75 bp
50 bp after end
tRNA Start
tRNA End
Position Relative to the Start of the tRNA
D
tC(GCA)B
TFIIIC
TFIIIB
wt, 30°C
wt, 37°C
wt, 30°C
wt, 37°C
E
TFIIIB
TFIIIC
TFIIIA
-60
A
C
RDN5
TFIIIC
TFIIIB
wt, 30°C
wt, 37°C
wt, 30°C
wt, 37°C

Figure 7.

Figure 7. RNAPII roadblock occurs at centromeres and RNAPIII genes.

A Aggregate plot of RNAPII occupancy (median reads count, PAR-CLIP data) around centromeres. Centromeres have been aligned on the beginning of the CDEI (top plot) or the CDEIII sequence (bottom plot) and a virtual centromere has been reconstituted by aligning the two plots based on the average length of the centromere. The 5′–3′ direction is indicated by a black arrow for each plot. The structure of the centromere and the interacting factors are schematically shown on the top.

B Snapshot showing the distribution of polymerases around CEN14. RNAPII CRAC distribution is shown for both wild-type and *rna15-2* cells at the permissive and non-permissive temperature for the mutant. A roadblock peak (red arrow) is observed upstream of CDEI in all conditions. Detection of the roadblock upstream of the CDEIII sequence requires increasing readthrough transcription at the upstream gene (*CIT1*) with the *rna15-2* mutation (bottom track). Cyan arrows indicate the direction of transcription.

C Metaprofile analysis of RNAPII distribution (median reads count, PAR-CLIP data) around tRNA genes. Genomic regions were aligned on the transcription start sites (top) or the transcription termination site (bottom) and the plots combined as for centromeres. A scheme of tRNA genes and associated factors is shown on the top of the plots. The 5′–3′ direction is indicated by a black arrow for each plot. Note that reads in the body of tRNAs have been removed because representing contamination from mature tRNAs (Schaughency *et al*, 2014).

D Snapshot of RNAPII distribution around tC(GCA)B. The profiles in wt cells grown at 30°C and 37°C have been shown as duplicates, as only minor differences are observed. Roadblock peaks are indicated by red arrows. The footprints of TFIIIB and TFIIIC (Nagarajavel *et al*, 2013) are shown for comparison. The direction of transcription is indicated by cyan arrows. Reads in the body of tRNAs are from contaminating tRNAs and should not be considered as *bona fide* RNAPII CRAC signals.

E RNAPII CRAC profile around the gene coding for 5S rRNA (*RDN5*) in wt cells at 30°C and 37°C as in (D). Roadblock peaks are indicated by red arrows. The footprints of TFIIIB and TFIIIC are shown. A scheme of the gene and the factors bound is shown in the top of the figure. As for tRNAs, the strong signal in the body of the gene should not be considered as a *bona fide* RNAPII signal, but contaminating 5S RNA.

(as if it were released by the NNS complex) but degraded, together with the RNA that is produced. This might look uneconomical, but the energetic balance might still be favorable in light of the evolutionary cost of developing highly efficient, error-proof termination processes. In this respect, the genomewide analyses reported here suggest that roadblock termination is likely devoted to controlling a relatively low fraction of polymerases that might nevertheless significantly affect the efficiency or robustness of neighboring processes.

Relationships between RB and the main pathways of termination in *Saccharomyces cerevisiae*

Many studies support the notion that pausing is a prerequisite for transcription termination (for a recent review, see Porrua *et al*, 2016). Slowing down the speed of the polymerase has been shown to promote earlier termination at NNS targets (Hazelbaker *et al*, 2012), and RNAPII pausing sites are preferential sites of Sen1-dependent termination *in vitro* (Porrua & Libri, 2013). This is thought to be due to a kinetic competition between RNAPII elongation and translocation on the RNA by Sen1, a concept that might also apply to CPF termination whereby "pursuing" of the polymerase is operated by the Rat1/XRN2 exonuclease (Fong *et al*, 2015). Whether pausing is induced by intrinsic components of the NNS or the CPF pathway, or by extrinsic factors, remains a poorly understood and important facet of termination.

We considered the possibility that roadblock pausing could be required for upstream termination, both CPF- and NNS-dependent. However, we did not detect the termination defects predicted by this model upon depletion of Rap1 or Reb1. In these conditions, readthrough was only observed downstream of the RB site, due to the release of a low fraction of polymerases that had accumulated at the RB. Transcription beyond the site of RB in the absence of Rap1 or Reb1 might produce RNAs that are more stable than the ones derived from RB- or NNS-dependent termination, as we suggest occurring at the *SNR8*, *SNR161*, and *SNR39B* sites. In the absence of the RB, the levels of these transcripts do not match the actual levels of transcriptional readthrough, which might lead to overestimating the impact of the RB on termination, possibly explaining the discrepancies with the model proposed by Roy *et al* (2016).

Although it remains possible that in the absence of Reb1 or Rap1, other roadblock events take over to slow down the RNAPII and promote termination, we favor the notion that pausing is induced by components of the NNS or CPF complexes, or depends on specific sequences, the nature of which remains elusive.

Widespread transcriptional readthrough at gene terminators in *Saccharomyces cerevisiae*

It is generally accepted that pervasive transcription is mainly generated by the leaky control imposed on transcription initiation by the structure of chromatin. Many studies have shown that altering the positioning or the modification status of nucleosomes (see for instance Churchman & Weissman, 2011; Marquardt *et al*, 2014; Venkatesh *et al*, 2016) significantly impacts the relative extent of initiation at divergent or cryptic promoters, generating pervasive transcription events. Our data strongly suggest that in addition to the leaky control on initiation, leaky termination generates pervasive transcription events, which might significantly impact gene expression and other cellular events. Extensive occurrence of transcriptional readthrough has been observed in the *B. subtilis* (Nicolas *et al*, 2012) and *E. coli* transcriptome (Stringer *et al*, 2014). In *S. pombe*, antisense transcripts derived from readthrough transcription are produced in wild-type cells and degraded by the exosome (Zofall *et al*, 2009) while in human cells readthrough transcripts induced by osmotic stress have been shown to represent a considerable fraction of pervasive transcription (Vilborg *et al*, 2015). However, the direct demonstration of widespread and constitutive transcriptional readthrough in wild-type cells was not attempted in these studies. Detection of these events in the present study is facilitated by the sensitivity and resolution of RNAPII CRAC and by the analysis of RB sites, where polymerases escaping termination accumulate. Because the genes used for the metasite analyses in Fig 3B were solely selected based on Reb1 or Rap1 downstream binding, they can be reasonably considered as a random sampling for their efficiency of termination. The average detection of transcriptional readthrough at these terminators therefore strongly suggests that leaky termination occurs genomewide, widening the repertoire and the potential impact of pervasive transcription. This conclusion is supported by the direct detection of significant readthrough signals

in intergenic regions downstream of many genes, which do not reflect the occurrence of independent intergenic initiation (see for instance Appendix Fig S2C).

Extensive occurrence of transcriptional readthrough might confer additional evolutionary advantages over the generation of *ex novo* genes from pervasive initiation (Carvunis *et al*, 2012; Wu & Sharp, 2013). Readthrough transcripts might evolve to generate new functions from existing modules, leading for instance to the fusion of contiguous ORFs or the generation of protein extensions.

The non-quantitative feature of termination also brings about a large potential for regulation, allowing anticorrelated expression of tandem genes and possibly its modulation by alterations in the efficiency of termination.

Roadblock occurs at many genomic sites

We show here that although the elongation complex is armed to progress relatively efficiently through nucleosomes *in vivo,* it is significantly affected by the presence of many other factors bound to the DNA. Two notable examples are centromeres and tRNAs.

Roadblock peaks for RNAPII were observed upstream of both centromere edges, where they are roadblocked presumably by Cbf1 and the CBF3 complex, binding, respectively, the CDEI site and the CDEIII sequence, which is consistent with early observations (Doheny *et al*, 1993). Very little, if any, RNAPII CRAC signal can be detected within the centromeric DNA in general and particularly in CDEII, suggesting that centromeres use intrinsic roadblock barriers to prevent trans-centromere transcription. This might underlie a requirement for maintaining the identity of the centromeric nucleosome, containing a specific variant of histone H3, Cse4, the occupancy of which might be affected by through transcription, and is consistent with the notion that directing strong transcription toward a centromere generally inactivates it (Hill & Bloom, 1987; Doheny *et al*, 1993). In contrast to these conclusions, earlier studies have proposed that a relatively moderate level of transcription through centromeres is actually required for function, which was supported by the detection of trans-centromeric transcripts by RT-qPCR, and by genetic experiments (Ohkuni & Kitagawa, 2011). Although we might have failed to detect RNAPII CRAC signals corresponding to very low levels of these trans-centromeric transcripts, we do not fully understand the basis of this discrepancy, and future work is required to elucidate the role of transcription at the point centromeres of *S. cerevisiae*.

We also show that RNAPII is roadblocked at the 5′ and 3′ ends of tRNA genes. The existence of a 5′ end roadblock is consistent with earlier studies on the tV(UAC)D locus, proposing a role for TFIIIB in preventing upstream intergenic transcription from entering the tRNA gene body (Korde *et al*, 2014; see also Roy *et al*, 2016). Here we extend this finding to a genomewide perspective, and additionally demonstrate the existence of an additional roadblock barrier that prevents antisense RNAPII transcription from crossing tRNA genes. The 3′ RB is aligned to the tRNA terminators, to which, in turn, is also aligned the trailing edge of TFIIIC footprint (Nagarajavel *et al*, 2013). However, the binding of TFIIIC alone at ETC sites is not sufficient for inducing a RB, which might indicate that additional factors (presumably TFIIIB) must be present to stabilize the interaction of TFIIIC with the DNA and induce a RB. Alternatively, it is possible that the specific topology of these transcription units, which are believed to be circularized for a more efficient transcription re-initiation (Dieci *et al*, 2013), or the general high persistence of RNAPIII at these sites might underlie the formation of the 3′ end RB.

In conclusion, our results suggest that the transcriptional landscape is modeled to a large extent by non-histone proteins bound to the DNA, which has a considerable impact in the partitioning of DNA-linked activities and in the control of pervasive transcription events. Aside from operating a quality control mechanism on the efficiency of termination at canonical sites, roadblock pausing (and termination) of polymerases has a large potential for shaping and regulating the transcriptome, in yeast and most likely many other organisms.

Materials and Methods

Yeast strains and plasmids

Yeast strains used in this study are listed in Appendix Table S1. Plasmids are listed in Appendix Table S3. The reporter construct used for selecting the Rap1-dependent terminators was previously described (Porrua *et al*, 2012). The full sequences of the selected terminators are available upon request. Rap1 constructs were expressed from the *RAP1* gene promoter to avoid growth defects due to the overexpression of Rap1 derivatives. The whole Rap1 coding sequence and 598 nt of the upstream region were cloned in a pCM185 backbone (Garí *et al*, 1997) in which the *TRP1* marker was replaced with the *S. pombe HIS5* gene. The Tet promoter and hybrid transactivator of pCM185 were deleted. The Rap1-DBD construct was obtained by replacing the Rap1 coding sequence with a fragment of the gene coding for amino acids 358–601.

RNA analyses

Northern blot analyses were performed as previously described (Colin *et al*, 2014). RT–qPCR was performed with standard procedures, using the primers listed in Appendix Table S2. Amplification efficiencies were calculated for every primer pair in each amplification reaction.

Growth conditions and preparation of cells for CRAC

Two liters of yeast cells expressing Rpb1-HTP tag (Granneman *et al*, 2009) were grown at 30°C to OD_{600} = 0.6 in CSM-Trp medium. For nuclear depletion of Reb1 and Rap1, rapamycin was added to anchor away strains or control untagged wild type for two hours to a final concentration of 1 μg/ml. *rna15-2* or wild-type cells were grown at 30°C to OD_{600} = 0.6; 1 volume of media preheated at 30 or 37°C was added and cultures were incubated at 30 or 37°C for 1.5 h. Nrd1 was depleted with the auxin degron system (Nishimura *et al*, 2009) by adding IAA (indole-3-acetic acid, Sigma) 100 μM to Nrd1-AID cells for 60 min before crosslinking.

Cells were submitted to UV crosslink using a W5 UV crosslinking unit (UVO3 Ltd) for 50 s, harvested by centrifugation, washed in cold PBS, and resuspended in TN150 buffer (50 mM Tris pH 7.8, 150 mM NaCl, 0.1% NP-40, and 5 mM beta-mercaptoethanol, 2.4 ml/g of cells) supplemented with protease inhibitors (Complete™, EDTA-free Protease Inhibitor Cocktail). The suspension was flash frozen in droplets, and cells were mechanically broken with a Mixer Mill MM 400 (5 cycles of 3 min at 20 Hz).

CRAC

The CRAC protocol used in this study is derived from Granneman *et al* (2009) with a few modifications. Cell powders were thawed and the resulting extracts were treated for one hour at 25°C with DNase I (165 U/g of cells) to solubilize chromatin and then clarified by centrifugation (20 min at 20,000 × *g* at 4°C).

IgG purification was performed with M-280 tosylactivated dynabeads coupled with rabbit IgG (15 mg of beads per sample). The complexes were eluted with TEV protease and treated with 0.2 U of RNase cocktail (RNace-IT, Agilent) to reduce the size of the nascent RNA. High salt washes for both purification steps were done at 1 M NaCl for increased stringency. The dephosphorylation step required for cleaving the 2′–5′ cyclic phosphate left by RNase treatment (Granneman *et al*, 2009) was omitted to enrich for nascent transcripts, the 3′ end of which being protected from RNase treatment.

After overnight binding on Ni-NTA column (Qiagen, 100 µl of slurry per sample), sequencing adaptors were added on the RNA as described in the original procedure. Adaptors were modified for sequencing from the 3′ end. The 3′ ligation was realized with T4 rnl 2 truncated K227Q enzyme (NEB) instead of classical T4 RNA ligase.

RNA–protein complexes were eluted in 400 µl of elution buffer (50 mM Tris pH 7.8, 50 mM NaCl, 150 mM imidazole, 0.1% NP-40, 5 mM beta-mercaptoethanol). Eluates were concentrated with Vivacon® ultrafiltration spin columns 30-kDa MWCO to a final volume of 120 µl. The protein fractionation step was performed with a Gel Elution Liquid Fraction Entrapment Electrophoresis (GelFree) system (Expedeon). Rpb1-containing fractions were treated with 100 µg of proteinase K in a buffer containing 0.5 % SDS. RNAs were purified and reverse-transcribed using reverse transcriptase Superscript IV (Invitrogen).

The absolute concentration of cDNAs in the reaction was estimated by quantitative PCR using a standard of known concentration. Amplifications were performed separately in 25 µl reactions containing each 2 µl of cDNA for typically 7–9 PCR cycles (LaTaq, Takara). The PCRs from all the samples were pooled and treated for 1 h at 37°C with 200 U/ml of Exonuclease I (NEB). The DNA was purified using NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) and sequenced using Illumina technology.

Dataset processing

CRAC samples were demultiplexed using the pyBarcodeFilter script from the pyCRACutility suite (Webb *et al*, 2014). Subsequently, the 5′ adaptor (Appendix Table S2, read at the 3′ end of reads) was clipped with Cutadapt ({m 10}, Martin, 2011) and the resulting insert quality-trimmed from the 3′ end using Trimmomatic rolling mean clipping (Bolger *et al*, 2014) (window size = 5, minimum quality = 25). At this stage, the pyCRAC script pyFastqDuplicateRemover was used to collapse PCR duplicates using a 6-nucleotide random tag included in the 3′ adaptor (Appendix Table S2, read at the 5′ end of reads). 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/) and mapped to the R64 genome (Cherry *et al*, 2012) with bowtie2 (using "-N 1" option) (Langmead & Salzberg, 2012).

RNAseq samples were demultiplexed by the sequencing platform with bcl2fastq2 v2.15.0; adaptor trimming of standard Illumina TruSeq adaptors was performed with cutadapt 1.9.1. Samples were subsequently quality-trimmed with trimmomatic (see above) and mapped to the R64 genome with bowtie2 (default options).

Metagene analyses and boxplots

For each feature included in the analysis, we extracted the polymerase occupancy values at every position around the feature and plotted the mean or median over all the values for that position in the final aggregate plot. To limit the influence of outliers on the final plot when using the mean to summarize the data, we excluded from the analysis every value at each site that was above the mean + 5 standard deviations.

To assess the occurrence of a termination defect at genes upstream of Rap1 RB sites upon Rap1 depletion or in *rna15-2* cells, we selected a subset of genes whose termination region (operationally defined as the region around the strongest site of polyadenylation) was within 300 bp upstream of a Rap1 site. To avoid interference with the signal at the RB site, we then calculated the average polymerase occupancy in the early segment of the termination region, that is, in a window 100 nucleotides immediately before the major site of polyadenylation. This value was then divided by the average polymerase occupancy signal across the whole body of the gene. A decline in the RNAPII signal in wt cell and a significantly higher signal in *rna15-2* cells in this region confirmed that loss of RNAPII indeed starts occurring in this region, presumably associated with multiple sites of 3′ end processing. The overall distribution of these ratios for several datasets was represented with boxplots and the statistical significance assessed with paired *t*-tests in order to account for different termination efficiencies at each polyadenylation site.

Dataset availability

All datasets used in this study are available under GEO numbers GSE97913 and GSE97915.

Expanded View for this article is available online.

Acknowledgments

We would like to thank F. Feuerbach, T. H. Jensen, and T. Villa for critical reading of the manuscript. D. Shore and R. Morse for the kind gift of the Rap1-AA and the *rap1-2* strains, respectively. We also thank D. Tollervey and S. Granneman for help with the CRAC protocol. This work was supported by the Centre National de la Recherche Scientifique (C.N.R.S.), the Fondation pour la Recherche Medicale (F.R.M., programme équipe 2013), l'Agence Nationale pour la Recherche (A.N.R., grants ANR-08-Blan-0038-01, ANR-12-BSV8-0014-01, and ANR-16-CE12-0022-01), and the Fondation Bettencourt (prix Coup d'Elan 2009) and the LABEX Who am I. J.B.B., T.C., and D.C. have been supported by fellowships from the French Ministry of Research. This work has benefited from the facilities and expertise of the high-throughput sequencing core facility of I2BC (Centre de Recherche de Gif—http://www.i2bc-saclay.fr/).

Author contributions

All authors designed and performed experimental work. DL and JC supervised the work. DL wrote the paper. JC, DC, TC, and OP reviewed and edited the draft. TC curated the data and generated software for analysis. DL was responsible for funding acquisition.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Arigo JT, Eyler DE, Carroll KL, Corden JL (2006) Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol Cell* 23: 841 – 851

Arimbasseri AG, Maraia RJ (2016) RNA polymerase III advances: structural and trna functional views. *Trends Biochem Sci* 41: 546 – 559

Azad GK, Tomar RS (2016) The multifunctional transcription factor Rap1: a regulator of yeast physiology. *Front Biosci (Landmark Ed)* 21: 918 – 930

Biggins S (2013) The composition, functions, and regulation of the budding yeast kinetochore. *Genetics* 194: 817 – 846

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114 – 2120

Carvunis A-R, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charloteaux B, Hidalgo CA, Barbette J, Santhanam B, Brar GA, Weissman JS, Regev A, Thierry-Mieg N, Cusick ME, Vidal M (2012) Proto-genes and de novo gene birth. *Nature* 487: 370 – 374

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M *et al* (2012) Saccharomyces genome database: the genomics resource of budding yeast. *Nucleic Acids Res* 40: D700 – D705

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

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM (2006) A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci USA* 103: 5320 – 5325

Dieci G, Bosio MC, Fermi B, Ferrari R (2013) Transcription reinitiation by RNA polymerase III. *Biochim Biophys Acta* 1829: 331 – 341

van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Né P, Loeillet S, Nicolas A, Thermes C, Morillon A (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475: 114 – 117

Doheny KF, Sorger PK, Hyman AA, Tugendreich S, Spencer F, Hieter P (1993) Identification of essential components of the *S. cerevisiae* kinetochore. *Cell* 73: 761 – 774

Fermi B, Bosio MC, Dieci G (2017) Multiple roles of the general regulatory factor Abf1 in yeast ribosome biogenesis. *Curr Genet* 63: 65 – 68

Fong N, Brannan K, Erickson B, Kim H, Cortazar MA, Sheridan RM, Nguyen T, Karp S, Bentley DL (2015) Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. *Mol Cell* 60: 256 – 267

Garí E, Piedrafita L, Aldea M, Herrero E (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast* 13: 837 – 848

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 USA* 106: 9613 – 9618

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* 431: 99 – 104

Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS (2016) Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. *Cell Rep* 15: 2147 – 2158

Hartley PD, Madhani HD (2009) Mechanisms that specify promoter nucleosome location and identity. *Cell* 137: 445 – 458

Haruki H, Nishikawa J, Laemmli UK (2008) The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol Cell* 31: 925 – 932

Hazelbaker DZ, Marquardt S, Wlotzka W, Buratowski S (2012) Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Mol Cell* 49: 55 – 66

Hill A, Bloom K (1987) Genetic manipulation of centromere function. *Mol Cell Biol* 7: 2397 – 2405

Knight B, Kubik S, Ghosh B, Bruzzone MJ, Geertz M, Martin V, Dénervaud N, Jacquet P, Ozkan B, Rougemont J, Maerkl SJ, Naef F, Shore D (2014) Two distinct promoter architectures centered on dynamic nucleosomes control ribosomal protein gene transcription. *Genes Dev* 28: 1695 – 1709

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

Kubik S, Bruzzone MJ, Jacquet P, Falcone J-L, Rougemont J, Shore D (2015) Nucleosome stability distinguishes two different promoter types at All protein-coding genes in yeast. *Mol Cell* 60: 422 – 434

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357 – 359

Lardenois A, Liu Y, Walther T, Chalmel F, Evrard B, Granovskaia M, Chu A, Davis RW, Steinmetz LM, Primig M (2011) 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 USA* 108: 1058 – 1063

MacIsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E (2006) An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* 7: 113

Malabat C, Feuerbach F, Ma L, Saveanu C, Jacquier A (2015) Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife* 4: e06722

Marquardt S, Escalante-Chong R, Pho N, Wang J, Churchman LS, Springer M, Buratowski S (2014) A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* 157: 1712 – 1723

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17: 10

Milligan L, Huynh-Thu VA, Delan-Forino C, Tuck A, Petfalski E, Lombraña R, Sanguinetti G, Kudla G, Tollervey D (2016) Strand-specific, high-resolution mapping of modified RNA polymerase II. *Mol Syst Biol* 12: 874

Moqtaderi Z, Struhl K (2004) Genome-wide occupancy profile of the RNA polymerase III machinery in *Saccharomyces cerevisiae* reveals loci with incomplete transcription complexes. *Mol Cell Biol* 24: 4118 – 4127

Nagarajavel V, Iben JR, Howard BH, Maraia RJ, Clark DJ (2013) Global 'bootprinting' reveals the elastic architecture of the yeast TFIIIB-TFIIIC transcription complex *in vivo*. *Nucleic Acids Res* 41: 8135 – 8143

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

Nicolas P, Mäder U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer A, Hansen A *et al* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* 335: 1103 – 1106

Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M (2009) An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* 6: 917 – 922

Ohkuni K, Kitagawa K (2011) Endogenous transcription at the centromere facilitates centromere activity in budding yeast. *Curr Biol* 21: 1695 – 1703

Ohkuni K, Kitagawa K (2012) Role of transcription at centromeres in budding yeast. *Transcription* 3: 193 – 197

Pelechano V, Wei W, Jakob P, Steinmetz LM (2014) Genome-wide identification of transcript start and end sites by transcript isoform sequencing. *Nat Protoc* 9: 1740 – 1759

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

Porrua O, 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, Libri D (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* 16: 190 – 202

Porrua O, Boudvillain M, Libri D (2016) Transcription termination: variations on common themes. *Trends Genet* 32: 508 – 522

Rhee HS, Pugh BF (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 147: 1408 – 1419

Roberts DN, Stewart AJ, Huff JT, Cairns BR (2003) The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. *Proc Natl Acad Sci USA* 100: 14695 – 14700

Roy K, Gabunilas J, Gillespie A, Ngo D, Chanfreau GF (2016) Common genomic elements promote transcriptional and DNA replication roadblocks. *Genome Res* 26: 1363 – 1375

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

Steinmetz EJ, Conrad NK, Brow DA, Corden JL (2001) RNA-binding protein Nrd1 directs poly(A)-independent 3′-end formation of RNA polymerase II transcripts. *Nature* 413: 327 – 331

Stringer AM, Currenti S, Bonocora RP, Baranowski C, Petrone BL, Palumbo MJ, Reilly AA, Zhang Z, Erill I, Wade JT (2014) Genome-scale analyses of *Escherichia coli* and Salmonella enterica AraC reveal noncanonical targets and an expanded core regulon. *J Bacteriol* 196: 660 – 671

Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D (2006) Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol Cell* 23: 853 – 864

Venkatesh S, Li H, Gogol MM, Workman JL (2016) Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. *Nat Commun* 7: 13610

Vilborg A, Passarelli MC, Yario TA, Tycowski KT, Steitz JA (2015) Widespread inducible transcription downstream of human genes. *Mol Cell* 59: 449 – 461

Webb S, Hector RD, Kudla G, Granneman S (2014) PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol* 15: R8

Westermann S, Drubin DG, Barnes G (2007) Structures and functions of yeast kinetochore complexes. *Annu Rev Biochem* 76: 563 – 591

Wilkening S, Pelechano V, Järvelin AI, Tekkedil MM, Anders S, Benes V, Steinmetz LM (2013) An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res* 41: e65

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

Wu X, Sharp PA (2013) Divergent transcription: a driving force for new gene origination? *Cell* 155: 990 – 996

Wyers F, Rougemaille M, Badis G, Rousselle J-C, Dufour M-E, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, Libri D, Jacquier A (2005) Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121: 725 – 737

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

Zofall M, Fischer T, Zhang K, Zhou M, Cui B, Veenstra TD, Grewal SIS (2009) Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. *Nature* 461: 419 – 422

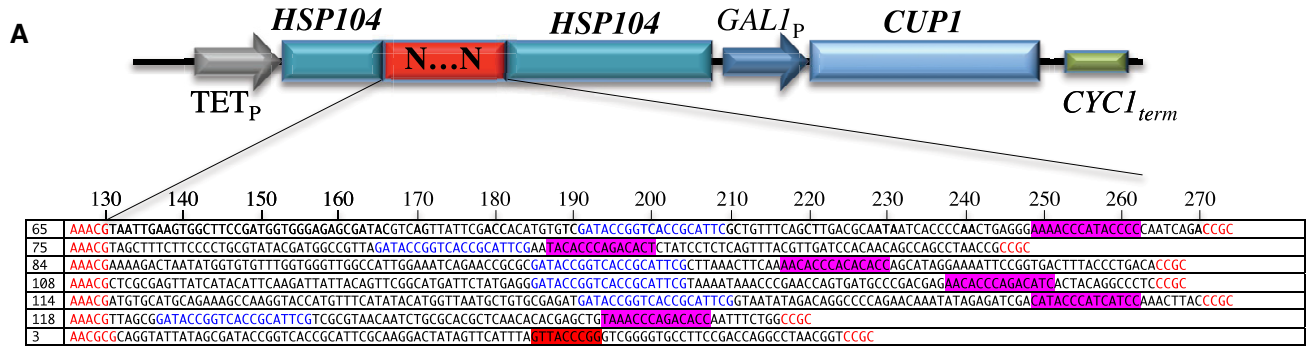
Expanded View Figures

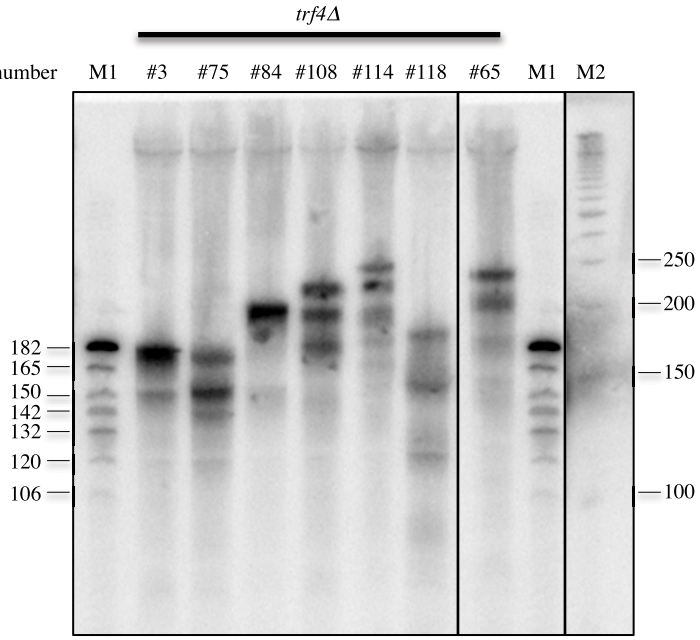
Figure EV1. 3′ end mapping of Rap1-terminated transcripts.

A Top: Schematic drawing of the reporter system used for selecting the RB terminators with the position of the insertion and the sequence of the selected clones containing a Rap1 site (purple). In blue the sequence of a constant linker used for constructing the pool. Flanking HSP104 sequences are indicated in red. Bottom: PAGE-northern blot analysis of RNAs produced by the different constructs after oligonucleotide-directed RNaseH cleavage at −130 nt from the start of the insertion. The position from the RNase H cleavage point is indicated above the sequences. The presence of several shorter RNAs might reveal the occurrence of termination at sites of RNAPII piling up. All analyses were done in an *trf4*Δ strain to detect unstable transcripts. All lanes are derived from the same gel; marker M1 is shown twice for clarity.

B Snapshots showing the RNAPII CRAC and RNAseq signal at intron-containing genes, illustrating the co-transcriptional nature of the CRAC signal.

Source data are available online for this figure.





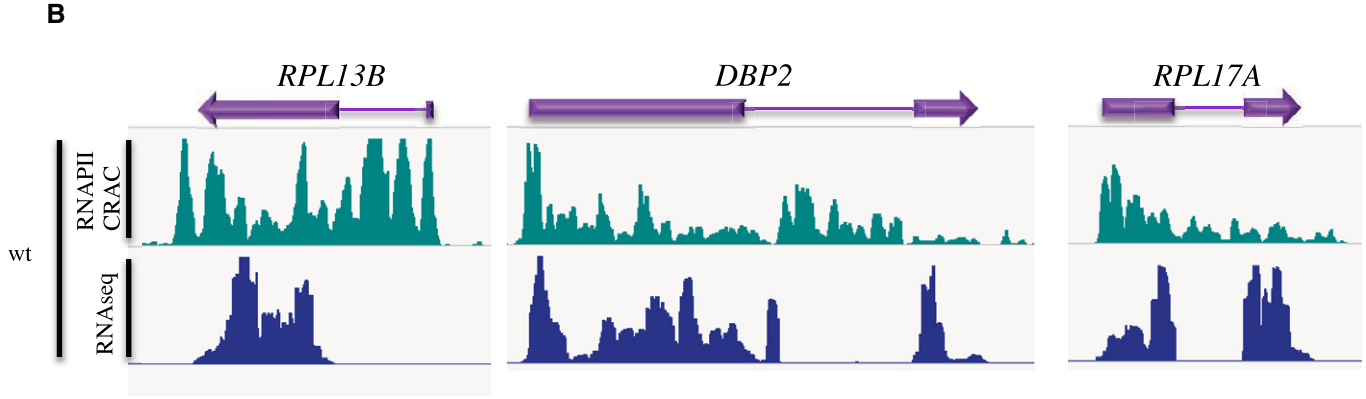


Figure EV1.

Figure EV2. Roadblock termination functions as a fail safe mechanism to limit constitutive readthrough.

A Aggregate plot showing the average RNAPII CRAC profile at sites of Reb1 occupancy located within 300 nt downstream of genes terminated by the CPF pathway as in Fig 3B. The plot demonstrates that in striking contrast to alteration of the CPF pathway, affecting NNS termination by nuclear depletion of Nrd1 has no significant effects on the accumulation of RNAPII at the site of roadblock. Sites used in these analyses are listed in Dataset EV1.

B Snapshots illustrating the presence of significant levels of intergenic RNAPII CRAC signals at three tandem gene loci. The transcription initiation sites (TSS, Malabat *et al*, 2015) detected in the regions shown are indicated on top of the RNAPII CRAC tracks. In all these cases, intergenic initiation cannot be detected, indicating that the intergenic RNAPII signal derives from the constitutive readthrough at the CPF-dependent terminator of the upstream gene. Note that the levels of the readthrough signal in these cases are comparable to the levels of transcription of the downstream gene.

C Aggregate plot showing the average RNAPII CRAC profile at sites of Abf1 occupancy located within 300 nt downstream of genes terminated by the CPF pathway as in Fig 3B. The plots show the average RNAPII occupancy in the wild-type and in *rna15-1* cells grown for 1.5 hours at the non-permissive temperature. Sites used in these analyses are listed in Dataset EV1.

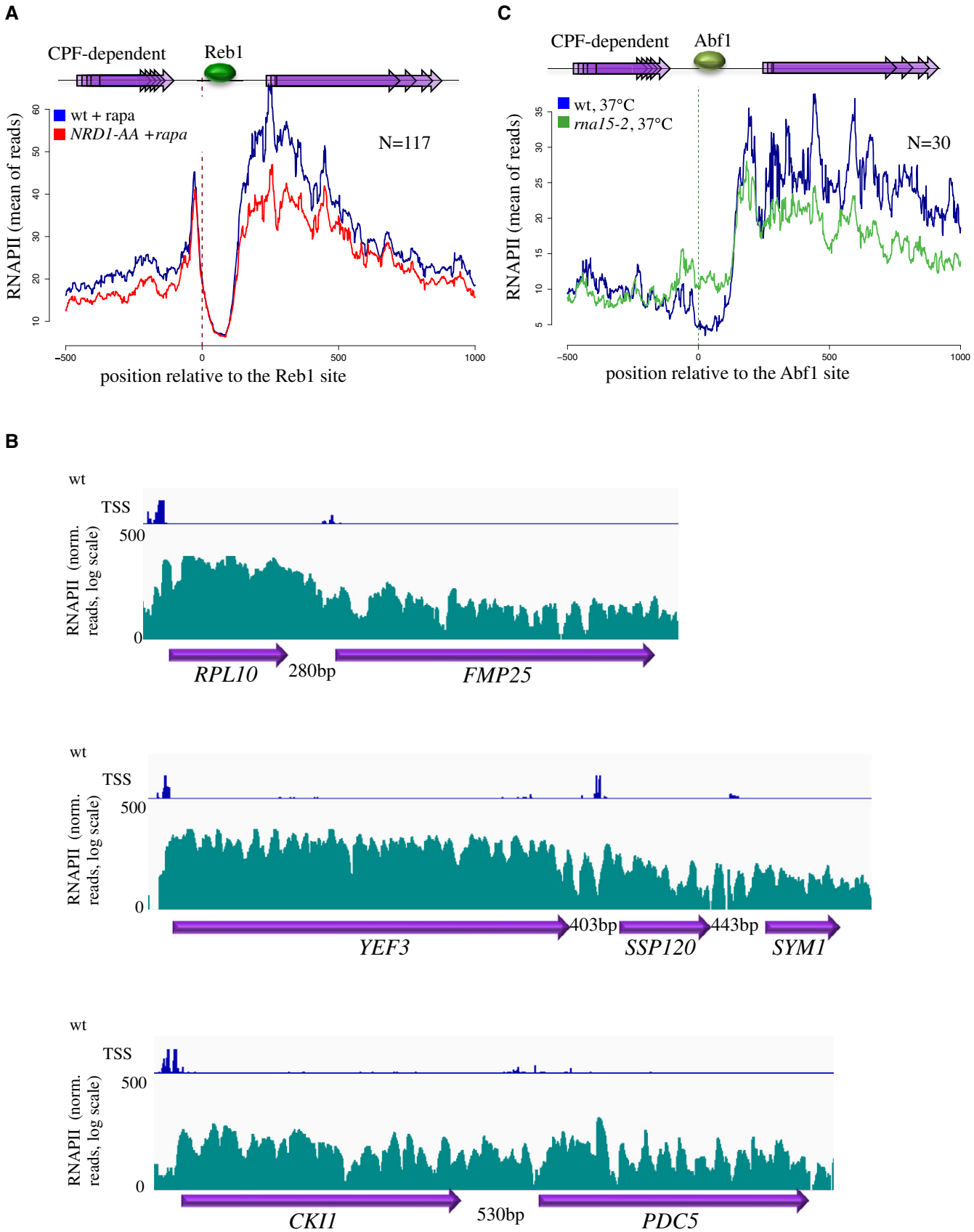


Figure EV2.

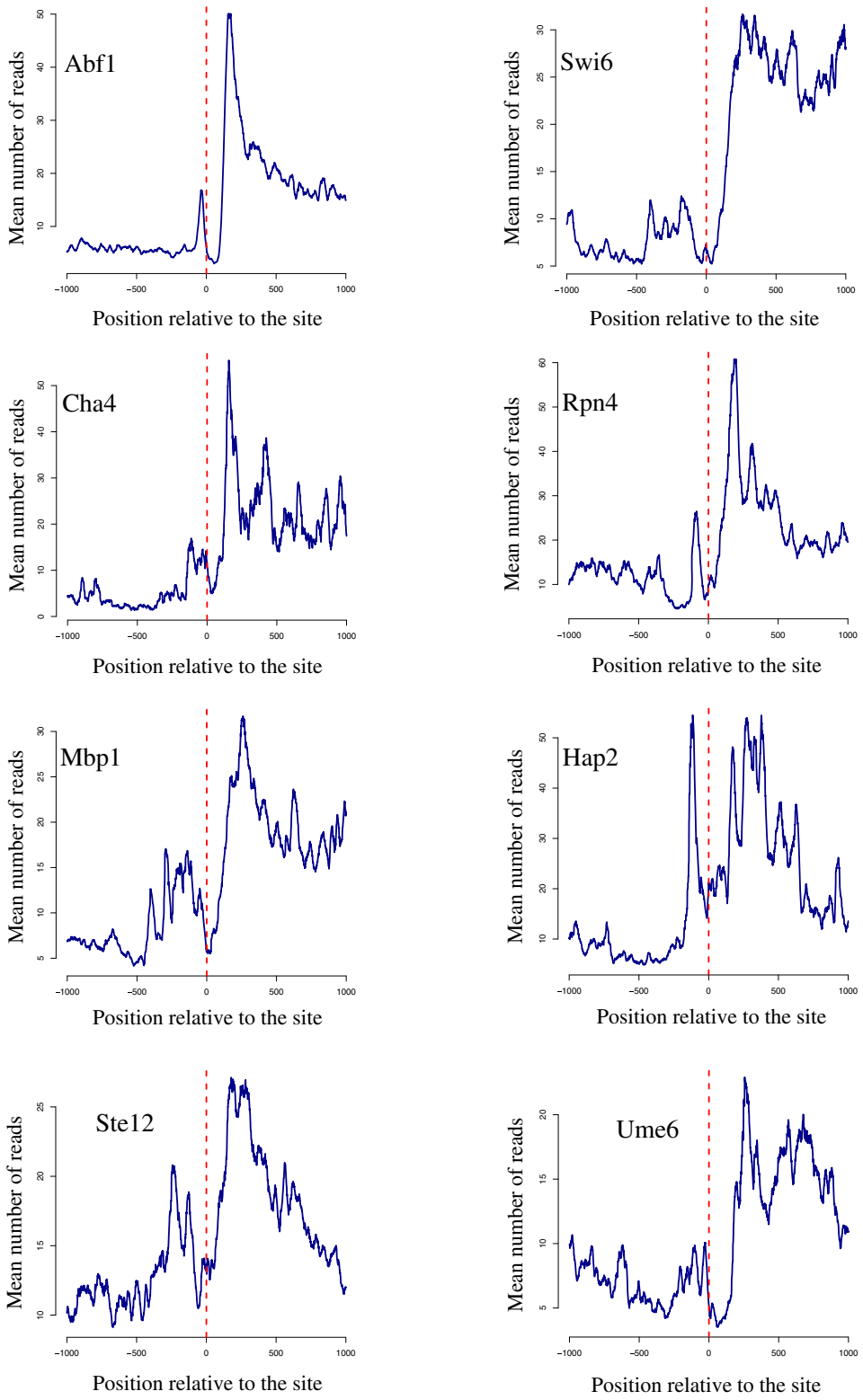


Figure EV3. Metasite analyses illustrating the profile of RNAPII CRAC signal at various transcription factor binding sites.

Aggregate plots showing the profile of RNAPII CRAC signal around sites of binding for the transcription factors indicated. The large peak after each binding site corresponds to downstream events of transcription initiation as for the plots in Fig 3A. The roadblock peak precedes the position of aligned binding sites. Sites used in these analyses are listed in Dataset EV1.

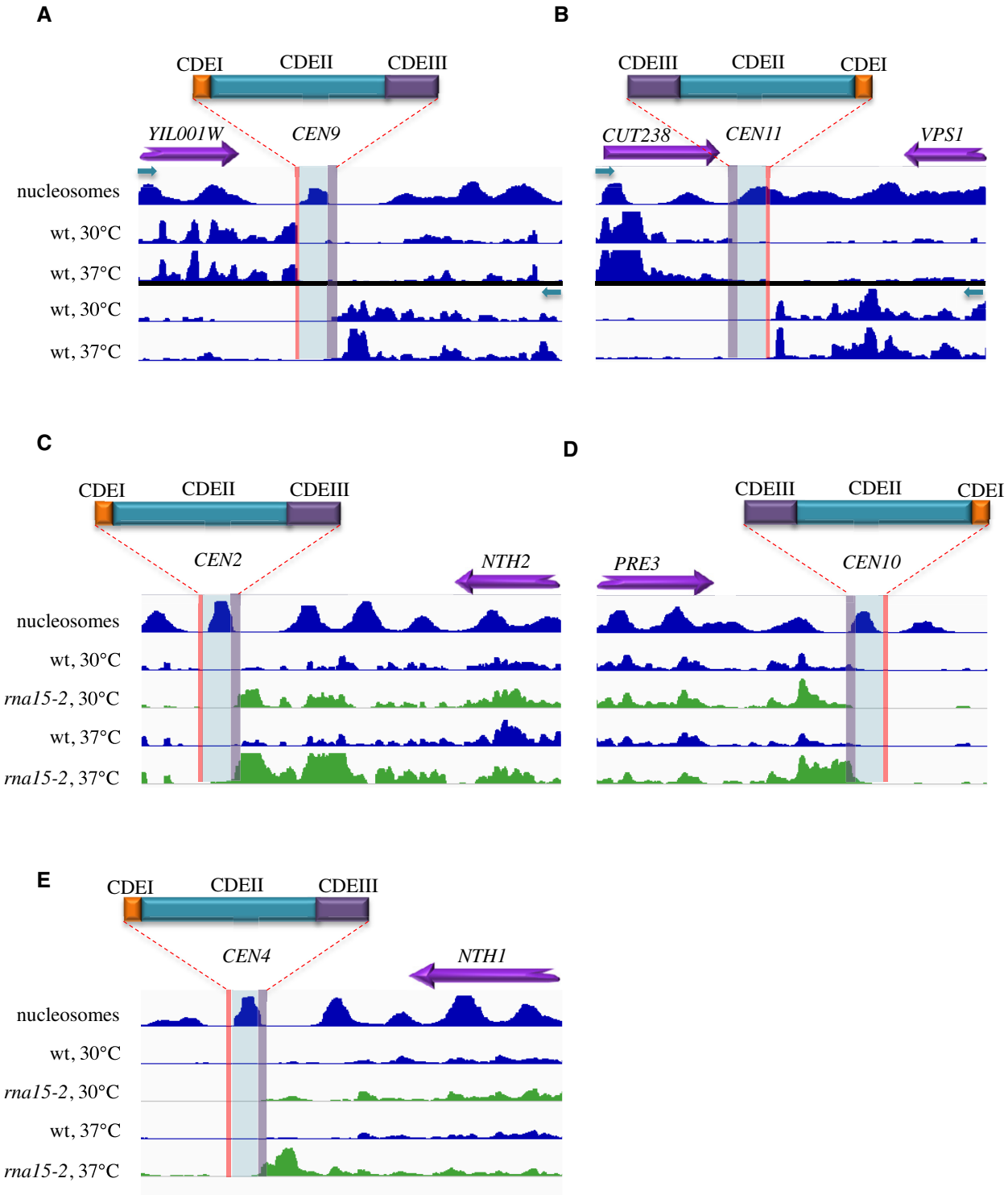


Figure EV4. Snapshots showing the RNAPII CRAC signal around representative centromeres.

A–E The structure of each centromere is schematically shown on top and reported, scaled, on the tracks. In (A, B), both strands are shown and the direction of transcription is indicated by a green arrow. In (C–E), only one strand is shown as no significant levels of transcription could be detected on the other strand. RNAPII CRAC tracks derived from *rna15-1* cells are shown at the permissive and non-permissive temperature to illustrate the effect of increased readthrough transcription at convergent genes, which leads to the accumulation of RNAPII signal at the roadblock.

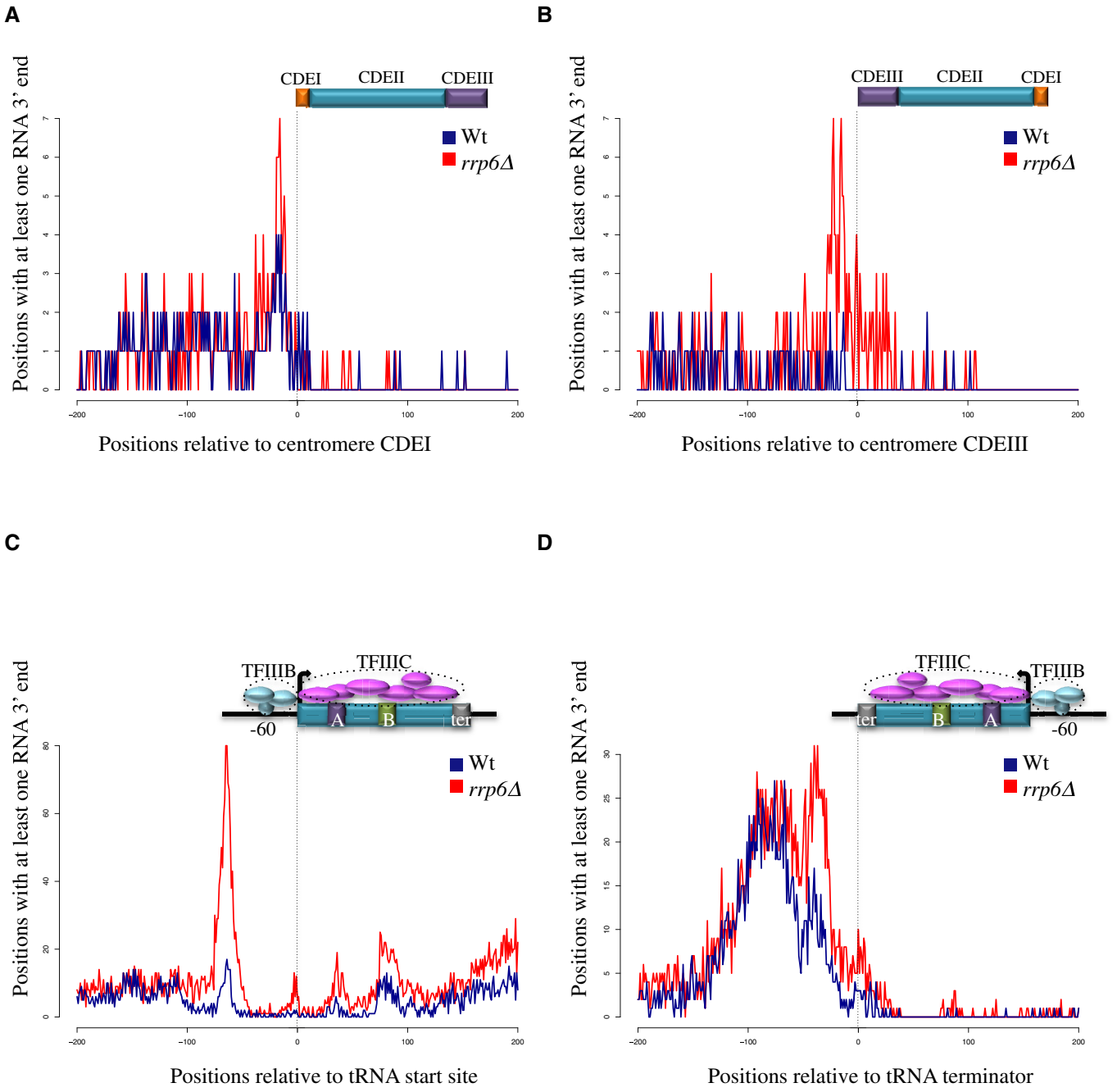


Figure EV5. Metasite analysis of RNA 3′ ends around centromeres and tRNAs.

A–D Aggregate plots showing the distribution of RNA 3′ ends around centromeres and tRNAs in wt and *rrp6*Δ cells to detect unstable RNAs. The orientation of centromeres and tRNA relative to transcription is shown by schematic drawings. Note the presence of 3′ ends peaks indicating the occurrence of transcription termination at sites of RB.

Appendix

Table of contents

Supplementary Figures

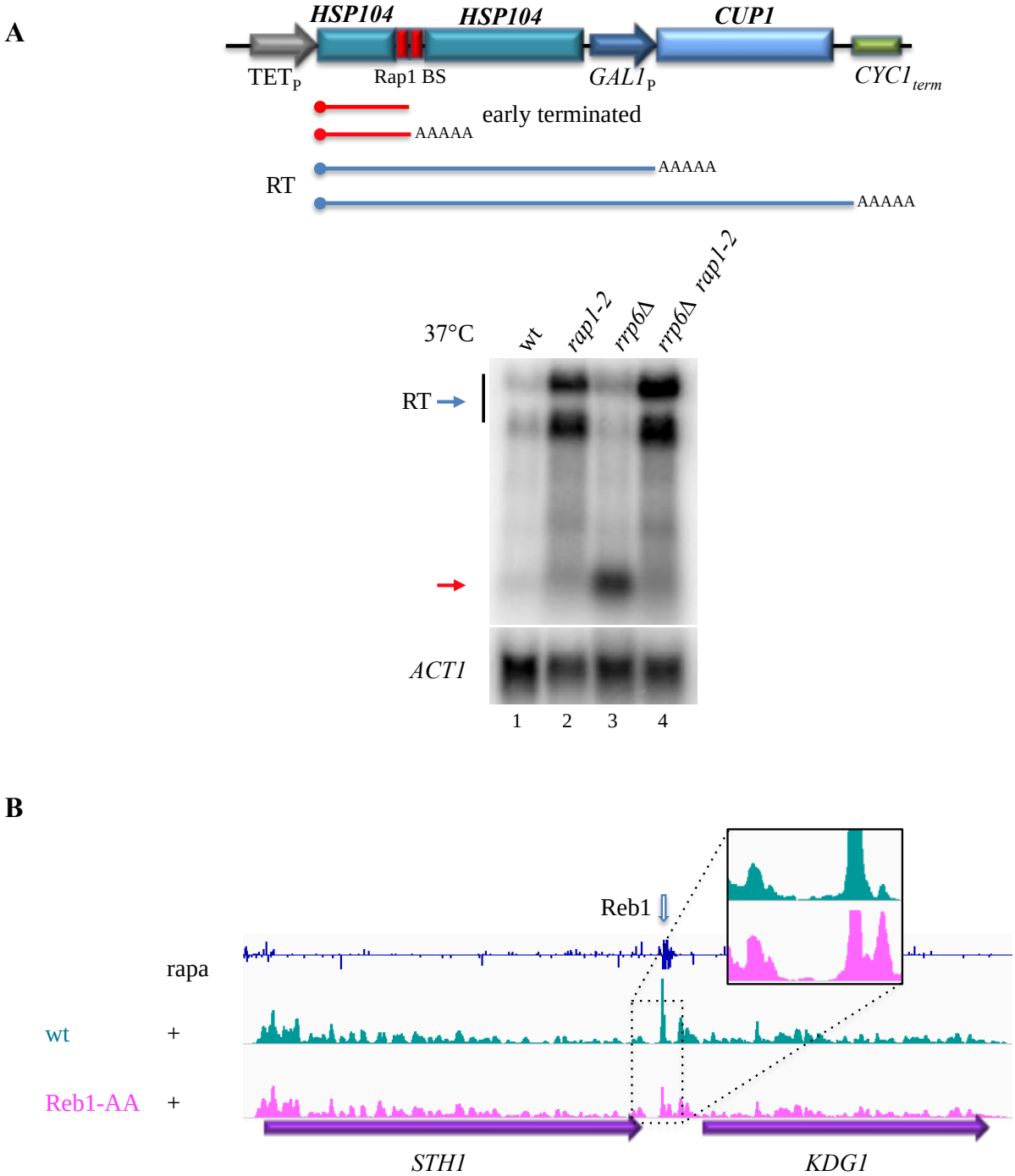


Figure S1. A. Northern blot analysis of transcripts derived from a construction containing the two Rap1 sites upstream of the *HYP2* gene in the context of the TET$_P$-*HSP104-GAL1*$_p$-*CUP1* reporter (scheme shown on top). Strains have been grown at permissive temperature (25°C) and shifted to 37°C for 4 hour to inactivate the *rap1-2* thermosensitive mutant. The transcript derived from termination at the Rap1 sites is indicated by a red arrow. Because of its instability, this transcript is best detected in *rrp6*Δ cells (lane 2). **B.** Snapshot showing the RNAPII CRAC signal at a site of Reb1-dependent transcriptional roadblock. The roadblock peak decreases significantly upon nuclear depletion of Reb1 by the addition of rapamycin for 1 hour to the Reb1-AA strain. As a control, the parental strain (containing untagged Reb1) was also treated with rapamycin for the same time. The inset contains a magnification of the region of the roadblock illustrating the appearance of a readthrough signal upon Reb1 depletion.

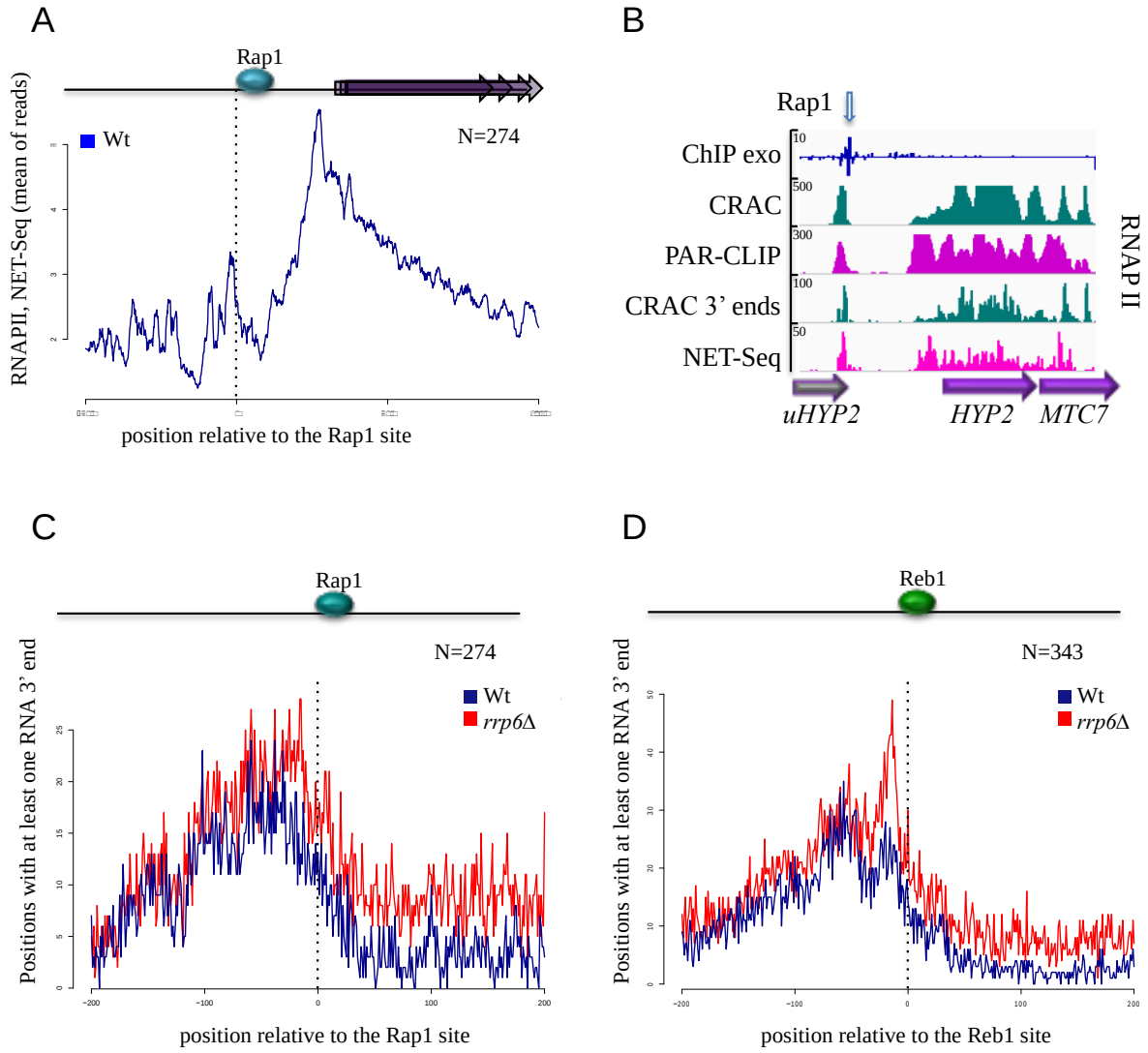


Figure S2. A. Metasite analysis analogous to the one shown in Figure 3A, but using NET-Seq, instead of CRAC data. **B.** Comparison of the CRAC, NET-Seq and PAR-CLIP signals at a site of Rap1 RB. For better comparison between CRAC and NET-Seq only the read 3' ends have been plotted in the last to tracks. Note the prominent presence of an RNAPII pausing peak revealed with all techniques. **C-D.** Aggregate distribution of RNA 3'-ends in wt or *rrp6*Δ cells as indicated upstream of Rap1 and Reb1 binding sites. The presence of these RNAs testifies to the occurrence of termination events at sites of roadblock.

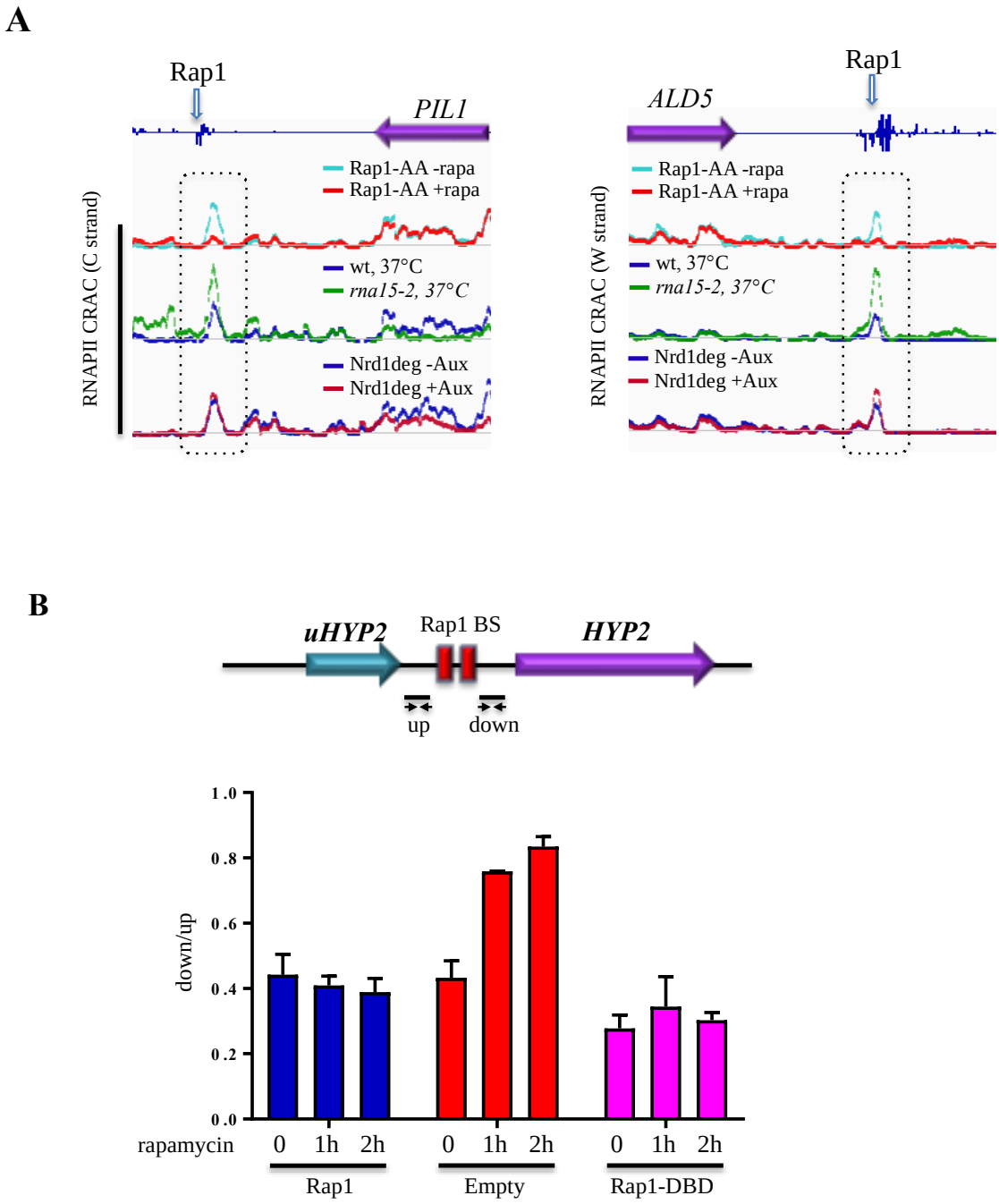


Figure S3. A. Representative snapshots illustrating the impact of the NNS complex on roadblock termination. The RNAPII CRAC signal at sites of roadblock termination downstream of *PIL1* and *ALD5* is shown (dotted rectangle). In both cases, the roadblock peak is insensitive to depletion of Nrd1 (bottom track, compare blue and red lines). As a control, the RB peak is strongly diminished by depletion of Rap1 (top track) and strongly increased upon mutation of the CPF pathway (middle track) because polymerases that fail to terminate at the CPF terminator accumulate at the RB site. **B.** Expression of the Rap1 DNA binding domain alone induces roadblock termination upstream of the *HYP2* locus. RT-qPCR analyses of transcripts around the Rap1 sites (n= 2, error bars indicate the half difference between the replicates). The ratio between the RT-qPCR signal detected before (up) and after (down) the roadblock (down/up) is used as a measure of roadblock efficiency upon depletion of Rap1 in the presence of ectopic Rap1, the Rap1 DNA binding domain (Rap1-DBD) or an empty vector as indicated. Data have been corrected for the different efficiency of amplification of the two amplimers. A schematic representation of the *HYP2* locus is shown on the top. The amplimers used for the RT-qPCR analysis are located 46nt before the first Rap1 site ("up") and immediately after the second Rap1 site ("down").

Table S1. Strains used in this study

Library #	Name	Genotype	Reference
DLY671	BMA64	*MAT**a** ura3-1; ade2-1; his3-11,15; leu2-3,112; can1-100; trp1*Δ	F. Lacroute
DLY678	*trf4*Δ *rrp6*Δ	As BMA, *MAT**a**; trf4::KAN; rrp6::URA*	Libri Lab
DLY815	*rrp6*Δ	As BMA, *MAT**a**; rrp6::KAN*	Libri Lab
DLY2241	W303	*ade2-1; ura3-1; his3-11,15; trp1-1; leu2-3,112; can1-100*	(Thomas & Rothstein, 1989)9)
DLY2207	*rap1-2*	As W303 *MAT**a**; rap1-2*	(Kurtz & Shore, 1991)1)
DLY2242	*rsp5-1*	As W303 *rsp5-1::HIS*	(Harreman *et al*, 2009)
DLY2547	HHY212 Anchor away loxed	As W303, *MAT**a** ; tor1-1; fpr1::loxP*; *RPL13A-2xFKBP12::loxP*; *ura3-1*	(Haruki *et al*, 2008)
DLY2568	Reb1-AA	as DLY2547, *REB1-FRB::KAN*	this study
DLY2570	Rap1-AA	as DLY2547, *RAP1-FRB::KAN*	this study
DLY2571	Rpb1-HTP	As BMA, *MAT**a**; RPB1-HTP::TRP1Kl*	this study
DLY2736	Reb1-AA Rpb1-HTP	As DLY2568, *MAT**a**; RPB1-HTP::TRP1Kl*	this study
DLY2754	Rpb1-HTP *rna15-2*	As DLY2571, *MAT rna15-2*	this study
DLY2838	Rpb1-HTP Anchor away	As W303; *MAT RPB1-HTP::URAKl, tor1-1; fpr1::NAT; RPL13A-2xFKBP12::loxP-TRP1-loxP*	this study
DLY2840	Rap1-AA Rpb1-HTP	As DLY2838, *MAT RAP1-FRB-RAP1::LEU*	this study
DLY2859	WT AA Rpb1-HTP *rrp6*Δ	As DLY2838, *MAT rrp6::HIS5Sp*	this study
DLY2860	Reb1AA Rpb1-HTP *rrp6*Δ	As DLY2736, *MAT**a** rrp6::HIS5Sp*	this study
DLY2861	Rap1 AA Rpb1-HTP *rrp6*Δ	As DLY2840, *MAT* ; *rrp6::HIS5Sp*	this study
DLY2867	Rpb1-HTP Nrd1-AID	*As 2571, MAT**a**, RPB1-HTP::TRP1kl NRD1-3Flag-mAID, KAN::OsTIR1*	this study
DLY3080	*rap1-2 rrp6::KAN*	As DLY2207, *rrp6::KAN*	this study

Table S2. Oligonucleotides used in this study

Library #	Sequence (5'-3')	Use
DL190	TTGAGCCAACGTCAAAATCGTTAGAGCCCTTTCTGTAAATTGCGTTTGGTCGTTCAT	Northern blot probe, against HSP104
DL2627	ATTCAAAAGCGAACACCGAATTGACCATGAGGAGACGGTCTGGTTTAT	Northern blot probe, U4
DL3248	AGCGTCCAGCTACAGCGT	RT-Q-PCR (uHYP2, up)
DL3249	AACGGGAACGGCGACTTG	RT-Q-PCR (uHYP2, up)
DL3198	TGTCGCCTCACACGGACC	RT-Q-PCR (uHYP2, down)
DL3199	CCTCGATGTATTCCGTAG	RT-Q-PCR (uHYP2,down
L3-6N-GA	/5rApp/GCTtcNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3ddC/	3'-adapter
L3-6N-GU	/5rApp/GCTacNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3ddC/	3'-adapter
L3-6N-AC	/5rApp/GCTgtNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3ddC/	3'-adapter
L3-6N-UC	/5rApp/GCTgaNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3ddC/	3'-adapter
L5miRCat	5-/5InvddT/CTTGrGrCrArCrCrCrGrArGrArArUrUrCrCrA-3	5'-adapter
RT L3-2	5-ACACTCTTTCCCTACACGACGCTCTTCCG-3	RT primer
P5_3prime	5-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3	PCR
miRCat_PCR2	5-CAAGCAGAAGACGGCATACGAgatcCTTGGCACCCGAGAAT-3	PCR

Table S3. Plasmids used in this study

Library	Name	Description	Reference
pDL431	pCM190(TRP)-TET-HSP104-X3-HSP104-GAL1-LACZ	Reporter containing a Reb1-dependent terminator	(Colin *et al*, 2014)
DL435	pCM190(TRP)-TET-HSP104-X118-HSP104-GAL1-LACZ	Reporter containing a Rap1-dependent terminator	This study
DL468	pCM190(URA)-TET-HSP104-X118-HSP104-GAL1-CUP1	Reporter containing a Rap1-dependent terminator	This study
DL436	pCM190(TRP)-TET-HSP104-X118-ΔRap1 BS-HSP104-GAL1-LACZ	Same as DL435, containing a precise deletion of the Rap1 binding site	This study
DL878	pcM185(HIS)-P_{RAP1} *RAP1*	Plasmid expressing full length Rap1 under control of the *RAP1* promoter	This study
DL879	pcM185(HIS)- P_{RAP1} -*RAP1-DBD*$_{358-601}$	Plasmid expressing the DNA binding domain of Rap1 (aa. 358-601) under control of the *RAP1* promoter	This study

SUPPLEMENTAL REFERENCES

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

Harreman M, Taschner M, Sigurdsson S, Anindya R, Reid J, Somesh B, Kong SE, Banks CAS, Conaway RC, Conaway JW & Svejstrup JQ (2009) Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. *Proc. Natl. Acad. Sci. U. S. A.* **106** 20705–20710

Haruki H, Nishikawa J & Laemmli UK (2008) The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol. Cell* **31** 925–932

Kurtz S & Shore D (1991) RAP1 protein activates and silences transcription of mating-type genes in yeast. *Genes Dev.* **5** 616–628

Malabat C, Feuerbach F, Ma L, Saveanu C & Jacquier A (2015) Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife* **4**

Thomas BJ & Rothstein R (1989) Elevated recombination rates in transcriptionally active DNA. *Cell* **56** 619–630

II - General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation

GRFs participate to the formation of nucleosome depleted regions at active promoters and are also essential for controlling gene expression. In the course of our previous study on roadblock termination, we demonstrated that the presence of GRFs is also crucial to limit the progression of RNAPII that fail to terminate at canonical sites, thus ensuring gene expression integrity. Careful analysis of the distribution of the polymerase revealed a striking and unexpected effect of Rap1 depletion. Indeed, we observed that the absence of Rap1 is associated with a major increase in the RNAPII CRAC signal within many Rap1-bound promoters, at odd with the notion of Rap1 being a transcriptional activator. Therefore, we wondered *what could be the mechanism underlying this phenomenon of ectopic transcription and how it could affect gene expression and transcription fidelity*.

In order to better understand the function of Rap1 in controlling transcription initiation, we mapped transcription start sites and nucleosomes in presence and absence of the latter. We have determined that Rap1 is not only important to correctly position nucleosomes around promoter regions but also to ensure transcription fidelity by preventing ectopic and pervasive transcription events to fire from alternative and spurious sites. The appearance of ectopic initiation in Rap1-depleted cells correlates with altered nucleosome positioning and is independent of chromatin remodelers. Interestingly, the occurrence of ectopic initiation and the mis-localisation of nucleosomes upon nuclear depletion of Rap1 can be partially reversed by expressing a truncated version of Rap1 carrying the DNA-binding domain alone, suggesting that Rap1, and probably other GRFs, promote NDRs formation and ensure transcription fidelity by a steric mechanism.

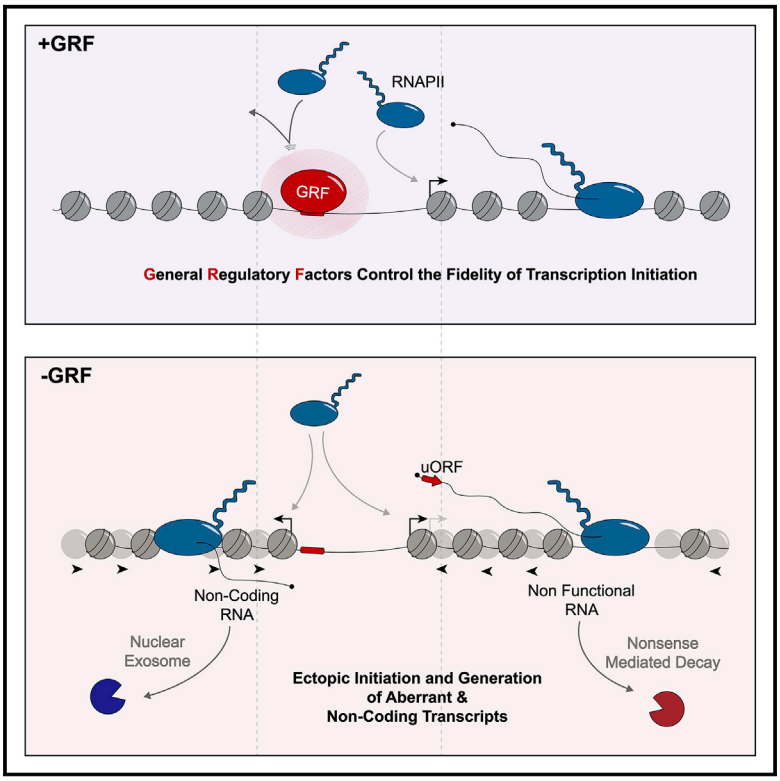
Detailed contribution: D.L supervised the project. D.L and D.S were responsible for funding acquisition. D.L and D.C wrote the original draft. D.L, D.C, D.S and S.K reviewed and edited the original draft. D.L and D.C conceptualized the project. The software design was performed by M.B and T.C. CRAC experiments were performed by D.C. TSS-Seq and MNase-seq were performed by D.C with the help of F.F and S.K respectively. Western blots and RNA-Seq were carried out by H.G and C.B under the supervision of D.C. F.F performed the northern blot experiments. Strains and plasmids were constructed and designed by D.C, F.F, S.K and H.G. Bioinformatic analysis were performed by D.L and M.B.

Article

Molecular Cell

General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation

Graphical Abstract



Highlights

- GRFs control transcription initiation fidelity and suppress pervasive transcription
- Ectopic initiation in the absence of Rap1 has variegated effects on gene expression
- Ectopic transcription initiation correlates with altered nucleosome positioning
- Rap1 suppresses transcription initiation by a steric hindrance mechanism

Authors

Drice Challal, Mara Barucco, Slawomir Kubik, ..., Chaima Benaksas, David Shore, Domenico Libri

Correspondence

domenico.libri@ijm.fr

In Brief

Challal et al. show that Rap1 and other general regulatory factors are required for the fidelity of transcription initiation. In their absence, initiation occurs at ectopic sites leading to the production of many novel RNAs with different stability and coding potential. Rap1 also represses non-coding transcription, thus controlling pervasive transcription.

Challal et al., 2018, Molecular Cell 72, 955–969
December 20, 2018
https://doi.org/10.1016/j.molcel.2018.11.037

CellPress

Molecular Cell
Article

General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation

Drice Challal,[1,2] Mara Barucco,[1,6] Slawomir Kubik,[3,6] Frank Feuerbach,[4] Tito Candelli,[5] Hélène Geoffroy,[1] Chaima Benaksas,[1] David Shore,[3] and Domenico Libri[1,7,*]

[1]Institut Jacques Monod, Centre National de la Recherche Scientifique, UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, 75205 Paris, France
[2]Université Paris Saclay, Ecole doctorale Structure et Dynamique des Systèmes Vivants, 91190 Gif sur Yvette, France
[3]Department of Molecular Biology and Institute of Genetics and Genomics of Geneva (iGe3), 30 quai Ernest-Ansermet, 1211 Geneva 4, Switzerland
[4]Institut Pasteur, Centre National de la Recherche Scientifique, UMR3525 Paris, France
[5]Princess Máxima Center for Pediatric Oncology, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands
[6]These authors contributed equally
[7]Lead Contact
*Correspondence: domenico.libri@ijm.fr
https://doi.org/10.1016/j.molcel.2018.11.037

SUMMARY

The fidelity of transcription initiation is essential for accurate gene expression, but the determinants of start site selection are not fully understood. Rap1 and other general regulatory factors (GRFs) control the expression of many genes in yeast. We show that depletion of these factors induces widespread ectopic transcription initiation within promoters. This generates many novel non-coding RNAs and transcript isoforms with diverse stability, drastically altering the coding potential of the transcriptome. Ectopic transcription initiation strongly correlates with altered nucleosome positioning. We provide evidence that Rap1 can suppress ectopic initiation by a "place-holder" mechanism whereby it physically occludes inappropriate sites for pre-initiation complex formation. These results reveal an essential role for GRFs in the fidelity of transcription initiation and in the suppression of pervasive transcription, profoundly redefining current models for their function. They have important implications for the mechanism of transcription initiation and the control of gene expression.

INTRODUCTION

Active promoters are generally depleted in nucleosomes (nucleosome depleted regions, NDRs), which is thought to provide access to DNA binding proteins required for transcription activation (for recent reviews, see Lai and Pugh, 2017; Lieleg et al., 2015). Promoter NDRs are bordered by two well-positioned nucleosomes, one of which is referred to as the +1 nucleosome because it is the first of the array of genic nucleosome that are also generally well positioned and regularly spaced. The concept of the upstream −1 nucleosome is more ambiguous, because the latter can also be the +1 nucleosome of a divergently transcribed gene. This distinction is not merely semantic, because the bona fide +1 nucleosome is characterized by a different histone composition as it typically contains the H2AZ histone variant instead of H2A. Importantly, the +1 nucleosome is associated with the transcription start site (TSS), which is generally located 12–15 nucleotides downstream of its upstream border in yeast (Hughes et al., 2012; Lee et al., 2007; Rhee and Pugh, 2012; Tsankov et al., 2010; Whitehouse et al., 2007). Formation of the NDR depends on many factors, the weight of which can vary from case to case (for recent reviews, see Lai and Pugh, 2017; Lieleg et al., 2015). One of these factors is the sequence of the DNA that is wrapped around the histone octamer, which defines the most thermodynamically favorable position for nucleosomes on naked DNA. Many studies have, however, demonstrated that DNA-histone interactions alone do not determine proper nucleosome positioning and that *trans*-acting factor are required (Struhl and Segal, 2013). Among these, chromatin remodelers (CRs) and general regulatory factors (GRFs) play important roles. The CR SWI/SNF and RSC complexes use the energy of ATP to displace nucleosomes at promoters (Hartley and Madhani, 2009; Parnell et al., 2008; Ryan et al., 1998; Shivaswamy and Iyer, 2008). Other remodelers, like ISWI and INO80, are believed to be important for positioning the +1 nucleosome and the packing of nucleosomes along transcription units (Krietenstein et al., 2016; Lai and Pugh, 2017). GRFs (including Rap1, Reb1, Abf1, and Tfb1) contain a related DNA binding domain and are required for the expression of several classes of genes, encoding for example ribosomal proteins, glycolytic enzymes, and snoRNAs. GRFs bind DNA at specific sites, generally within NDRs, and have been shown to be important for excluding nucleosomes from these regions (Badis et al., 2008; Hartley and Madhani, 2009; Preti et al., 2010; Ganapathi et al., 2011; Hughes et al., 2012; Kubik et al., 2015, 2018). These proteins do not possess ATPase activity, and therefore they must

act on nucleosomes by different mechanisms, either directly or indirectly, for instance, by recruiting CRs.

Rap1 (Repressor Activator Protein 1) was originally described as an activator and a repressor of gene expression at silent mating-type loci (for a recent review, see Azad and Tomar, 2016). Gene activation was shown to depend on a C-terminal domain of the protein that can activate transcription alone when fused to a DNA binding domain with altered specificity (Johnson and Weil, 2017). This region of Rap1 interacts with the PIC (pre-initiation complex) components TFIID and TFIIA (Garbett et al., 2007; Johnson and Weil, 2017; Papai et al., 2010). Rap1 can also repress expression when bound to promoters, possibly by directly interacting with and inhibiting TBP binding to the DNA (Bendjennat and Weil, 2008). However, the mechanism underlying the bimodal role of Rap1 at promoters is not fully understood.

One of the salient features of promoters is their intrinsic bi-directionality (Jin et al., 2017; Neil et al., 2009; Rhee and Pugh, 2012; Xu et al., 2009). Bi-directional transcription allows expression of divergent genes but can also generate non-coding and non-functional transcripts in the opposite direction of a functional gene. These transcripts are often unstable in wild-type cells and are degraded either in the nucleus by the RNA exosome or in the cytoplasmic by the nonsense-mediated decay (NMD) pathway (for a review, see Porrua and Libri, 2015). Cells have evolved strategies to favor transcription toward functional coding regions and limit the extent of what is called pervasive, non-coding transcription (Jin et al., 2017). The nucleosomal architecture of NDRs appears to be important for this control as mutants in CRs or modifiers have been shown to increase the extent of pervasive antisense transcription events (Marquardt et al., 2014; Whitehouse et al., 2007). Limiting pervasive transcription is essential for the cell, and spurious transcription events that have escaped control at the level of initiation are terminated by several mechanisms (Porrua and Libri, 2015).

In this study, we have used a combination of high-resolution genome-wide analyses to address the effects of rapid depletion of Rap1 and other GRFs. We studied the changes in occupancy of RNAPII after rapid Rap1 depletion and correlated these results to changes in transcription initiation, stability of the transcripts vis-à-vis nuclear and NMD degradation pathways, and changes in the nucleosome architecture of NDRs. We demonstrate a massive change in the pattern of transcription initiation, which generates transcripts of diverse stability and coding potential. This translates into variegated effects on gene expression, from activation to repression and leads to the generation of different protein isoforms, drastically changing current models of GRF action. We show that Rap1, Abf1, and Reb1 have crucial roles in limiting pervasive transcription at the level of initiation since many novel non-coding RNAs are generated when any of these factors is depleted.

Together, these results support a model whereby Rap1 participates in orchestrating the appropriate pattern of transcription initiation by controlling the position of neighboring nucleosomes and by actively preventing spurious transcription initiation within the NDR. This provides a unified view of how Rap1 controls both gene activation and repression, and the quality of the transcripts produced in terms of coding potential. Because we show that neither RNA polymerase occupancy nor overall transcript levels can be considered *a priori* as faithful predictors of gene expression, our data have important and general implications for the modeling of transcriptional networks.

RESULTS

Rap1 Depletion Promotes Ectopic and Pervasive Transcription Initiation Events

In the course of a previous study aimed at describing the function of Rap1 in transcription termination (Candelli et al., 2018), we generated high-resolution transcription maps using a modified version of the CRAC technique (crosslinking analysis of cDNA; Granneman et al., 2009; Candelli et al., 2018), which allows detecting the position of RNAPII position by sequencing the nascent RNA. We first generated CRAC RNAPII transcription maps under conditions of transient depletion of Rap1 from the nucleus with the Anchor Away technique (Candelli et al., 2018; Haruki et al., 2008).

Upon nuclear depletion of Rap1 for 2 hr, we observed the expected effects of up- and downregulation of RNAPII occupancy at genes containing a Rap1 site within 500 nt upstream of the TSS (Tables S1 and S2, 334 genes, Figure 1A). These alterations are specific because they were not observed for a set of randomly chosen, expression-matched genes that did not contain a Rap1 site in their promoter region (Figure 1B; Table S1). In a large number of cases (34%, false discovery rate [FDR] <0.05), gene expression was downregulated in the absence of Rap1, and only a small set of genes (1.8%, FDR <0.05) showed statistically significant upregulation (Figure S1A; Table S2).

Surprisingly, a closer look revealed that the depletion of Rap1 is associated with a major global increase in the RNAPII signal in the promoter regions of Rap1 targets (Figure 1A, left of TSS). This increase was observed by a conservative estimate at 31% of all the analyzed Rap1 targets (FDR <0.05, see STAR Methods; Table S3) but not for control genes and is best illustrated by sorting the same set of genes based on the CRAC signal upstream of the TSS (Figures 1C and 1D).

We considered that some of these effects could be indirect or due to the co-sequestration of Rap1-interacting factors in the cytoplasm. Therefore, we generated genome-wide RNAPII distribution maps after 10 and 20 min of Rap1 depletion using the auxin degron system (Figure S1B; Nishimura et al., 2009). Comparison of RNAPII differential profiles sorted as in Figure 1A revealed similar effects on gene expression even at very early time points of Rap1 depletion (Figure S1C), with 85% of overlap between downregulated features after 2 hr (anchor away) or 20 min (auxin degron) of Rap1 depletion (see Table S2).

Importantly, increased RNAPII signals in gene promoters appeared as early as 10 min after inducing Rap1 depletion and were more pronounced at later time points (Figures S1C).

Analysis of individual genes (Figures 1E–1H and S1D) confirmed that depletion of Rap1 leads to the appearance of RNAPII signals within the NDR, reflecting the unexpected emergence of novel, ectopic transcription start sites (eTSS). Ectopic initiation was found to arise in one (Figures 1E–1G) or both (Figure 1H) directions of transcription relative to the site of Rap1 binding and in a manner that is not dependent on the orientation of the non-palindromic Rap1 site (data not shown). The presence of eTSS was often associated (39% of total events, FDR <0.05) with a decreased RNAPII

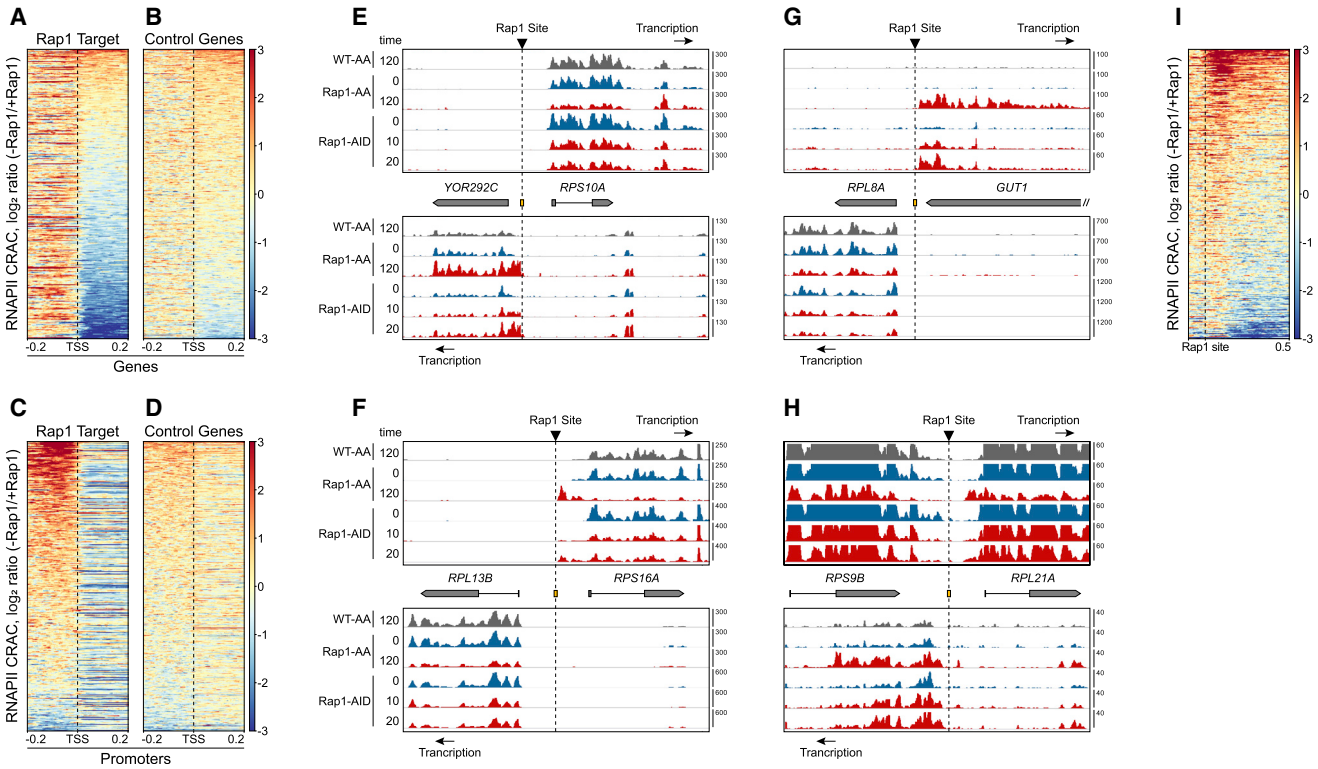


Figure 1. RNAPII CRAC Analysis Reveals Early Appearance of Ectopic Transcription upon Rap1 Depletion

(A–D) Heatmaps illustrating the distribution of the RNAPII CRAC signal change (log_2 ratio) at Rap1 target (334) or control genes (424) upon Rap1 depletion by the anchor-away technique for 2 hr. Features are aligned on the TSS and sorted by decreasing average signal within the first 200 nucleotides of the gene body (A and B) or the 200 nt upstream of the TSS (C and D).

(E) Snapshot showing the RNAPII CRAC signal at the *RPS10A-YOR292C* locus upon Rap1 depletion by the anchor away (Rap1-AA) or the auxin (Rap1-AID) methods for the times indicated. WT-AA indicates an anchor away strain containing a non-tagged Rap1. The position of the Rap1 binding site and the direction of transcription are indicated.

(F) The same as in (E) for the locus *RPL13B-RPS16A*.

(G) The same as in (E) for the locus *RPL8A-GUT1*.

(H) The same as in (E) for the locus *RPS9B-RPL21A*.

(I) Heatmap illustrating the RNAPII CRAC signal change (log_2 ratio) upon Rap1 depletion around Rap1 sites that do not have an annotated coding gene within the downstream 500 nt. Genomic regions are aligned on the Rap1 site and sorted by decreasing signal.

See also Figure S1.

CRAC signal in the body of the downstream gene (Figures 1F, 1H, and S1D, top) but was also observed concomitantly with the apparent upregulation of downstream transcriptional activity observed at a minority (1.8% of total) of Rap1 targets (Figures 1E and S1D, middle), prefiguring diverse effects of ectopic initiation on gene expression (see below).

In many cases, we also observed the appearance of non-coding transcription in the opposite direction from a Rap1 target gene, indicating that Rap1 restricts the intrinsic bidirectionality of many promoters and limits pervasive transcription at the level of initiation (Figures 1G, 1H, and S1D, bottom). To assess the generality of this finding, we aligned all the Rap1 sites that lack an annotated coding gene within the downstream 500 nt in either one of the two possible orientations (n = 304) and profiled the log_2 ratio of the RNAPII CRAC signals observed in the absence or presence of Rap1. As shown in Figure 1I, many novel non-coding, transcription events are generated upon Rap1 depletion, generally initiating immediately downstream of the Rap1 site. We estimated that non-coding ectopic initiation occurs in at least 25% of Rap1-bound loci that do not have a downstream annotated gene (FDR <0.05, Table S4). Finally, alterations in gene expression were not exclusively associated with ectopic initiation since some genes displayed decreased transcription initiation from canonical sites in the absence of an eTSS (e.g., *RPL8A* and *RPL13B* in Figures 1E and 1F; see below).

Together, these results demonstrate that Rap1 represses ectopic transcription initiation generating non-coding RNAs and 5′-extended RNA isoforms. Many novel transcripts are synthesized extremely rapidly after Rap1 depletion, underlying the appearance of an unexpectedly complex landscape of effects on gene expression.

Diverse Effects of Rap1 Depletion on Gene Expression

Ectopic initiation is expected to generate transcripts with different stabilities and coding potential, which can have multiple consequences for gene expression.

To disentangle overlapping transcription events derived from different initiation sites we mapped the 5′ end of transcripts by TSS sequencing (TSS-seq) (Malabat et al., 2015) after nuclear depletion of Rap1 for 60 min, which we estimated to be a good compromise for reliably detecting eTSSs while minimizing secondary effects. To detect transcripts that might be unstable, we also performed TSS-seq in cells defective for nuclear quality control (*rrp6Δ*) or NMD (*upf1Δ*), which degrades RNAs containing premature stop codons. As expected, many novel sites of transcription initiation were specifically detected at Rap1-bound loci upon depletion of Rap1, but not at control genes under the same conditions or after addition of rapamycin to a control strain in which Rap1 was not FKBP12-rapamycin-binding (FRB)-tagged (Figures 2A, 2B, S2A, and S2B).

eTSSs (see Table S5 for a set of verified sites) were distributed over a range of several hundred nucleotides around Rap1 sites (Figures S2A and 2C), with a large fraction (42%) within the 150 nt surrounding the site (Figure 2C). Differential analyses of the TSS-seq signal in *upf1Δ* and *rrp6Δ* cells (see STAR Methods) revealed that many eTSSs belong to RNAs that are significantly sensitive to NMD (42%, FDR <0.05; Table S5; Figure 2D), while only roughly 6% are sensitive to nuclear degradation (FDR <0.05, Table S5, Figure S2C).

The cases described below illustrate the variegated and sometimes complex impact on gene expression ensuing from this deregulation in TSS selection fidelity.

Ectopic Initiation and Downregulation of Gene Expression

A marked downregulation of canonical initiation was observed at many Rap1 target genes (Figure 2A), which mirrors the downregulation observed by RNAPII CRAC (Figures 1A–1D). To assess to what extent this downregulation is associated to the occurrence of ectopic initiation, we calculated the signal ratio change at the canonical TSS of genes containing an upstream eTSS (Table S5). The distribution of these values is shown in Figure 2E, together with the distribution of TSS signal changes at non-Rap1 targets. Roughly 65% of the canonical TSSs of genes containing at least one upstream eTSS are significantly downregulated (FDR <0.05), which testifies to the strong correlation between ectopic initiation and gene downregulation.

The paradigmatic case of the *PRE2* gene is shown in Figures 2F and 2G. Ectopic initiation generates a transcript starting roughly 140 nt upstream of the canonical *PRE2* TSSs, which contains two upstream open reading frames (ORFs) and is thereby sensitive to NMD (Figure 2F, cf. tracks –Rap1 and –Rap1/*upf1Δ*; Figure 2G, cf. lanes 2 and 3 to 5 and 6). The appearance of this novel 5′-extended RNA is accompanied by a marked decrease in transcription starting at the canonical *PRE2* TSS (Figures 2F and 2G) and a consistent decrease in the production of the Pre2 protein, as shown by western blot (Figure 2H). Thus, lower expression from the canonical TSS and production of an unstable RNA explain the downregulation of *PRE2* expression in the absence of Rap1.

Apparent Upregulation of Gene Expression

Besides cases of downregulation, we observed apparent gene upregulation upon selection of an upstream eTSS. One example is shown in Figure 2F for the *NAT3* locus (see also *YOR292C*, Figure 4F). In the absence of Rap1, an upstream TSS is selected at levels markedly higher than the natural TSS, leading to increased transcription levels from an altered start site. A consistent fraction of these RNAs are degraded in the cytoplasm by the NMD pathway as they contain upstream ORFs (uORFs) (Figure 2F), but the molecules that escape degradation are still more abundant than the correctly initiated *NAT3* RNAs. Thus, although both transcription and steady-state RNA levels for *NAT3* (and *YOR292C*) appear to be upregulated in the absence of Rap1 (potentially qualifying Rap1 as a repressor) the RNAs produced do not have the same coding potential as the RNA initiated at the canonical TSS.

An interesting variant of apparent upregulation is represented by the *RXT3* locus (Figures 2F–2H). In this case, transcription is naturally started from two TSS clusters, one that is internal to the ORF, and a second upstream of the natural ATG. Use of the internal site leads to the production of a truncated protein, lacking the first 51 aa. Depletion of Rap1 induces strong repression of the natural TSS, which is accompanied both by the strong induction of the internal TSS and by the selection of an additional upstream eTSS generating NMD-sensitive transcripts. This leads to an overall increased transcriptional and steady-state RNA signal associated with the *RXT3* locus, which, however, translates to the increased production of a truncated isoform and to decreased levels of the normal protein (Figure 2H).

These results strongly suggest that many cases of apparent transcriptional upregulation in the absence of Rap1 in reality hide a constellation of scenarios that generally converge on downregulation of gene expression.

Bona Fide Upregulation following Rap1 Depletion

Alterations in TSS usage in the absence of Rap1 can also lead to a bona fide increase in gene expression as illustrated by the case of *APS2* (Figures 2F–2H). In the presence of Rap1, transcription initiates at two sites: the ATG proximal site is responsible for the expression of the functional *APS2* RNA, whereas a preferentially used distal site leads to the expression of a transcript that contains premature stop codons and is subject to NMD (Figures 2F and 2G, cf. lanes 1 and 4). Upon Rap1 depletion, the ATG proximal site is favored over the upstream site (Figures 2F and 2G, lanes 3 and 6), leading to the increased production of a stable *APS2* RNA and a functional Aps2 protein (Figure 2H). In this case, Rap1 functions as a bona fide repressor of *APS2* expression, but through a non-canonical mechanism that favors the usage of a non-functional TSS.

Rap1 Inhibits Transcription Initiation in a Heterologous Context

A large fraction of ectopic initiation events arise predominantly around the Rap1 binding site (Figure 2C), suggesting that Rap1 controls ectopic initiation by physically hindering access to these cryptic transcription initiation sites. To test this hypothesis, we inserted a Rap1 binding site or a mutated site at various distances from the TSS of a non-coding RNA (*NEL025C*) driven by the *DLD3* bidirectional promoter. Initiation at the *NEL025C* TSS allows expression of the *CUP1* gene, which confers copper-resistant growth to *cup1Δ* yeast cells. The presence of a

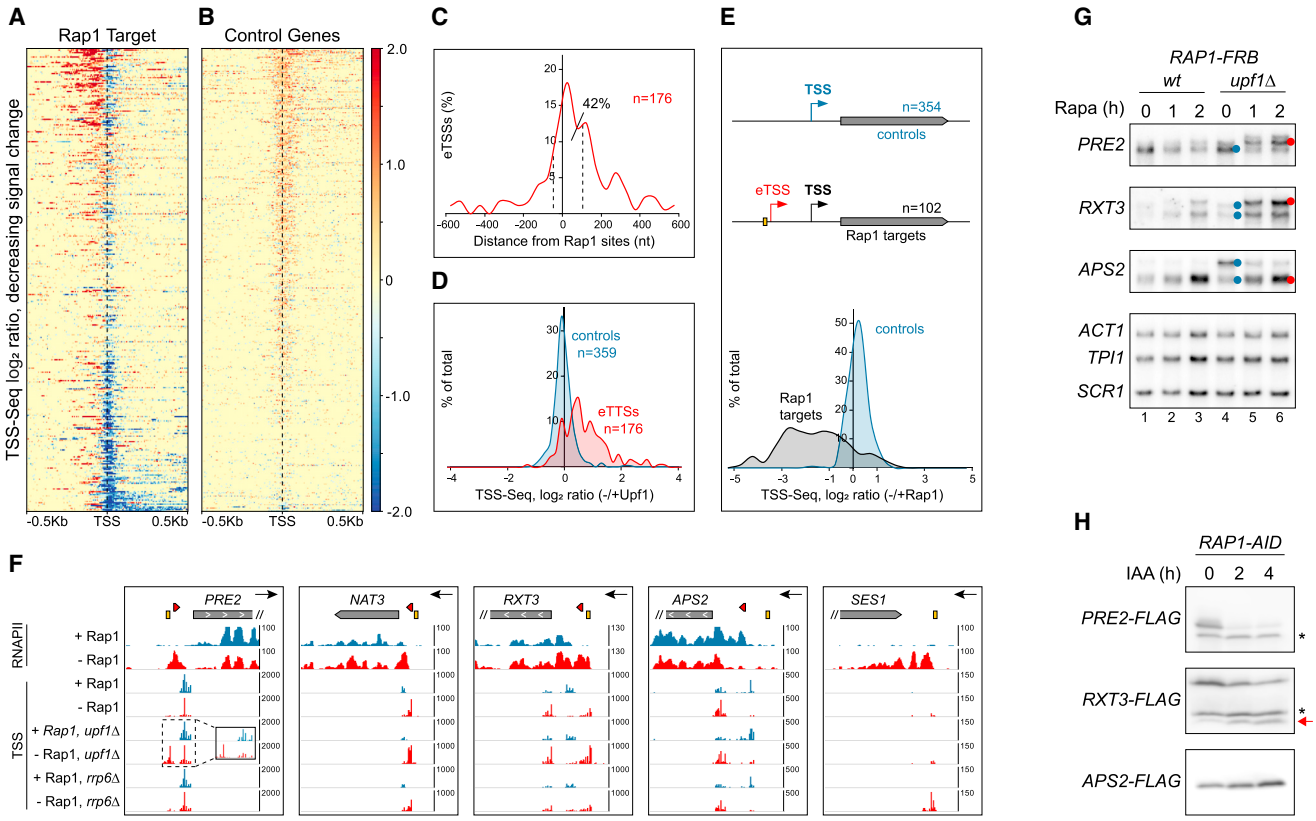


Figure 2. Diverse Impact of Ectopic Transcription Initiation on Gene Expression

(A and B) Heatmaps showing the TSS-seq signal change (log_2 ratio) upon nuclear depletion of Rap1 at target genes (A) or controls (B). Features are sorted for decreasing TSS-seq signal change and aligned on the canonical TSS of target (A) or control (B) genes.

(C) Distribution of ectopic initiation events relative to Rap1 binding sites. The fraction of total eTSSs in 50-nt bins is plotted against the distance from the Rap1 site. 42% of all eTSSs are located within the window indicated by the dashed lines.

(D) A large fraction of eTSSs generate RNAs that are degraded by NMD. TSS-seq signals at ectopic initiation sites have been computed in the absence or presence of Upf1. The binned distribution of the log_2 ratio of these signals (red) is shown in comparison to the same analysis performed on the TSS of control genes (non-Rap1 targets) that are not expected to be affected by NMD, blue. The eTSS population is significantly upregulated in *upf1Δ* cells relative to controls ($p = 4.1E^{-43}$, two-tailed Student's t test).

(E) The occurrence of ectopic initiation strongly correlates with downregulation of normal transcription initiation. Analysis performed on features containing an eTSS in their upstream region (red arrow in the scheme, Table S5). The distribution of TSS-seq signal change (log_2 ratio –Rap1/+Rap1) at the canonical TSS of these genes (black arrow in the scheme) is compared to the distribution of TSS-seq signal change at control genes (blue). In the presence of ectopic initiation, transcription initiated at a canonical downstream site is significantly downregulated ($p = 6.3E^{-21}$, two-tailed Student's t test).

(F) Snapshots illustrating the alterations in TSS usage upon depletion of Rap1. The RNAPII CRAC signal is shown in parallel for comparison. Track colors for RNAPII CRAC and TSS-seq as in Figure 1. The presence of a red arrowhead in the scheme indicates the existence of a small uORF, which induces NMD sensitivity. Only the relevant strand is shown here (black arrow), but complete snapshots are in Figures 3 and S4.

(G) Northern blot analysis of transcripts produced at the indicated loci upon Rap1 depletion, in different genetic backgrounds. Transcripts produced in the presence of Rap1 are indicated by a blue dot in the *upf1Δ* series of samples. Transcripts that are specifically produced or upregulated in the absence or Rap1 are indicated by a red dot.

(H) Western blot analysis of proteins produced from HA-tagged *PRE2*, *RXT3*, and *APS2* loci during Rap1 depletion for the indicated times. A red arrow indicates the position of the N-ter truncated isoform of Rxt3 produced when Rap1 is depleted. The asterisk indicates a cross-reacting protein that can be used as a loading control. See also Figure S2.

Rap1 binding site prevents normal expression of *CUP1* when located at 20 or 85 nt upstream of the TSS but is permissive for expression when located 295 nt upstream (Figures S2E and S2F). A similar result was obtained using the strong *ACT1* promoter. Inhibition was dependent on the sequence of the Rap1 binding site, and on the presence of Rap1 because depletion of the latter led to restoration of transcription initiation as detected by RT-qPCR both for pACT1 and p20 constructs (Figure S2G). From this experiment, we conclude that Rap1 can inhibit gene expression when binding close to the TSS even in a heterologous context.

Altered Nucleosomal Architecture Is Associated with Spurious Transcription Initiation

It has been shown that depletion of Rap1 is linked to the appearance of micrococcal nuclease (MNase)-resistant segments in

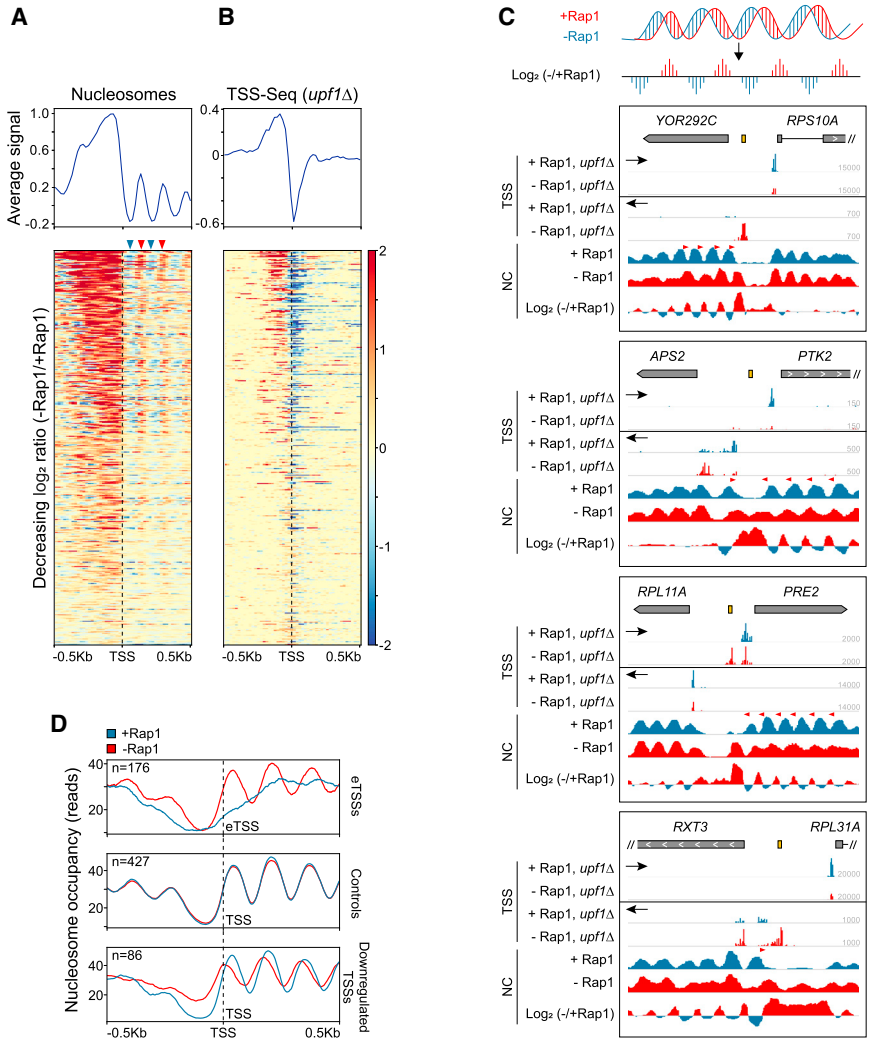


Figure 3. Nucleosome and TSS Selection Changes upon Depletion of Rap1

(A) Heatmap showing the MNase sequencing (MNase-seq) signal change (log_2 ratio) upon Rap1 depletion at Rap1 target genes aligned by their TSS. The MNase-seq signals in the presence or absence of Rap1 at the same genes are shown in Figure S3. Features are sorted by decreasing signal change. Alternate stripes of red and blue downstream of the alignment point (red and blue arrowheads) indicate changes in the phasing of the genic nucleosome array at many genes when Rap1 is depleted (see also C).

(B) Genes are aligned and sorted as in (A), but the log_2 ratio of the TSS-seq signal change upon Rap1 depletion is shown, indicating that changes in TSS usage strongly correlate with nucleosome positioning changes.

(C) Individual examples of the correlations between changes in the pattern of initiation and the altered MNase-seq profile. The movement of nucleosomes in Rap1-depleted cells (red tracks) is indicated by small red arrowheads in the wild-type tracks (blue) and by the characteristic sigmoidal pattern of the log_2 ratio due to nucleosome dephasing (see scheme on top). The latter is responsible for the "striped" pattern seen in (A).

(D) Top: aggregate plots illustrating the position of ectopic TSSs relative to newly positioned nucleosomes in the absence of Rap1. MNase-seq signals have been aligned to ectopic TSSs (eTSSs) upon Rap1 depletion (red) or in the presence of Rap1 as a control (blue). Middle: MNase-seq signals in the absence or presence of Rap1 have been aligned to the TSS of genes that are not Rap1 targets. Bottom: nucleosomes have been aligned to the TSS of genes that are downregulated in the absence of Rap1 but that do not contain ectopic initiation sites in their upstream NDR.

See also Figure S3.

many NDRs (Badis et al., 2008; Hartley and Madhani, 2009; Knight et al., 2014; Kubik et al., 2015). Increased nucleosome occupancy (or decreased MNase accessibility) in NDRs is generally linked to inhibition of transcription initiation (Jiang and Pugh, 2009; Lieleg et al., 2015; Shivaswamy et al., 2008), which is seemingly at odds with the appearance of novel TSSs. We considered the possibility that ectopic initiation and altered nucleosome occupancy might occur independently at distinct loci in the absence of Rap1. Therefore, we analyzed the occurrence of ectopic initiation at sites of high nucleosome occupancy change upon depletion of Rap1.

Consistent with earlier reports (Knight et al., 2014; Kubik et al., 2015), depletion of Rap1 led to a markedly increased nucleosome occupancy in roughly half of the NDRs bound by Rap1 (Figures S3A and S3B, regions aligned on TSSs; Figures S3E and S3F, regions aligned on the Rap1 site) but not at NDRs not bound by Rap1 (Figures S3C and S3D). The differential (log_2 ratio) MNase resistance signal also revealed a striped pattern (blue and red arrowheads, Figure 3A) indicating that in many instances at least part of the nucleosomal array bordering the NDR is phase-shifted relative to the wild-type pattern and moves toward the unoccupied Rap1 site (see individual cases in Figure 3C). Thus, increased nucleosome occupancy in Rap1-dependent NDRs can result from +1 nucleosome shifting, the addition of extra nucleosomes, or both.

Comparison of nucleosome occupancy and TSS-seq changes at Rap1-bound NDRs (Figures 3A and 3B) demonstrates a clear correlation between altered nucleosome occupancy and changes in TSS usage (both emergence of eTSS and decreased use of normal TSS). This indicates that in the absence of Rap1 the two events occur concomitantly, which is also illustrated by the examples shown in Figure 3C (see also Figure S3H).

Ectopic initiation generally occurs in association with newly positioned nucleosomes as observed at individual genes (see, for instance, *YOR292C*, *PRE2*, and *RXT3* in Figure 3C). This is more generally illustrated by the nucleosome metaprofile at 176 eTSSs (Figure 3D, top), showing that eTSSs are generally positioned on the 5′ edge of the nucleosome, mirroring the position of canonical TSSs relative to the +1 nucleosome (Figure 3D, compare top and middle).

Downregulation of the canonical TSS is frequently associated with the upstream shift of the +1 nucleosome (see, for instance, *RXT3*, *APS2*, and *RPL11A*) in the absence of Rap1, which presumably hinders the site of initiation as previously suggested (Reja et al., 2015; Shivaswamy et al., 2008). This was also observed for a set of genes downregulated in the absence of Rap1 that do not contain eTSSs in their promoter region (Table S5; Figure 3D, bottom).

Finally, in roughly 14% of cases, ectopic initiation occurs within gene coding regions (Table S5). These events are also associated with changes in nucleosome positioning. For instance, at the *RXT3* and *YMR027W* loci (Figures 3C and S3H) the upstream shift of the +1 nucleosome enlarges a small NDR between the +1 and +2 nucleosomes, which presumably favors the use of the internal TSS coding for N-terminal truncated isoforms. At the *PEX12* locus (Figure S3H), the upstream shift of four nucleosomes is associated with increased internal initiation between the +4 and +5 nucleosomes.

Taken together, the genome-wide analyses and the single examples illustrate a strong correlation between the modified nucleosomal architecture and the massive alterations in the selection of the transcription initiation site observed in the absence of Rap1.

The DNA Binding Domain of Rap1 Restores Nucleosome Positioning and Transcription Initiation at Many Rap1 Binding Sites

The mechanism of Rap1 action on nucleosomes and on gene activation and repression is still not well understood in spite of a wealth of studies. Besides the DNA binding domain (DBD), Rap1 contains regions involved in the interaction with chromatin remodeling complexes (e.g., SWI/SNF) and general transcription factors (TFIID), notably in the C-terminal region (Reid et al., 2000; Garbett et al., 2007; Tomar et al., 2008; Johnson and Weil, 2017)

To assess whether the role of Rap1 in controlling the fidelity of initiation is related to the recruitment or function of TFIID or SWI/SNF, we depleted Rap1 from the nucleus of cells ectopically expressing only the DBD of Rap1 (aa 358–601, $Rap1_{DBD}$), the wild-type protein or containing an empty vector as controls. $Rap1_{DBD}$ has been shown to strongly bind DNA *in vitro* (Gilson et al., 1993) and *in vivo*, because it could induce transcription termination at a site where Rap1 roadblocks RNAPII (Candelli et al., 2018). Under these conditions, expression of $Rap1_{DBD}$ alone did not support viability and affected normal yeast growth even in the presence of wild-type Rap1. Such a dominant-negative phenotype is expected if $Rap1_{DBD}$ competes with the wild-type protein for DNA binding.

The growth defects caused by expression of $Rap1_{DBD}$ precluded a reliable analysis of RNAPII distribution by CRAC. However, in spite of their sensitivity to NMD, many 5′-extended transcripts produced in the absence of Rap1 could still be detected in otherwise wild-type cells (Figures 2 and 4). Therefore, we restricted our analyses to these RNA species as a proxy for ectopic initiation and analyzed in parallel the nucleosomal architecture in the same conditions. To our surprise, $Rap1_{DBD}$ suppresses to a very significant extent some of the nucleosome positioning phenotypes of Rap1-deficient cells (Figures S3E–S3G, 4A, and 4B). When sorted based on the strongest effect of Rap1 depletion (Figures 4A and S3E), four main NDR clusters can be identified (Table S6): cluster 1 and 2, containing regions with large (400–600 nt) NDRs and eccentric Rap1 binding; cluster 3, with smaller NDRs and central Rap1 binding; and cluster 4, containing regions generally insensitive to Rap1 depletion. A clear suppression of the nucleosome positioning phenotype was observed in clusters 1–3, which was, however, not complete and most prominently observed in the region immediately surrounding the Rap1 binding site (cf. the differential heatmap profiles, Figures 4A and 4B, and the aggregate plots for clusters 1 and 2, and 3, Figures 4C and 4D). This indicates that a nucleosomal architecture that is close to normal can be maintained around the Rap1 binding site without domains of Rap1 involved in transcriptional activation.

Significantly, partial suppression of ectopic transcription initiation by $Rap1_{DBD}$ was also observed, generally correlating with restored nucleosome positioning. For example, at the *PRE2* and *YOR292C* loci (Figures 4E and 4F), $Rap1_{DBD}$ restores an apparently normal position of nucleosomes as well as the normal site of transcription initiation and the level of average steady-state RNA (which is higher in the wild-type for *PRE2* and lower for *YOR292C*). Conversely, nucleosome positioning is not restored at the *PTK2* locus (Figure 4G), and ectopic initiation leading to the expression of a 5′-extended RNA is maintained upon expression of $Rap1_{DBD}$. At the *RPL21A* and *RPS23B* loci, containing large NDRs with an eccentrically positioned Rap1 site, $Rap1_{DBD}$ only restores normal position of nucleosomes and transcription initiation around its binding site (Figures S4A and S4B).

$Rap1_{DBD}$ suppression occurred genome-wide, as shown by the comparison of the RNA sequencing (RNA-seq) signal change around eTSSs in the absence of Rap1 or in the presence of $Rap1_{DBD}$. The strong peak in the $\log_2$ signal ratio for –Rap1/Rap1 was considerably reduced when $Rap1_{DBD}$ was expressed (Figure 4H, cf. Rap1/Rap1 to DBD/Rap1), both for eTSS upstream of coding (Figure S4C) or non-coding (Figure S4D) genes. As seen at individual loci, suppression was generally restricted to eTSSs that are proximal to the Rap1 site (i.e., included in a 200-nt window centered on the site, Figures 4I–4J).

For a more quantitative assessment, we plotted the distributions of RNA-seq signals in the presence of Rap1, its absence or the presence of $Rap1_{DBD}$ at the most affected promoters (Figure S4E), regions of non-coding transcription (Figure S4F), and genes (Figure S4G). In all instances, the differences in the distribution of RNA-seq signals were supportive of a partial suppression of the ectopic initiation by $Rap1_{DBD}$ with a strong statistical significance ($p < 6.5E^{-4}$).

Together these results demonstrate that expression of only the DBD of Rap1 is sufficient to restore the canonical position of nucleosomes and suppress ectopic initiation at a significant number of Rap1 binding sites.

The Control on Initiation Fidelity by Rap1 Is Not Mediated by SWI/SNF, RSC, or INO80 CR Complexes

Although $Rap1_{DBD}$ can partially suppress nucleosome positioning phenotypes due to the absence of Rap1, we cannot formally exclude that Rap1 displaces nucleosomes by recruiting CRs. If CRs were downstream effectors of Rap1 action, affecting

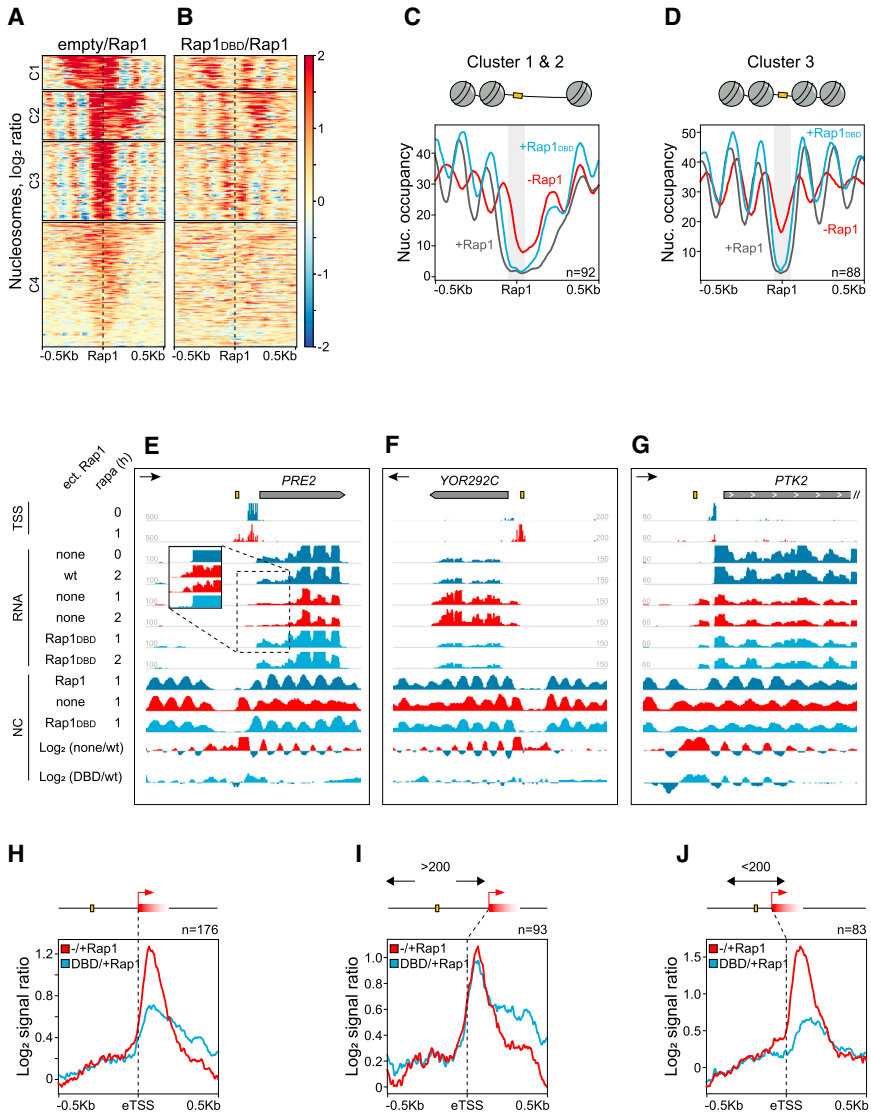


Figure 4. Rap1$_{DBD}$ Partially Suppresses the Gene Expression Phenotypes Linked to Rap1 Depletion

(A and B) Differential map of nucleosome occupancy (log_2 ratio) in the absence of Rap1 (A) or in the presence of the Rap1 DBD (Rap1$_{DBD}$, B) after nuclear depletion of Rap1. Regions are aligned to Rap1 sites and ordered by k-means clustering and decreasing signal change in (A). Clusters 1 and 2 contain large NDRs that have an eccentric Rap1 site relative to flanking nucleosomes.

(C and D) Aggregate plots illustrating the extent of suppression of the nucleosome positioning defect by Rap1$_{DBD}$ in clusters 1 and 2 (C) and 3 (D). For a better visualization, all the regions in cluster 1 have been inversed so that the closest nucleosome lies on the left of the Rap1 site as in cluster 2 (see scheme on top), and the signals from the two clusters have been combined. The most efficient suppression of nucleosome positioning by Rap1$_{DBD}$ occurs in the proximity of Rap1 sites (shaded region).

(E–G) Individual snapshots are shown to illustrate cases of full (E and F) or less-efficient (G) suppression, which correlate with nucleosome positioning changes. Additional examples are shown in Figure S4.

(H) Aggregate plots illustrating the partial suppression of ectopic initiation by Rap1$_{DBD}$ after nuclear depletion of Rap1. The log_2 ratio of the RNA-seq signals detected in the absence or presence of Rap1 (red plot, –/+ Rap1) around ectopic TSSs is compared to the signal detected in the presence of Rap1$_{DBD}$ (blue plot, DBD/+Rap1). The signal is maximum close to the aligned eTSSs where ectopic transcription arises in all regions and is significantly decreased in the presence of Rap1$_{DBD}$.

(I and J) Same as (H) but using only eTSSs distal to the Rap1 site (i.e., located outside of a 200-nt window centered on the Rap1 site, I) or proximal (i.e., located within the 200-nt window, J).

See also Figure S4.

their function should lead to a similar and possibly more general phenotype as depletion of Rap1. We therefore assessed the effect on transcription initiation of depleting the catalytic subunits of the SWI/SNF or RSC complexes (Snf2 and Sth1, respectively). We also co-depleted Isw2 and Ino80, which are ATPases involved in chromatin remodeling that have been implicated, possibly redundantly, in the positioning of the +1 nucleosome (Krietenstein et al., 2016; Whitehouse et al., 2007). As for Rap1 depletion, experiments were also performed in an *upf1Δ* background. The distribution of TSSs at Rap1 targets, sorted according to decreasing levels of nucleosomal changes in Rap1-deficient cells, is shown in Figures 5B–5E. Individual snapshots are shown in Figures 5F–5I, where we also show the changes in nucleosome occupancy in these conditions (S.K., D.C., R. Dreos, S. Mattarocci, M.J. Bruzzone, P. Bucher, D.L., and D.S., unpublished data). The levels and distribution of ectopic initiation were strikingly different in Sth1, Snf2, or Ino80/Isw2-deficient cells compared to cells lacking Rap1. Although novel TSSs were observed in NDRs upon Sth1 and Snf2 depletion, their distribution was qualitatively different from that of Rap1-repressed eTSSs, which strongly argues against an epistatic relationship (Figures 5F–5I). Of note, upon co-depletion of Ino80/Isw2 the pattern of ectopic transcription initiation was complementary and less defined relative to that observed in Rap1-depleted cells, with eTSSs frequently present within genes (Figure 5).

From these experiments, we conclude that Rap1 suppresses ectopic transcription initiation independently of CRs that function at its target sites.

Reb1 and Abf1 Suppress Ectopic Transcription Initiation

Reb1 and Abf1 are GRFs that have similar roles to Rap1 in terms of nucleosome exclusion at NDR (Badis et al., 2008; Ganapathi et al., 2011; Hartley and Madhani, 2009; Kubik et al., 2015). We analyzed the role of both factors in suppressing ectopic initiation at their target genes by profiling RNAPII occupancy by CRAC.

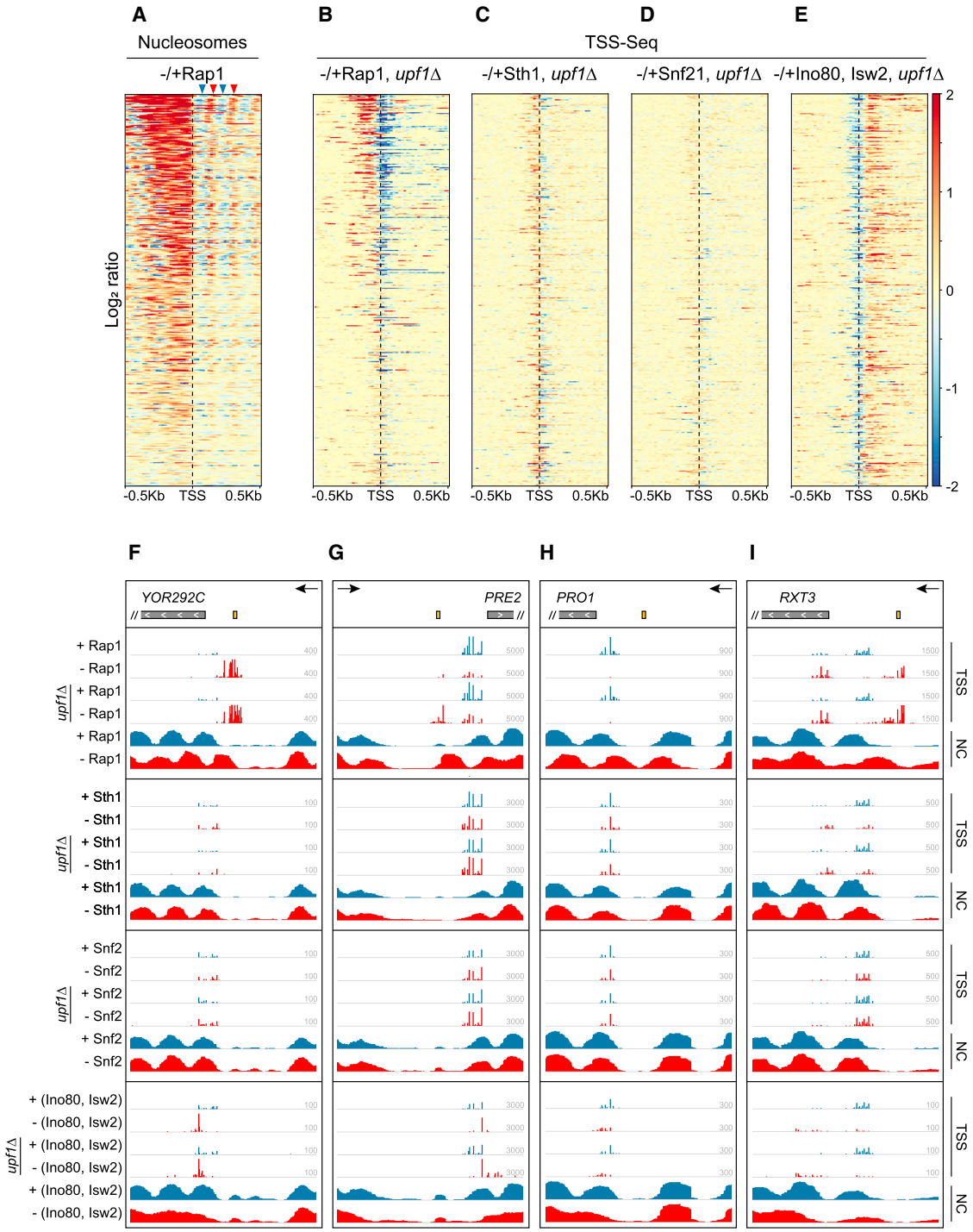


Figure 5. Depletion of CRs Has Distinct Effects on TSS Usage Than on Rap1 Depletion

(A–E) In all heatmaps, features are sorted according to changes in MNase-seq signal upon Rap1 depletion (A) as in Figure 3A. Changes in TSS usage (log_2 ratio) upon depletion of the indicated CR subunits are shown in (C)–(E) and compared to signal changes upon Rap1 depletion (B, same as Figure 3B, shown here for ease of comparison). Only signals derived from cells defective for NMD (*upf1Δ*) are shown, but similar results have been obtained with NMD-proficient cells.

(F) Snapshot comparing TSS and nucleosome changes at the *YOR292C* locus upon depletion of Rap1 or the indicated chromatin remodeler subunits. Data from depleted cells are shown in red. The position of the Rap1 site is indicated.

(G) The same as in (F) for the *PRE2* locus.

(H) The same as in (F) for the *PRO1* locus.

(I) The same as for (F) for the *RXT3* locus.

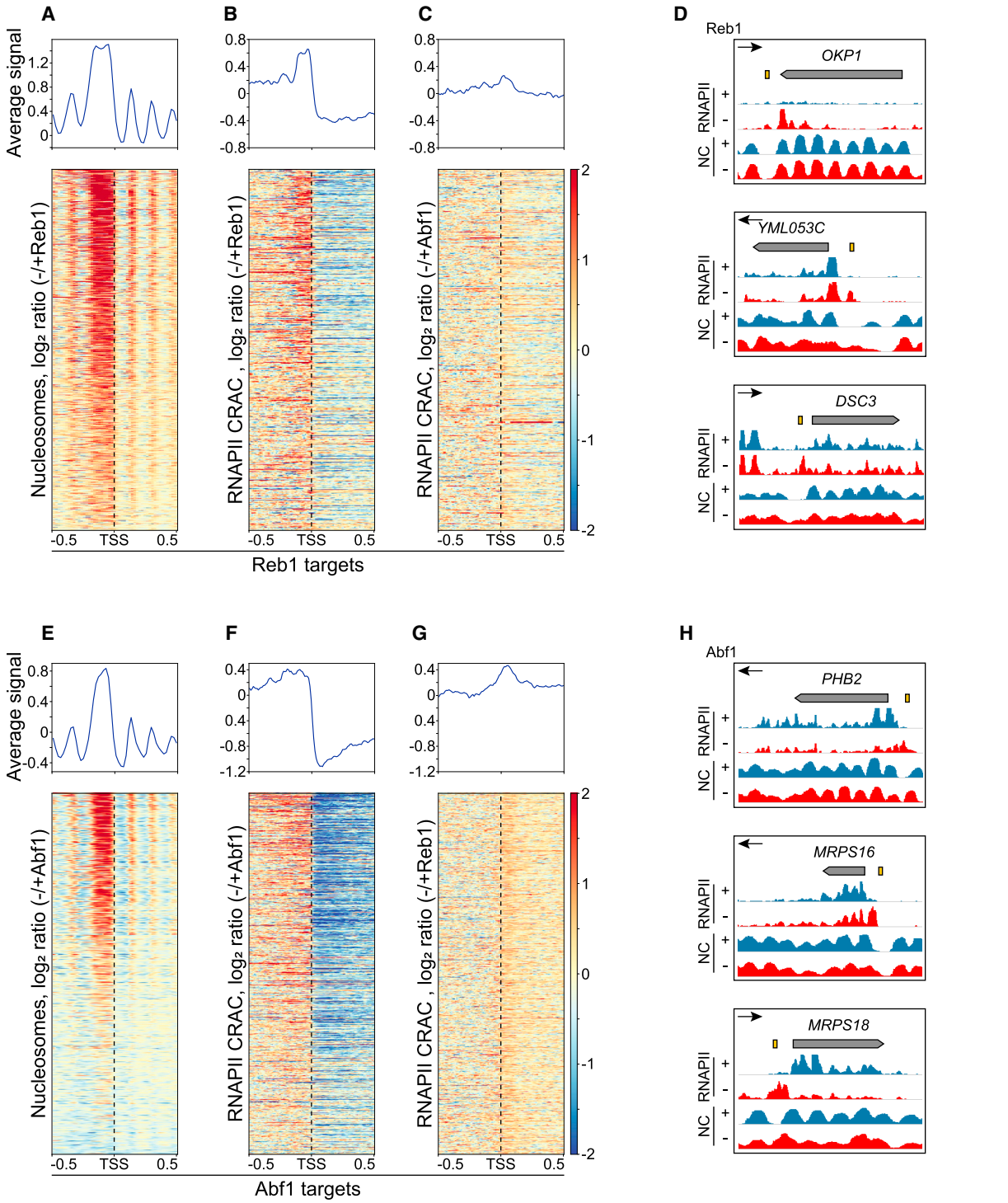


Figure 6. Nucleosome Occupancy and TSS Selection Changes upon Depletion of Reb1 and Abf1

(A) Genes containing a Reb1 site within the upstream 300 nt (n = 650) were aligned to their TSSs and sorted according to decreasing nucleosome changes.

(B) RNAPII CRAC signal changes upon Reb1 depletion. The same set of genes are aligned to canonical TSSs and sorted as in (A).

(C) Same as (B), but the RNAPII CRAC signal change in Abf1-deficient cells is plotted instead of Reb1-deficient cells (note that this is a specificity control for Abf1 depletion). The control for Reb1 depletion is shown in (G).

(D) Individual snapshots illustrating the occurrence of ectopic initiation at Reb1 targets upon Reb1 depletion. Top: depletion of Reb1 induces non-coding transcription antisense to OKP1 is shown. Middle and bottom: the absence of Reb1 is associated to changes in nucleosome positioning and ectopic initiation upstream of *YML053C* and *DSC3* is shown.

(legend continued on next page)

Depletion of either factor led to increased nucleosome occupancy in NDRs (Figures 6A and 6E; data from Kubik et al., 2015), as previously reported. We also observed a very frequent upstream shift of the intragenic nucleosomal array relative to the canonical TSS, as observed for Rap1 depletion, responsible for the characteristic ''striped'' pattern in the differential signal (Figures 6A and 6E).

Heatmaps reporting the changes in the RNAPII CRAC signal upon depletion of Reb1 and Abf1 (Figures 6B and 6F, respectively) clearly demonstrate the occurrence of transcription initiation in the NDR upstream of Reb1 and Abf1 targets, which is generally associated with gene downregulation (see also the summary plots associated to each heatmap). As a control, Reb1 and Abf1 depletion did not induce ectopic initiation at the non-cognate targets (respectively, Abf1 dependent in Figure 6G, and Reb1 dependent in Figure 6C). As for Rap1, changes in nucleosome positioning induced by the absence of Reb1 or Abf1 strongly correlate with the appearance of novel eTSSs and the downregulation of initiation at the wild-type site. In many instances, we observed effects on gene expression that were not previously noticed based on steady-state RNA-level changes, most likely because they were masked by the overlapping with 5′-extended transcripts (data not shown).

Together these results extend the essential role of Rap1 in controlling the fidelity of transcription initiation to two other GRFs and suggest that other transcription factors may behave similarly.

DISCUSSION

In this study, we have addressed the effects on transcription and on gene expression of depleting Rap1 and two other GRFs, Abf1 and Reb1. These factors have been known for decades to affect the expression of many highly expressed genes, yet, their mode of action in gene activation or repression has remained relatively obscure. We show that Rap1 and other GRFs control the fidelity of transcription initiation, preventing inappropriate and non-coding transcription events from taking place within the NDRs to which they bind (Figure 7).

Rap1 and Nucleosome Positioning

Many earlier studies have shown that Rap1 is important for the size and nucleosome occupancy of the NDRs to which it binds (Badis et al., 2008; Ganapathi et al., 2011; Hartley and Madhani, 2009; Kubik et al., 2015), which we also have observed. One important facet of our work is the demonstration that the DBD of Rap1 alone can restore nucleosome positioning at many sites of Rap1 binding, although restoration is generally restricted to the region around the binding site (Figures 4A–4D and S3E–S3G). We propose that Rap1 constrains nucleosomes at least in part by a steric hindrance mechanism, consistent with earlier results from the Morse lab (Yu et al., 2001). *In vitro* reconstitution experiments with purified factors have shown that GRFs are not intrinsically required for NDR formation but are important for positioning the +1 nucleosome in the presence of ISW2/ISW1 remodelers (Krietenstein et al., 2016). By extension, we propose that Rap1 constitutes a physical barrier against which genic nucleosomes are ''pushed'' by ISW2/ISW1, a ''place-holder'' function for which in many instances only DNA binding is absolutely required.

We observed that, when the Rap1 binding site is eccentric in large NDRs (e.g., clusters 1 and 2 in Figures 3A–3C), $Rap1_{DBD}$ can only constrain proximal nucleosomes, which implies that the domains missing in $Rap1_{DBD}$ are important for maintaining the integrity of the NDR at some distance from its binding site. This might be related to the role of additional factors associated with Rap1, such as Fhl1/Ifh1, Sfp1 (FIS), and Hmo1 (Knight et al., 2014; Reja et al., 2015). These factors are generally associated with Rap1 in large NDRs, notably upstream of ribosomal protein genes, and are highly enriched in clusters 1 and 2 (Figure S5). This notion is fully consistent with earlier results showing that the absence of Hmo1 induces an upstream shift of the +1 nucleosome that inhibits gene expression (Kasahara et al., 2011; Reja et al., 2015) and would imply that components of the FIS and/or Hmo1 are not bound, or cannot restrict nucleosome displacements when $Rap1_{DBD}$ is bound to the NDR (Figure 7).

The Chicken and Egg Issue of Nucleosomes and Transcription Initiation

The presence of newly positioned nucleosomes within the NDR and ectopic initiation might be fully independent events. However, we observed that displaced nucleosomes and eTSSs are frequently associated, suggesting the existence of a causal connection. Ectopic initiation occurs roughly 12–15 nt downstream of the upstream border of the closest newly positioned nucleosome (Figure 3D, top and middle), which is similar to the relative position of canonical initiation and the +1 nucleosome (Hughes et al., 2012; Lee et al., 2007; Tsankov et al., 2010; Whitehouse et al., 2007). Whether it is initiation that specifies the exact position of the +1 nucleosome or the latter that directs the position of initiation is still a matter of debate (for recent reviews, see Jiang and Pugh, 2009; Lieleg et al., 2015; Struhl and Segal, 2013).

Some support for the latter hypothesis might come from the changes in the TSS distribution after the double depletion of Ino80 and Isw2 (Figures 5F–5I). In this case, it can be envisioned that the primary consequence of Ino80 and Isw2 depletion is to induce a downstream shift of the +1 nucleosome. If initiation were fully independent of the presence of a +1 nucleosome, it should occur at the wild-type position even when the nucleosome is misplaced. Rather, we generally observe a concomitant downstream shift of initiation that appears to ''accompany'' the shifted nucleosome. However, a direct role of Ino80 and Isw2 in specifying the position of the TSS independently of nucleosomes cannot be excluded.

(E) Abf1 gene targets (n = 781) are aligned to their TSSs and sorted according to decreasing nucleosome changes as in (A).
(F) RNAPII CRAC signal changes at Abf1 target genes upon Abf1 depletion as in (B).
(G) Same as (F), but the changes in RNAPII CRAC signal upon Reb1 depletion is shown as a control for the effect of Reb1 depletion at unrelated genes. Genes containing both a Reb1 and Abf1 binding site have been excluded.
(H) Individual snapshots illustrating the occurrence of ectopic initiation at Abf1 targets upon Abf1 depletion.

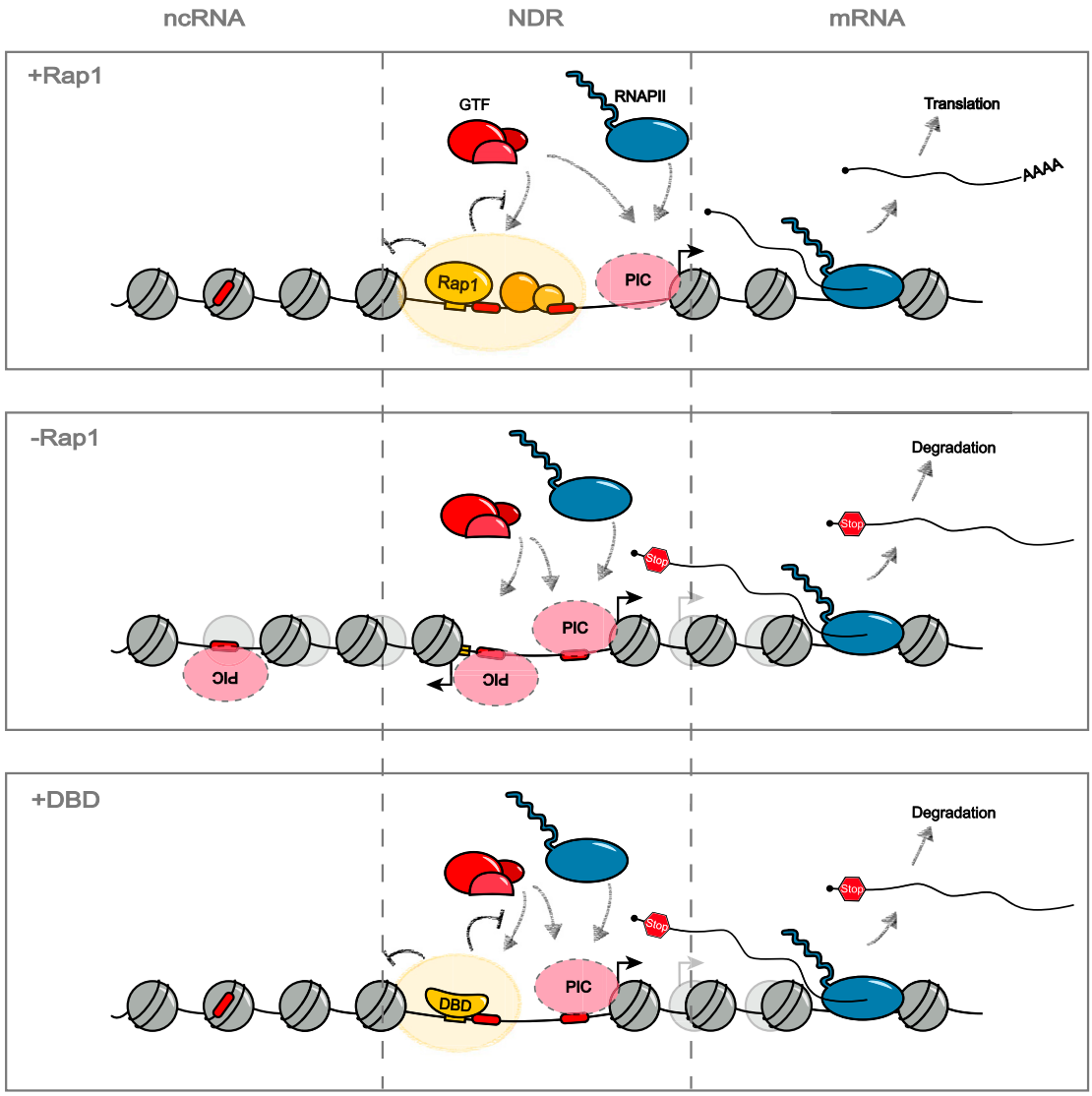


Figure 7. Model Illustrating the Role of Rap1 and GRFs in the Suppression of Ectopic Initiation

In wild-type cells (top), Rap1 (and other GRFs) participates in the positioning of nucleosomes that are excluded from the NDR, at least in part by a steric mechanism. In the case of Rap1, the associated FIS (Fhl1, Ifh1, and Sfp1) and Hmo1 could also be required for excluding nucleosomes from the NDR. Rap1 or other GRFs (yellow ellipse) also prevent general transcription factors (GTFs) from accessing the DNA at cryptic sites (red boxes) and generate ectopic initiation events. In the absence of Rap1 (middle), these cryptic PIC formation sites become accessible, leading to the production of non-coding (leftward) or miscoding RNAs (rightward). When only the DBD of Rap1 is expressed (bottom), in many instances, nucleosome positioning is restored, mainly in the vicinity of the site, together with the suppression of cryptic PIC formation and eTSSs. In this example, we show one of the large, Rap1-dependent NDRs in which restoration only occurs on the side of the Rap1 binding site (yellow box).

One interesting possibility is that the position of nucleosomes and TSSs influence each other in the absence of Rap1 and possibly in normal conditions. Nucleosomes might encroach on the NDR when released from the steric control of Rap1 or other GRFs, but their precise position would be influenced by the occurrence and position of ectopic transcription initiation.

Rap1 Controls the Fidelity of Transcription Initiation

The massive occurrence of novel initiation events observed in the absence of Rap1 was unexpected. In many instances, generation of these novel TSSs must imply the formation (or the activation) of a novel PIC. Because there is no reason to hypothesize the existence of some evolutionary pressure to maintain these ectopic PICs, we must conclude that they form promiscuously in the absence of a specific negative control mechanism.

Our results indicate that Rap1 and other GRFs exert such a negative control when bound to their sites. The observation that $Rap1_{DBD}$ preferentially suppresses eTSSs at proximal sites (Figures 4E–4J, S4A, and S4B) suggests that, akin to its effect

on proximal nucleosomes, Rap1 sterically hinders use of these cryptic sites of initiation.

Approximately half of the eTSSs that we detected are located more than 100 nt from the site of Rap1 binding and are generally not, or not efficiently, suppressed by $Rap1_{DBD}$. This raises the question of how wild-type Rap1 suppresses ectopic initiation at a distance from its binding site. It has been shown that Rap1 can bind with lower affinity to DNA sequences that do not have a recognizable Rap1 recognition sequence (Feldmann and Galletto, 2014), and it is possible that suppression of distal eTSSs depends on non-canonical Rap1 binding to these sites in a manner that is not detectable by chromatin immunoprecipitation (ChIP) experiments.

However, we favor the hypothesis (Figure 7) that the negative control on ectopic initiation is exerted distally by a similar mechanism as the control of distal nucleosomes in large NDRs, i.e., via the intervention of factors whose binding depends on Rap1 such as FIS, and Hmo1. Consistent with this hypothesis, it was shown that the absence of Hmo1 generates 5′-extended RNAPII signals and ectopic binding of TFIIB, both of which are strong indicators of ectopic upstream initiation (Reja et al., 2015). Importantly, because binding of Rap1 cannot inhibit transcription initiation at a distance in a heterologous context, it must be postulated that Rap1-dependent NDRs additionally contain specific features (e.g., binding sites for some of the aforementioned factors; MacIsaac et al., 2006; Badis et al., 2008; Knight et al., 2014; Reja et al., 2015) that make them susceptible to Rap1 negative control.

In a parallel study based on the analysis of two non-coding transcription units, upstream of the *MLP1* and *IME1* genes, Wu et al. (2018) propose an alternative model according to which the DBD is not sufficient to suppress ectopic non-coding transcription initiation. It is unclear to what extent this result can be extended to a genome-wide scale, but in our experimental setup we observed eTSS suppression by $Rap1_{DBD}$ at the *MLP1* and *IME1* loci (data not shown). The rational for these differences is unclear but might relate to the fact that the DBD construct used in the Wu et al. study is slightly different from our own and is fused to a tag and a nuclear localization signal, which might lead to a weaker binding and poor suppression at some sites. This would also be consistent with the finding in the Wu et al. study that expression of the Rap1 DBD does not induce the same dominant-negative phenotype that we observed (Wu et al., 2018). However, we cannot exclude that at some sites the binding of Rap1 alone is not sufficient and that additional factors are required.

Impact of Novel TSSs on Gene Expression

One important implication of this study is that neither the assessment of RNAPII levels nor the steady-state levels of RNA within genes can be used as a reliable proxy for inferring gene expression levels, unless information concerning the position of the TSS is also integrated into the equation. This implies that the notion of positively or negatively regulated genes has to be generally revisited for the GRFs that we have studied, but also potentially for all factors that might affect the selection of transcription initiation sites.

We found several examples in which overlapping transcription signals initiated at a canonical and an ectopic TSS leads to an apparent increase in gene expression, or an apparent lack of effect. However, in the majority of cases deconvolution of the two signals using TSS analysis demonstrates a likely decrease in the ultimate outcome of gene expression because the novel transcripts have a different and generally lower coding potential. This already high level of indetermination is further complicated by some uncertainty concerning the fate of the new transcripts that contain premature stop codons or uORFs. Although we show that these RNAs are generally strongly sensitive to NMD, some of them can still be detected in wild-type cells. It is unclear why they escape NMD, but it is possible that some undergo frameshifting during translation and produce proteins with similar composition to the normal gene product.

A similar discrepancy between RNA abundance and protein levels was the starting point of a recent study demonstrating the regulation of a meiotic network by the use of alternative TSSs producing isoforms with different coding potential (Cheng et al., 2018).

Our findings have important implications for the mathematical modeling of complex gene expression networks. Such holistic approaches are often based on the assumption that RNA abundance or RNAPII occupancy directly translate into the expression of a gene product, which, in turn, can influence the network. This is, however, not that frequently verified, and many physiological and non-physiological perturbations might bring about changes in transcription patterns similar to the ones that we describe here, which could be source of inaccuracy for gene network modeling.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Yeast strains, plasmids and oligonucleotides
  - RNA and protein analysis
  - RNAPII CRAC
  - TSS sequencing
  - MNase-seq
  - Dataset processing and data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and seven tables and can be found with this article online at https://doi.org/10.1016/j.molcel.2018.11.037.

ACKNOWLEDGMENTS

We wish to thank K. Adelman, U. Aiello, A. Oldfield, J. Gros, S. Marcand, O. Porrua, M. Werner, and T. Villa for critical reading of the manuscript, K. Adelman, A. Oldfield, R. Jothi, A. Wu, and F. van Werven for sharing results before publications, and C. Maufrais for help with TSS-seq analyses. We also wish to thank Yan Jaszczyszyn for expert technical help in preparing CRAC libraries for sequencing. This work was supported by the Centre National de la

Recherche Scientifique (CNRS), the Fondation pour la Recherche Medicale (FRM, programme équipes 2013), l'Agence National pour la Recherche (ANR, grants ANR-12-BSV8-0014-01 and ANR-16-CE12-0022-01 to D.L.), the Labex Who Am I? (ANR-11-LABX-0071 et Idex ANR-11-IDEX-0005-02 to D.L.), and the Swiss National Fund (grant no. 31003A_170153 to D.S.). D.C. and T.C. have been supported by fellowships from the French Ministry of Research. D.C. was also supported by a fellowship from the Ligue contre le Cancer. This work has benefited from the facilities and expertise of the high-throughput sequencing core facility of I2BC (Centre de Recherche de Gif: http://www.i2bc-saclay.fr/).

AUTHORS CONTRIBUTION

Conceptualization, D.L., D.C., and S.K.; Methodology, T.C., M.B., and D.C.; Software, M.B. and T.C.; Formal Analysis, M.B., D.L., and D.C.; Investigation, D.C., S.K., F.F., C.B., and H.G.; Writing – Original Draft, D.C. and D.L.; Writing – Review and Editing, D.C., D.L., S.K., and D.S.; Funding Acquisition, D.L. and D.S.; Supervision, D.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 6, 2018
Revised: October 29, 2018
Accepted: November 29, 2018
Published: December 20, 2018

REFERENCES

Azad, G.K., and Tomar, R.S. (2016). The multifunctional transcription factor Rap1: A regulator of yeast physiology. Front. Biosci. *21*, 918–930.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., et al. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol. Cell *32*, 878–887.

Bendjennat, M., and Weil, P.A. (2008). The transcriptional repressor activator protein Rap1p is a direct regulator of TATA-binding protein. J. Biol. Chem. *283*, 8699–8710.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Candelli, T., Challal, D., Briand, J.-B., Boulay, J., Porrua, O., Colin, J., and Libri, D. (2018). High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast. EMBO J. *37*. Published online February 15, 2018. https://doi.org/10.15252/embj.201797490.

Cheng, Z., Otto, G.M., Powers, E.N., Keskin, A., Mertins, P., Carr, S.A., Jovanovic, M., and Brar, G.A. (2018). Pervasive, coordinated protein-level changes driven by transcript isoform switching during meiosis. Cell *172*, 910–923.e16.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: The genomics resource of budding yeast. Nucleic Acids Res. *40*, D700–D705.

Feldmann, E.A., and Galletto, R. (2014). The DNA-binding domain of yeast Rap1 interacts with double-stranded DNA in multiple binding modes. Biochemistry *53*, 7471–7483.

Ganapathi, M., Palumbo, M.J., Ansari, S.A., He, Q., Tsui, K., Nislow, C., and Morse, R.H. (2011). Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. Nucleic Acids Res. *39*, 2032–2044.

Garbett, K.A., Tripathi, M.K., Cencki, B., Layer, J.H., and Weil, P.A. (2007). Yeast TFIID serves as a coactivator for Rap1p by direct protein-protein interaction. Mol. Cell. Biol. *27*, 297–311.

Gilson, E., Roberge, M., Giraldo, R., Rhodes, D., and Gasser, S.M. (1993). Distortion of the DNA double helix by RAP1 at silencers and multiple telomeric binding sites. J. Mol. Biol. *231*, 293–310.

Granneman, S., Kudla, G., Petfalski, E., and 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. USA *106*, 9613–9618.

Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell *137*, 445–458.

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: Rapid, conditional establishment of yeast mutant phenotypes. Mol. Cell *31*, 925–932.

Hughes, A.L., Jin, Y., Rando, O.J., and Struhl, K. (2012). A functional evolutionary approach to identify determinants of nucleosome positioning: A unifying model for establishing the genome-wide pattern. Mol. Cell *48*, 5–15.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: Advances through genomics. Nat. Rev. Genet. *10*, 161–172.

Jin, Y., Eser, U., Struhl, K., and Churchman, L.S. (2017). the ground state and evolution of promoter region directionality. Cell *170*, 889–898.e10.

Johnson, A.N., and Weil, P.A. (2017). Identification of a transcriptional activation domain in yeast repressor activator protein 1 (Rap1) using an altered DNA-binding specificity variant. J. Biol. Chem. *292*, 5705–5723.

Kasahara, K., Ohyama, Y., and Kokubo, T. (2011). Hmo1 directs pre-initiation complex assembly to an appropriate site on its target gene promoters by masking a nucleosome-free region. Nucleic Acids Res. *39*, 4136–4150.

Knight, B., Kubik, S., Ghosh, B., Bruzzone, M.J., Geertz, M., Martin, V., Dénervaud, N., Jacquet, P., Ozkan, B., Rougemont, J., et al. (2014). Two distinct promoter architectures centered on dynamic nucleosomes control ribosomal protein gene transcription. Genes Dev. *28*, 1695–1709.

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., and Korber, P. (2016). Genomic nucleosome organization reconstituted with pure proteins. Cell *167*, 709–721.e12.

Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.-L., Rougemont, J., and Shore, D. (2015). Nucleosome stability distinguishes two different promoter types at all protein-coding genes in yeast. Mol. Cell *60*, 422–434.

Kubik, S., O'Duibhir, E., de Jonge, W.J., Mattarocci, S., Albert, B., Falcone, J.-L., Bruzzone, M.J., Holstege, F.C.P., and Shore, D. (2018). Sequence-directed action of RSC remodeler and general regulatory factors modulates +1 nucleosome position to facilitate transcription. Mol. Cell *71*, 89–102.e5.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat. Rev. Mol. Cell Biol. *18*, 548–562.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat. Genet. *39*, 1235–1244.

Lieleg, C., Krietenstein, N., Walker, M., and Korber, P. (2015). Nucleosome positioning in yeasts: Methods, maps, and mechanisms. Chromosoma *124*, 131–151.

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.

Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mRNA decay. eLife *4*, 4.

Marquardt, S., Escalante-Chong, R., Pho, N., Wang, J., Churchman, L.S., Springer, M., and Buratowski, S. (2014). A chromatin-based mechanism for limiting divergent noncoding transcription. Cell *157*, 1712–1723.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal *17*. https://doi.org/10.14806/ej.17.1.200.

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.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat. Methods *6*, 917–922.

Papai, G., Tripathi, M.K., Ruhlmann, C., Layer, J.H., Weil, P.A., and Schultz, P. (2010). TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. Nature *465*, 956–960.

Parnell, T.J., Huff, J.T., and Cairns, B.R. (2008). RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. EMBO J. *27*, 100–110.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: Why, where and how to stop. Nat. Rev. Mol. Cell Biol. *16*, 190–202.

Preti, M., Ribeyre, C., Pascali, C., Bosio, M.C., Cortelazzi, B., Rougemont, J., Guarnera, E., Naef, F., Shore, D., and Dieci, G. (2010). The telomere-binding protein Tbf1 demarcates snoRNA gene promoters in Saccharomyces cerevisiae. Mol. Cell *38*, 614–620.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Reid, J.L., Iyer, V.R., Brown, P.O., and Struhl, K. (2000). Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. Mol. Cell *6*, 1297–1307.

Reja, R., Vinayachandran, V., Ghosh, S., and Pugh, B.F. (2015). Molecular mechanisms of ribosomal protein gene coregulation. Genes Dev. *29*, 1942–1954.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.

Ryan, M.P., Jones, R., and Morse, R.H. (1998). SWI-SNF complex participation in transcriptional activation at a step subsequent to activator binding. Mol. Cell. Biol. *18*, 1774–1782.

Shivaswamy, S., and Iyer, V.R. (2008). Stress-dependent dynamics of global chromatin remodeling in yeast: Dual role for SWI/SNF in the heat shock stress response. Mol. Cell. Biol. *28*, 2221–2234.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol. *6*, e65.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. *20*, 267–273.

Tomar, R.S., Zheng, S., Brunke-Reese, D., Wolcott, H.N., and Reese, J.C. (2008). Yeast Rap1 contributes to genomic integrity by activating DNA damage repair genes. EMBO J. *27*, 1575–1584.

Tsankov, A.M., Thompson, D.A., Socha, A., Regev, A., and Rando, O.J. (2010). The role of nucleosome positioning in the evolution of gene regulation. PLoS Biol. *8*, e1000414.

Webb, S., Hector, R.D., Kudla, G., and Granneman, S. (2014). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. Genome Biol. *15*, R8.

Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature *450*, 1031–1035.

Wu, A.C.K., Patel, H., Chia, M., Moretto, F., Frith, D., Snijders, A.P., and van Werven, F. (2018). Repression of divergent noncoding transcription by a sequence-specific transcription factor. Mol. Cell *72*, this issue, 942–954.

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.

Yu, L., Sabet, N., Chambers, A., and Morse, R.H. (2001). The N-terminal and C-terminal domains of RAP1 are dispensable for chromatin opening and GCN4-mediated HIS4 activation in budding yeast. J. Biol. Chem. *276*, 33257–33264.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-Rap1	D. Shore lab	N/A
Rat anti-Tubulin	Millipore	Cat# MAB1864; RRID: AB_2210391
Rabbit Peroxidase Anti-Peroxidase	Sigma-Aldrich	Cat# P1291; RRID: AB_1079562
Mouse anti Flag	Sigma-Aldrich	Cat# F1804; RRID: AB_262044
Goat anti-rabbit IgG-HRP	Santa Cruz	Cat# sc-2004; RRID: AB_631746
Goat anti-mouse IgG-HRP	Santa Cruz	Cat# sc-2005; RRID: AB_631736
IgG from rabbit serum	Sigma-Aldrich	Cat# I5006; RRID: AB_1163659
Chemicals, Peptides, and Recombinant Proteins		
cOmplete EDTA-free protease inhibitor cocktail tablets	Sigma-Aldrich (Roche)	Cat# 11873580001
Pefabloc SC-Protease-Inhibitor	Carl Roth	Cat# A154.3
DNase I recombinant, RNase-free	Sigma-Aldrich (Roche)	Cat# 04716728001
Dynabeads M-280 Tosylactivated	Thermo Fisher Scientific	Cat# 14204
Recombinant GST-TEV protease	This paper; Granneman et al., 2009	N/A
RNace-It Ribonuclease Cocktail	Agilent	Cat# 400720
Guanidine hydrochloride	Sigma-Aldrich	Cat# G4505
Ni-NTA Agarose	Qiagen	Cat# 30230
Imidazole	Sigma-Aldrich	Cat# I0125
RNaseOUT Recombinant Ribonuclease Inhibitor	Thermo Fisher Scientific	Cat# 10777019
T4 RNA Ligase 2, truncated KQ	NEB	Cat# M0373L
T4 Polynucleotide Kinase	NEB	Cat# M0201L
T4 RNA Ligase 1 (ssRNA Ligase)	NEB	Cat# M0204L
Proteinase K, recombinant, PCR grade	Sigma-Aldrich (Roche)	Cat# 03115887001
SuperScript IV Reverse Transcriptase	Thermo Fisher Scientific	Cat# 18090050
RNase H	NEB	Cat# M0297S
Exonuclease I	NEB	Cat# M0293S
LA Taq	Takara	Cat# RR002M
Phusion High-Fidelity DNA Polymerase	NEB	Cat# M0530S
GoTaq Flexi DNA Polymerase	Promega	Cat# M8291
M-MLV Reverse Transcriptase	Thermo Fisher Scientific	Cat# 28025013
Oligo $d(T)_{25}$ Magnetic Beads	NEB	Cat# S1419S
Streptavidin Magnetic Beads	NEB	Cat# S1420S
FastAP Thermosensitive Alkaline Phosphatase	Thermo Fisher Scientific	Cat# EF0654
Cap-Clip Acid Pyrophosphatase	Tebu-bio	Cat# C-CC15011H
Glycogen	Thermo Fisher Scientific	Cat# AM9510
Revertaid Premium Reverse Transcriptase	Thermo Fisher Scientific	Cat# EP0732
NucleoMag® NGS Clean-up and Size Select	Macherey-Nagel	Cat# 744970.50
Agencourt RNAClean XP Kit	Beckman Coulter	Cat# A63987
Critical Commercial Assays		
LightCycler FastStart DNA Master SYBR Green I	Roche	Cat# 12239364001
LightCycler 480 SYBR Green I Master	Roche	Cat# 04887352001
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel	Cat# 740609
Pierce Spin Columns - Snap Cap	Thermo Fisher Scientific	Cat# 69725

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Vivacon 500	Sartorius	Cat# VN01H22
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific (Invitrogen)	Cat# Q32851
Amersham Hybond-N+	GE Healthcare Life Sciences	Cat# RPN203B
Deposited Data		
Raw and analyzed data	This paper	GEO: GSE1145
Experimental Models: Organisms/Strains		
DLY2736 Reb1-AA Rpb1-HTP	(Candelli et al., 2018)	As W303; *MATa; tor1-1; fpr1::loxP*; *Rpl13A-2xFKBP12::loxP*; *Rpb1-HTP::TRP1KI*; *Reb1-FRB::KAN*
DLY2838 Rpb1-HTP Anchor away	(Candelli et al., 2018)	As W303; *MATα; tor1-1; fpr1::NAT; Rpl13A-2xFKBP12::loxP-TRP1-loxP; Rpb1-HTP::URAKI*
DLY2840 Rap1-AA Rpb1-HTP	(Candelli et al., 2018)	As W303; *MATα; tor1-1; fpr1::NAT; Rpl13A-2xFKBP12::loxP-TRP1-loxP or Rpl13A-2xFKBP12::TRP1; Rpb1-HTP::URAKI*; *Rap1-FRB-Rap1::LEU2*
DLY3085 Rap1-AID Rpb1-HTP	This study	As W303; *Rpb1-HTP::TRP; Rap1-AID-Rap1; Padh-Os.Tir1::URA3; ADE2*
DLY2973 Rap1-AA Rpb1-HTP ΔUpf1	This study	As W303; *MATα; tor1-1; fpr1::NAT; Rpl13A-2xFKBP12::loxP-TRP1-loxP or Rpl13A-2xFKBP12::TRP1*; *Rap1-FRB-Rap1::LEU2; Rpb1-HTP::URAKI; Upf1::KAN*
DLY3066 Rap1-AA Rpb1-HTP ΔRrp6	This Study	As W303; *MATα; tor1-1; fpr1::NAT; Rpl13A-2xFKBP12::loxP-TRP1-loxP or Rpl13A-2xFKBP12::TRP1*; *Rap1-FRB-Rap1::LEU2; Rpb1-HTP::URAKI; Rrp6::HIS5KI; Rrp6::HIS5KI*
DLY3131 AA background Rpb1-HTP ΔUpf1 ΔSet2	This Study	*MATα; ade2-1; can1-100; leu2-3,112; his3-11,15; GAL; psi+; tor1-1; fpr1::loxP-LEU2-loxP; Rpl13A-2xFKBP12::TRP1; Rpb1-HTP::URAKI, Upf1::HIS5KI; Set2::KAN*
DLY3133 Rap1-AA Rpb1-HTP ΔUpf1 ΔSet2	This Study	*MATα; ade2-1; can1-100; leu2-3,112; his3-11,15; GAL; psi+; tor1-1; fpr1::NAT, Rpl13A-2xFKBP12::TRP1; Rap1-FRB-Rap1::LEU2; Rpb1-HTP::URAKI, Upf1::HIS5KI; Set2::KAN*
DLY3136 AA background Rpb1-HTP ΔUpf1 ΔSet3	This Study	*MATα; ade2-1; can1-100; leu2-3,112; his3-11,15; GAL; psi+; tor1-1; fpr1::loxP-LEU2-loxP; Rpl13A-2xFKBP12::TRP1; Rpb1-HTP::URAKI, Upf1::HIS5KI; Set3::KAN*
DLY3138 Rap1-AA Rpb1-HTP ΔUpf1 ΔSet3	This Study	*MATα; ade2-1; can1-100; leu2-3,112; his3-11,15; GAL; psi+; tor1-1; fpr1::NAT, Rpl13A-2xFKBP12::TRP1; Rap1-FRB-Rap1::LEU2; Rpb1-HTP::URAKI, Upf1::HIS5KI; Set3::KAN*
DLY3229 Rap1-AID Pre2-FLAG	This Study	As W303, *MATa; ADE2; HIS3; Rap1-AID-Rap1; Padh-Os.Tir1::URA3; Pre2-FLAG::NAT*
DLY3231 Rap1-AID Rxt3-FLAG	This Study	As W303, *MATa; ADE2; HIS3; Rap1-AID-Rap1; Padh-Os.Tir1::URA3; Rxt3-FLAG::NAT*
DLY3232 Rap1-AID APS2-FLAG	This Study	As W303, *MATa; ADE2; HIS3; Rap1-AID-Rap1; Padh-Os.Tir1::URA3; Aps2-FLAG::NAT*
DLY3199 Sth1-AA	This Study	As W303, *tor1-1; fpr1::NAT, Rpl13A-2xFKBP12::TRP1; Sth1-FRB::KAN*

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DLY3200 Snf2-AA	This Study	As W303, *tor1-1; fpr1::NAT, Rpl13A-2xFKBP12::TRP1; Snf2-FRB::KAN*
DLY3201 Isw2-AA Ino80-AID	This Study	As W303, *tor1-1; fpr1::NAT, Rpl13A-2xFKBP12::TRP1; Padh-Os.Tir1::URA3; Isw2-FRB::HIS5KI; Ino80-AID-myc::HYG*
DLY3213 Sth1-AA ΔUpf1	This Study	As DLY3199, *Upf1::LEU2Cg*
DLY3214 Snf2-AA ΔUpf1	This Study	As DLY3200, *Upf1::LEU2Cg*
DLY3215 Isw2-AA Ino80-AIDΔUpf1	This Study	As DLY3201, *Upf1::LEU2Cg*
DLY3128 Abf1-AA Rpb1-HTP	This Study	*MATα, ade2-1; can1-100; leu2-3,112; his3-11,15; GAL; psi+; tor1-1; fpr1::LEU; Rpl13A-2xFKBP12::TRP1, Rpb1-HTP::URAKI, Abf1-FRB::HIS5Sp*
Oligonucleotides		
See Table S7		
Recombinant DNA		
DL878 pcM185(HIS)-p_{RAP1} *RAP1*	(Candelli et al., 2018)	Plasmid expressing full length Rap1 under control of the *RAP1* promoter
DL879 pcM185(HIS)- p_{RAP1} – RAP1-$DBD_{358-601}$	(Candelli et al., 2018)	Plasmid expressing the DNA binding domain of Rap1 (aa. 358-601) under control of the *RAP1* promoter
DL828 pCM190 pDLD3-NEL CUP1 Rap1-20	This Study	Plasmid expressing *CUP1* under the control of p*DLD3-NEL025C*. A Rap1 binding site has been inserted 20nt upstream of the *NEL025C* TSS.
DL829 pCM190 pDLD3-NEL CUP1 Rand-20	This Study	Plasmid expressing *CUP1* under the control of p*DLD3-NEL025C*. A random sequence has been inserted 20nt upstream of the *NEL025C* TSS.
DL830 pCM190 pDLD3-NEL CUP1 Rap1-85	This Study	Plasmid expressing *CUP1* under the control of p*DLD3-NEL025C*. A Rap1 binding site has been inserted 85nt upstream of the *NEL025C* TSS.
DL831 pCM190 pDLD3-NEL CUP1 Rand-85	This Study	Plasmid expressing *CUP1* under the control of p*DLD3-NEL025C*. A random sequence has been inserted 85nt upstream of the *NEL025C* TSS.
DL832 pCM190 pDLD3-NEL CUP1 Rap1-295	This Study	Plasmid expressing *CUP1* under the control of p*DLD3-NEL025C*. A Rap1 binding site has been inserted 295nt upstream of the *NEL025C* TSS.
DL833 pCM190 pDLD3-NEL CUP1 Rand-295	This Study	Plasmid expressing *CUP1* under the control of p*DLD3-NEL025C*. A random sequence has been inserted 295nt upstream of the *NEL025C* TSS.
DL834 pCM190 pACT1 CUP1	This Study	Plasmid expressing *CUP1* under the control of pACT1.
DL835 pCM190 pACT1 CUP1 Rap1-5	This Study	Plasmid expressing *CUP1* under the control of pACT1. A Rap1 binding site has been inserted 5nt upstream of the *ACT1* TSS.
DL836 pCM190 pACT1 CUP1 Rand-5	This Study	Plasmid expressing *CUP1* under the control of pACT1. A random sequence has been inserted 5nt upstream of the *ACT1* TSS.
Software and Algorithms		
pyCRAC v1.2.2.7	Webb et al., 2014	http://sandergranneman.bio.ed.ac.uk/Granneman_Lab/pyCRAC_software.html
cutadapt v1.5	Martin, 2011	https://cutadapt.readthedocs.io/en/stable/#
Trimmomatic v0.33	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic
Fastx toolkit v0.0.13	Hannon Lab	http://hannonlab.cshl.edu/fastx_toolkit/
Bowtie2 v2.2.3	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
peakCcall	This Study	Available upon request to D.L. or M.B.

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
Qubit Fluorometer	Thermo Fisher Scientific (Invitrogen)	Cat# Q32857
Gelfree 8100 Fractionation Station	Expedeon	Cat# 48100
Gelfree 8100 5% Tris Acetate Cartridge Kit	Expedeon	Cat# 42104
"Megatron" W5 UV crosslinking unit	UVO3 Ltd	www.uvo3.co.uk
Mixer Mill MM 400	Retsch	Cat# 20.745.0001

CONTACT FOR REAGENT AND RESOURCE SHARING

Please contact D.L. (domenico.libri@ijm.fr) for reagents and resources generated in this study.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Yeast (*S.cerevisiae*) is the experimental model used in this study.

METHOD DETAILS

Yeast strains, plasmids and oligonucleotides

Yeast strains, plasmids, oligonucleotides used in this study are described in Key Resources Table.

RNA and protein analysis

RNAs were prepared by the hot acid phenol method. Briefly, yeast cells were spinned and resuspended in 400μl of 50 mM Sodium acetate (pH 5.5), 10 mM EDTA, 1% SDS. An equal volume of water-saturated phenol was added and the samples were incubated for 30min at 65°C with shaking in a Thermomixer (Eppendorf). The aqueous phase was recovered. The phenol extraction was repeated once with water saturated phenol and once with chloroform. The RNAs were ethanol precipitated. For Northern blot analysis, 10 μg of RNA were separated by agarose gel electrophoresis and transferred to a Hybond Nylon N+, membrane (GE Helthcare) by capillarity. Hybrydization was performed in UltraHyb buffer (Ambion). For RT-qPCR, 4 μg of RNAs were reverse transcribed using oligo d(T) and random primers and the resulting cDNAs were analyzed by quantitative PCR (Lightcycler 480, Roche).

For protein analysis, 5 OD_{600} of cells from exponential cultures were harvested, washed with water and resuspended in 200 μl of 0.1 N NaOH solution. After 5 min at room temperature, cells were pelleted and resuspended in 100 μl of 1X Laemmli buffer. After 5 min at 95°C, the cell debris pellets were discarded and 10 to 15 μl of supernatants were loaded on acrylamide gels. Proteins were separated by 6 to 10% PAGE and transferred to a nitrocellulose membrane for further analysis.

RNAPII CRAC

RNAPII CRAC data upon nuclear depletion of Rap1 and Reb1 have been generated in a separate study (Candelli et al., 2018). For nuclear depletion of Abf1 rapamycin was added to the Abf1 anchor away strain for 30 and 90 minutes and data from the 90 minutes time point was used for the analyses shown in Figure 6. Rap1 was also depleted using the auxin degron system (Nishimura et al., 2009) by adding IAA (Indole-3′-Acetic Acid, Sigma) 500 μM to Rap1-AID cells for 10 or 20 min before crosslinking.

The CRAC protocol used in this study is derived from Granneman et al. (2009), modified as described in Candelli et al. (2018). Briefly, 2 L of yeast cells expressing Rpb1-HTP tag were grown at 30°C to OD_{600} = 0.6 in CSM-Trp medium before addition of rapamycin or IAA for the times required. Cells were UV crosslinked using a W5 UV crosslinking unit (UVO3 Ltd) for 50 s, harvested by centrifugation, washed in cold PBS and resuspended in TN150 buffer (50 mM Tris pH 7.8, 150 mM NaCl, 0.1% NP-40 and 5 mM beta mercaptoethanol, 2.4 ml/g of cells) supplemented with protease inhibitors (Complete, EDTA-free Protease Inhibitor Cocktail). The suspension was flash frozen in droplets and cells were mechanically broken with a Mixer Mill MM 400 (5 cycles of 3 minutes at 20 Hz). Extracts were treated for one hour at 25°C with DNase I (165 U/g of cells) to solubilize chromatin and then clarified by centrifugation (20 min at 20000 g at 4°C). The complexes were purified by a two-step procedure, the second one under denaturing conditions as described (Candelli et al., 2018). High salt washes for both purification steps were done at 1 M NaCl for high stringency. The dephosphorylation step required for cleaving the 2′-5′ cyclic phosphate left by RNase treatment (Granneman et al., 2009) was omitted to enrich for nascent transcripts, the 3′ end of which being protected from RNase treatment. The protein fractionation step was performed with a Gel Elution Liquid Fraction Entrapment Electrophoresis (GelFree) system (Expedeon). Rpb1-containing fractions were treated with 100 μg of proteinase K in a buffer containing 0.5% SDS. RNAs were purified and reverse transcribed using reverse transcriptase Superscript IV (Invitrogen).

The concentration of cDNAs in the reaction was estimated by quantitative PCR using a standard of known concentration. PCR amplification were performed in separate 25 μl reactions containing each 2 μl of cDNA for typically 7-9 PCR cycles (LA Taq, Takara). The PCR reactions from all the samples were pooled and treated for 1 hour at 37°C with 200 U/ml of Exonuclease I (NEB). The DNA was purified using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel) and sequenced using Illumina technology.

TSS sequencing

The TSS sequencing protocol has been described in Malabat et al. (2015). Yeast cells were grown to mid-exponential phase in YPD-rich media. After treatment of cells with Rapamycin, Auxin, both or none, *Schizosaccharomyces pombe* cells were added for spiking purposes at a ratio of 1 to 10. Cells were then harvested and pellets were frozen in liquid nitrogen. Total RNA was extracted using two successive hot phenol steps and one chloroform. After ethanol precipitation, RNA pellet was treated with DNase and extracted again with phenol chloroform 5:1 pH4.5.

Polyadenylated transcripts were purified from 75 μg of total RNAs using oligo $d(T)_{25}$ magnetic beads (New England Biolabs). RNAs were dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher) and treated with Cap-Clip Acid Pyrophosphatase (Tebu-bio). RNAs were then ligated overnight at 16°C to the biotinylated 5′ adaptor (oligonucleotide 3365, 50 pmol) using T4 RNA ligase I (10 units, New England Biolabs) and ATP at a final concentration of 1 mM. After 5′ ligation, the RNA was fragmented for 5′ at 70°C in fragmentation buffer (10 mM $ZnCl_2$, 10 mM Tris pH7.5). The reaction was stopped by adding 75 μl of a cold solution containing 1 μl of EDTA 0.5 M. Ligated RNA molecules were purified on streptavidin magnetic beads from New England Biolabs (50 μl of slurry). Binding was performed at 37°C for 10', beads were washed and RNA were eluted in 20 μl of water at 95°C for 5′. Reverse transcription was performed with 50pmoles of primer 3018 and 300 units of RevertAid reverse transcriptase (ThermoFisher) in 30 μl. cDNAs were purified by adding 1.8 volumes of Agencourt RNAClean XP beads (Beckman Coulter) and eluted in 50 μl of water at room temperature for 10'.

4 independent PCR reactions were performed in a final volume of 25 μl with 5 pmoles each of primer 1 and Illumina multiplexing PCR primer, and 0.25 μl of LA Taq DNA polymerase (Takara). The four PCR reactions were pooled, purified on NucleoMag NGS Clean-up and Size Select dynabeads (Macherey-Nagel) and the DNA eluted in 30 μl of water.

MNase-seq

MNase-seq has been performed based on the procedure describe in Kubik et al. (2015). Briefly, 120 mL of Rap1-AA yeast cells ectopically expressing $Rap1_{DBD}$, wild-type Rap1 or containing an empty plasmid were grown at 30°C in CSM-His media to OD_{600} ~0.15. Rapamycin was added for one hour to a final concentration of 1μg/ml and cells were crosslinked for 5min in 1% formaldehyde at room temperature. After crosslinking the procedure was carried out as described (Kubik et al., 2015)

Dataset processing and data analysis

CRAC

CRAC datasets were analyzed as described (Candelli et al., 2018). The pyCRAC script pyFastqDuplicateRemover was used to collapse PCR duplicates using a 6 nucleotides random tag included in the 3′ adaptor (see Key Resources Table). The resulting sequences were reverse complemented with Fastx reverse complement (part of the fastx toolkit, http://hannonlab.cshl.edu/fastx_toolkit/) and mapped to the R64 genome (Cherry et al., 2012) with bowtie2 (-N 1) (Langmead and Salzberg, 2012).

RNA-seq and TSS-Seq samples were demultiplexed by the sequencing platform with bcl2fastq2 v2.15.0 and illumina trueseq adaptors were trimmed with cutadapt 1.9. Sequencing reads were quality trimmed with trimmomatic and mapped to the R64 genome with bowtie2 (default options).

TSS Seq dataset processing

Cleavage of the 5′ barcode and demultiplexing were performed with the script pyBarcodeFilter from the pyCRACutility suite (Webb et al., 2014). Filtering the adaptors on the 3′ end (AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCG TATCATT, minlength = 10) was performed with cutadapt (Martin, 2011) and the 3′ end was trimmed using Trimmomatic (window size = 5, minimum quality = 25, Bolger et al., 2014).

The obtained sequences were collapsed with the script pyFastqDuplicateRemover from the pyCRACutility suite (Webb et al., 2014) in order to remove PCR duplicates. The resulting sequences were mapped to the *S.cerevisiae* R64 genome (Cherry et al., 2012) with bowtie2 (Langmead and Salzberg, 2012). Only the position of the first 5′ end base was kept for each read.

We cleaned our data from two types of noise: the random noise and the systematic noise due to the TSS Seq technique. The random noise was filtered out by considering the signal at each position only when present in at least two out of three replicates (and saving the sum of the signals). The systematic noise was mainly due to signals derived from non-capped molecules (for instance caused by failure in the dephosphorylation step) and was detected by sequencing samples for which we had omitted the decapping step. Because of the sparse nature of the signal, we aggregated signals closer than 50nt. The regions of aggregated signals in the control samples are potentially noisy regions. Whenever these regions overlapped regions of aggregated signal in the test sample we evaluated the background/signal ratio in the segments of overlap and excluded all the segments in which the ratio is > 1.

Normalization among samples was done using the *S.pombe* spike-in. *S. cerevisiae* signals were divided by total number of per million reads longer than 40 nucleotides that map only on *S.pombe*.

TSS-Seq peak calling with peakCcall and eTSS identification

The TSSs are typically detected as small clusters of signals that reveal heterogeneity in the precise position of transcription initiation. In order to define the most likely position of initiation, we created a peak-calling pipeline for clustering signals that correspond to the same starting event (peak cluster call, peakCcall, tool). We first smoothed the signal with a Gaussian kernel to obtain a pseudo-continuous function enabling the evaluation of the (discrete) second derivative of this function.

This smoothing process consists in computing the convolution of the TSS-Seq signal with a Gaussian with a defined standard deviation σ. This has the effect of redistributing and combining signals derived from clusters. Given $h_{rs}(x_i)$ the height of the TSS-Seq profile at position xi, the height of the smoothed profile (h_{ss}) at each position x_i is:

$$h_{ss}(x_i) = h_{rs} \bigstar \mathcal{N}_{\sigma}(x_i) = \sum_j \frac{1}{\sigma\sqrt{2\pi}} h_{rs}(x_j) \cdot e^{\frac{-(x_i - x_j)^2}{2\sigma^2}}$$

The value of σ defines the resolution of the analysis. We empirically observed that using σ equals to 15 bp allowed separating peaks distant more than 100 bp but combined in the same peak signals closer than 30 bp.

To automatically detect TSSs we first defined as Peak Cluster Regions (PCRs) those regions included between two inflection points of h_{ss} in which the second derivative is negative; this isolates a unique local maximum of the function in each region.

For each PCR we defined the weighted average TSS position (wTSS) using the raw signal as follows:

$$wTSS_{pos} \approx \frac{\sum_{i\ \epsilon PCR} h_{rs}(x_i) \cdot x_i}{\sum_{i\ \epsilon PCR} h_{rs}(x_i)}$$

We assigned scores to each PCR, which allowed evaluating the signals in the different conditions. The following score, which was validated by a principal component analysis, gave the best results in terms of scoring:

$$PK3 = \frac{H_{rs}^2}{s} \cdot \frac{H_{ss}^2}{w}$$

where $H_{rs} = \max_{i\ \epsilon PCR} h_{rs}(x_i)$ is the maximal height of the TSS Seq profile in the PCR; $H_{ss} = \max_{i\ \epsilon PCR} h_{ss}(x_i)$ is the maximal height of the smoothed profile in the PCR; $s = \sum_{i\ \epsilon PCR} h_{rs}(x_i)$ is the sum of the values in the PCR, and w is the width of the PCR i.e., the distance between the two inflection points.

The score was used to generate a list of putative eTSS, that was validated manually and used for subsequent analyses.

Data obtained from RNAPII CRAC, TSS-Seq, RNA-Seq and MNase-seq were analyzed using deeptools 2.0 (Ramírez et al., 2014) on the Roscoff (http://galaxy3.sb-roscoff.fr/root/login?redirect=%2F) and Freiburg (http://deeptools.ie-freiburg.mpg.de/) Galaxy platforms.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical significance of expression changes for Rap1 targets has been calculated using the distribution of $\log_2$ ratio values for control genes (i.e., genes that do not contain a Rap1 binding site). FDR q-values have been calculated from p values according to the Benjamini and Hochberg procedure (Benjamini and Hochberg, 1995).

To assess the occurrence of ectopic transcription in intergenic regions, the $\log_2$ ratio (-Rap1/+Rap1) per nucleotide has been calculated in the 200nt upstream of the TSS (for 5′ extended transcripts) or around (−100 to +500nt) Rap1 sites that are not followed by a canonical gene (for transcription leading to production of non-coding RNA). The number of positions displaying high signal changes (i.e., two standard deviations above the average detected at non Rap1 targets) was scored per each genomic locus. Loci with a statistically significant number of high signal positions relative to control genes (FDR < 0.05) were identified as positive. This conservative procedure allowed identifying regions of sparse ectopic transcription because these signals are not averaged out over the whole intergenic region.

For the statistical analysis of ectopic initiation, we used a list of 176 eTSSs generated by peakCcall tool and manually curated. This list was used for most of the subsequent analyses, e.g., to generate the distribution of distances to Rap1 sites (Figure 2C) and to evaluate the sensitivity of eTSSs to the *upf1Δ* (Figure 2D) and *rrp6Δ* mutation (Figure S2C). For the latter analyses we evaluated the $\log_2$ ratio of the TSS-Seq signals detected in the mutant versus the wt strains (*upf1Δ*/wt and *rrp6Δ*/wt) in the absence of Rap1 in a window of 40nt around the eTSS. The statistical significance of the scores obtained was evaluated relative to the same analysis performed on a set of 358 randomly chosen control genes (non-Rap1 targets) that are not expected be sensitive to the absence of Upf1 or Rrp6 (distributions shown in blue in Figures 2D and S2C). A similar procedure was used to model the impact of eTSS on the use of the canonical TSS of Rap1 targets: the TSS-Seq signal was scored as previously in a window of 40nt around the canonical TSS of Rap1 target genes containing at least one ectopic TSS in their promoter. Statistical significance was assessed by the same analysis performed on control genes as above.

Statistical significance for the differences in the distributions of RNA-Seq values shown in Figures S4E–S4G where obtained using a two-tailed Student's t test, with paired samples.

DATA AND SOFTWARE AVAILABILITY

The accession number for the data reported in this paper is GEO: GSE114589. RNAPII CRAC Data for Reb1-AA and Rap1-AA and RNA-Seq data for the Rap1DBD series have been generated in a previous study (Candelli et al., 2018). The PeakCcall tool will be described in a separate report, but is available upon request to D.L. or M.B. Source datasets have been deposited into Mendeley and can be retrieved using the following link: The original Northern and western blot have been deposited in Mendeley (https://doi.org/10.17632/rd2xsvsvkz.1).

Molecular Cell, Volume *72*

Supplemental Information

General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation

Drice Challal, Mara Barucco, Slawomir Kubik, Frank Feuerbach, Tito Candelli, Hélène Geoffroy, Chaima Benaksas, David Shore, and Domenico Libri

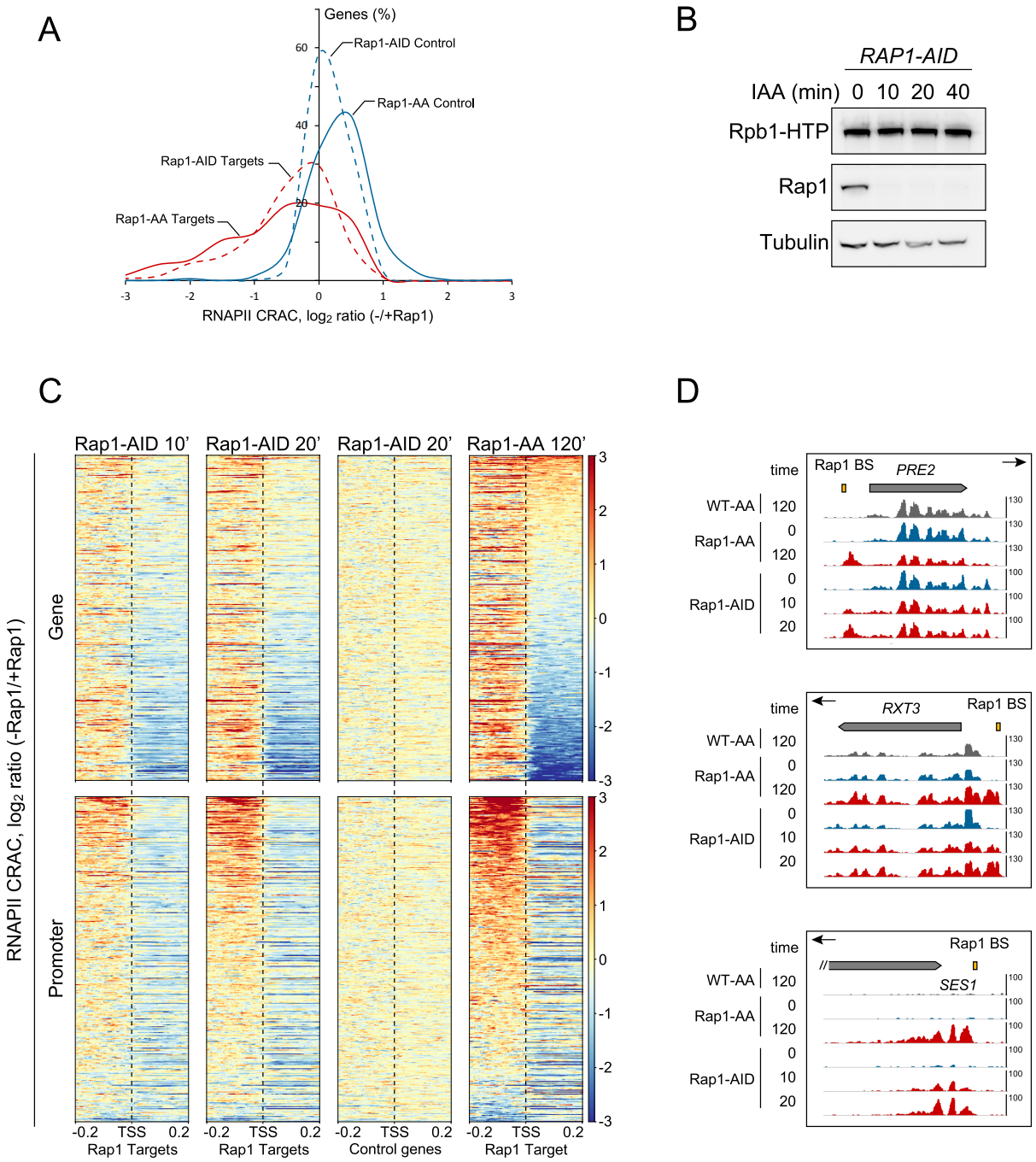


Figure S1. related to Figure 1. **A.** Binned distributions of RNAPII CRAC signal changes at Rap1 target genes (red) or controls (blue) upon Rap1 depletion by the anchor away (solid lines, 120 min) or auxin-degron method (dashed lines, 20min time point). The data illustrates the general downregulation of gene expression that is seen rapidly after Rap1 depletion. **B.** western blot indicating the effectiveness of Rap1 depletion by the auxin-degron method. Rap1 is detected by a polyclonal anti-Rap1 antibody after addition of IAA for the indicated times. Tubulin or the HTP-tagged Rpb1 expressed in the same cells are used as loading controls. **C.** Heatmaps showing the RNAPII CRAC signal change upon fast depletion of Rap1 by the auxin-degron method for the times indicated. Rap1 target genes have been sorted as in Figure 1, i.e. for decreasing signal change in genes or in promoters in Figures 1A and 1C respectively (shown also here for comparison). Control genes are as in Figure 1. **D.** Additional snapshots showing a 5'-extended RNAPII CRAC signal at the *PRE2* and *RXT3* loci upon Rap1 depletion. Note also the appearence of a prominent, non-coding transcription signal antisense to *SES1*.

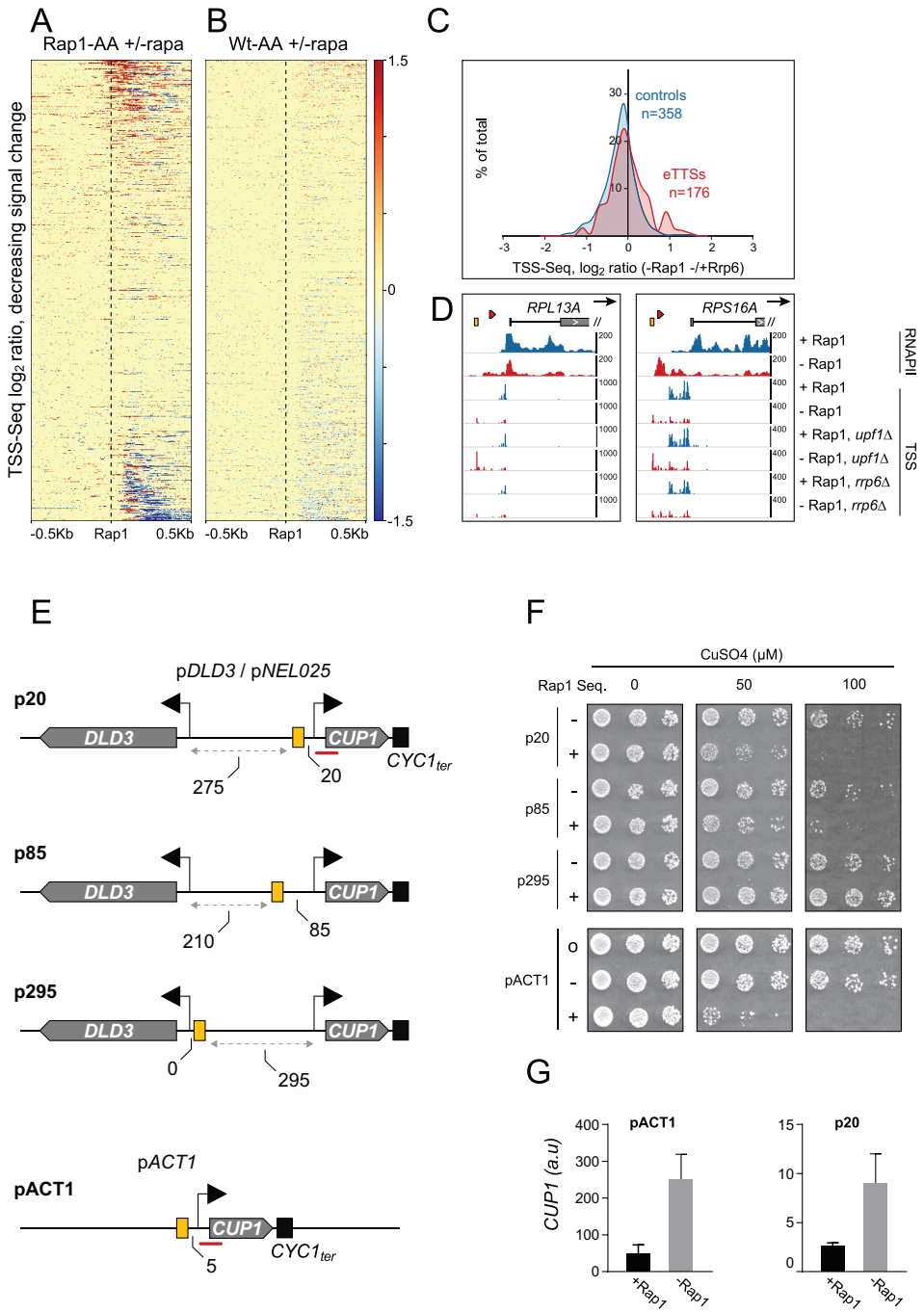


Figure S2. related to Figure 2. **A-B**. Heatmaps illustrating the TSS-Seq signal change (log_2 ratio) around Rap1 sites. A. Regions are aligned on the Rap1 site and sorted by decreasing signal change (log_2 ratio) after addition of rapamycin to a Rap1-FRB strain. For a control on non-specific effects of rapamycin, the same analysis was performed using an anchor away strain expressing an untagged Rap1 (no nuclear depletion of Rap1 is expected, B). **C.** A small fraction of eTSSs are sensitive to Rrp6-dependent nuclear degradation. TSS-Seq signals at ectopic initiation sites have been computed in the absence or presence of Rrp6 (log_2 ratio (*rrp6Δ/RRP6*)). The binned distribution of these signals (red) is shown in comparison with the distribution of log_2 ratios at the transcription start site of control genes (not generally sensitive to Rrp6 degradation, blue). **D.** Snapshots illustrating the downregulation of two ribosomal protein genes in association with the occurrence of ectopic initiation. **E-G.** Inhibition of transcription initiation by Rap1 in a heterologous context. **E.** schemes of the different constructions containing either the bi-directional promoter between *DLD3* and the CUT *NEL025C* (p20, p85 and p295) or the *ACT1* promoter driving expression of the *CUP1* gene (p*ACT1*). Each construct contains a Rap1 site or a mutated derivative (yellow box) inserted at different positions from the TSSs as indicated. A derivative of p*ACT1* without insertion was also constructed. A red line indicates the approximate position of qPCR primer for the quantification shown in G. **F.** Panel illustrating the growth of yeast cells containing the indicated constructs at the different concentrations of copper as indicated. The presence of a Rap1 binding site (+), the mutated sequence (-) or no insertion (o) is indicated. **G.** RT-qPCR quantification of the *CUP1* RNAs produced from p20 and p*ACT1* constructs containing a Rap1 binding site in the presence of Rap1 or after Rap1 nuclear depletion for one hour. The p20 and pACT1 constructs were introduced in the Rap1 anchor away strain and Rap1 was depleted by addition of rapamycin for 1 hour. The oligonucleotides used anneal in the *CYC1* promoter and the *CUP1* gene to avoid detection of endogenous *CUP1* RNAs. Average of 3 experiments. Error bars indicate the standard deviation.

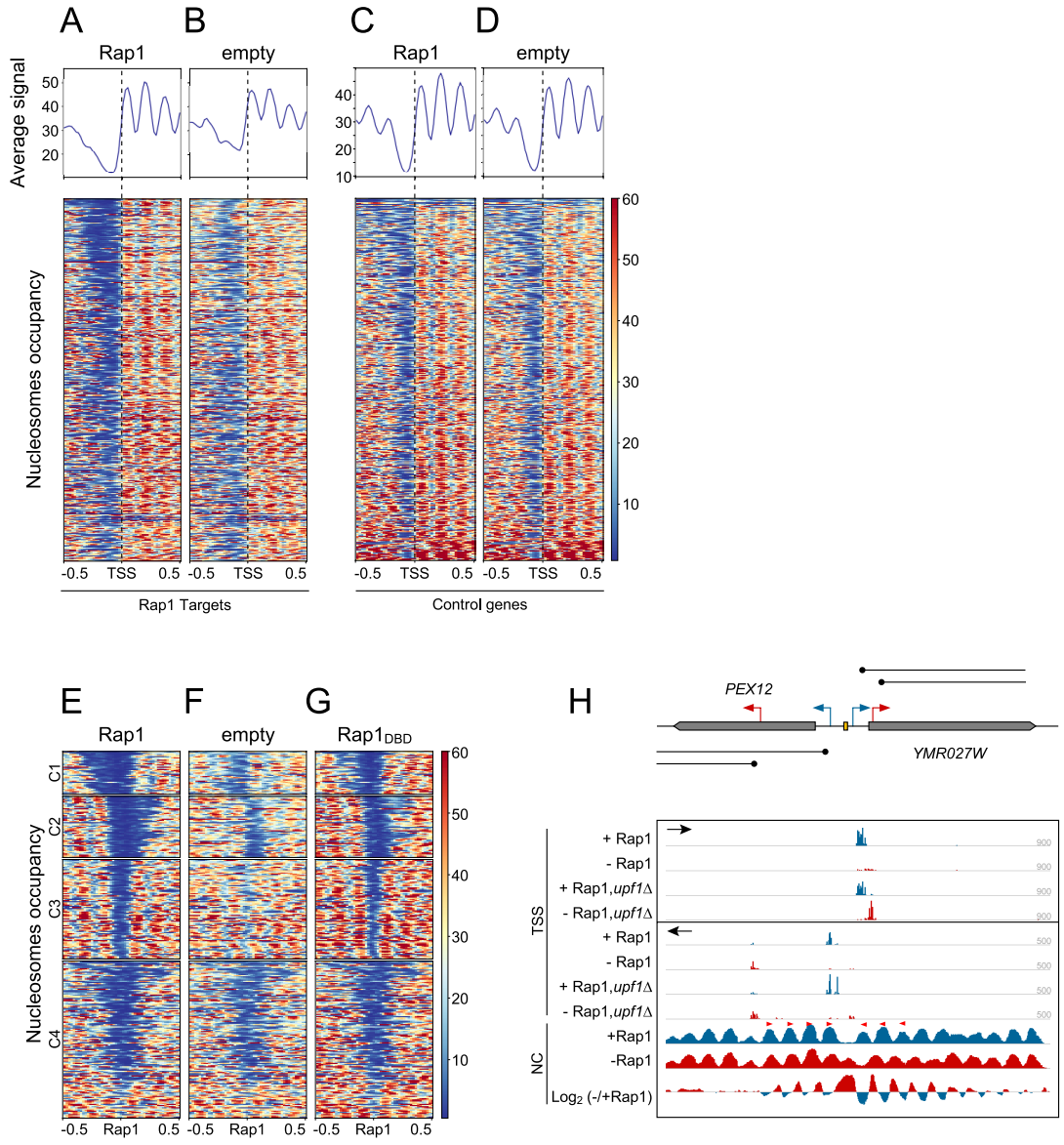


Figure S3. related to Figure 3. **A-B**. Nucleosome positioning upon depletion of Rap1. Regions are aligned on the TSS of genes containing a Rap1 site in the upstream NDR and sorted by increasing MNase-seq signal in A. Endogenous Rap1 was depleted in cells containing a plasmid expressing wild type Rap1 (A) or an empty control (B) as indicated. The $\log_2$ ratio of signal shown in A and B is presented in Figure 3. **C-D.** The same analysis was performed on a set of control genes, not affected by Rap1. **E-G.** Heatmaps illustrating the position of nucleosomes at genomic regions aligned on Rap1 sites and sorted by k-means clustering based on signals in figure 4A. Depletion of Rap1 was performed in cells containing a plasmid expressing wild type Rap1 (E), an empty plasmid (F) or the DNA binding domain of Rap1 (G) ($Rap1_{DBD}$). Clusters 1 and 2 contain large NDRs that have an eccentric Rap1 site relative to flanking nucleosomes. **H.** Snapshot illustrating two examples of internal initiation linked to a complex pattern of nucleosome shifting. The +1 nucleosome of *YMR027W* and 4 nucleosomes from the divergent *PEX12* move upstream, close the original NDR and open two novel NDRs, one between the +1 and +2 nucleosomes of *YMR027W* and a second one, smaller, between the +4 and +5 nucleosomes of *PEX12*. This is coupled with two internal eTSS (red arrows), one associated with the original +2 nucleosome of *YMR027W* and another associated with the +5 nucleosome of *PEX12*. Note also the existence of an upstream eTSS for *PEX12*, associated with the original +1 nucleosome of *YMR027W*.

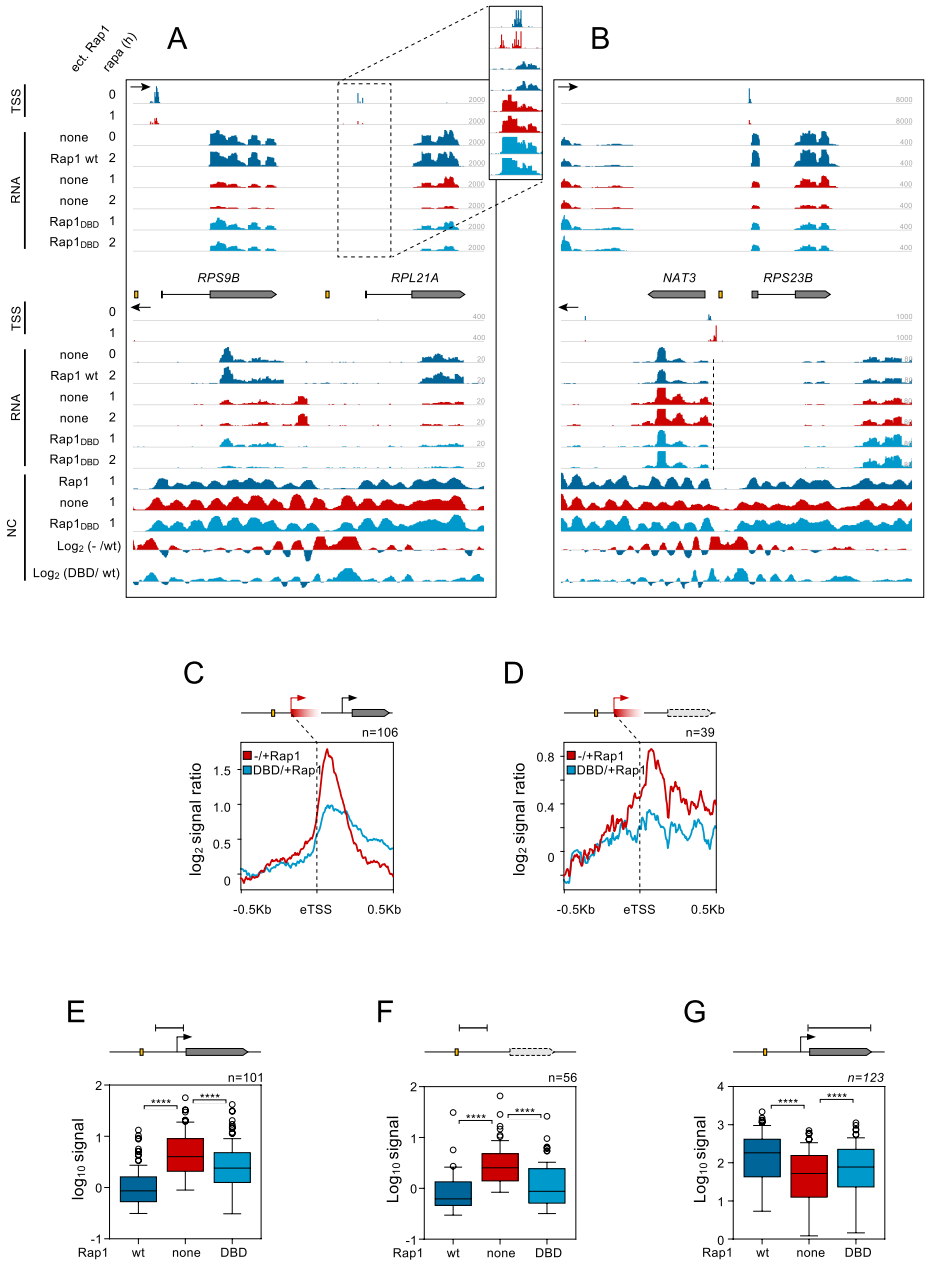


Figure S4. related to Figure 4. **A.** Snapshot illustrating the partial suppression of nucleosome positioning and ectopic initiation by $Rap1_{DBD}$ in Rap1 depleted cells. Suppression only occurs for the non-coding RNA antisense of *RPS9B*, but not on the opposite side for *RPL21A* (compare RNA-Seq tracks for the relative conditions). Note that the eTSS for the non-coding RNA is only seen in this experiment in the absence of Rrp6 (data not shown). See inset at the appropriate scale for visualizing the eTSS upstream of *RPL21A*. Nucleosomes positioning is also only restored on the side of the non-coding RNA. Note that *RPS9B* is itself under control of another Rap1 site (yellow rectangle) around which nucleosome positioning is almost completely restored by $Rap1_{DBD}$ (not shown) leading to partial restoration of *RPS9B* expression (see RNA-Seq tracks for the corresponding strand). **B.** The snapshot illustrates another example of partial suppression of *NAT3* gene expression that is paralleled by a partial suppression of the nucleosome positioning defect upon Rap1 depletion. Note that the nucleosome positioning and expression of the divergent *RPS23B* gene are not suppressed. **C-D.** Aggregate plots illustrating the partial suppression of ectopic initiation by $Rap1_{DBD}$ as in figure 4H but profiling the differential RNA-Seq profile at eTSSs upstream of canonical genes (indicated by a grey rectangle, C) or defining non-coding RNA ectopic transcription (D, white rectangle). **E-G.** Boxplots comparing the distribution of RNA-Seq signals at promoters of canonical genes (E), in regions of non-coding ectopic transcription (F) or within genes (G) downstream of Rap1 binding sites (yellow rectangle). Distributions are calculated for wild type cells, Rap1-deficient cells and cells depleted for Rap1 and expressing $Rap1_{DBD}$ as indicated. The approximate regions for which the signals were calculated are shown on the top of each plot. For a more quantitative assessment, we selected the features most affected by Rap1 depletion. For transcription in promoters and regions of non-coding transcription (E and F) we selected regions with an RNA-Seq signal change (log_2 ratio -/+ Rap1) >1. For genes (G) we selected features the expression of which is decreased at least by a factor of two. Partial suppression by $Rap1_{DBD}$ in all instances was statistically significant (E: $p=2{,}7*10^{-5}$; F: $p=6.5*10^{-4}$; G: $p=2*10^{-6}$). Statistical significance based on a two-tailed Student t-test, paired samples, same variance.

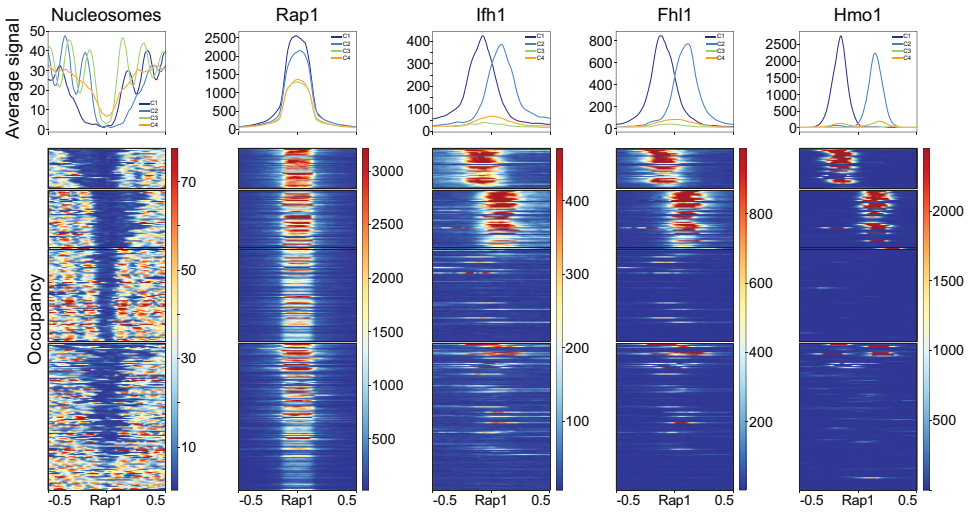


Figure S5. Related to figure 4. Distribution of Rap1, Ifh1, Fhl1 and Hmo1 in the clusters defined in Figures 3 and S3. Regions aligned on Rap1 sites and clusters defined as in Figure 4A and S3E. ChIP data from Knight et al. (Knight et al., 2014)

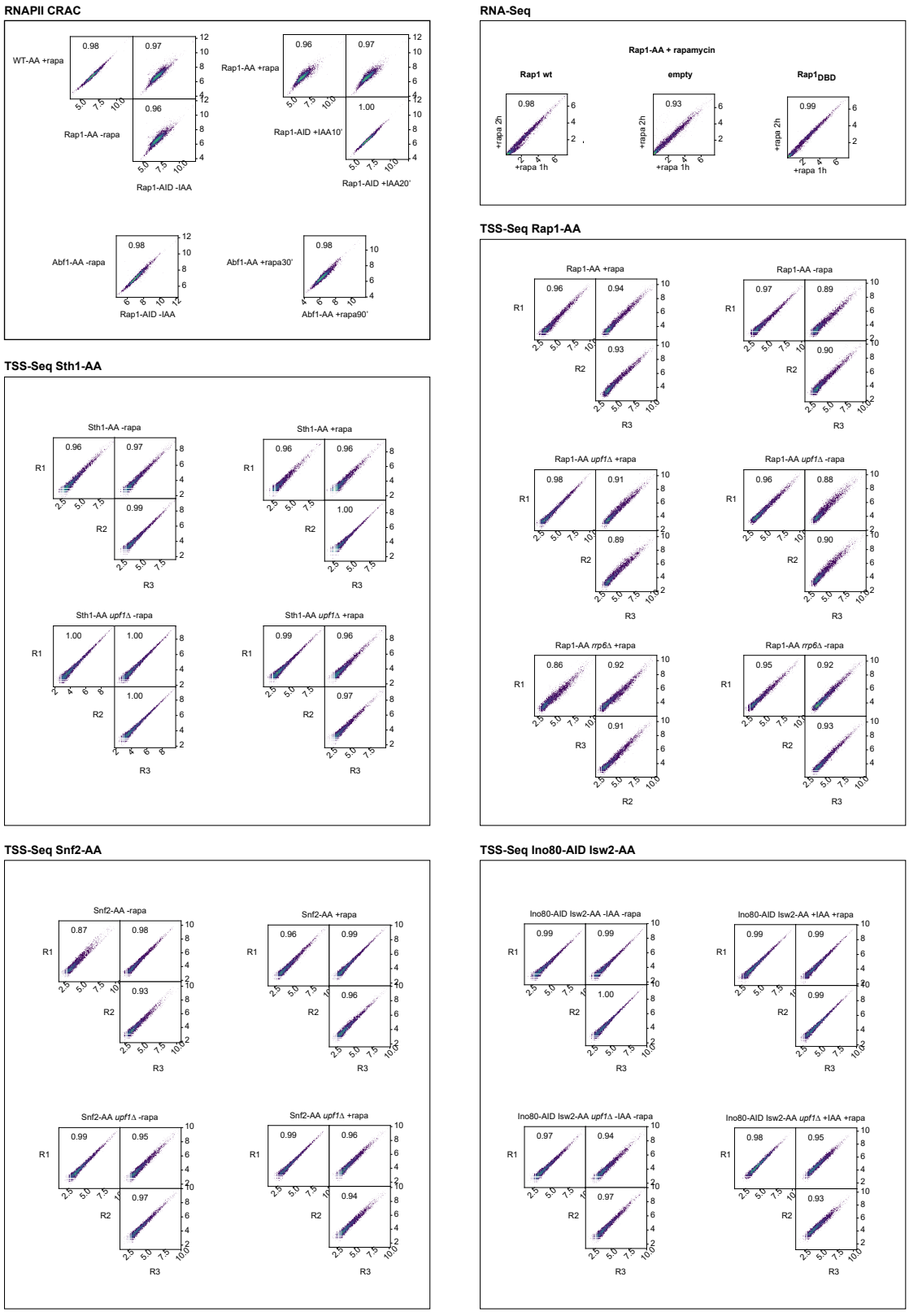


Figure S6. Related to figures 1-6**.** Correlation plots of the genomic data presented in this study. Dots represent 1kb bins, plots were generated with the appropriate Galaxy tool. For the RNAPII CRAC the datasets produced in this study using the Rap1 degron strain were also compared to the dataset produced in Candelli et al. (Candelli et al., 2018) using the Rap1 anchor away strain. The data compared in the RNA-Seq panel are also from Candelli et al. (Candelli et al., 2018). We compared replicates (TSS-Seq) or similar conditions (different time points for RNAPII CRAC and RNA-Seq data). For each comparison, the Pearson correlation coefficient is indicated.

III - Opposing Chromatin Remodelers Control Transcription Initiation Frequency and Start Site Selection

Chromatin remodelers act with GRFs to promote NDRs formation at promoters. In this context, many different studies have highlighted the cooperative or opposing activity of remodelers. However, these studies often lack a more global vision on the possible redundancy of remodelers thus underestimating their function. This study aims at understanding *how chromatin remodelers activities are intertwined in vivo and how their collective action is integrated into the control of gene expression and NDR formation.*

By considering the occupancy of chromatin remodelers and the alteration of nucleosome positioning, we could demonstrate that the correct positioning of the +1 nucleosome is the result of the combined action of multiple remodelers. Interestingly, we have shown that the re-organization observed upon depletion of a single remodeler is often due to the antagonizing activity of other factors. Supporting this notion, the position of the +1 remains unchanged when enzymes acting in an opposite-manner (i.e. sliding the +1 nucleosome in opposite direction with respect to the NDR) are conjointly depleted. Finally, we have assessed the role of chromatin remodelers in transcription initiation and have reported that the establishment of NDRs by the latter influences the occupancy of TBP and TSS selection.

Detailed contribution: This work was mainly operated by and within the Shore laboratory. Our laboratory took part of the project for the following aspects: D.L supervised the project with D.S and P.B and was responsible for funding acquisition together with D.S. D.L and D.C reviewed and edited the original draft with members of the Shore lab. D.L and D.C were involved in the conceptualization with S.K and D.S. D.C constructed some of the strains and performed the TSS-Seq experiments.

Opposing chromatin remodelers control transcription initiation frequency and start site selection

Slawomir Kubik[1], Maria Jessica Bruzzone[1*], Drice Challal[2*], René Dreos[3], Stefano Mattarocci[1], Philipp Bucher[3], Domenico Libri[2], and David Shore[1, 4]

[1] Department of Molecular Biology and Institute of Genetics and Genomics of Geneva (iGE3), 30 quai Ernest Ansermet, Geneva 4, CH-1211 Switzerland

[2] Institut Jacques Monod, CNRS-Université Paris Diderot, 15 rue Hélène Brion, 75205 Paris cedex 13, France

[3] Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, CH-1015 Switzerland

* These authors made equal contributions

[4] Corresponding and lead author: David.Shore@unige.ch

Abstract

Precise nucleosome organization at eukaryotic promoters is thought to be generated by multiple chromatin remodeler (CR) enzymes and to affect transcription initiation. Using an integrated analysis of chromatin remodeler binding and nucleosome occupancy change following rapid remodeler depletion, we investigate the interplay between these enzymes and their impact on transcription in budding yeast. We show that many promoters are acted upon by multiple CRs that operate either in concert or in opposition to position the key transcription start site-associated +1 nucleosome. We demonstrate that nucleosome movement following CR inactivation usually results from the activity of another CR and that in the absence of any remodeling activity +1 nucleosomes maintain their positions. Finally, we present functional assays suggesting that +1 nucleosome positioning often reflects a trade-off between maximizing RNA Polymerase II recruitment and minimizing transcription initiation at incorrect sites. Our results provide a detailed picture of fundamental mechanisms linking promoter nucleosome architecture to transcription initiation.

Introduction

The availability of information encoded in eukaryotic genomes is restricted by wrapping of the DNA helix in nucleosomes, the basic units of chromatin. Regulation of the accessibility of chromosomal DNA to transcription factors (TFs) and the transcriptional machinery itself is believed to play an important role in keeping certain genes silenced while permitting transcription initiation at precisely defined positions at active genes. Such tight and selective regulation is reflected by canonical patterns of promoter nucleosome architecture typically followed by phased arrays of nucleosomes over the downstream gene bodies [1]. Thus, at promoters of active genes, the transcription start site (TSS) is typically located upstream of the dyad axis of a well-positioned nucleosome termed "+1" which is followed by regularly spaced genic nucleosomes (+2, +3, etc.). Directly upstream of the +1 nucleosome one typically finds a region of accessible chromatin, termed the nucleosome-depleted region (NDR), whose size is gene-dependent and which at some promoters can be occupied by an unstable nucleosome-like particle termed a "fragile nucleosome" (FN) [2-7]. The position of the +1 nucleosome and the existence of the NDR are crucial for recruitment of the transcriptional machinery and initiation of transcription [8,9].

ATP-dependent chromatin remodelers (CRs), multi-subunit molecular machines that utilize the energy of ATP hydrolysis to slide, eject or modify nucleosomes [10], have emerged as major factors shaping the chromatin landscape at promoters [1]. There are four main CR subfamilies – SWI/SNF, ISWI, CHD and INO80 – all of which are conserved from yeast to humans. Each subfamily displays unique biochemical activities and is associated with specific roles in the cell [10]. SWI/SNF subfamily members, in budding yeast represented by the essential remodeler RSC and the canonical SWI/SNF remodeler, slide nucleosomes to the edge of linear DNA templates in vitro, maximizing the amount of nucleosome-free DNA [11,12]. They are also able to displace histone octamers from the DNA template [13-16]. In vivo, RSC and SWI/SNF participate in the generation of NDRs at promoters by at least two mechanisms: sliding the +1 and -1 nucleosomes away from each other [5,17-20] and destabilizing or ejecting promoter nucleosomes [5,13,21,22]. ISWI type (represented by ISW1 and ISW2 in budding yeast) and CHD type (CHD1) CRs equalize the length of the naked DNA on both sides of a nucleosome in vitro [23]. In vivo they have a predominant role in setting the spacing of intragenic nucleosomes, with CHD1 acting mostly at genes with shorter linkers than ISWI [24-27]. Curiously, despite acting mostly within gene bodies, binding by these CRs is detected predominantly at gene promoters [28]. Finally, members of the INO80 family, SWR-C and INO80, are implicated in deposition and removal of histone H2A.Z, respectively [29-32], although the latter function is currently controversial [33,34]. In vitro, INO80 but not SWR-C can slide nucleosomes

similarly to ISWI [35] and recently was also shown to be able to move a significant number of +1 nucleosomes to in vivo-like positions on a reconstituted yeast chromatin template [19]. The actual in vivo chromatin state likely results from the interplay between these diverse remodeling activities but the links between them are just starting to emerge.

Several studies, in both yeast and mammalian cells, point to concordant or opposing activities of certain CRs [20,24,36-40]. It is thus of interest to learn how the activities of the complete set of these enzymes are intertwined. Since RSC is the only yeast CR essential for cell viability, the activities of all other CRs in live cells have been studied predominantly by the use of deletion mutants. However, this approach carries the risk of underestimating the role of individual CRs due to compensating effects [41]. Recent analysis with in vitro-reconstituted chromatin and purified remodeling complexes [19] suggests that nucleosome positioning in live cells is achieved by the combined action of CRs and specific transcription factors. However, since many cellular processes such as transcription or replication were not reconstituted in vitro, and the concentrations of proteins used do not necessarily reflect the physiological state, this model awaits rigorous testing in vivo.

By integrating the analysis of novel remodeler binding data with nucleosome occupancy and position changes upon conditional depletion of these complexes we obtained insights into their functionality and the interplay between them. We show that promoter nucleosome arrangements are the net result of combined activities of collaborating and opposing CRs. We demonstrate that the majority of nucleosome rearrangements observed in the absence of a remodeler are caused by the antagonizing activity of other enzymes. As a consequence, the in vivo position of +1 nucleosomes is often determined by the activities of two opposing groups – “pushers” and “pullers” – and has a significant effect on both RNA polymerase II (RNAPII) initiation rates and TSS selection. Remarkably, removal of all “pushers” and “pullers” leaves +1 nucleosome positions largely unaffected, in contrast to removal of only one type of remodeler. Our results provide a detailed picture of mechanisms leading to the establishment of promoter nucleosome architecture and the functional significance of +1 nucleosome position.

Results

Chromatin remodelers bind intergenic regions in specific combinations

To investigate links between the activities of different CRs we performed parallel measurements of remodeler binding, using chromatin endogenous cleavage (ChEC)-seq [42,43], and nucleosome occupancy changes upon conditional depletion of each remodeler, by MNase-seq (Figure 1a). ChEC-seq signals for individual remodeler subunits were normalized to the signal obtained in a strain expressing "free" MNase (see Methods). Since SWR-C is known to lack nucleosome sliding activity [35] we did not investigate this remodeling complex. Strains used for ChEC analysis, which carried MNase tags at endogenous CR subunits genes, did not display any growth defects (Supplementary Fig. 1a).

A distinct MNase-cleavage pattern was observed for each remodeler (RSC, SWI/SNF, ISW2, INO80, ISW1 and CHD1), whilst a common feature noted was the predominant location of signal peaks in intergenic regions, with the majority found at gene promoters (Fig. 1b and Supplementary Fig. 1b). Rsc8 binds at a large number of gene promoters (n=3,702) whereas Swi3, Isw2 and Ino80 bind at a smaller subset (n=466, 1,802, and 1,646, respectively; Supplementary Fig. 1c). Isw1 and Chd1 displayed lower ChEC-seq signals at a large number of promoters (n=2,236 and 2,927, respectively) and a higher signal within the coding regions than other CRs. Due to a frequent overlap between regions bound by various CRs we decided to assemble a single common list of all regions bound by at least one remodeler (see Methods) and calculated normalized ChEC signal of every remodeler in each of these regions. By comparing these normalized ChEC signals at all bound regions, we observed a relatively strong correlation between the Isw2 and Ino80 signals as well as between Isw1 and Chd1, whereas Swi3 binding was anti-correlated with that of Isw1 and Chd1 (Supplementary Fig. 1d)**.**

To further investigate remodeler co-occurrence at distinct intergenic regions we performed k-means clustering of ~5,000 gene promoters for the CRs known to affect promoter nucleosomes (RSC, SWI/SNF, ISW2, and INO80) using CR binding data that were converted to binary form (see Methods for details and Supplementary Table 1). ISW1 and CHD1 were not included in the clustering analysis since their promoter binding signals were relatively low and these complexes act predominantly in coding regions [24,26,27]. This analysis identified 8 clusters, with RSC binding observed at a majority of clusters (I-V) representing over two-thirds of all promoters (Fig. 1c and Supplementary Fig. 1e). Slightly less than 8% of RSC-bound promoters (n=294) are also bound

by SWI/SNF (cluster I). RSC-bound promoters more frequently display binding by ISW2 and/or INO80 (n=2,147; clusters I-IV) whereas SWI/SNF-bound promoters display a slight bias for ISW2 co-binding compared to INO80 (cluster VI). Cluster V is bound by RSC but no other CR, whereas cluster VII is bound most prominently by ISW2, sometimes together with INO80. Cluster VIII did not show a clear signal for either RSC, SWI/SNF, ISW2 or INO80. Promoters in different clusters display diverse nucleosome arrangements with SWI/SNF-bound clusters I and VI having the broadest NDRs and clusters VII-VIII, not bound by either RSC or SWI/SNF, the narrowest (Supplementary Fig. 1f). SWI/SNF-bound clusters are also associated with the highest transcription rate, as measured either by RNAPII ChIP-seq [44] or NET-seq [45], whereas cluster VIII is associated with the lowest (Fig. 1g). Moreover, SWI/SNF-bound clusters (I, VI and VII) contain promoters bearing the canonical TATA-box more frequently than other clusters (Supplementary Fig. 1h). Finally, several clusters are enriched for specific functional categories based upon gene ontology (GO) term analysis (Supplementary Table 2), suggesting that their unique remodeler configurations might play some role in their co-regulation. In summary, clustering analysis points to the existence of a limited number of remodeler combinations present at particular genomic locations.

Effects of remodeler depletion define three distinct remodeler classes

To reveal how the activities of CRs interact to establish genomic nucleosome patterns we compared nucleosome occupancy changes upon conditional depletion of chromatin remodeler catalytic subunits at different genomic regions. Rapid depletion was achieved by the anchor-away method [42] or, in cases where the FRB tag required for anchoring conferred a slight growth defect, using the AID* degron system [46,47]. For RSC we used our previously published Sth1 anchor-away data [9]. We confirmed nuclear depletion of FRB-tagged remodeler subunits (Sth1, Snf2, Isw2 and Chd1) by fluorescence microscopy on fixed cells following 60 min of rapamycin treatment (Supplementary Fig. 2a). Similarly, for AID experiments we observed extensive degradation of the targeted protein by 10 or 15 min following auxin addition (Supplementary Fig. 2b). We also showed that the endogenous tags (FRB or AID) had little or no effect on growth in the absence of the inducing agent (Supplementary Fig. 2c). Finally, as an additional control for our set of CR depletion strains and sequencing library preparation procedures, we compared the MNase-seq profiles generated from all of the strains following mock depletion (treatment with vehicle alone) and found an very high correlation (typically >0.95) in all pairwise comparisons (Supplementary Fig. 2d).

Each remodeler depletion caused distinct changes to nucleosome occupancy patterns (Fig. 2a). RSC depletion caused shrinkage of the NDR due to upstream movement of the +1 nucleosome and downstream movement of the -1 (Fig. 2b; [5,17,18,38,48]). Similarly, depletion of the catalytic subunit of the related SWI/SNF complex (Snf2) resulted in an upstream +1 nucleosome shift (Fig. 2b) but the number of affected regions was much lower than for RSC (n=3,010 versus 137). Interestingly, both RSC and SWI/SNF depletion led to the stabilization of FNs at 114 and 84 promoters, respectively (Supplementary Fig. 2e,f). We also noted that genes displaying SWI/SNF-mediated nucleosome rearrangements are characterized by an unusually large NDR (500-750 bp; defined as the distance between the +1 dyad and that of the first stable nucleosome upstream) often occupied by more than one FN, whereas RSC-affected genes show a bi-modal distribution of NDR sizes with a predominant peak at ~350 bp (Fig. 2c).

Depletion of ISW2 and INO80 both yielded similar +1 nucleosome repositioning downstream (i.e. away from the NDR; Fig. 2b) at distinct subsets of genes, with INO80 having an effect on a larger set of genes than ISW2 (n=617 versus 117). Curiously, the absence of these CRs also led to destabilization of a number of nucleosomes throughout the genome, some of which were +1 nucleosomes of annotated genes (15%, n=95 for ISW2 and 57%, n=342 for INO80), as defined by a strong loss of signal at high levels of MNase digestion and a displacement of the dyad axis of >73 bp (Supplementary Fig. 2g,h). Although we do not know whether nucleosome destabilization following either ISW2 or INO80 depletion is a consequence of transcription itself, we found no significant difference in either RNAPII or TBP levels between INO80-affected promoters where the +1 nucleosome was either shifted or destabilized (data not shown). Furthermore, we noted previously [5] that the vast majority of destabilized nucleosomes (FNs) are unaffected by RNAPII depletion. Nevertheless, the action of puller CRs may represent a special case that would need to be addressed by simultaneous depletion of puller CRs together with RNAPII.

Depletion of ISW1 and CHD1 did not result in any notable changes in +1/-1 nucleosome position (Fig. 2b). Rather, genes displayed more disordered intragenic nucleosomes (more variation in peak-to-peak distance and decreased peak heights), consistent with previous studies utilizing deletion mutants ([24]; Fig. 2a). The previously ascribed role of ISW2 in spacing of genic nucleosomes [24] appears to stem from its role in setting the position of the +1 barrier, since the nucleosomes of genes with an unaffected +1 following ISW2 depletion did not display any change in spacing or peak height (data not shown).

Based on the observed effects at promoters and gene bodies, we defined three main groups of nucleosome-repositioning complexes: (1) “pushers” (RSC and SWI/SNF) that can shift nucleosomes away from the NDR and are able to destabilize nucleosomes, (2) “pullers” (ISW2 and INO80) that can shift the +1 (and potentially other intragenic nucleosomes) in the direction of the NDR and (3) “spacers” (ISW1 and CHD1) that control the distance between intragenic nucleosomes without affecting +1 nucleosome position.

Next, we examined the extent to which remodeler binding correlates with nucleosome occupancy changes upon remodeler depletion. To this end, we grouped sites displaying strong remodeler binding signal (>3-fold enrichment over background) for each remodeler and measured nucleosome occupancy changes in the surrounding regions upon depletion of the remodeler. These values were then compared to occupancy changes at sites where the remodeler signal was low (<1.5-fold enrichment). For each “pusher” or “puller” CR we observed significantly higher changes in nucleosome occupancy at remodeler-bound regions (Fig. 2d). For “spacers” we observed no such trend (data not shown) since the effects of their depletion are prominent in gene bodies despite apparent binding of the complexes at promoters [28]. Nevertheless, genes whose promoters were strongly bound by the “spacers” displayed better nucleosome phasing in gene bodies than genes displaying a weak signal (Supplementary Fig. 2i,j). Curiously, though, both bound and unbound genes displayed a similar loss of phasing upon depletion of the respective CR (Supplementary Fig. 2i,j).

Concordant and opposing activities of multiple chromatin remodelers determine nucleosome positions

Existing evidence suggests that the interplay between CRs might include both additive and opposing interactions [20,37-40]. We formulated the following hypotheses which we subsequently tested: (i) CRs with similar activities can act redundantly; (ii) the activity of a remodeler can be suppressed by that of an opposing remodeler(s). The implication of these two propositions would be that the depletion of a remodeler might not yield a measurable effect on nucleosome occupancy/position near its binding site due to compensation by synergistic remodeler(s) or in cases where an opposing remodeler is simultaneously depleted or simply not present. Indeed, nucleosome occupancy changes measured upon depletion of RSC or SWI/SNF at sites strongly bound by both CRs (see Methods) were weaker than at sites bound by just one of them (Supplementary Fig. 3a). Similarly, the effects of ISW2 depletion were stronger at clusters bound by ISW2 only than at clusters bound by ISW2 and INO80 (Supplementary Fig. 3b). In contrast,

the effects of INO80 depletion were stronger when it bound together with ISW2 rather than alone (Supplementary Fig. 3c). However, binding of INO80 was weaker when it bound alone relative to cases where it bound together with ISW2 (Supplementary Fig. 3d).

To test CR redundancy more directly we simultaneously depleted both “pushers” or both “pullers” and compared the resulting nucleosome rearrangements to those of the corresponding single depletions. Double depletion of RSC and SWI/SNF led to stronger changes at sites bound by both CRs than either single depletion (Fig. 3a,b), consistent with a recent report [40]. Interestingly, RSC-SWI/SNF double depletion also had a stronger effect at sites bound by SWI/SNF but not RSC (cluster VI; Fig. 3b). This might indicate that upon Snf2 depletion, RSC is recruited to SWI/SNF targets and partially substitutes for this remodeler. We did not observe similar behavior of SWI/SNF at sites bound by RSC alone (clusters III-V). ISW2-INO80 redundancy was already evident at the level of cell growth: the double depletion is lethal, contrary to either single depletion (Fig. 3c). Consistent with this finding, in every cluster bound by ISW2 or INO80 the double depletion had stronger effects than either single depletion (Fig. 3d), leading to widespread aberrations in nucleosome patterns (downstream shifts and/or destabilization of +1 nucleosomes, and impaired phasing in gene bodies) that were qualitatively similar to single deletions but amplified in magnitude and number of affected genes (Fig. 3e,f). Similarly, simultaneous depletion of both “spacers” led to a stronger loss of nucleosome phasing in gene bodies than either single depletion, consistent with observations from the corresponding deletion mutants ([24]; Supplementary Fig. 3e). In summary, our analysis indicates a considerable degree of redundancy between RSC and SWI/SNF and between ISW2 and INO80. The same is probably true for ISW1 and CHD1.

+1 nucleosomes maintain their positions in the absence of opposing remodeling activities

It is not known what causes the nucleosome repositioning frequently observed upon depletion of a remodeler. Given the opposing activities of “pushers” and “pullers”, we hypothesized that the effects observed upon depletion of one type of remodeler might result from the activity of an enzyme of the other type. Consistent with this idea, the effects of RSC and SWI/SNF depletion were strongest at sites where these CRs co-bound with ISW2 (Supplementary Fig. 4a). Conversely, the effects of INO80 depletion were strongest at sites where this remodeler binds alongside both RSC and SWI/SNF compared to sites where it binds with just one “pusher”. Furthermore, the weakest effect of INO80 depletion was observed at sites where neither “pusher” binds nearby (Supplementary Fig. 4b).

Our results thus indicate that promoter nucleosome arrangement results from the net activity of specific combinations of CRs and that the absence of a given remodeler might result in nucleosome changes caused by the activity of the remaining remodeling complexes. In order to test this proposition more directly we performed experiments in which we simultaneously depleted “pushers” and “pullers”. Based on our binding data (Fig. 1c) RSC is often predicted to be opposed by either ISW2 or INO80, whereas SWI/SNF would mostly be counteracted by ISW2. We concentrated on RSC and ISW2/INO80-bound sites due to the high number of regions having these combinations and the fact that depletion of RSC cannot be compensated by SWI/SNF. If nucleosome rearrangements observed upon RSC depletion – upstream shift of the +1 nucleosome and stabilization of FNs – resulted from the activity of ISW2 or INO80 we would expect to observe milder effects upon simultaneous depletion of RSC with ISW2 or INO80. We first measured nucleosome occupancy centered on the +1 nucleosome of genes associated with RSC and ISW2, in a wild-type strain and in strains where one or both complexes were depleted (Fig. 4a). Interestingly, the nucleosome pattern obtained upon simultaneous depletion of RSC and ISW2 was similar to the one obtained upon depletion of RSC alone (Fig. 4a; compare blue versus green plots). We performed a similar analysis for RSC- and INO80-bound +1 nucleosomes (Fig. 4b). Here, the effect of double depletion was slightly weaker than depletion of RSC alone, as the upstream +1 shift was somewhat less pronounced. Still, the +1 nucleosome clearly shifted upstream in the absence of both CRs and the changes upon double depletion generally resembled RSC depletion alone. These observations might result from (i) redundant activity of the “pullers”, (ii) activity of “spacer” CRs that is more pronounced in the absence of “pullers”, or (iii) an inherent preference of nucleosomes for positions that are more upstream than those observed in wild-type cells. To distinguish between these possibilities, we performed experiments in which we simultaneously depleted RSC together with the two opposing CRs and compared the results to RSC alone or ISW2/INO80 double depletion. At promoters bound by RSC and ISW2, or RSC and INO80, we observed only minor changes in +1 nucleosome position in the absence of all three CRs (Fig. 4c,d). These observations lead to an important notion, namely that when chromatin remodeling is shut down by multiple CR depletion +1 nucleosomes tend to remain relatively close to their positions under wild-type conditions.

Although ISW1 and CHD1 act predominantly within gene bodies, they could influence the position of the +1 nucleosome in a way that might be masked by other CRs. To test this idea, we simultaneously depleted “pullers” and “spacers” and compared the results to depletion of only the two “pullers”. We found that the change in +1 nucleosome position is very similar under these two conditions, in either the presence of RSC (Fig. 4e,f) or in its absence (Fig. 4g,h), arguing against

an additional role for “spacers” in +1 positioning. Nevertheless, genic nucleosome phasing was more strongly disrupted upon depletion of “spacers” and “pullers” compared to “spacers” alone, presumably due to the altered position of the +1 nucleosome, proposed to act as a barrier against which downstream genic nucleosome are phased through a “statistical positioning” mechanism [49,50]. In summary, these results indicate that while most +1 nucleosomes remain robustly positioned in the absence of remodeling activity, genic nucleosomes significantly change their positions when “spacers”, or “spacers” plus “pullers” are depleted.

We also noted that nucleosomes destabilized by either ISW2 or INO80 depletion (i.e. ones that became FNs; Supplementary Fig. 2g,h) were much less affected upon simultaneous depletion of RSC and both “pullers” (Supplementary Fig. 4c,d). This suggests that their destabilization results from a destructive activity of RSC. Moreover, the FNs that became stable upon RSC depletion were not stabilized when RSC was co-depleted with both “pullers” (Supplementary Fig. 4e). Co-depletion of RSC with only one “puller” did not prevent FN stabilization indicating that ISW2 and INO80 are also redundant with respect to this aspect of chromatin organization. Taken together, these results imply that nucleosome destabilization in the absence of “pullers” is due to the destructive activity of RSC. Conversely, substitution of an FN with a stable nucleosome following RSC depletion is mediated by “pullers”.

Combined remodeler action at +1 nucleosome influences TBP binding and TSS selection

We showed recently that the role of RSC in +1 nucleosome placement is crucial for TBP binding, a key step in RNAPII recruitment [9]. However, it is still unknown which steps of gene activation are affected by SWI/SNF, or by the absence of the “pullers”, which should generally increase the accessibility of the TBP binding site. To explore the role of these CRs in gene expression we measured binding of TBP (Spt15) and the RNAPII catalytic subunit (Rpb1) by ChIP-seq following CR depletion.

Genes down-regulated in the absence of SWI/SNF (n=128, at least 1.5-fold decrease in RNAPII levels) displayed a prominent decrease in TBP binding at their promoters that was linked to an upstream shift of the +1 nucleosome and an increase in nucleosome occupancy at the TATA element (Fig. 5a). The genome-wide correlation between TBP and RNAPII occupancy change upon SWI/SNF depletion was strong (Pearson R=0.62; Supplementary Fig. 5a), consistent with a causal relationship. Nevertheless, the anti-correlation between nucleosome occupancy change (in a 300 bp window centered on the TATA element) and TBP signal at the promoters of these down-regulated genes was more modest (Pearson R=-0.20; Supplementary Fig. 5b) compared to that

associated with RSC depletion (R=-0.39; [9]). These results indicate that SWI/SNF, like RSC, facilitates transcription at least in part by promoting TBP binding, but suggests that other mechanisms are also at work. A smaller number of genes (n=39) actually displayed an increase in RNAPII association and a modest increase in TBP signal upon SWI/SNF depletion, consistent with previous studies showing that SWI/SNF can act as a repressor [51]. However, these effects were not associated with major nucleosome occupancy changes (Supplementary Fig. 5c).

Consistent with ISW2-INO80 redundancy with respect to +1 positioning, we observed relatively few genes where their single depletion had a pronounced effect on either TBP binding or RNAPII levels (Supplementary Fig. 5d-f). As expected, double depletion of these two "puller" CRs produced more significant effects, with many more genes displaying increased RNAPII levels (n=1,592, versus 45 and 104 for ISW2 and INO80 single depletions, respectively) and only a few with lower levels (n=52) (Fig. 5b,c). For the up-regulated genes we observed an anti-correlation between nucleosome occupancy changes and TBP binding, albeit moderate (R=-0.25; Supplementary Fig. 5g). Amongst the group of most strongly up-regulated genes that also displayed the most prominent +1 nucleosome shifts (n=735) GO-term analysis identified "response to stimulus" as the most over-represented identifier (n=253; p<1.0e-8). Nevertheless, and as was the case for INO80 depletion alone, the genome-wide RNAPII – TBP binding correlation upon double "puller" depletion was not strong (Supplementary Fig. 5h). Regarding the "spacers", we found that depletion of ISW1 led to both decreases (n=503) and increases (n=332) in RNAPII levels within gene bodies without any significant changes in +1 nucleosome position or promoter TBP binding yet a modest loss of genic nucleosome positioning (Supplementary Fig. 5i,j). In contrast, the absence of CHD1 caused an increase in RNAPII levels at only 116 genes (with none showing a decrease), accompanied by negligible changes in genic nucleosome positioning and no change in TBP binding (Supplementary Fig. 5k).

"Puller" depletion affects TSS selection by facilitating TBP binding at cryptic downstream TATA elements

In addition to affecting RNAPII initiation rates, nucleosome repositioning can also influence TSS selection at individual genes [52,53], possibly giving rise to non-coding transcription [27] or altered levels of potentially function transcripts. In order to test how CRs affect transcription initiation events we first performed a genome-wide rapid amplification of 5' cDNA ends (5'-RACE) analysis [54] in strains depleted for the "pusher" CRs. Depletion of RSC resulted predominantly in an upstream shift of the +1 nucleosome and a decrease of initiation events at genes, as expected

([5,9,18]; Fig, 6a,b), but no marked change in TSS selection. Similarly, SWI/SNF depletion often led to a decrease in transcription initiation at genes where nucleosomes were rearranged (Fig. 6c,d). Four genes displayed additional strong signals 3' of the annotated +1 nucleosome (visible as prominent peaks in the average plot) which were also suppressed in the absence of SWI/SNF (Fig. 6d). In rare cases, though, RSC or SWI/SNF depletion led to either a shift in the TSS or a change in TSS distribution in situations where two or more predominant initiation sites were observed (Supplementary Fig. 6a,b).

Notably, simultaneous "puller" depletion had a dramatic effect on TSS selection at a large number of genes, often leading to a decrease of initiation at the wild type TSS and the appearance of high levels of novel initiation events located downstream (Fig. 6e,f and Supplementary Fig. 6c). When we plotted 5'-RACE signals at genes with increased RNAPII levels upon "puller" depletion we often observed increases in the signal downstream of the TSS without prominent decreases at the original TSS position (Supplementary Fig. 6d), which was not evident at those genes where transcription decreased (Supplementary Fig. 6e).

To investigate in more detail possible links between +1 nucleosome shifts and the appearance of novel TSSs we first identified the single, strongest TSS at each gene following ISW2/INO80 co-depletion and retained those having a normalized signal of at least 150 reads (n=3524 genes). We then measured signal at these TSSs in wild-type conditions and sorted all genes according to the ratio of the "puller" double-depletion to wild-type signal. This identified a large number of sites (n=1372) where transcription clearly initiated more frequently following "puller" depletion (Fig. 7a, "increase", top) but also a smaller number of sites (n=377) where the initiation events became less frequent ("decrease", bottom). Most of the genes with up-regulated novel start sites have well-annotated TSSs in wild-type cells (n=1210 out of 1372; [48], and in most cases (n=946) the novel prominent TSSs following "puller" double depletion were more than 20 bp downstream from the wild-type TSS (Fig. 7b). In only 73 cases was the novel TSS more than 20 bp upstream from the wild-type site. Nevertheless, both cases were associated with very similar downstream shifts of the +1 nucleosome (Fig. 7c and Supplementary Fig. 7).

The observations outlined above begged the question of why genes with similar nucleosome rearrangements following "puller" depletion would vary so much in the intensity and position of TSS changes. The likely answer became apparent when we plotted the distribution of TATA box motifs for both up-regulated and down-regulated genes (Fig. 7c). At sites where transcription initiated more frequently after "puller" depletion we often found a canonical TATA-box peak within 150 bp of the strongest new TSS (n=661 genes), at a position where +1

nucleosome occupancy decreased following “puller” depletion. This was not the case at genes where a decrease in initiation frequency was observed, suggesting that increased downstream initiation following “puller” depletion is often due to the exposure of a “cryptic” TATA box normally occluded in wild-type cells, at least under the conditions of growth employed here. This interpretation is consistent with the increased downstream TBP binding observed at these genes upon ISW2/INO80 double depletion (Fig. 5b). These findings indicate that ISW2 and INO80 act together to repress transcription at a large number of genes through a mechanism linked to upstream movement of the +1 nucleosome and occlusion of a TATA element otherwise capable of driving PIC assembly and transcription.

Discussion

Concordant and opposing remodeler activities establish promoter nucleosome landscapes

We describe here a comprehensive examination of chromatin remodeler binding and action at promoters in living cells. Our approach of measuring nucleosome positions by MNase-seq immediately following rapid depletion of individual or multiple CRs is likely to reveal the direct action of these enzymes, while avoiding secondary effects that might arise from transcriptional changes that occur in gene deletion strains. Comparing nucleosome occupancy changes following remodeler depletion to ChEC analysis of remodeler localization provides additional insights into remodeler targeting and redundancy. Our results imply a central role for CRs in determining transcription initiation rates but also reveal an unanticipated role for these factors in determining precisely where transcription starts at individual genes.

Our findings establish two different remodeler groups with respect to positioning of the canonical +1/-1 promoter nucleosomes, which we refer to as the “pushers” and “pullers”, the former acting to expand the NDR, the latter to contract it. We identify RSC as the most pervasive promoter-directed remodeler, acting as a “pusher” at a majority of protein-coding genes, consistent with previous reports [5,9,17,18,22,38,55,56]. We show that the second “pusher”, SWI/SNF, has a much more restricted set of target genes, primarily working in conjunction with RSC at a set of highly-transcribed genes [40], but also alone at a smaller set of stress responsive genes (based upon their high TATA box frequency and most significant GO-term enrichment). Conversely, both ISW2 and

INO80 function as promoter-specific “pullers” that act to reduce the NDR due to movement of the +1 or -1 nucleosome (or both) towards its center, consistent with previous work [19,22,27,39,57].

A picture that emerges from our study is that of a competition between “pushers” (RSC and SWI/SNF) and “pullers” (ISW2 and INO80), the result of which leads to precise positioning the +1 nucleosome (Fig. 7d). This insight followed from our ability to simultaneously deplete cells of both ISW2 and INO80. Unexpectedly, we found that these two CRs are largely redundant at a significant number of genes. Thus, either ISW2 or INO80 alone are largely sufficient to counteract the effect of the “pushers”, whereas cells depleted of both “pullers” display a significant broadening of the NDR. Our results thus suggest that these two CRs probably constitute the main force counteracting the destructive effect of RSC on the nucleosomes flanking NDRs.

We imagine that both “pushers” and “pullers” act immediately following replication fork passage to re-establish promoter nucleosome architecture, which in yeast appears to occur in a matter of minutes [58-60]; reviewed in [61]). Since our studies were carried out on populations of unsynchronized cells and show that depletion of one remodeler often leads to nucleosome movement dependent upon another, they suggest that opposing CRs might be continuously acting upon promoter nucleosomes, thus maintaining+1 nucleosome position within a highly limited range. This “spring trap” state appears to play a key role in determining the probability of TBP binding and PIC assembly and may poise the +1 nucleosome to change its position whenever an additional factor (e.g. TF binding) shifts the balance in favor of one of the two opposing activities. One example of such dynamic remodeler regulation could be the +1 nucleosome movement that occurs at hundreds of genes during cell state changes in the yeast metabolic cycle [62].

Our finding that +1 and -1 nucleosomes remain relatively stable upon simultaneous depletion of both “pusher” and both “puller” CRs suggests that neither thermal motion nor some additional active process is sufficient to cause a major alteration in the preferred positioning of these nucleosomes. We imagine that under these conditions relatively few cells passage through S phase, and that the few that do might be largely responsible for the minimal displacements observed. Interestingly, genic nucleosomes, which are likely to be subjected to disruption by RNAPII, display a massive loss of positioning upon multiple CR depletion.

Global and local control of remodeler action

As pointed out previously [20], detection of CR binding by ChIP is notoriously problematic, making it difficult to determine the extent to which these factors are targeted to specific promoters.

We showed here that the application of ChEC-seq to remodeler subunits reveals patterns of remodeler binding specificity that correlate well with nucleosome occupancy changes caused by depletion of the corresponding remodeler. Interestingly, these correlations are stronger for the "pushers" (RSC and SWI/SNF) than for the "pullers" (INO80 and ISW2), which, despite their limited binding overlap, are often functionally redundant. This discrepancy between INO80 and ISW2 binding and action may indicate that one factor can rapidly replace the other when it is depleted. Alternatively, ChEC may not capture certain functional remodeler-promoter interactions, perhaps due to either their short half-life or because of limited MNase access upon remodeler binding at certain promoters. Interestingly, we note that RSC can often substitute for SWI/SNF when the latter is depleted, whereas the converse is not the case. This may reflect more stringent co-factor requirements for SWI/SNF binding or activity.

The "pushers" were shown to slide nucleosomes off the edge of DNA templates in vitro, in effect maximizing the size of the linker between nucleosomes [11,12]. The same principle seems to apply at NDRs in living cells – RSC and SWI/SNF act to increase the length of linker DNA separating the +1 and -1 nucleosomes. Although the precise mechanism(s) by which these two CRs act is still unclear, it is important to note that both RSC and SWI/SNF appear to engulf the nucleosomes upon which they act [2,63,64]. Given that these CRs move the +1 and -1 nucleosomes in different directions, it would seem likely that they orient their direction of action with respect to some landmark feature(s) of the NDR that separates these nucleosomes. This could result from the inherent length of relatively nucleosome-free DNA in this region due to the presence of nucleosome-disfavoring poly(dA:dT) tracts or the binding of TFs.

The remaining CRs that we examined (ISW2, INO80, ISW1 and CHD1) all possess similar activities in vitro – they slide nucleosomes towards the central position on a DNA template, thus equilibrating linker length on both sides of a core particle [23,35]. However, their in vivo roles vary, with "spacers" exclusively affecting genic nucleosomes (+2, +3, etc.) and "pullers" acting on the +1 nucleosome (Fig. 2). What could explain this dichotomy? ISW2 and INO80 form larger complexes than ISW1 and CHD1 which might make it more difficult for them to act on densely packed genic nucleosome arrays due to steric hindrance. Moreover, accumulating evidence suggests that "pullers" are targeted to promoters through direct interactions with general or more gene-specific TFs [65], though the only well-documented example to date is that of ISW2 recruitment by Ume6 [66]. Interestingly, though neither ISW1 nor CHD1 are known to associate with promoter-specific factors, fusing CHD1 to the DNA-binding domain of Ume6 leads to nucleosome repositioning at Ume6 binding sites qualitatively similar to that normally carried out by ISW2 [67].

Taken together, these findings suggest that promoter-targeted “pullers” recognize NDRs as extremely long linker DNAs of the +1 or -1 nucleosome, which they shorten, whereas “spacers” primarily scan genic regions where they act to equalize linker lengths.

We also observed nucleosome destabilization upon depletion of the “pullers” caused by the destructive activity of RSC or SWI/SNF. This might result from the collision of a mobilized “pusher”-bound nucleosome with an adjacent core particle as shown by in vitro studies of SWI/SNF [68,69]. “Pullers” might act to directly protect the vulnerable particle or to mediate its re-deposition.

Significance of +1 nucleosome positioning for transcription

We have shown recently that RSC facilitates gene transcription by globally increasing the accessibility of TBP binding sites [9]. Data presented here show that the related chromatin remodeler SWI/SNF has a similar role but that its action is limited predominantly to genes possessing a canonical TATA box in their promoter. The increase in promoter nucleosome occupancy observed in the absence of RSC and SWI/SNF leads to impaired transcription initiation events which become either less frequent or occur at altered positions. Therefore, the “pushers” not only create a “landing spot” for the transcriptional machinery by generating wide NDRs but also participate in accurate TSS selection, consistent with a recent report [22]. Interestingly, general regulatory factors known to influence promoter nucleosome occupancy (e.g. Rap1, Abf1 and Reb1; [5,9,18]) have also recently been shown to suppress spurious initiation events [52,70].

Previous studies showed that NDR expansion in the absence of ISW2 leads to an increase of ncRNA synthesis [27,71]. Studies on the SWI/SNF family CR called esBAF, carried out in mouse embryonic stem cells, suggest that this might be a general feature of at least some CRs [72]. We expand this picture considerably by showing that “puller” double depletion in yeast causes widespread activation of novel downstream TSSs, most likely driven by cryptic TATA elements that become functional upon +1 nucleosome re-positioning. Importantly, these novel TSSs are likely to produce functional transcripts in many cases, suggesting that ISW2/INO80 may have a regulatory role.

In conclusion, we provide a comprehensive view of the effect of CRs on promoter nucleosome positioning in a simple eukaryote (budding yeast). Our results reveal a complex interplay between these factors that plays an important role in determining not only transcription initiation rates but also the precise site of initiation. Results and methods established here will provide a basis for future studies to explore the role of CRs in controlling gene expression under variable growth

conditions. Finally, since the CRs as well as the general features of promoter nucleosome organization are highly conserved in metazoans, we anticipate that our general findings will be relevant to unraveling promoter function in these more complex systems.

Acknowledgments

We thank Mylène Docquier and the iGE3 Genomics Platform (https://ige3.genomics.unige.ch/) at the University of Geneva for high throughput sequencing services, Nicolas Roggli for expert assistance with data presentation and artwork, Benjamin Albert for help with fluorescence microscopy, and all members of the Shore lab for comments and discussions throughout the course of this work. M.J.B. was supported in part by an iGE3 Ph.D. student fellowship. D.S. acknowledges funding from the Swiss National Science Foundation (grant no. 31003A_170153) and the Republic and Canton of Geneva.

Author Contributions

Conceptualization – S.K., D.C., D.L. and D.S.; Formal Analysis – S.K., D.C., M.J.B., R.D., P.B. and D.L.; Investigation – S.K., D.C., M.J.B. and S.M.; Data Curation – S.K., D.C., M.J.B. and R.D.; Writing – Original Draft, S.K. and D.S.; Funding Acquisition – D.S., D.L. and P.B.; Resources – D.S. and D.L.; Supervision, D.S., D.L., and P.B.

Declaration of Interests

The authors declare no competing interests.

References

1. Lai, W.K.M. & Pugh, B.F. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat Rev Mol Cell Biol* **18**, 548-62 (2017).

2. Brahma, S. & Henikoff, S. RSC-Associated Subnucleosomes Define MNase-Sensitive Promoters in Yeast. *Mol Cell* **73**, 238-49 e3 (2018).

3. Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M. & Henikoff, S. Epigenome characterization at single base-pair resolution. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 18318-23 (2011).

4. Kent, N.A., Adams, S., Moorhouse, A. & Paszkiewicz, K. Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. *Nucleic Acids Res* **39**, e26 (2011).

5. Kubik, S. et al. Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. *Mol Cell* **60**, 422-34 (2015).

6. Weiner, A., Hughes, A., Yassour, M., Rando, O.J. & Friedman, N. High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome research* **20**, 90-100 (2010).

7. Xi, Y., Yao, J., Chen, R., Li, W. & He, X. Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. *Genome research* **21**, 718-24 (2011).

8. Rhee, H.S. & Pugh, B.F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295-301 (2012).

9. Kubik, S. et al. Sequence-Directed Action of RSC Remodeler and General Regulatory Factors Modulates +1 Nucleosome Position to Facilitate Transcription. *Mol Cell* **71**, 89-102 e5 (2018).

10. Clapier, C.R., Iwasa, J., Cairns, B.R. & Peterson, C.L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**, 407-422 (2017).

11. Flaus, A. & Owen-Hughes, T. Dynamic properties of nucleosomes during thermal and ATP-driven mobilization. *Mol Cell Biol* **23**, 7767-79 (2003).

12. Kassabov, S.R., Zhang, B., Persinger, J. & Bartholomew, B. SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol Cell* **11**, 391-403 (2003).

13. Boeger, H., Griesenbeck, J., Strattan, J.S. & Kornberg, R.D. Nucleosomes unfold completely at a transcriptionally active promoter. *Mol Cell* **11**, 1587-98 (2003).

14. Clapier, C.R. et al. Regulation of DNA Translocation Efficiency within the Chromatin Remodeler RSC/Sth1 Potentiates Nucleosome Sliding and Ejection. *Mol Cell* **62**, 453-61 (2016).

15. Lorch, Y., Griesenbeck, J., Boeger, H., Maier-Davis, B. & Kornberg, R.D. Selective removal of promoter nucleosomes by the RSC chromatin-remodeling complex. *Nat Struct Mol Biol* **18**, 881-5 (2011).

16. Lorch, Y., Maier-Davis, B. & Kornberg, R.D. Chromatin remodeling by nucleosome disassembly in vitro. *Proc Natl Acad Sci U S A* **103**, 3090-3 (2006).

17. Badis, G. et al. A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol Cell* **32**, 878-87 (2008).

18. Ganguli, D., Chereji, R.V., Iben, J.R., Cole, H.A. & Clark, D.J. RSC-dependent constructive and destructive interference between opposing arrays of phased nucleosomes in yeast. *Genome Res* **24**, 1637-49 (2014).

19. Krietenstein, N. et al. Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell* **167**, 709-721 e12 (2016).

20. Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T. & Pugh, B.F. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* **149**, 1461-73 (2012).

21. Floer, M. et al. A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. *Cell* **141**, 407-18 (2010).

22. Klein-Brill, A., Joseph-Strauss, D., Appleboim, A. & Friedman, N. Dynamics of Chromatin and Transcription during Transient Depletion of the RSC Chromatin Remodeling Complex. *Cell Rep* **26**, 279-292 e5 (2019).

23. Stockdale, C., Flaus, A., Ferreira, H. & Owen-Hughes, T. Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. *J Biol Chem* **281**, 16279-88 (2006).

24. Gkikopoulos, T. et al. A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science* **333**, 1758-60 (2011).

25. Ocampo, J., Chereji, R.V., Eriksson, P.R. & Clark, D.J. The ISW1 and CHD1 ATP-dependent chromatin remodelers compete to set nucleosome spacing in vivo. *Nucleic Acids Res* **44**, 4625-35 (2016).

26. Tirosh, I., Sigal, N. & Barkai, N. Widespread remodeling of mid-coding sequence nucleosomes by Isw1. *Genome Biol* **11**, R49 (2010).

27. Whitehouse, I., Rando, O.J., Delrow, J. & Tsukiyama, T. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**, 1031-5 (2007).

28. Zentner, G.E., Tsukiyama, T. & Henikoff, S. ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. *PLoS Genet* **9**, e1003317 (2013).

29. Kobor, M.S. et al. A Protein Complex Containing the Conserved Swi2/Snf2-Related ATPase Swr1p Deposits Histone Variant H2A.Z into Euchromatin. *PLoS Biol* **2**, E131 (2004).

30. Krogan, N.J. et al. A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Mol Cell* **12**, 1565-76 (2003).

31. Mizuguchi, G. et al. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**, 343-8 (2004).

32. Papamichos-Chronakis, M., Watanabe, S., Rando, O.J. & Peterson, C.L. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**, 200-13 (2011).

33. Wang, F., Ranjan, A., Wei, D. & Wu, C. Comment on "A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme". *Science* **353**, 358 (2016).

34. Watanabe, S. & Peterson, C.L. Response to Comment on "A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme". *Science* **353**, 358 (2016).

35. Udugama, M., Sabri, A. & Bartholomew, B. The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor. *Mol Cell Biol* **31**, 662-73 (2011).

36. Mohd-Sarip, A. et al. DOC1-Dependent Recruitment of NURD Reveals Antagonism with SWI/SNF during Epithelial-Mesenchymal Transition in Oral Cancer Cells. *Cell Rep* **20**, 61-75 (2017).

37. Morris, S.A. et al. Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. *Nat Struct Mol Biol* **21**, 73-81 (2014).

38. Parnell, T.J., Schlichter, A., Wilson, B.G. & Cairns, B.R. The chromatin remodelers RSC and ISW1 display functional and chromatin-based promoter antagonism. *Elife* **4**, e06073 (2015).

39. Tomar, R.S., Psathas, J.N., Zhang, H., Zhang, Z. & Reese, J.C. A novel mechanism of antagonism between ATP-dependent chromatin remodeling complexes regulates RNR3 expression. *Mol Cell Biol* **29**, 3255-65 (2009).

40. Rawal, Y. et al. SWI/SNF and RSC cooperate to reposition and evict promoter nucleosomes at highly expressed genes in yeast. *Genes Dev* **32**, 695-710 (2018).

41. El-Brolosy, M.A. & Stainier, D.Y.R. Genetic compensation: A phenomenon in search of mechanisms. *PLoS Genet* **13**, e1006780 (2017).

42. Haruki, H., Nishikawa, J. & Laemmli, U.K. The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Molecular cell* **31**, 925-32 (2008).

43. Zentner, G.E., Kasinathan, S., Xin, B., Rohs, R. & Henikoff, S. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat Commun* **6**, 8733 (2015).

44. Bruzzone, M.J., Grunberg, S., Kubik, S., Zentner, G.E. & Shore, D. Distinct patterns of histone acetyltransferase and Mediator deployment at yeast protein-coding genes. *Genes Dev* **32**, 1252-1265 (2018).

45. Churchman, L.S. & Weissman, J.S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368-73 (2011).

46. Morawska, M. & Ulrich, H.D. An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast* **30**, 341-51 (2013).

47. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* **6**, 917-22 (2009).

48. van Bakel, H. et al. A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. *PLoS Genet* **9**, e1003479 (2013).

49. Hughes, A.L. & Rando, O.J. Mechanisms underlying nucleosome positioning in vivo. *Annu Rev Biophys* **43**, 41-63 (2014).

50. Kornberg, R.D. & Stryer, L. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res* **16**, 6677-90 (1988).

51. Shivaswamy, S. & Iyer, V.R. Stress-dependent dynamics of global chromatin remodeling in yeast: dual role for SWI/SNF in the heat shock stress response. *Mol Cell Biol* **28**, 2221-34 (2008).

52. Challal, D. et al. General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation. *Mol Cell* **72**, 955-969 e7 (2018).

53. Dreos, R., Ambrosini, G. & Bucher, P. Influence of Rotational Nucleosome Positioning on Transcription Start Site Selection in Animal Promoters. *PLoS Comput Biol* **12**, e1005144 (2016).

54. Malabat, C., Feuerbach, F., Ma, L., Saveanu, C. & Jacquier, A. Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *Elife* **4**(2015).

55. Hartley, P.D. & Madhani, H.D. Mechanisms that specify promoter nucleosome location and identity. *Cell* **137**, 445-58 (2009).

56. Parnell, T.J., Huff, J.T. & Cairns, B.R. RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. *EMBO J* **27**, 100-10 (2008).

57. Whitehouse, I. & Tsukiyama, T. Antagonistic forces that position nucleosomes in vivo. *Nat Struct Mol Biol* **13**, 633-40 (2006).

58. Fennessy, R.T. & Owen-Hughes, T. Establishment of a promoter-based chromatin architecture on recently replicated DNA can accommodate variable inter-nucleosome spacing. *Nucleic Acids Res* **44**, 7189-203 (2016).

59. Vasseur, P. et al. Dynamics of Nucleosome Positioning Maturation following Genomic Replication. *Cell Rep* **16**, 2651-2665 (2016).

60. Yadav, T. & Whitehouse, I. Replication-Coupled Nucleosome Assembly and Positioning by ATP-Dependent Chromatin-Remodeling Enzymes. *Cell Rep* **15**, 715-723 (2016).

61. Ramachandran, S., Ahmad, K. & Henikoff, S. Capitalizing on disaster: Establishing chromatin specificity behind the replication fork. *Bioessays* **39**(2017).

62. Nocetti, N. & Whitehouse, I. Nucleosome repositioning underlies dynamic gene expression. *Genes Dev* (2016).

63. Chaban, Y. et al. Structure of a RSC-nucleosome complex and insights into chromatin remodeling. *Nat Struct Mol Biol* **15**, 1272-7 (2008).

64. Dechassa, M.L. et al. Architecture of the SWI/SNF-nucleosome complex. *Mol Cell Biol* **28**, 6010-21 (2008).

65. Bowman, G.D. & McKnight, J.N. Sequence-specific targeting of chromatin remodelers organizes precisely positioned nucleosomes throughout the genome. *Bioessays* **39**, 1-8 (2017).

66. Goldmark, J.P., Fazzio, T.G., Estep, P.W., Church, G.M. & Tsukiyama, T. The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. *Cell* **103**, 423-33. (2000).

67. McKnight, J.N., Tsukiyama, T. & Bowman, G.D. Sequence-targeted nucleosome sliding in vivo by a hybrid Chd1 chromatin remodeler. *Genome Res* **26**, 693-704 (2016).

68. Dechassa, M.L. et al. SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent nucleosomes. *Mol Cell* **38**, 590-602 (2010).

69. Engeholm, M. et al. Nucleosomes can invade DNA territories occupied by their neighbors. *Nat Struct Mol Biol* **16**, 151-8 (2009).

70. Wu, A.C.K. et al. Repression of Divergent Noncoding Transcription by a Sequence-Specific Transcription Factor. *Mol Cell* **72**, 942-954 e7 (2018).

71. Yadon, A.N. et al. Chromatin remodeling around nucleosome-free regions leads to repression of noncoding RNA transcription. *Mol Cell Biol* **30**, 5110-22 (2010).

72. Hainer, S.J. et al. Suppression of pervasive noncoding transcription in embryonic stem cells by esBAF. *Genes Dev* **29**, 362-78 (2015).

73. David, F.P. et al. HTSstation: a web application and open-access libraries for high-throughput sequencing data analysis. *PLoS One* **9**, e85879 (2014).

74. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* **46**, D260-D266 (2018).

75. Lerdrup, M., Johansen, J.V., Agrawal-Singh, S. & Hansen, K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nat Struct Mol Biol* **23**, 349-57 (2016).

Methods

Yeast strains

All experiments presented in this study were performed using budding yeast *Saccharomyces cerevisiae* as the model system. Yeast strains used in this study are listed in Supplementary Table 3. For ChIP-seq of Rpb1 and TSS-seq experiments crosslinked chromatin obtained from fission yeast *Schizosaccharomyces pombe* was used as a spike-in control. In a typical experiment, saturated overnight cultures were diluted to OD_{600} = 0.1, grown in YPAD medium at 30°C. Cells were collected for analysis at OD_{600} ≈ 0.35.

Protein depletion experiments

Anchor-away of FRB-tagged protein was induced by the addition of rapamycin (1 mg/ml of 90% ethanol/10% Tween 20 stock solution) to the culture media to a final concentration of 1 µg/ml for 1h [42]. Degradation of AID*-tagged proteins was obtained by addition of IAA to a final concentration of 0.5 µM for 30 min. In experiments in which anchor-away and degron were used simultaneously the cells were treated with rapamycin, after 30 min IAA was added to the culture and cells were grown for another 30 min before harvesting.

The efficiency of protein depletion was monitored by fluorescence microscopy of cells bearing FRB-GFP-tagged fusion proteins. Briefly, cells fixed with cold methanol by a 6-min incubation at -20°C, centrifugated, resuspended in PBS+DAPI solution (20 ng/ml final DAPI concentration), incubated for 5 min, washed once and resuspended in PBS for microscopy (Molecular Devices ImageXpress Micro XL). Degradation of AID*-tagged proteins was monitored by western blotting.

ChEC-seq

ChEC-seq experiments were performed essentially as described [9,43]. A strain expressing “free” MNase under the control of the *REB1* promoter was used as a control. Briefly, cells were washed and resuspended in buffer A (15 mM Tris 7.5, 80 mM KCl, 0.1 mM EGTA, 0.2 mM spermine, 0.5 mM spermidine, 1xRoche EDTA-free mini protease inhibitors, 1 mM PMSF) with 0.1% digitonin and incubated for 5 min at 30°C. Calcium chloride was added to the final concentration of 2 mM to induce MNase activity. Reactions were stopped after 1 min by adding EGTA to a final concentration of 50 mM. DNA was purified using MasterPure Yeast DNA purification Kit (Epicentre) and small DNA fragments were preserved by purification with AMPure beads

(Agencourt) as described [9]. Libraries were prepared using NEBNext kit (New England Biolabs) as described before [9] and sequenced using HiSeq 2500 in single-end mode. Reads were mapped to the genome (sacCer3 assembly) using bowtie2 through HTSStation [73] and the positions of the 5'-most base of each read were used as the positions of MNase cut sites. All densities were normalized to 10M reads.

MNase-seq

Experiments were performed as described before (Kubik et al., 2018). Yeast cultures were crosslinked, spheroplasted and treated with a range of concentrations of MNase (0.1 to 2.5U) for 45 min at 37°C. Reactions were stopped by addition of 30 mM EDTA and the samples were de-crosslinked by overnight incubation at 65°C in the presence of SDS (0.5%) and proteinase K (0.5 mg/ml). DNA was purified by ethanol precipitation and treated by RNase. Samples chosen for library preparation included one "low MNase" sample where the density of the mono- and di-nucleosomal bands visualized on an agarose gel were approximately equal and one "high MNase" sample where the density of the mono-nucleosomal band was ~90% of total DNA. Sequencing libraries were prepared as described (Kubik et al., 2018). The libraries were sequenced using a HiSeq 2500 in paired-end mode. Mapping of the sequencing data to the sacCer3 genome assembly was performed using bowtie2 through HTSStation [73]. Mapped reads were trimmed by 15 bp from each side when calculating densities to better visualize individual nucleosome peaks. All densities were derived from read counts normalized to the total number of reads for each experiment and displayed as a value per 10M reads. We therefore refer to these values as "normalized reads".

ChIP-seq

ChIP-seq was performed essentially as described before [9]. Crosslinked cells were lysed by bead-beating, chromatin was sonicated, and the soluble fraction was incubated with the appropriate antibody and magnetic beads for 3h. For RNAPII ChIP-seq 5% (v/v) of crosslinked, sonicated *S. pombe* chromatin was added as a spike-in control prior to antibody addition. The beads were washed, and DNA was eluted, de-crosslinked and purified using High Pure PCR Cleanup Micro Kit (Roche). The libraries were prepared using TruSeq ChIP Sample Preparation Kit (Illumina) according to manufacturer's instructions. The libraries were sequenced using HiSeq 2500 and the reads were mapped to sacCer3 genome assembly using HTSStation [73] (read densities calculated using shift=100 bp, extension=50 bp). All densities were normalized to 10M reads.

TSS-seq

The experiments were performed as described before [52,54]. Total RNA was extracted from the cells using phenol and chloroform and precipitated with ethanol, DNA was digested with DNase I and RNA was extracted and precipitated again. Polyadenylated transcripts were purified using oligo d(T)25 magnetic beads (New England Biolabs). RNA was dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher) and treated with Cap-Clip Acid Pyrophosphatase (Tebu-bio). RNA was then ligated to the biotinylated 5' adaptor and fragmented for 5 min at 70°C in fragmentation buffer (10mM ZnCl2, 10mM Tris pH7.5). The reaction was stopped with 1 µl of 0.5 M EDTA. Ligated RNA molecules were purified using streptavidin magnetic beads (New England Biolabs). Reverse transcription was performed with RevertAid reverse transcriptase (ThermoFisher) and cDNAs were purified with Agencourt RNAClean XP beads (Beckman Coulter). DNA was amplified with LA Taq DNA polymerase (Takara) and purified with NucleoMag NGS Clean-up and Size Select (Macherey-Nagel). The resulting libraries were sequenced in single-end mode and the results were mapped to sacCer3 genome assembly.

ChEC-seq signal normalization

In our previous work [9], ChEC-seq was normalized by calculating the ratio between the ChEC-seq tag counts at a position (i.e. Rsc8-MNase cuts sites) and the tag counts of free MNase at the same site. Although this approach is generally correct and robust for Rsc8 it has a serious disadvantage: regions of low cut frequency tend to have high variation in signal ratio that might not reflect true binding events but instead result from a random fluctuation of the sequencing signal. This increases the noise in the data and reduces the possibility of finding true binding events, particularly for weak sites. Smoothing the ratio by calculating an average of ratio values in neighboring sites partially reduces the noise but at the same time reduces the precision of the technique.

We turned to a non-parametric normalization method for ChEC-seq data that reduces the noise without reducing the precision. Our method uses an empirical Bayesian estimation of the prior distribution (in this case, the ratio between the ChEC-seq signal for the tested protein and for free-MNase) to increase the signal to noise ratio by reducing the effect of random fluctuations in low coverage areas. As a result, at low-coverage regions the ratio is decreased to the genome-wide average. Empirical Bayes estimation uses signal ratios (scaled between 0 and 1) as the prior. The scaling was done by dividing the number of ChEC seq reads in a 10 bp window by the total number of reads in that window (i.e. ChEC-seq + MNase-seq + 1). Distribution of the ratios calculated

genome-wide is fit to the beta distribution (as the observed distribution was unimodal) and the α and β parameters of the distribution are used to adjust the signal ratio according to the equation $\hat{R}=(Ti+\alpha)/(Mi+\alpha+\beta)$, where Ti is the signal in the test sample and Mi is the signal for control (free MNase). Such adjusted ratio was used in all subsequent analysis of ChEC signal.

ChEC peak calling and clustering

Peaks of protein binding signal were determined from genome-wide normalized ChEC ratio (see above) using the peak-finding algorithm described in [5] and available at https://gitlab.unige.ch/JLFalcone/peakmatic with the minimal normalized signal threshold of 5 and the window size of 100 bp. Peaks determined for different remodelers were pooled and all regions found within 150 bp of each other were merged. This common list of all chromatin remodeler binding sites was used to calculate the average normalized signal for each remodeler +/- 75 bp from the midpoint of each region. For analysis of promoters, signal was calculated in the region spanning -250 to -100 bp from the dyad of the +1 nucleosome for every gene with a well annotated TSS [48]. In the next step, each region displaying a signal of at least 2 was assigned the value of 1 and below 2 was assigned the value of 0. For Swi3, due to significantly higher peak signals, the threshold was set to 6. The list was then k-means clustered according to the 0/1 values with k=8, excluding data for remodelers whose depletion did not significantly affect promoter nucleosomes (i.e. ISW1 and CHD1). The k value was chosen empirically and validated by tabulating the occupancy for all possible combinations of 4 remodelers (n=16) present at the promoters and counting the number of occurrences of each group. The most abundant combinations represented individual clusters in our analysis while the less abundant ones, displaying similar occupancy of <4 remodelers, were merged by the clustering.

Nucleosome occupancy and stability change

Nucleosome occupancy change (either positive or negative) was calculated in 10 bp windows as the log2 ratio of read counts in remodeler-depleted cells compared to mock-treated cells (using high concentration MNase-seq data; Fig. 4d). To quantify the average overall magnitude of nucleosome occupancy change at promoter regions, absolute values for read count differences between CR-depleted and mock-treated cells were used (Fig. 2c and 3b,d).

We considered +1 nucleosome occupancy as changed if the absolute log2 ratio of occupancy, calculated in the region spanning -/+ 150 bp from +1 dyad, was higher than 0.7. To estimate nucleosome stability changes nucleosome occupancy was calculated in a region -/+50 bp from

each nucleosome dyad in remodeler-depleted and mock-treated cells. A fragile nucleosome was considered to become stabilized by remodeler depletion if its average occupancy in the high MNase assay increased by at least 15 normalized reads and its dyad was found within 50 bp of its original position (from [5]). A nucleosome was considered to be destabilized by remodeler depletion if its average occupancy in the high MNase assay decreased by at least 15 normalized reads and its dyad was not found within 73 bp of its original position (determined in mock-treated cells).

ChIP-seq spike-in normalization and quantification

RNAPII ChIP-seq signal in *S. cerevisiae* was normalized using a *S. pombe* spike-in control as described before [44]. TBP binding was calculated in regions spanning 200 bp centered on all TATA and TATA-like sites, taken from [8]. RNAPII binding signal was calculated in the transcribed region of all genes with well determined TSSs and TTSs (based on [48]) and in the ORF for all other genes. RNAPII signals that decreased/increased by 1.5-fold, and where the average signal was at least 30 normalized reads/bp in the lowest case (treated or mock, respectively) were considered as up-regulated/down-regulated genes, respectively. Genes were considered as not affected if the log2 change in RNAPII signal was in the range >-0.1 and <0.1 and the average signal in the mock-treated sample was at least 30 normalized reads/bp.

TSS determination

TSS signals from three replicates of each experiment were averaged separately for the Watson and the Crick strand. For the analysis shown in Fig. 7 all TSSs in “puller”-depleted cells were found with a minimum signal of 150 normalized reads. For each peak the nearest ATG on the respective strand was found (at a maximum distance of 500 bp) and then a single, strongest TSS was identified for each gene. Signals were calculated for each of these TSSs in the wild-type and “puller”-depleted conditions. Signals were considered as decreased or increased following CR depletion when signal log2 ratio (depletion/untreated) bypassed -/+1, respectively. Regions displaying artefactually high signal (e.g. found near rDNA) were removed from the analysis.

TATA-box search

All putative TATA-box sites were searched for by first looking for matches to the canonical TATAWAWR motif using FIMO from the MEME Suite with a threshold of $p<0.001$. Searches were

also performed for motifs with up to 2 substitutions in the consensus or using the frequency matrix determined for TBP (Spt15) binding [74]. All three types of searches yielded very similar motif frequencies for both up- and down-regulated gene classes shown in Fig. 7.

Plots and statistics

Fig. 1c, 4d and Supplementary Figs. 1b,c, 5a,f and 7 were made using EaSeq [75]. For box-and-whisker plots center line, box limits and whiskers indicate median, upper and lower quartiles, and 1.5x interquartile range, respectively. Statistic test were applied where indicated.

Data and software availability

All sequencing data generated in this study were submitted to the GEO database as Series GSE115412 (for ChEC-seq, MNase-seq and ChIP-seq) and Series GSE114589 (TSS-seq). Peak-calling software is available at https://gitlab.unige.ch/JLFalcone/peakmatic.

Figure 1

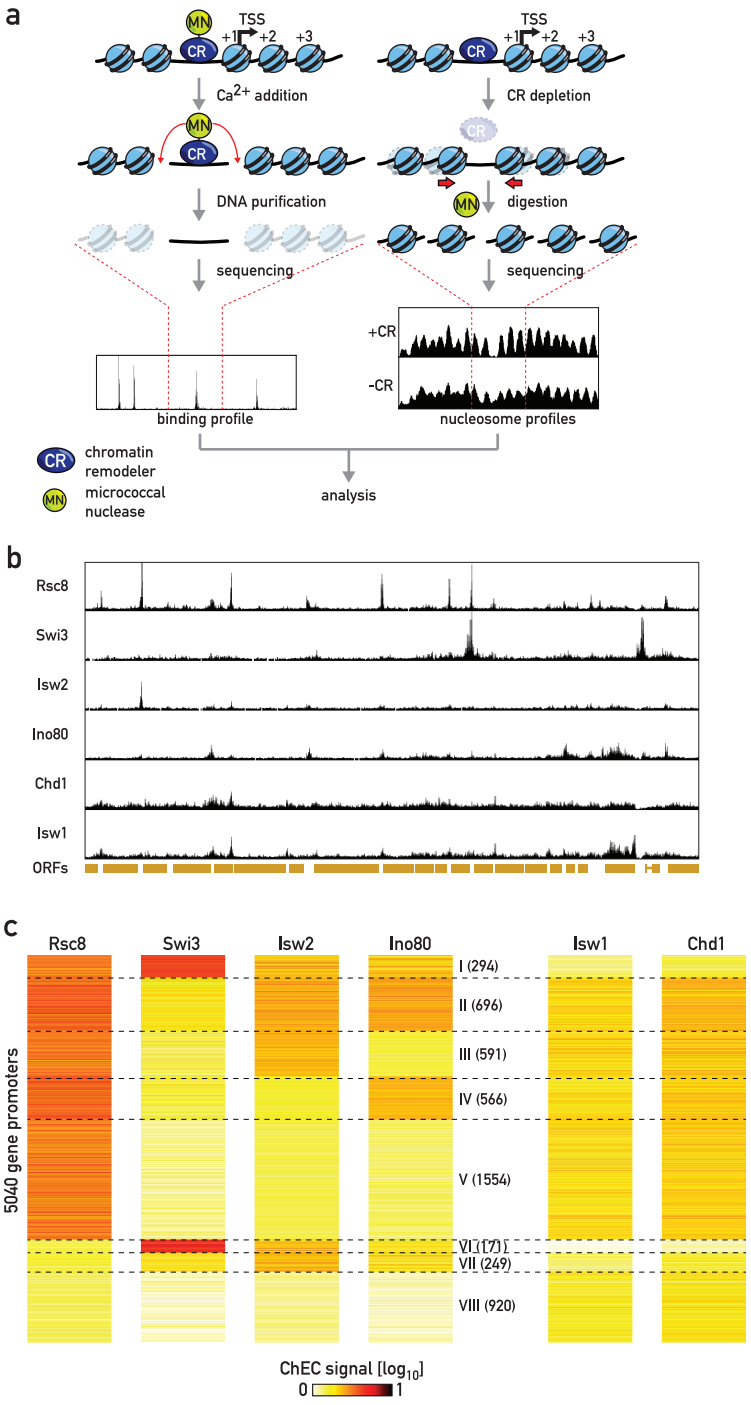

a
MN
CR
TSS
+1
+2
+3
Ca2+ addition
CR depletion
DNA purification
digestion
sequencing
+CR
-CR
binding profile
nucleosome profiles
chromatin remodeler
micrococcal nuclease
analysis
b
Rsc8
Swi3
Isw2
Ino80
Chd1
Isw1
ORFs
c
5040 gene promoters
I (294)
II (696)
III (591)
IV (566)
V (1554)
VI (17)
VII (249)
VIII (920)
ChEC signal [log10]
0
1

Fig. 1 | Chromatin remodelers (CRs) bind in defined combinations. a, Schematic representation of the experimental setup. CR binding was measured by ChEC-seq (left). Remodeler activity was evaluated as the local change in nucleosome occupancy measured by MNase-seq (right) following conditional depletion of the catalytic subunit of a CR. **b**, Snapshot of a genomic region displaying normalized ChEC-seq signal for each CR. **c**, Heatmap representing normalized remodeler ChEC signal at gene promoters clustered by k-means (k=8).

Figure 2

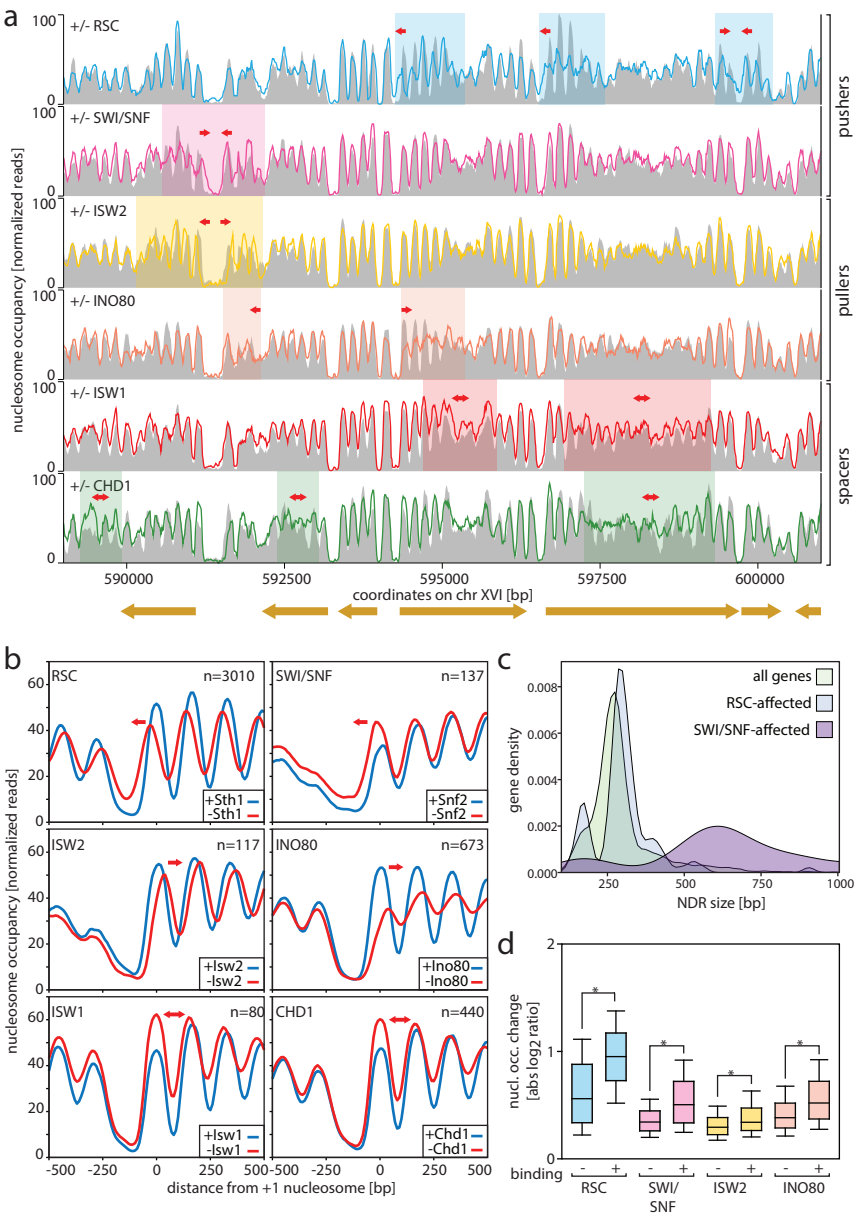

a
+/- RSC
+/- SWI/SNF
+/- ISW2
+/- INO80
+/- ISW1
+/- CHD1
nucleosome occupancy [normalized reads]
pushers
pullers
spacers
590000
592500
595000
597500
600000
coordinates on chr XVI [bp]
b
RSC
n=3010
SWI/SNF
n=137
ISW2
n=117
INO80
n=673
ISW1
n=80
CHD1
n=440
+Sth1
-Sth1
+Snf2
-Snf2
+Isw2
-Isw2
+Ino80
-Ino80
+Isw1
-Isw1
+Chd1
-Chd1
nucleosome occupancy [normalized reads]
distance from +1 nucleosome [bp]
c
all genes
RSC-affected
SWI/SNF-affected
gene density
NDR size [bp]
d
nucl. occ. change [abs log2 ratio]
binding
RSC
SWI/ SNF
ISW2
INO80

Fig. 2 | CRs display three broad types of activity. a, Snapshot of a genomic region displaying nucleosome occupancy in wild-type (grey area) or CR-depleted cells (colored lines); red arrows and shaded areas indicate regions and directions of strong rearrangements for each remodeler depletion. **b**, Nucleosome occupancy at promoters of genes displaying significant changes upon CR depletion in regions centered on the +1 nucleosome dyad. **c**, Density of genes plotted as a function of their NDR size for all genes, or those whose promoter nucleosome stability decreased upon RSC or SWI/SNF depletion. **d,** Boxplot comparing nucleosome occupancy changes upon depletion of RSC, SWI/SNF, ISW2 or INO80 at sites displaying low or high ChEC signal of each depleted remodeler; asterisk indicates significant difference ($p<0.05$, Mann-Whitney test).

Figure 3

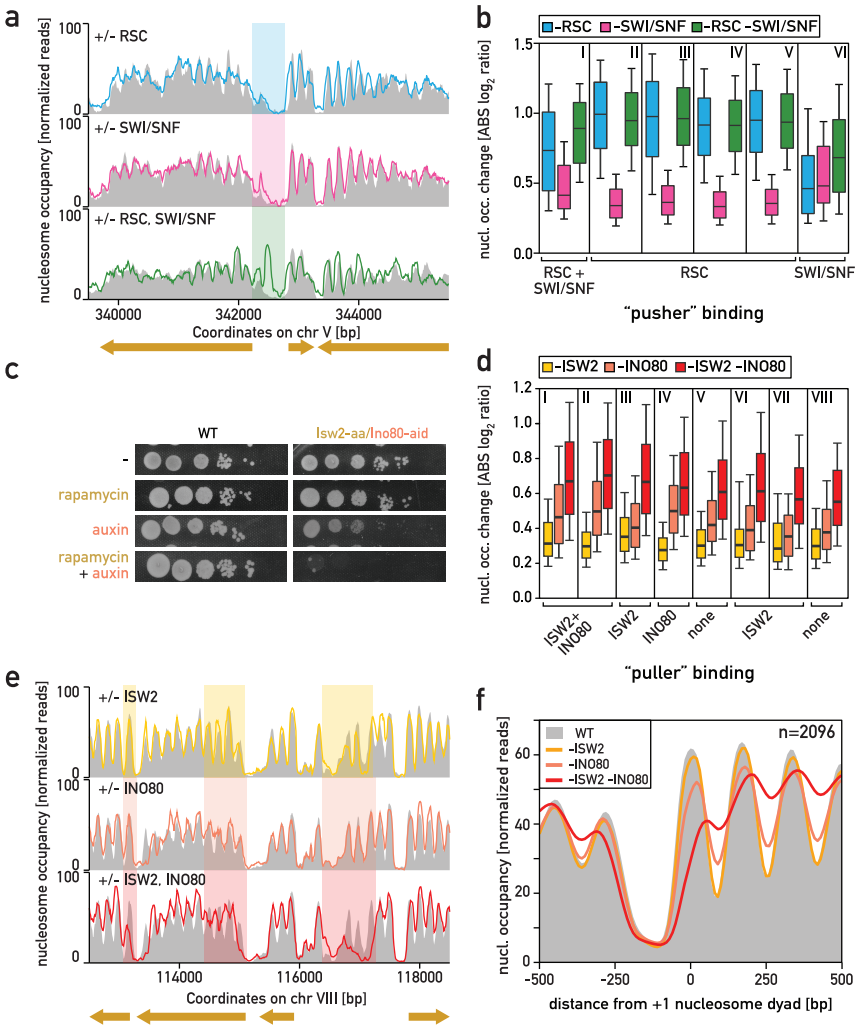

a
nucleosome occupancy [normalized reads]
+/- RSC
+/- SWI/SNF
+/- RSC, SWI/SNF
Coordinates on chr V [bp]
b
-RSC -SWI/SNF -RSC -SWI/SNF
nucl. occ. change [ABS log2 ratio]
RSC + SWI/SNF
RSC
SWI/SNF
"pusher" binding
c
WT
Isw2-aa/Ino80-aid
rapamycin
auxin
rapamycin + auxin
d
-ISW2 -INO80 -ISW2 -INO80
nucl. occ. change [ABS log2 ratio]
ISW2+ INO80
ISW2
INO80
none
"puller" binding
e
+/- ISW2
+/- INO80
+/- ISW2, INO80
Coordinates on chr VIII [bp]
f
WT
-ISW2
-INO80
-ISW2 -INO80
n=2096
nucl. occupancy [normalized reads]
distance from +1 nucleosome dyad [bp]

Fig. 3 | CRs with similar activities act redundantly. a, Snapshot of a sample genomic region displaying stronger nucleosome occupancy change upon simultaneous RSC and SWI/SNF depletion compared to depletion of individual complexes. **b**, Nucleosome occupancy changes upon depletion of RSC, SWI/SNF or both CRs simultaneously calculated in each cluster binding these complexes. **c**, Spot assay of wild-type yeast strain (left) and strain in which Isw2 was tagged with FRB in order to deplete it with rapamycin and Ino80 was tagged by AID* for auxin-mediated depletion (right) plated on medium containing rapamycin, auxin or both chemicals simultaneously. **d**, Nucleosome occupancy changes upon depletion of ISW2, INO80 or both CRs simultaneously calculated for all clusters. **e**, Snapshot of a representative genomic region displaying a stronger nucleosome occupancy change upon simultaneous ISW2 and INO80 depletion compared to depletion of either individual complex. **f**, Average nucleosome occupancy plot for wild-type cells and cells depleted of ISW2, INO80 or both CRs simultaneously, averaged over all genes where a significant change of +1 nucleosome occupancy was observed upon simultaneous depletion.

Figure 4

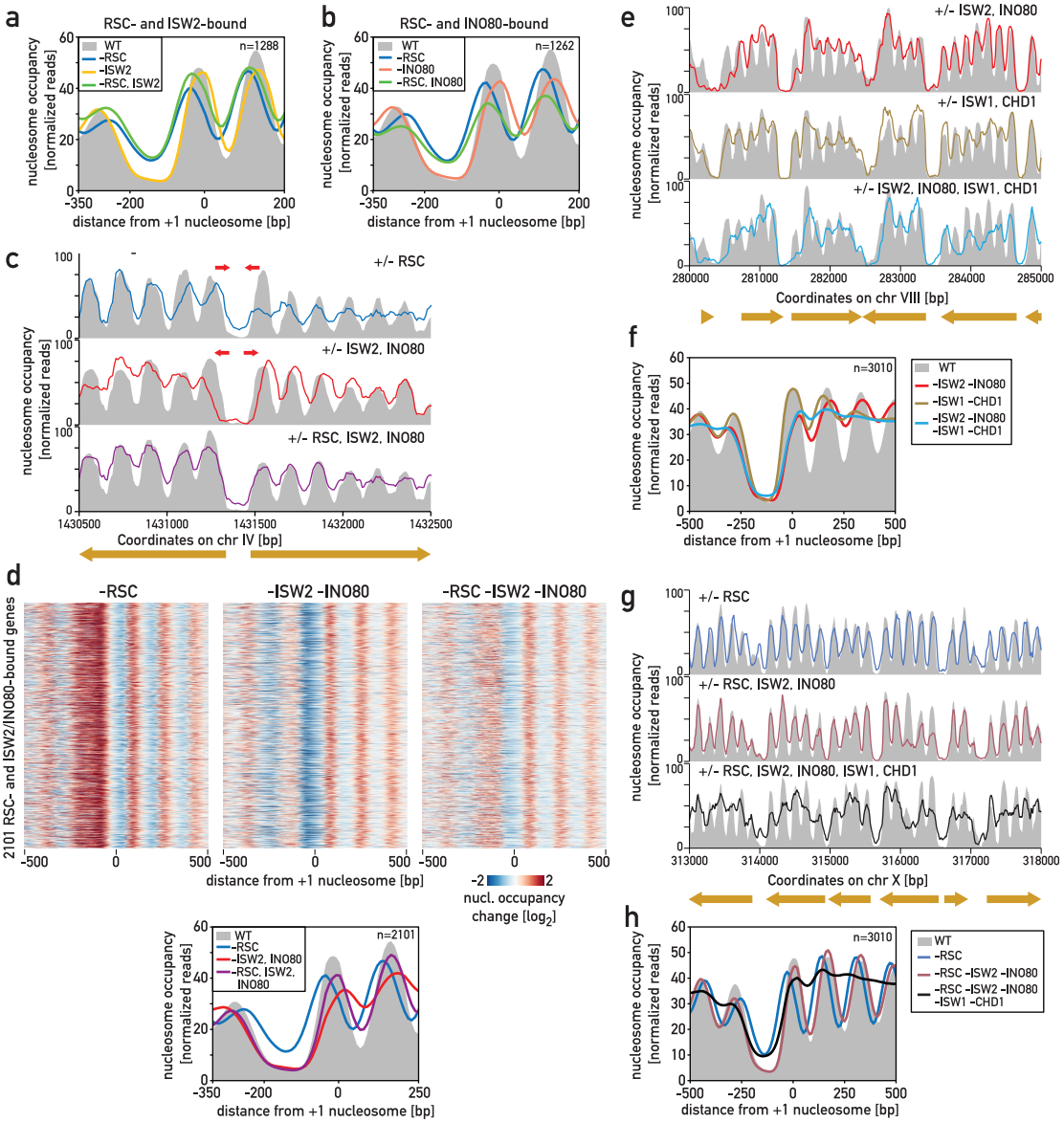

a
RSC- and ISW2-bound
WT
-RSC
-ISW2
-RSC, ISW2
n=1288
nucleosome occupancy [normalized reads]
distance from +1 nucleosome [bp]
b
RSC- and INO80-bound
WT
-RSC
-INO80
-RSC, INO80
n=1262
nucleosome occupancy [normalized reads]
distance from +1 nucleosome [bp]
c
+/- RSC
+/- ISW2, INO80
+/- RSC, ISW2, INO80
nucleosome occupancy [normalized reads]
Coordinates on chr IV [bp]
d
-RSC
-ISW2 -INO80
-RSC -ISW2 -INO80
2101 RSC- and ISW2/INO80-bound genes
distance from +1 nucleosome [bp]
nucl. occupancy change [log2]
WT
-RSC
-ISW2, INO80
-RSC, ISW2, INO80
n=2101
nucleosome occupancy [normalized reads]
distance from +1 nucleosome [bp]
e
+/- ISW2, INO80
+/- ISW1, CHD1
+/- ISW2, INO80, ISW1, CHD1
nucleosome occupancy [normalized reads]
Coordinates on chr VIII [bp]
f
n=3010
WT
-ISW2 -INO80
-ISW1 -CHD1
-ISW2 -INO80 -ISW1 -CHD1
nucleosome occupancy [normalized reads]
distance from +1 nucleosome [bp]
g
+/- RSC
+/- RSC, ISW2, INO80
+/- RSC, ISW2, INO80, ISW1, CHD1
nucleosome occupancy [normalized reads]
Coordinates on chr X [bp]
h
n=3010
WT
-RSC
-RSC -ISW2 -INO80
-RSC -ISW2 -INO80 -ISW1 -CHD1
nucleosome occupancy [normalized reads]
distance from +1 nucleosome [bp]

Fig. 4 | Position of +1 nucleosome results from the net activity of multiple cooperating and opposing CRs. a, Average nucleosome occupancy upon depletion of RSC, ISW2 and both CRs simultaneously at all genes displaying binding of RSC and ISW2. **b,** As (**a**) but for RSC and INO80-bound promoters. **c,** Snapshot of a representative genomic region displaying strong opposing changes in +1 nucleosome position upon depletion of RSC or the two “pullers” (ISW2 and INO80) and only minor changes upon simultaneous depletion of all three CRs. **d,** Heatmaps of nucleosome occupancy change (top three panels) and average plots of nucleosome occupancy (bottom panel) for cells depleted of either “pullers” or RSC and “pullers” simultaneously, at genes bound by these CRs. **e**, Snapshot of a sample genomic region displaying nucleosome occupancy changes shown for depletion of the “pullers” (ISW2, INO80), the “spacers” (ISW1, CHD1) and all four CRs simultaneously. **f**, Nucleosome occupancy in wild-type cells and cells depleted of both “pullers”, both “spacers” and all four CRs simultaneously, averaged over all genes displaying +1 nucleosome occupancy changes upon RSC depletion. **g,** Snapshot of a representative genomic region displaying nucleosome occupancy changes following depletion of RSC (top), RSC and “puller” (RSC, ISW2, INO80; middle), and RSC, “puller” and “spacer” simultaneously (bottom). **h**, Average nucleosome occupancy plots for wild-type cells and cells depleted of RSC, RSC and both “pullers” and RSC, “pullers” and “spacers” simultaneously.

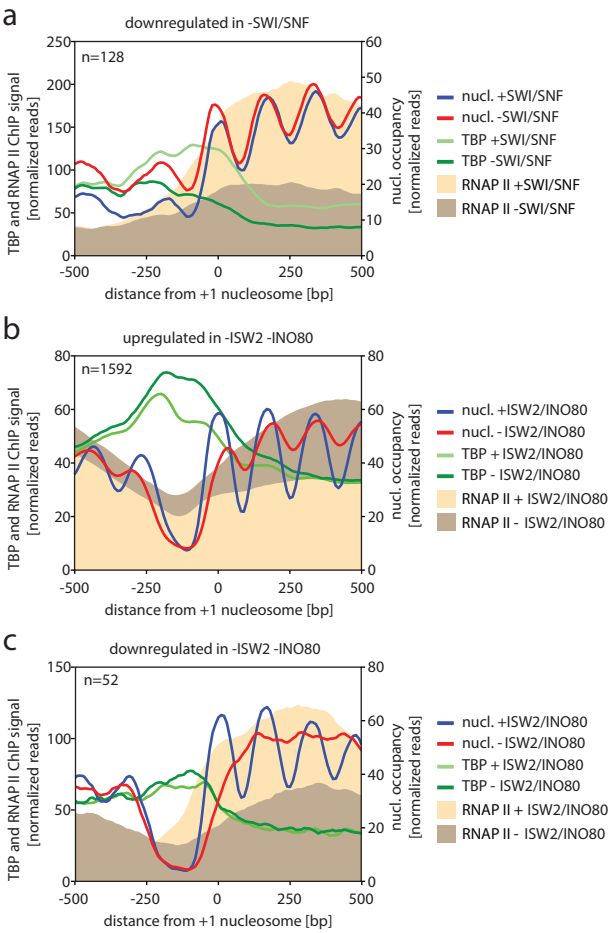

a
downregulated in -SWI/SNF
n=128
TBP and RNAP II ChIP signal [normalized reads]
nucl. occupancy [normalized reads]
distance from +1 nucleosome [bp]
nucl. +SWI/SNF
nucl. -SWI/SNF
TBP +SWI/SNF
TBP -SWI/SNF
RNAP II +SWI/SNF
RNAP II -SWI/SNF
b
upregulated in -ISW2 -INO80
n=1592
TBP and RNAP II ChIP signal [normalized reads]
nucl. occupancy [normalized reads]
distance from +1 nucleosome [bp]
nucl. +ISW2/INO80
nucl. - ISW2/INO80
TBP + ISW2/INO80
TBP - ISW2/INO80
RNAP II + ISW2/INO80
RNAP II - ISW2/INO80
c
downregulated in -ISW2 -INO80
n=52
TBP and RNAP II ChIP signal [normalized reads]
nucl. occupancy [normalized reads]
distance from +1 nucleosome [bp]
nucl. +ISW2/INO80
nucl. - ISW2/INO80
TBP + ISW2/INO80
TBP - ISW2/INO80
RNAP II + ISW2/INO80
RNAP II - ISW2/INO80

Fig. 5 | Changes in +1 nucleosome occupancy are linked to transcriptional down- and up-regulation. a, Plots of nucleosome occupancy, RNAPII and TBP ChIP signals, in the presence and absence of SWI/SNF, at genes displaying a significant decrease in RNAPII level upon SWI/SNF depletion. **b**, Plots of nucleosome occupancy, RNAPII and TBP ChIP signals, in the presence and absence of ISW2 and INO80, at genes displaying a significant increase in RNAPII level upon simultaneous depletion of ISW2 and INO80. **c**, As (**b**) but for down-regulated genes.

Figure 6

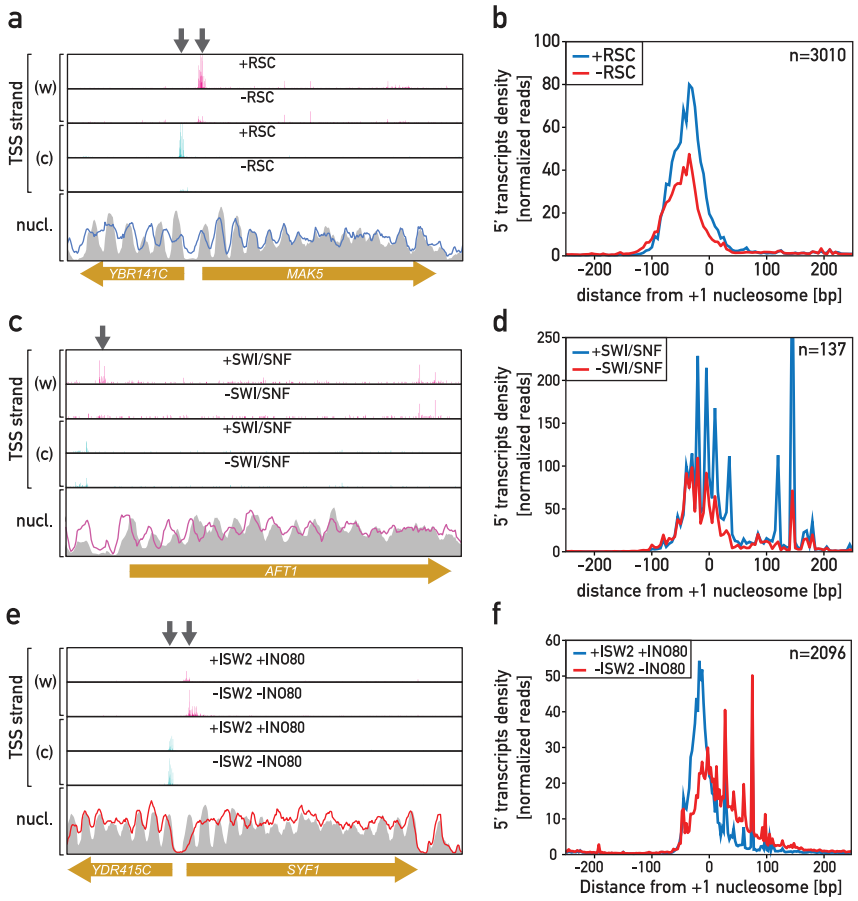

a
TSS strand
(w)
(c)
+RSC
-RSC
+RSC
-RSC
nucl.
YBR141C
MAK5
b
+RSC
-RSC
n=3010
5' transcripts density [normalized reads]
distance from +1 nucleosome [bp]
c
+SWI/SNF
-SWI/SNF
+SWI/SNF
-SWI/SNF
nucl.
AFT1
d
+SWI/SNF
-SWI/SNF
n=137
5' transcripts density [normalized reads]
distance from +1 nucleosome [bp]
e
+ISW2 +INO80
-ISW2 -INO80
+ISW2 +INO80
-ISW2 -INO80
nucl.
YDR415C
SYF1
f
+ISW2 +INO80
-ISW2 -INO80
n=2096
5' transcripts density [normalized reads]
Distance from +1 nucleosome [bp]

Fig. 6 | +1 nucleosome shift interferes with transcription start site selection. a, Snapshot of genomic region showing 5'RACE signal ("TSS") for the Watson (w) and the Crick (c) strands as well as nucleosome occupancy in the presence (grey background) and absence of RSC (colored line). **b**, Average plot showing 5'RACE signal, in the presence and absence of RSC, for all genes displaying significant occupancy changes at their +1 nucleosome upon RSC depletion. **c**, As in (**a**) but for depletion of SWI/SNF. **d**, As in (**b**) but for depletion of SWI/SNF. **e**, As in (**a**) but for simultaneous depletion of ISW2 and INO80. **f**, As in (**b**) but for simultaneous depletion of ISW2 and INO80.

Figure 7

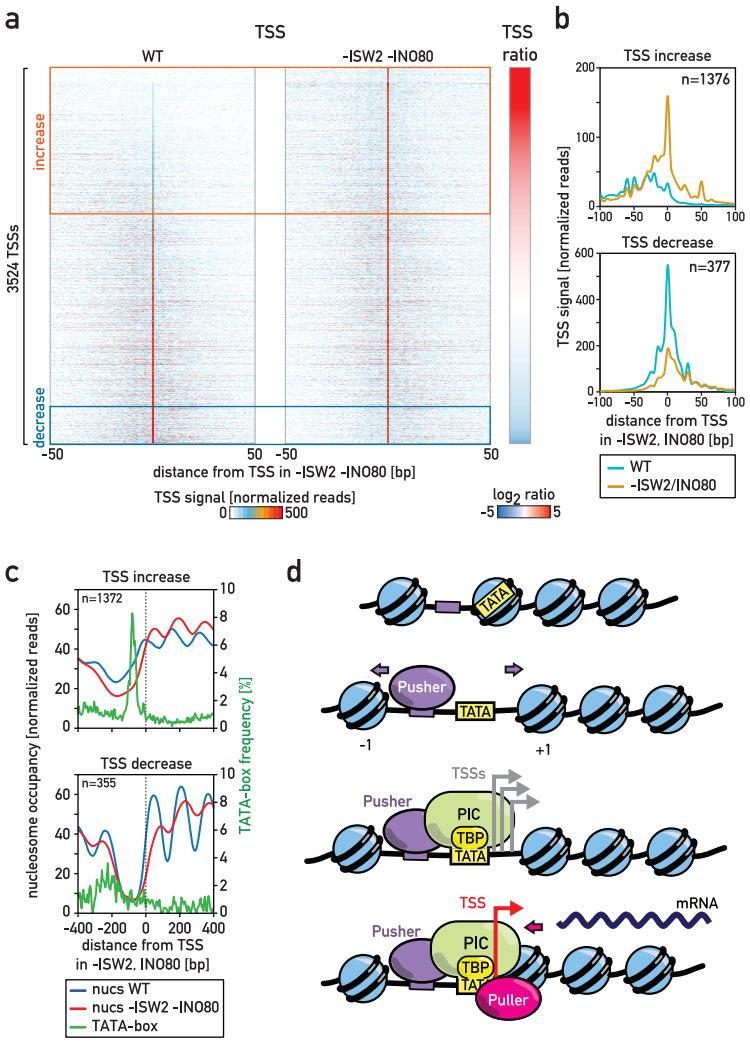
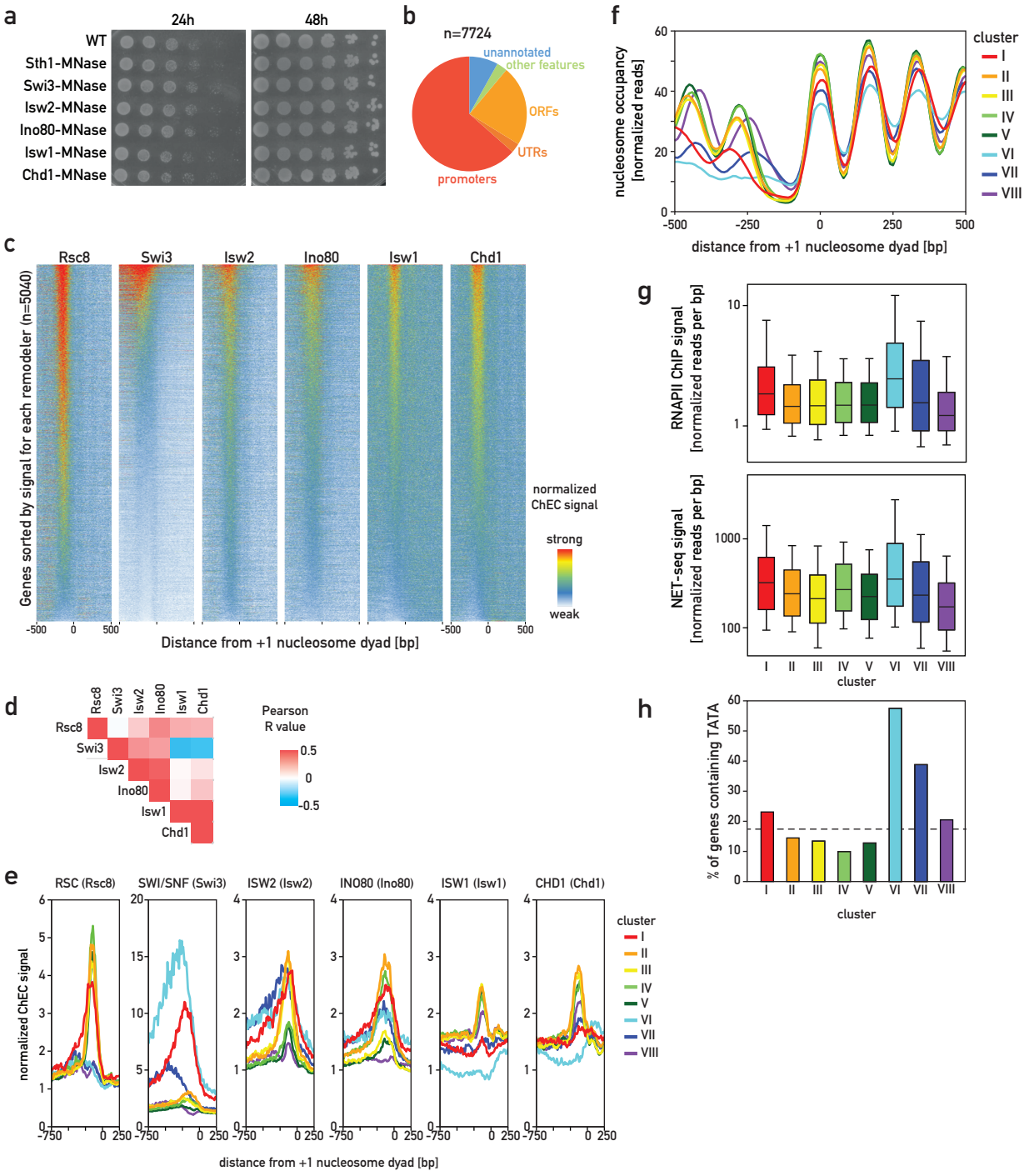

a
TSS
WT
-ISW2 -INO80
TSS ratio
increase
decrease
3524 TSSs
-50
50
-50
50
distance from TSS in -ISW2 -INO80 [bp]
TSS signal [normalized reads]
0
500
log2 ratio
-5
5
b
TSS increase
n=1376
TSS decrease
n=377
TSS signal [normalized reads]
distance from TSS
in -ISW2, INO80 [bp]
WT
-ISW2/INO80
c
TSS increase
n=1372
TSS decrease
n=355
nucleosome occupancy [normalized reads]
TATA-box frequency [%]
distance from TSS
in -ISW2, INO80 [bp]
nucs WT
nucs -ISW2 -INO80
TATA-box
d
TATA
Pusher
-1
+1
TSSs
PIC
TBP
TSS
mRNA
Puller

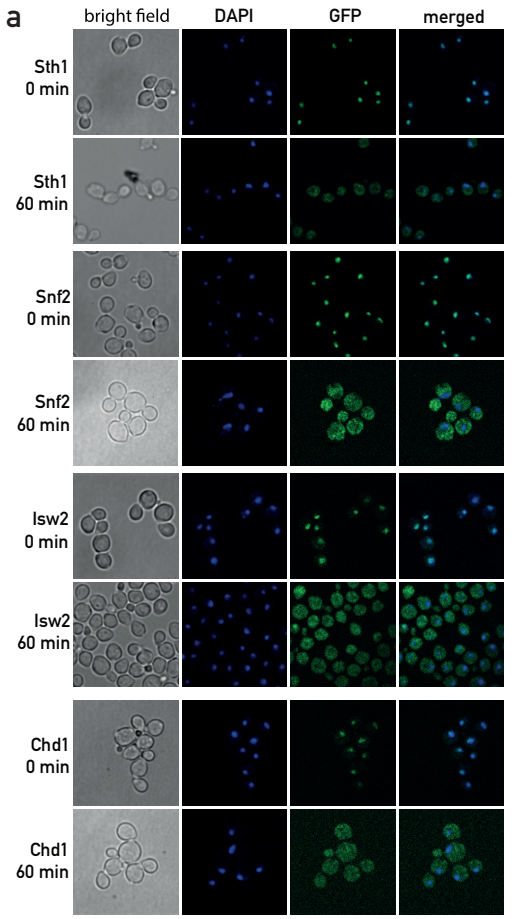
Fig. 7 | a, Heatmap showing 5'RACE signal in wild-type cells (left) and cells depleted of ISW2 and INO80 centered on predominant TSS site determined in the absence of these CRs for each gene. **b**, Average 5'RACE signal for genes displaying most significant increase (top) or decrease (bottom) in the signal. **c**, Plots displaying nucleosome occupancy in the presence (blue) and absence (red) of ISW2 and INO80 as well as average frequency of the consensus TATA-box motif (green) for genes displaying most significant increase (top) or decrease (bottom) in the 5'RACE signal. **d**, Schematic representation of mechanisms determining +1 nucleosome position and TSS selection at active genes. Recruitment of "pushers" such as RSC might be guided by specific DNA motifs or TFs leading to creation/expansion of the NDR, exposition of TBP binding sites (TATA) and formation of the PIC; "pullers" reposition the +1 nucleosome to reduce NDR size and to restrict transcription initiation to the position observed in wild-type cells.

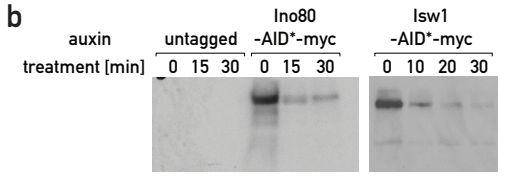
Supplementary Figure 1

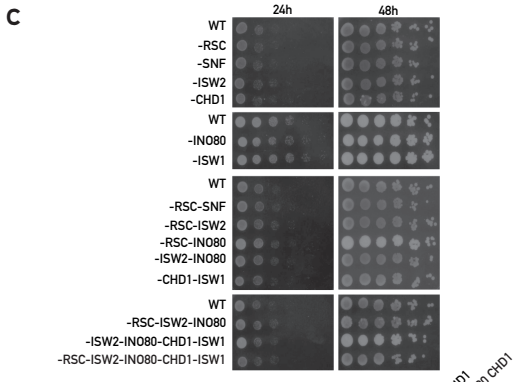


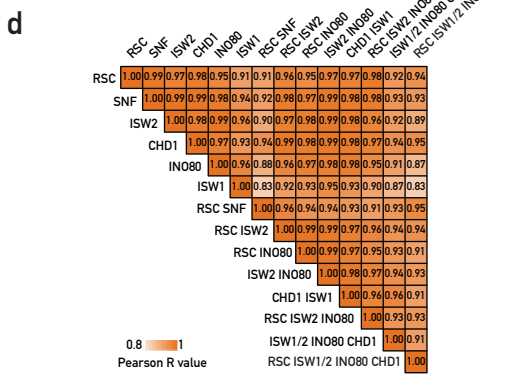
Supplementary Fig. 1. | Characterization of remodeler binding by ChEC-seq: experiments related to Fig. 1. a, Growth assays (serial dilution "spot assays") of the indicated MNase-tagged chromatin remodeler strains on YPAD medium, compared to the wild-type parent strain (WT). Plates were photographed following 24 or 48 hrs incubation at 30°C. **b,** Distribution of identified remodeler binding sites between different genomic features. **c**, Heatmaps displaying ChEC signal for every CR centered on the +1 nucleosome of 5,040 protein-coding genes, sorted (top to bottom) by decreasing signal calculated in a region -250 to -50 bp from +1 nucleosome dyad. **d**, Grid representing Pearson correlation coefficients between ChEC signal for different CRs calculated at all identified remodeler binding sites. **e**, Plots displaying average normalized ChEC signal for each CR calculated for all clusters. **f**, Average nucleosome occupancy in each cluster. **g**, Boxplots displaying expression of genes belonging to each cluster measured either as RNAPII ChIP-seq signal in the gene body or NET-seq signal. **h**, Fraction of genes in each cluster which contain a canonical TATA-box dashed line indicates genomic average (~17%).

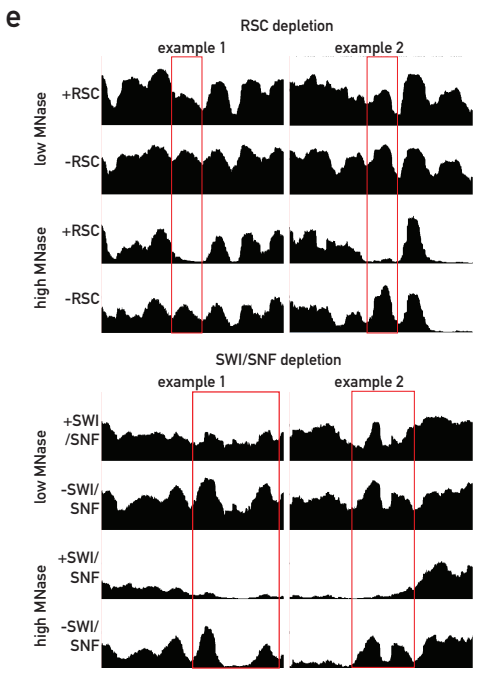
Supplementary Figure 2

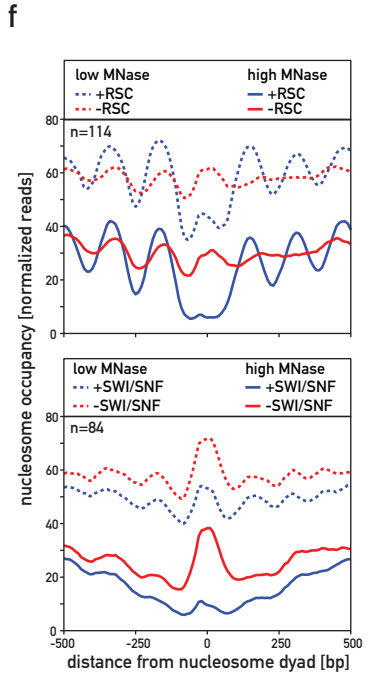


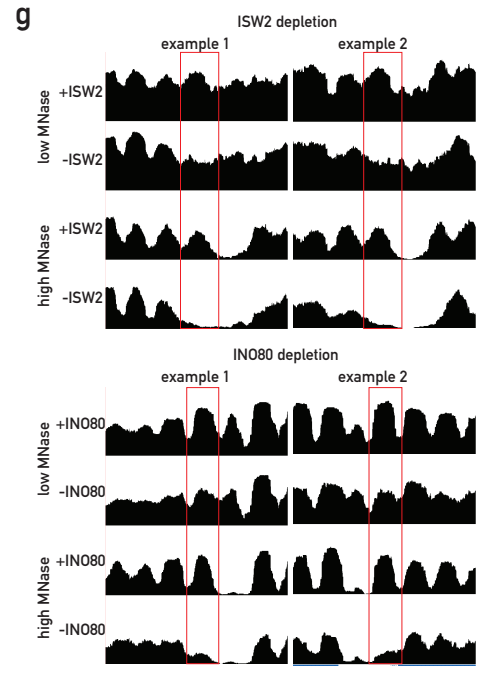


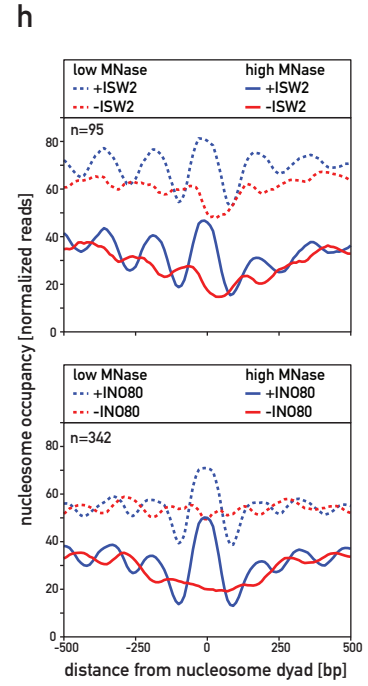


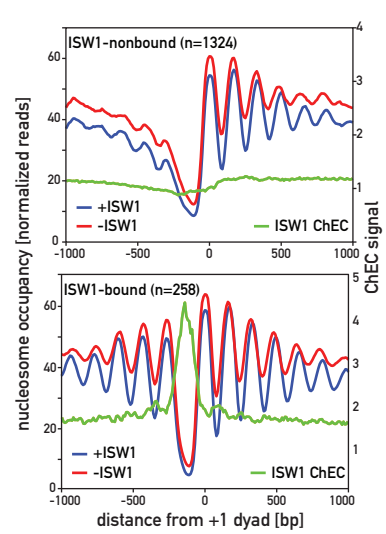


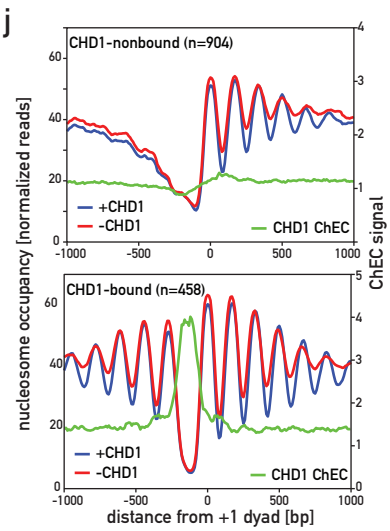






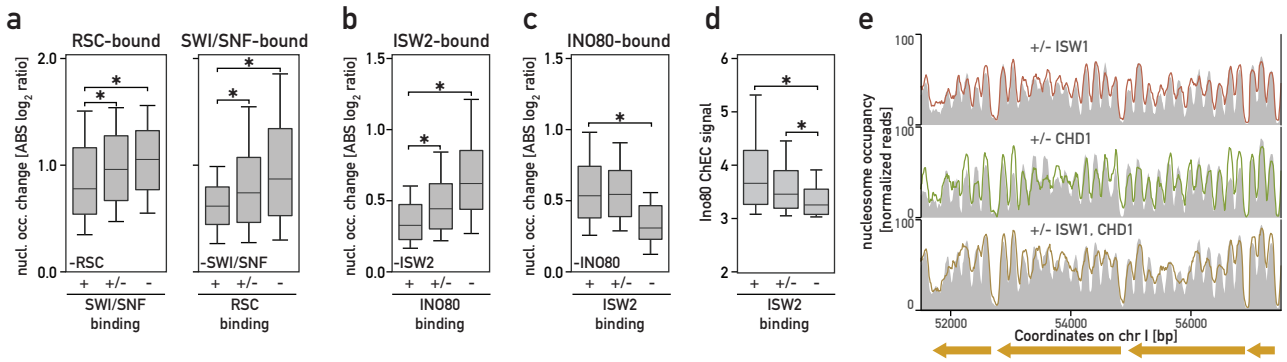






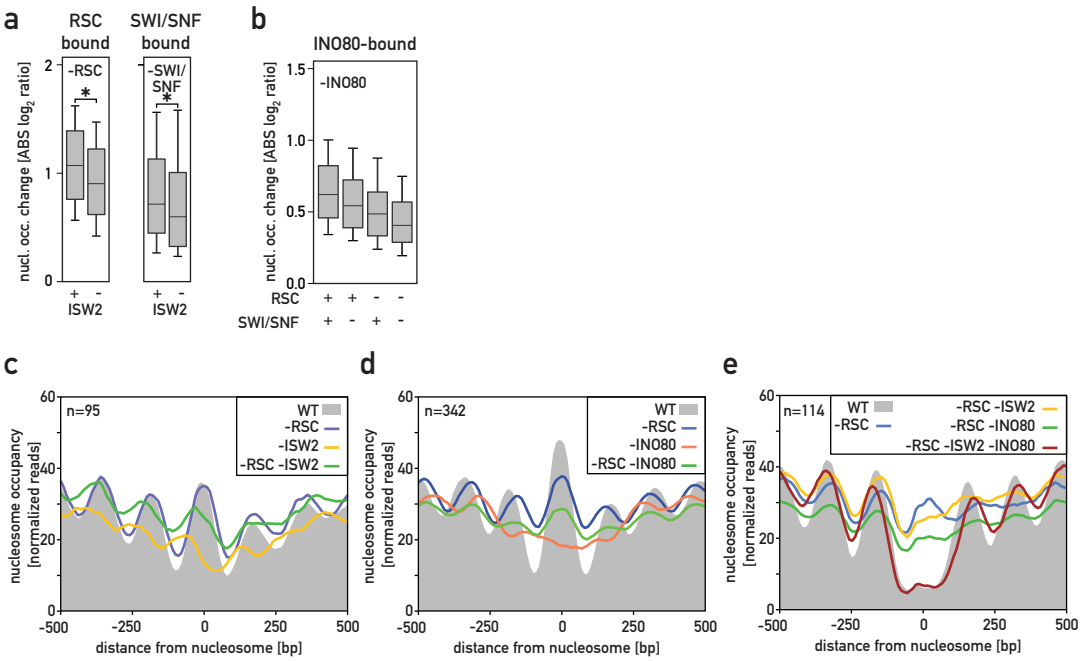
Supplementary Fig. 2 | Verification and characterization of remodeler depletion and effects on nucleosome occupancy and stability; related to Fig. 2. a, Fluorescence microscopy of cells bearing FRB-GFP fusions of Sth1, Snf2, Isw2 and Chd1; cells were treated with rapamycin for indicated times, fixed and stained with DAPI. **b**, Western blotting (anti-myc antibodies) of cell lysates from an untagged strain and strains bearing Ino80-AID*-myc and Isw1-AID*-myc fusions treated with auxin for indicated amount of time. **c**, Growth assays (serial dilution "spot assays", as in Supplementary Fig. 1a) of the indicated anchor-away or AID depletion strains on YPAD medium. "WT" indicates the parental anchor-away and/or AID strains background. **d**, Pearson correlations for all pairwise comparisons of genome-wide nucleosome occupancy change over 100 bp windows for the indicated CR depletion strains in the absence of depletion (mock-treated). **e**, Screenshots of sample regions in which nucleosomes become stabilized (marked with a red rectangle) upon depletion of RSC (top) or SWI/SNF (bottom). **f**, Average plots of nucleosome occupancy for all nucleosomes becoming stabilized upon depletion of RSC (top) or SWI/SNF (bottom). **g**, Screenshots of sample regions in which nucleosomes become destabilized (marked with a red rectangle) upon depletion of ISW2 (top) or INO80 (bottom). **h**, Average plots of nucleosome occupancy for all nucleosomes destabilized upon depletion of ISW2 (top) or INO80 (bottom). **i,j** Average plots of nucleosome occupancy with (red) or without (blue) depletion of ISW1 (i) or CHD1 (j), plotted separately for genes with the lowest (top) or highest (bottom) binding by the relevant CR (average binding profiles shown in green).

Supplementary Figure 3



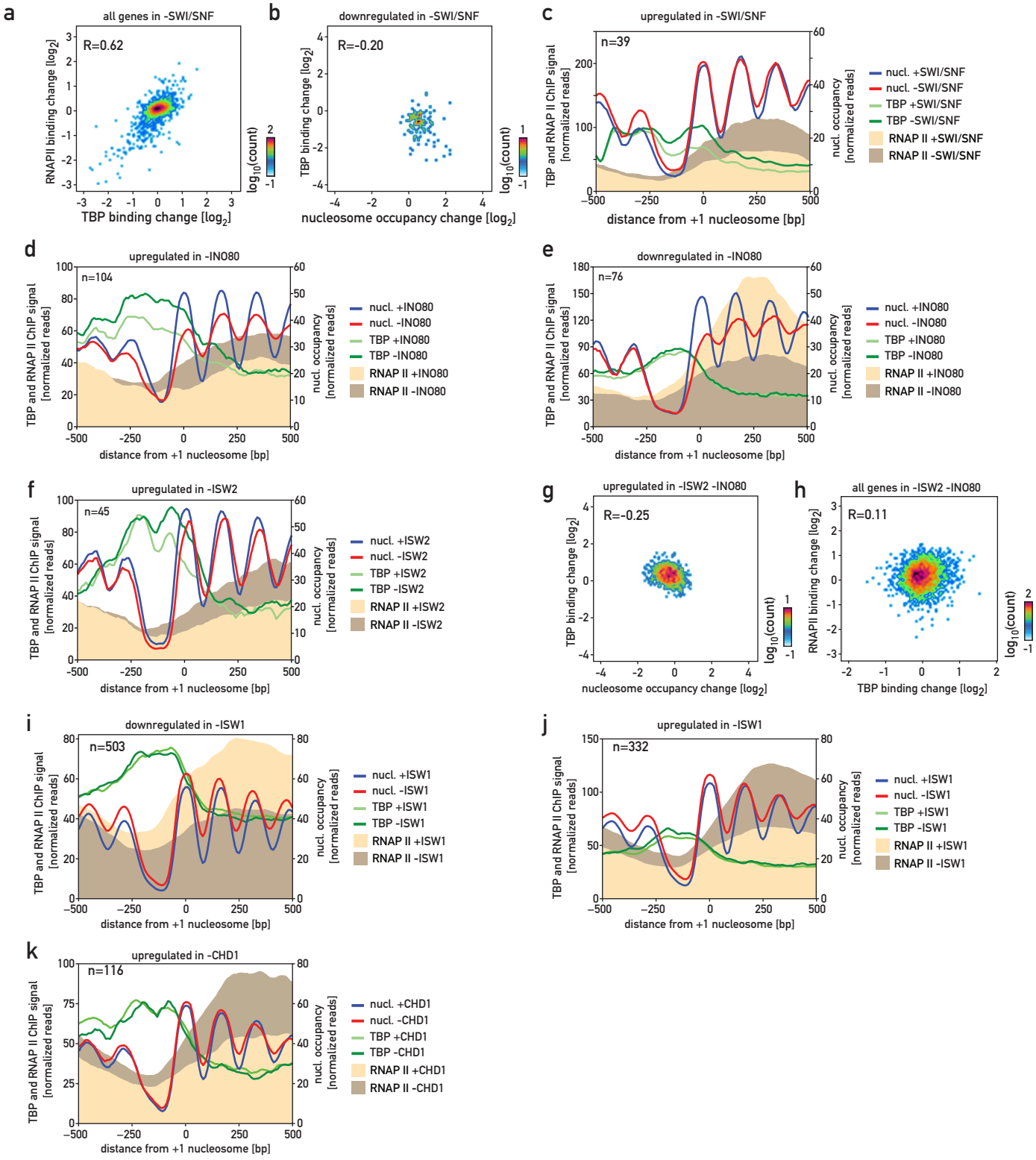
Supplementary Fig. 3 | **Remodeler redundancy in nucleosome positioning; related to Fig. 3. a,** Nucleosome occupancy change upon depletion of RSC (left) or SWI/SNF (right) at sites bound by each remodeler and displaying varying binding signal (+, +/-, -) of the other one. **b**, Nucleosome occupancy change upon depletion of ISW2 at sites bound by this remodeler and displaying varying binding signal of INO80 (as in (**a**)). **c**, Nucleosome occupancy change upon depletion of INO80 at sites bound by this remodeler and displaying varying binding signal of ISW2 (as in (a)). **d**, Boxplot of INO80 binding signal at INO80-bound sites displaying varying binding signal of ISW2. In (**a-d**), asterisks indicate significant differences ($p<0.05$, Mann-Whitney test). **e**, Snapshot of a sample genomic region displaying nucleosome occupancy change upon depletion of ISW1, CHD1 or both remodelers simultaneously.

Supplementary Figure 4



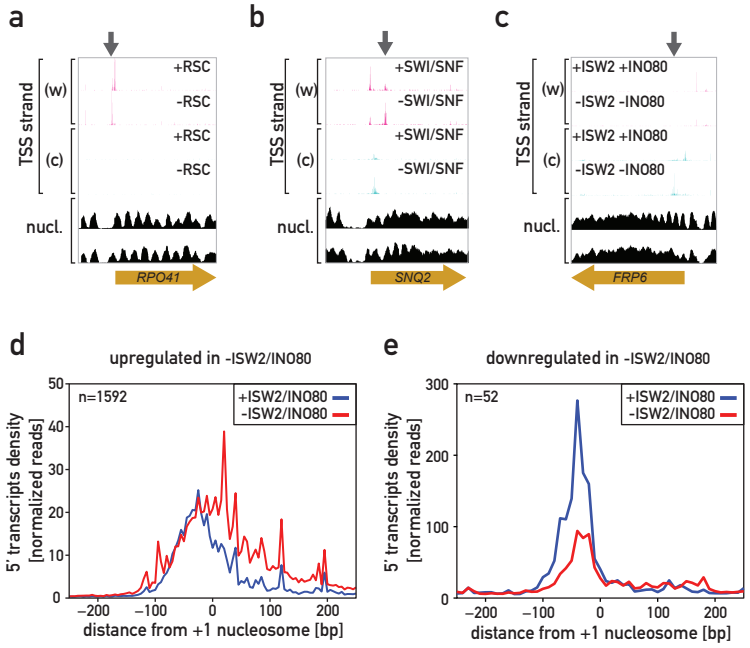
Supplementary Fig. 4 | Multiple concordant and opposing remodeler activities control promoter nucleosome occupancy; related to Fig. 4. a, Nucleosome occupancy change upon RSC depletion at sites bound by RSC (left) and change upon SWI/SNF depletion at sites bound by SWI/SNF (right). In both cases comparisons are made between sites co-bound by ISW2 (+) or not (-). Asterisks indicates significant difference ($p<0.05$, Mann-Whitney test). **b**, Nucleosome occupancy change upon INO80 depletion at sites bound by INO80 and co-bound by RSC and/or SWI/SNF or not, as indicated below. **c**, Average plots of nucleosome occupancy for all nucleosomes destabilized upon depletion of ISW2, comparing wild-type cells and cells depleted of RSC, ISW2, or both remodelers simultaneously. **d**, Average plots of nucleosome occupancy for all nucleosomes destabilized upon depletion of INO80, comparing for wild-type cells and cells depleted of RSC , INO80 , or both remodelers simultaneously. **e**, Average plots of nucleosome occupancy for all nucleosomes stabilized upon depletion of RSC, comparing wild-type cells and cells depleted of RSC , RSC and ISW2 , RSC and INO80, or all three remodelers simultaneously.

Supplementary Figure 5



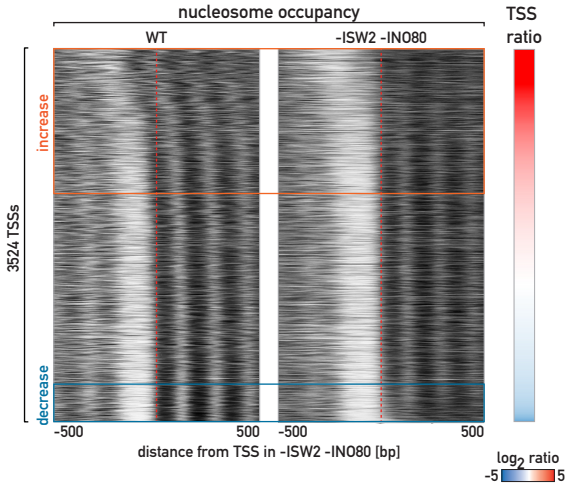
Supplementary Fig 5 | Links between nucleosome occupancy and transcriptional regulation; related to Fig. 5. a, Scatterplot showing relationship between TBP binding change at gene promoters and RNAPII binding change in corresponding gene bodies following SWI/SNF depletion, for all genes with a well-defined TSS; Pearson R value shown. **b**, Scatterplot showing relationship between TBP binding and nucleosome occupancy changes following SWI/SNF depletion at down-regulated genes. **c**, Average plots displaying nucleosome occupancy, RNAPII and TBP ChIP signals, in the presence and absence of SWI/SNF, for those genes up-regulated upon SWI/SNF depletion. **d-f**, Average plots displaying nucleosome occupancy together with RNAPII and TBP ChIP-seq signals for genes up-regulated by INO80 depletion (**d**), down-regulated by INO80 depletion (**e**), or up-regulated by ISW2 depletion (**f**), in each case with or without (+/-) the indicated CR. **g**, Scatter plot displaying the relationship between TBP binding and nucleosome occupancy changes following double "puller" depletion at genes where transcription was affected (see Fig. 5b,c). **h**, Scatterplot displaying the TBP - RNAPII ChIP-seq signal relationship at all genes following double "puller" depletion. **i-k**, Average plots displaying nucleosome occupancy together with RNAPII and TBP ChIP-seq signals for genes down-regulated upon ISW1 depletion (**i**), up-regulated upon ISW1 depletion (**j**), or up-regulated upon CHD1 depletion (**k**), in each case with or without (+/-) the indicated CR.

Supplementary Figure 6



Supplementary Fig. 6 | Effects of remodeler depletion on TSS selection; related to Fig. 6. a, Snapshot of genomic region showing 5'RACE signal for the Watson (w) and the Crick (c) strands as well as nucleosome occupancy in the presence and absence of RSC; upon RSC depletion the *RPO49* gene transcription initiates more upstream comparing to wild-type conditions. **b**, As in (**a**) but for SWI/SNF depletion; upon SWI/SNF depletion the *SNQ2* gene transcription initiates more frequently downstream comparing to wild-type conditions; additionally, there is more initiation events in the opposite strand just downstream from the upstream-most *SNQ2* TSS. **c**, As in (**a**) but for ISW2 and INO80 simultaneous depletion; upon "pullers" depletion the *FRP6* gene transcription initiates at a downstream position comparing to wild-type conditions. **d**, Average 5'RACE signal at genes upregulated upon simultaneous depletion of ISW2 and INO80. **e,** As in (**d**) but for downregulated genes.

Supplementary Figure 7



Supplementary Fig. 7 | Effect of "puller" remodelers on nucleosome occupancy; related to Fig. 7. Heatmaps showing nucleosome occupancy, centered at TSS positions determined in cells depleted of ISW2 and INO80, shown in wild type cells and cells depleted of both remodelers.

GENERAL DISCUSSION & PERSPECTIVES

Role of GRFs in Limiting Pervasive Transcription

GRFs have been extensively studied for their role in gene expression (activation and repression). My Ph.D. project mainly aimed at understanding the role of GRFs in limiting pervasive transcription in yeast. We provided clear evidence that the binding of GRFs is essential for transcription initiation fidelity by preventing the occurrence of spurious initiation events but also to restrain RNAPII elongation arising from upstream initiation sites. In this respect, my project unveiled completely new and unexpected functions for GRFs.

GRFs control pervasive transcription at the level of termination

In the course of a previous study, our lab demonstrated that Reb1, one of the most abundant GRFs in yeast (see II.1.1.3), is able to induce transcription termination by a roadblock mechanism (Colin et al., 2014). Subsequently we have shown that Rap1, another GRFs, also functions to limit RNAPII progression, most likely by a similar mechanism that involves the ubiquitin ligase Rsp5 (Candelli et al., 2018a). Importantly, the analyses of RNAPII distribution have revealed that Reb1 and Rap1-dependent termination events are widespread across the yeast genome.

During the course of my first project, we have shown that RNAPII often fails to efficiently terminate at canonical termination sites thus leading to readthrough transcription. Readthrough transcription also accounts for the production of a significant number of non-coding and unstable transcripts. In this context, we demonstrated that, along their progression across DNA, inefficiently terminated RNAPIIs encounter roadblock factors that trigger termination by acting as a failsafe mechanism. Importantly, because both the RNA and the polymerase that are terminated by this pathway are thought to be degraded (Candelli et al., 2018a; Colin et al., 2014; Roy et al., 2016), it is unlikely that roadblock serves as a productive pathway for upstream transcription events. Instead, we favour a model whereby roadblock would be important to protect downstream transcription by preventing invasion of promoters by RNAPII.

An important notion for GRF-mediated roadblock termination is that it does not depend on the regulatory domain of the factor. Instead, we have shown that the expression of Reb1-DBD (Colin et al., 2014) and Rap1-DBD (Candelli et al., 2018a) are, alone, able to ensure proper termination upstream of their respective binding site. Remarkably, we also demonstrated that the expression of the DBD can be sufficient to allow normal expression of a gene located downstream of a Rap1-dependent roadblock site (Candelli et al., 2018a; Colin et al., 2014).

The DBD construct is missing domains described as necessary for transcription activation (Azad and Tomar, 2016; Hartley and Madhani, 2009; Ozonov and van Nimwegen, 2013; Tomar et al., 2008), which suggested that Rap1-DBD restores normal gene expression by other means than transcription activation. We originally interpreted these results as supporting the notion that the DBD could function by preventing upstream polymerases to invade the downstream promoter and cause transcriptional interference. This interpretation should, however, be revisited in the light of our latest study demonstrating that the Rap1-DBD can also support gene expression by correctly positioning proximal nucleosomes (Challal et al., 2018). Although this does not exclude an important role for roadblock termination in the maintenance of robust gene expression, the two roles of Rap1 DNA-binding (in termination and nucleosome positioning) have to be distinguished when assessing the impact on gene expression.

Roadblock termination events occur upstream of many additional DNA-binding factors, including other GRFs such as Abf1, but also near the RNAPIII transcription machinery or around centromeres (Candelli et al., 2018a; Roy et al., 2016). More recently, the laboratory also provided strong evidence for roadblock termination events around Autonomously Replicating Sequences (ARSs), notably upstream of the ORC-binding factor (Candelli et al., 2018b). Because most roadblock factors are associated with important DNA associated events, it is fair to assume that this alternative termination pathway plays a global role in preventing RNAPII to invade crucial regions, thus insuring the integrity of these processes. This might be more relevant in species with compact genome such as *S. cerevisiae.* Finally, since the mechanism and proteins involved in roadblock are generally conserved, we anticipate that this termination pathway is also present in more complex eukaryotes.

GRFs control pervasive transcription at the level of initiation

Many studies have shown that the intrinsic bidirectionality of promoters constitutes an important source of pervasive transcription (Churchman and Weissman, 2011; Jin et al., 2017; Marquardt et al., 2014; Neil et al., 2009; Rhee and Pugh, 2012; Xu et al., 2009). It has also been recently proposed that species-specific elements (e.g. *cis*-elements and/or *trans*-acting factors) favour transcription towards the functional direction across evolution (Jin et al., 2017). Consistently with this notion, we proved that depletion of Rap1, and GRFs in general, can alter the bidirectional balance at many promoter regions (Challal et al., 2018). Notably, we found that the absence of Rap1 leads either to the appearance of transcription at silent regions, or to an increase of RNAPII signal at pre-existing non-coding transcription units. In some cases, this can also result in increased TSS usage, suggesting the presence of more efficient initiation sites close to canonical TSSs. Importantly, while chromatin remodelers have already been

shown to impact promoter bidirectionality (Whitehouse et al., 2007; Xue et al., 2017; Yadon et al., 2010), our study is the first to unveil a similar function of GRFs. Moreover, our data are consistent with a parallel work from the van Werven lab showing that Rap1 supresses divergent non-coding transcription (Wu et al., 2018). Because pervasive transcription can impact the expression of neighbouring genes or other DNA-associated phenomenon, the comprehension of mechanisms that limit the later is crucial.

GRFs and Transcription Fidelity: Impacts and Implications

As previously mentioned, transcription initiates about 15 bp downstream of the 5' edge of the +1 nucleosome in *S. cerevisiae* (see I.1.3). Upon Rap1 depletion, we have shown that the upstream shift of the +1 is not only associated with a decrease transcription efficiency from the canonical TSS, but also with the appearance of eTSSs located in a similar position relative to the newly positioned +1 nucleosome. For protein-coding genes, this is particularly important since aberrant initiation can lead to the production of mRNAs with premature stop codon and upstream ORFs, leading to the degradation of the later by NMD in the cytoplasm (Malabat et al., 2015). Supporting this notion, we have shown that most ectopic transcripts arising within promoter regions upon Rap1 depletion are sensitive to Upf1 (NMD pathway). The sensitivity to Upf1 constitutes an indirect, but yet strong, proof that these RNA molecules are exported in the cytoplasm where they are translated. Thus, at least two different mechanisms co-exist in the cell to restrict the production of 5' extended mRNAs: one occurring at the level of initiation by limiting their production and one in the cytoplasm that prevents the synthesis of aberrant peptides.

GRFs, and most particularly Rap1, are able to both activate and repress gene expression. Heatmap analyses confirm the increase in the RNAPII signal at coding regions of some genes upon Rap1 depletion, suggesting upregulation of these genes in the absence of Rap1 (Challal et al., 2018). However, careful analysis revealed that in most (if not all) cases, transcription initiation is in reality arising from upstream TSSs that are more efficiently used than the canonical site (see for instance *RXT3*, *YOR292C* or *NAT3* Figure 2F, 3C or 4F). This is also sometimes reflected at the RNA levels, notably when NMD does not degrade very efficiently the aberrant transcripts. Because upregulation is linked with the usage of a different TSS, this questions the real role of Rap1 in negatively regulating gene transcription. This notion is particularly relavant when considering large scale analysis and highlights the importance of considering not only the absolute signal within coding regions, but also the exact TSS from which the signal arises.

The notion that DNA-binding factors are important to control transcription fidelity is emerging in yeast but also in other organisms. In a recent study from the Brar lab, the authors reported that meiosis is characterized by an increase of mRNA production and a decrease of the total protein level. They demonstrated that this opposite effect is due to the production of 5' extended RNAs containing upstream ORFs and a poor translation efficiency. Also, they proposed that the modification of the TSS might result from a modification or switch in the binding of specific transcription factors (Cheng et al., 2018). The Reb1 GRF has also been shown in a previous study to be required for proper TSS selection although this work was performed at a single locus (Wang and Donze, 2016). In mammals, the DNA-binding factor NF-Y is also involved in transcription fidelity and TSS selection. NF-Y shares important similarity with Rap1. Notably, this protein is also a ubiquitously expressed transcription factor. Akin to Rap1, NF-Y has the ability to displace nucleosomes by binding upstream TSS regions. In their study (bioRxiv), Oldfield and colleagues reported that depletion of NF-Y leads to an upstream shift in TSS selection leading to the production of 5' extended RNA molecules. The correlation with nucleosome organisation has also been demonstrated, thus suggesting a similar model as the one proposed for Rap1. Finally, the same mechanism has been described in *D. melanogaster* where it involves the NLS-binding factor (Lam et al., 2019). Collectively, these studies all point out the crucial and evolutionary conserved role of promoter-binding factors in preventing ectopic initiation and in the maintenance of transcription initiation fidelity.

The effect of Rap1 on gene expression was expected since it has been demonstrated in many instances to bind promoters of hundreds of genes and control their expression (see II.1.1.1). The main unexpected observation is that transcription can still occur upon its depletion, albeit from a different initiation site. This is due to the fact that NDRs formation is not completely abolished even in the absence of GRFs although the reason why is still unclear. The persistence of short NDRs could result from at least three distinct but not mutually exclusive phenomena: (i) the presence of other DNA-binding proteins that either remain or associate with DNA upon Rap1 depletion (ii) the constant action of chromatin remodelers even in the absence of Rap1 (iii) the DNA sequence itself that has been shown to be involved in nucleosome exclusion, especially at AT-rich sequences present upstream of TSSs (Iyer and Struhl, 1995; Yuan, 2005). This latter hypothesis is favoured by the fact that NDR-like regions are observed along purified yeast genomic DNA even in the absence of any GRFs and chromatin remodelers, suggesting an intrinsic tendency for nucleosomes to be excluded from promoter regions (Kaplan et al., 2009; Krietenstein et al., 2016).

Mechanism of Gene Regulation by GRFs

Current models of gene regulation and NDRs formation by GRFs favour a mechanism whereby DNA-binding factors recruit chromatin remodelers in order to establish the correct position of the +1 and -1 nucleosomes thus promoting the assembly of the PIC. This model is supported, for instance, by pull-down assays showing the co-purification of remodelers with Rap1 (or truncated forms of Rap1) (Tomar et al., 2008). In addition, Reb1 and RSC have been shown to both affect similar NDRs *in vivo* (Hartley and Madhani, 2009). Finally, the fact that chromatin remodelers, but not GRFs, harbour a catalytic activity has reinforced the idea that they might be the main direct actors of NDRs formation.

Comparison of Rap1 depletion with RSC, SWI/SNF, INO80 and ISW2 depletion however reveals a distinct profile and effect on both TSS and nucleosome positioning (Challal et al., 2018). The clear difference between these factors strongly suggests that they function independently from one another. This is consistent with a recent publication from the Shore lab showing that the co-depletion of GRFs (Abf1 or Reb1) and RSC has a stronger effect on NDRs suppression as compared with either one of the two factors independently (Kubik et al., 2018). The absence of epistatic effects indicates that, although GRFs and chromatin remodelers act concomitantly at similar targets, they act independently to promote NDR formation.

The fact that GRFs and remodelers are in reality independent is an important result that raises the question of how GRFs can, on their own, provoke nucleosome exclusion at promoters. The observation that the DBD alone is sufficient to support normal chromatin structure at promoters, especially close to the Rap1-binding motif, at least partially answers this question (Challal et al., 2018). Indeed, we speculate that part of the nucleosome-displacing activity of GRFs is due to the establishment of a steric constraint at promoters, which prevents the invasion of nucleosomes and restores, in many instances, normal or quasi-normal gene expression. Our results are supported by one earlier study showing that Rap1-DBD is sufficient to promote activation of the *HIS4* gene (Yu et al., 2001) but are not in line, in this respect, with the parallel work published by Wu and colleagues (Wu et al., 2018) which might challenge our model. In the latter study, the authors reported that truncated forms of Rap1 containing the DBD are not sufficient to supress the occurrence of non-coding transcripts at the *MLP1* and *IME1* loci (note that this analysis was not performed genome-wide). As already mentioned in the published manuscript (Challal et al., 2018), we do not reproduce these results in our system and clearly observe suppression at both the *MLP1* and *IME1* loci by expressing Rap1-DBD.

Also, by using the system and strains construct from the van Werven lab, we could successfully reproduce our results at a few diagnostic natural cases (data not shown). We suspect that the constructs used in that study do not bind DNA as efficiently as ours, possibly because they contain extra amino acids (tags and nuclear localization signals). Because the predictive value of a positive result observed genome-wide (suppression by Rap1-DBD) is higher than a negative result that is limited to only a couple of examples, we are confident in the validity of our model.

What makes GRFs different from canonical transcription factors, notably in their ability to displace nucleosomes? In a recent study, Yan and colleagues developed an assay to identify proteins with nucleosome-displacing activity by inserting DNA-binding sites for various factors within a nucleosome-containing region (Yan et al., 2018). In this context, the authors verify the ability of hundreds of proteins to induce NDRs formation and could classify transcription factors into three main groups: strong, weak or no nucleosome-displacing activity. Unsurprisingly, Abf1, Reb1 and Rap1 as well as Cbf1, Orc1 and Mcm1 all belong to the first category of factors (i.e. highly capable of provoking NDRs formation). The authors could link this ability of the first category to their high abundance and binding affinity, thus defining two main features of nucleosome-displacing factors. Supporting the importance of the notion of “abundance”, factors from the second category (i.e. weak ability) behave similarly to the first category upon overexpression (Yan et al., 2018). Finally, the DBD of Ume6 (second category) is also sufficient to evict nucleosomes away from its binding site strengthening our model of steric occlusion.

When considering the organisation of nucleosomes at a more distal position from Rap1 sites, we found that the C-terminal and/or N-terminal domains are required for proper localisation. This is particularly true for clusters 1 and 2 (Challal et al., 2018) characterized by large NDRs. We speculate that this might result from the loss of Rap1-associated factors at promoters such as Hmo1 and the FIS complex (see II.1.1.1 and Challal et al., 2018 Figure S5 and discussion section). This idea is supported by the fact that, in a WT context, these factors are located between Rap1 and the eccentric +1 nucleosome and that depletion of Hmo1 has also been associated with upstream shift of both RNAPII and nucleosomes (Kasahara et al., 2011; Reja et al., 2015). This hypothesis could be confirmed by verifying the presence of Rap1-associated factors upon expression of the DBD alone, or by expressing a truncated version of Rap1 containing the interaction domains with the FIS complex.

Chromatin Remodelers Dictate TSS Decisions

Among the different chromatin remodelers, four complexes are particularly important to influence the position of the +1 nucleosome: RSC, SWI/SNF, INO80 and ISW2. By mapping the position of TSSs upon their depletion, we could confirm that accurate position of the +1 is intimately linked with TSS selection and gene expression (Kubik et al., 2019). Upon RSC depletion, the shift of the +1 nucleosome towards the NDRs is usually associated with a repressive effect on TSS intensity, suggesting that RSC is important to correctly expose the initiation site to the transcription machinery. In some instances, the upstream re-positioning of the +1 is coupled to a change in TSS selection. These data are in agreement with a recent study that also reported these two possible phenomena (i.e. complete repression or change of TSS usage) upon RSC depletion (Klein-Brill et al., 2019). Importantly however, the extent of the shift is very different from what is observed upon GRFs depletion.

More interestingly, a downstream shift of the +1 nucleosome observed upon INO80 and ISW2 double depletion is also linked with the selection of downstream TSSs, even though the PIC most likely assembles in the same position. This is particularly intriguing because in this configuration, the canonical TSS is fully accessible and could therefore be used by the polymerase scanning from the PIC. The more plausible hypothesis to explain the concomitant shift of the +1 and TSSs (also discussed in Challal et al., 2018) is that the +1 nucleosome is required for initiation *in vivo* and directly influences the position of transcription initiation. Although the correlation between the position of the +1 nucleosome and the TSS is well-established, whether it is the TSS that positions the +1 nucleosome or the reverse remains subject of debate. If the latter hypothesis is correct, we could speculate that the +1 nucleosome has specific features that favour initiation, or, more simply, that the first nucleosome encountered during scanning triggers initiation. However, it has to be noted that, despite the similarity of the transcription machinery, the distance between the +1 nucleosome and the TSS is not a conserved feature in eukaryotes (see I.1.3), suggesting that the role of the +1 nucleosome on initiation might not be conserved. The comprehension of the mechanisms that link the +1 nucleosome to TSS selection is a particularly challenging and interesting topic that remains to be elucidated.

The Ground State of Transcription Initiation among Eukaryotic Genomes

As previously mentioned, an important discovery of our study is that transcription initiation can occur from alternative and spurious sites upon Rap1 depletion. Although ectopic initiation is mainly arising within promoter regions upon Rap1 depletion, we also found cases of internal and more distal initiation events (Challal et al., 2018 see for instance Figure S3H). Similarly, spurious initiation has also been observed in various yeast mutants including *SPT6*, *SET2*, *INO80*, *CHD1* or *ISW1* and has been described in more complex eukaryotes (detailed in III.2.1). In most circumstances, ectopic initiation arises from promoter-like regions containing all the required elements for efficient initiation (Doris et al., 2018; Kaplan et al., 2003). The production of RNAs from alternative sites represents a potential danger for genome integrity as it can interfere with other DNA associated events and/or provoke synthesis of aberrant peptides. Yet, across evolution, the *cis*-elements leading to spurious initiation sites have not been counter selected. Instead, cells have developed mechanisms to restrict the usage of many possible alternative transcription start sites.

The crowded genome: a barrier against spurious initiation

Nucleosomes and DNA-binding proteins are often considered as “obstacles” against the progression of DNA and RNA polymerases that need to be removed in order to ensure the smooth progression of the different machineries. Yet, as indicated by various studies, aberrant displacement of nucleosomes can be associated with ectopic initiation, suggesting an important role for the later in ensuring genome integrity by occluding spurious transcription sites. At NDRs, we demonstrated that the binding of GRFs also plays a role in the control of TSS positioning, suggesting a similar function of GRFs and nucleosomes at least in their ability to limit alternative initiation. I envision that the presence of a well-positioned +1, together with the binding of regulatory factor, restrict the number of possible “spots” where PICs could otherwise assemble and therefore promote the binding at the most favourable sites. This appears to be essential considering the fact that the sequence required for transcription initiation is not based on a strong consensus, and is thereby probably found in many instances in the eukaryotic genome. Besides, in this context, it would be interesting to know how often an ectopically formed NDR could favour the firing of transcription in at least one direction.

Spurious initiation: a source of new genes?

One possible line of evolution could have been to increase the complexity of promoter regions, thus limiting the number of possible ectopic PICs and alternative initiation events. So why are loose sequences used as docking sites for transcription initiation? A plausible explanation is that it may represent a stock of new promoters and initiation sites buried under nucleosomes that could emerge with evolution to promote the formation of new genes. In this case, the appearance or modification of a binding site for a regulatory factor could potentially be sufficient to form a new and stable NDR and create a promoter region. In this model, two different strategies could be envisioned: On one hand, the modification of a pre-existing NDR could favour the production of 5'-extended RNAs (as is it the case upon Rap1 depletion) containing additional regulatory regions within 5' UTRs or giving rise to the production of longer proteins. On the other hand, the formation of an ectopic NDR could promote the synthesis of an RNA molecule containing a new ORF and therefore give rise to the emergence of a new poly-peptide. Despite the possible role of spurious promoters in generating new genes, it is however important to mention that the increased number of genes does not represent a major source of evolution of eukaryotic species, as even highly divergent organisms such as yeast and human contain a relatively similar number of coding units (~6000 and ~25000 respectively). Instead, other mechanisms have been shown to contribute more significantly to the acquisition of organism's complexity, including for instance alternative-splicing (Barbosa-Morais et al., 2012).

BIBLIOGRAPHY

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572–576.

Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., et al. (2013). The human cap-binding complex is functionally connected to the nuclear RNA exosome. Nature Structural & Molecular Biology *20*, 1367–1376.

Angus-Hill, M.L., Schlichter, A., Roberts, D., Erdjument-Bromage, H., Tempst, P., and Cairns, B.R. (2001). A Rsc3/Rsc30 Zinc Cluster Dimer Reveals Novel Roles for the Chromatin Remodeler RSC in Gene Expression and Cell Cycle Control. Molecular Cell *7*, 741–751.

Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006a). Termination of Cryptic Unstable Transcripts Is Directed by Yeast RNA-Binding Proteins Nrd1 and Nab3. Molecular Cell *23*, 841–851.

Arigo, J.T., Carroll, K.L., Ames, J.M., and Corden, J.L. (2006b). Regulation of Yeast NRD1 Expression by Premature Transcription Termination. Molecular Cell *21*, 641–651.

Arimbasseri, A.G., and Maraia, R.J. (2015). Mechanism of Transcription Termination by RNA Polymerase III Utilizes a Non-template Strand Sequence-Specific Signal Element. Molecular Cell *58*, 1124–1132.

Azad, G.K., and Tomar, R.G. (2016). The multifunctional transcription factor Rap1 a regulator of yeast physiology. Frontiers in Bioscience *21*, 918–930.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., et al. (2008). A Library of Yeast Transcription Factor Motifs Reveals a Widespread Function for Rsc3 in Targeting Nucleosome Exclusion at Promoters. Molecular Cell *32*, 878–887.

Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S., et al. (2017). Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. Molecular Cell *66*, 38-49.e6.

Baptista, T., Grünberg, S., Minoungou, N., Koster, M.J.E., Timmers, H.T.M., Hahn, S., Devys, D., and Tora, L. (2017). SAGA Is a General Cofactor for RNA Polymerase II Transcription. Molecular Cell *68*, 130-143.e5.

Barabino, S.M.L., Ohnacker, M., and Keller, W. (2000). Distinct roles of two Yth1p domains in 3′-end cleavage and polyadenylation of yeast pre-mRNAs. The EMBO Journal *19*, 3778–3787.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science *338*, 1587–1593.

Barilla, D., Lee, B.A., and Proudfoot, N.J. (2001). Cleavage–polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. 6.

Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. Cell *116*, 699–709.

Bataille, A.R., Jeronimo, C., Jacques, P.-É., Laramée, L., Fortin, M.-È., Forest, A., Bergeron, M., Hanes, S.D., and Robert, F. (2012). A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. Molecular Cell *45*, 158–170.

Beaudenon, S.L., Huacani, M.R., Wang, G., McDonnell, D.P., and Huibregtse, J.M. (1999). Rsp5 Ubiquitin-Protein Ligase Mediates DNA Damage-Induced Degradation of the Large Subunit of RNA Polymerase II in *Saccharomyces cerevisiae*. Molecular and Cellular Biology *19*, 6972–6979.

Beinoravičiūtė-Kellner, R., Lipps, G., and Krauss, G. (2005). In vitro selection of DNA binding sites for ABF1 protein from Saccharomyces cerevisiae. FEBS Letters *579*, 4535–4540.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science *301*, 1090–1093.

Bendjennat, M., and Weil, P.A. (2008). The transcriptional repressor activator protein Rap1p is a direct regulator of TATA-binding protein. J. Biol. Chem. *283*, 8699–8710.

Berroteran, R.W., Ware, D.E., and Hampsey, M. (1994). The sua8 suppressors of Saccharomyces cerevisiae encode replacements of conserved residues within the largest subunit of RNA polymerase II and affect transcription start site selection similarly to sua7 (TFIIB) mutations. Molecular and Cellular Biology *14*, 226–237.

Biggins, S. (2013). The Composition, Functions, and Regulation of the Budding Yeast Kinetochore. Genetics *194*, 817–846.

Bilaud, T., Koering, C.E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S.M., and Gilson, E. (1996). The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human. Nucleic Acids Res. *24*, 1294–1303.

Bortvin, A., and Winston, F. (1996). Evidence That Spt6p Controls Chromatin Structure by a Direct Interaction with Histones. Science *272*, 1473–1476.

Bosio, M.C., Fermi, B., and Dieci, G. (2017a). Transcriptional control of yeast ribosome biogenesis: A multifaceted role for general regulatory factors. Transcription *8*, 254–260.

Bosio, M.C., Fermi, B., Spagnoli, G., Levati, E., Rubbi, L., Ferrari, R., Pellegrini, M., and Dieci, G. (2017b). Abf1 and other general regulatory factors control ribosome biogenesis gene expression in budding yeast. Nucleic Acids Res. *45*, 4493–4506.

Braglia, P., Percudani, R., and Dieci, G. (2005). Sequence Context Effects on Oligo(dT) Termination Signal Recognition by *Saccharomyces cerevisiae* RNA Polymerase III. Journal of Biological Chemistry *280*, 19551–19562.

Brahma, S., and Henikoff, S. (2019). RSC-Associated Subnucleosomes Define MNase-Sensitive Promoters in Yeast. Molecular Cell *73*, 238-249.e3.

Brigati, C., Kurtz, S., Balderes, D., Vidali, G., and Shore, D. (1993). An essential yeast gene encoding a TTAGGG repeat-binding protein. Molecular and Cellular Biology *13*, 1306–1314.

Briggs, S.D. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes & Development *15*, 3286–3295.

Briggs, S.D., Xiao, T., Sun, Z.-W., Caldwell, J.A., Shabanowitz, J., Hunt, D.F., Allis, C.D., and Strahl, B.D. (2002). Trans-histone regulatory pathway in chromatin. Nature *418*, 498–498.

Brindle, P.K., Holland, J.P., Willett, C.E., Innis, M.A., and Holland, M.J. (1990). Multiple factors bind the upstream activation sites of the yeast enolase genes ENO1 and ENO2: ABFI protein, like repressor activator protein RAP1, binds cis-acting sequences which modulate repression or activation of transcription. Molecular and Cellular Biology *10*, 4872–4885.

Buck, M.J., and Lieb, J.D. (2006). A chromatin-mediated mechanism for specification of conditional transcription factor targets. Nat. Genet. *38*, 1446–1451.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549–561.

Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. Science *303*, 983–988.

Čabart, P., Újvári, A., Pal, M., and Luse, D.S. (2011). Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. PNAS *108*, 15786–15791.

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007). Antisense RNA Stabilization Induces Transcriptional Gene Silencing via Histone Deacetylation in S. cerevisiae. Cell *131*, 706–717.

Camblong, J., Beyrouthy, N., Guffanti, E., Schlaepfer, G., Steinmetz, L.M., and Stutz, F. (2009). Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. Genes Dev. *23*, 1534–1545.

Candelli, T., Challal, D., Briand, J., Boulay, J., Porrua, O., Colin, J., and Libri, D. (2018a). High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast. The EMBO Journal *37*, e97490.

Candelli, T., Gros, J., and Libri, D. (2018b). Pervasive transcription fine-tunes replication origin activity. ELife *7*.

Cardenas, M.E., Cutler, N.S., Lorenz, M.C., Di Como, C.J., and Heitman, J. (1999). The TOR signaling cascade regulates gene expression in response to nutrients. Genes Dev. *13*, 3271–3279.

Carey, M., Li, B., and Workman, J.L. (2006). RSC Exploits Histone Acetylation to Abrogate the Nucleosomal Block to RNA Polymerase II Elongation. Molecular Cell *24*, 481–487.

Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D., and Corden, J.L. (2004). Identification of cis Elements Directing Termination of Yeast Nonpolyadenylated snoRNA Transcripts. Molecular and Cellular Biology *24*, 6241–6252.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M.P., et al. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581–592.

Carvunis, A.-R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charloteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B., et al. (2012). Proto-genes and de novo gene birth. Nature *487*, 370–374.

Castelnuovo, 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. Nature Structural & Molecular Biology *20*, 851–858.

Challal, D., Barucco, M., Kubik, S., Feuerbach, F., Candelli, T., Geoffroy, H., Benaksas, C., Shore, D., and Libri, D. (2018). General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation. Molecular Cell *72*, 955-969.e7.

Chanfreau, G., Buckle, M., and Jacquier, A. (2000). Recognition of a conserved class of RNA tetraloops by Saccharomyces cerevisiae RNase III. 6.

Chasman, D.I., Lue, N.F., Buchman, A.R., LaPointe, J.W., Lorch, Y., and Kornberg, R.D. (1990). A yeast protein that influences the chromatin structure of UASG and functions as a powerful auxiliary gene activator. Genes Dev. *4*, 503–514.

Chen, W., and Struhl, K. (1985). Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. EMBO J *4*, 3273–3280.

Chen, H.-T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nature Structural & Molecular Biology *14*, 696–703.

Cheng, Z., Otto, G.M., Powers, E.N., Keskin, A., Mertins, P., Carr, S.A., Jovanovic, M., and Brar, G.A. (2018). Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis. Cell *172*, 910-923.e16.

Chereji, R.V., Ocampo, J., and Clark, D.J. (2017). MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers. Molecular Cell *65*, 565-577.e3.

Cheung, A.C.M., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249–253.

Cheung, A.C.M., and Cramer, P. (2012). A Movie of RNA Polymerase II Transcription. Cell *149*, 1431–1437.

Cheung, V., Chua, G., Batada, N.N., Landry, C.R., Michnick, S.W., Hughes, T.R., and Winston, F. (2008). Chromatin- and Transcription-Related Factors Repress Transcription from within Coding Regions throughout the Saccharomyces cerevisiae Genome. PLOS Biology *6*, e277.

Chinchilla, K., Rodriguez-Molina, J.B., Ursic, D., Finkel, J.S., Ansari, A.Z., and Culbertson, M.R. (2012). Interactions of Sen1, Nrd1, and Nab3 with Multiple Phosphorylated Forms of the Rpb1 C-Terminal Domain in Saccharomyces cerevisiae. Eukaryotic Cell *11*, 417–429.

Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: The exosome. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1829*, 552–560.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nature Reviews Molecular Cell Biology *18*, 407–422.

Close, P., Hawkes, N., Cornez, I., Creppe, C., Lambert, C.A., Rogister, B., Siebenlist, U., Merville, M.-P., Slaugenhaupt, S.A., Bours, V., et al. (2006). Transcription Impairment and Cell Migration Defects in Elongator-Depleted Cells: Implication for Familial Dysautonomia. Molecular Cell *22*, 521–531.

Colin, J., Candelli, T., Porrua, O., Boulay, J., Zhu, C., Lacroute, F., Steinmetz, L.M., and Libri, D. (2014). Roadblock Termination by Reb1p Restricts Cryptic and Readthrough Transcription. Molecular Cell *56*, 667–680.

Collin, P., Jeronimo, C., Poitras, C., and Robert, F. (2019). RNA Polymerase II CTD Tyrosine 1 Is Required for Efficient Termination by the Nrd1-Nab3-Sen1 Pathway. Molecular Cell *73*, 655-669.e7.

Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., and Corden, J.L. (2000). A Yeast Heterogeneous Nuclear Ribonucleoprotein Complex Associated With RNA Polymerase II. 15.

Davis, C.A., and Ares, M. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomycescerevisiae. PNAS *103*, 3262–3267.

Dengl, S., and Cramer, P. (2009). *Torpedo* Nuclease Rat1 Is Insufficient to Terminate RNA Polymerase II *in Vitro*. Journal of Biological Chemistry *284*, 21270–21279.

Dheur, S., Vo, L.T.A., Voisinet-Hakil, F., Minet, M., Schmitter, J.-M., Lacroute, F., Wyers, F., and Minvielle-Sebastia, L. (2003). Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation. EMBO J. *22*, 2831–2840.

Dichtl, B., and Keller, W. (2001). Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. The EMBO Journal *20*, 3197–3209.

Dichtl, B., Blank, D., Sadowski, M., Hubner, W., Weiser, S., and Keller, W. (2002). Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. The EMBO Journal *21*, 4125–4135.

van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114–117.

Dikstein, R. (2011). The unexpected traits associated with core promoter elements. Transcription *2*, 201–206.

Doris, S.M., Chuang, J., Viktorovskaya, O., Murawska, M., Spatt, D., Churchman, L.S., and Winston, F. (2018). Spt6 Is Required for the Fidelity of Promoter Selection. Molecular Cell *72*, 687-699.e6.

Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. (2002). Methylation of Histone H3 by COMPASS Requires Ubiquitination of Histone H2B by Rad6. Journal of Biological Chemistry *277*, 28368–28371.

Dvir, A., Conaway, R.C., and Conaway, J.W. (1997). A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. Proceedings of the National Academy of Sciences *94*, 9006–9010.

Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes & Development *32*, 127–139.

Egecioglu, D.E., Henras, A.K., and Chanfreau, G.F. (2006). Contributions of Trf4p- and Trf5p-dependent polyadenylation to the processing and degradative functions of the yeast nuclear exosome. RNA *12*, 26–32.

Egloff, S., Dienstbier, M., and Murphy, S. (2012a). Updating the RNA polymerase CTD code: adding gene-specific layers. Trends in Genetics *28*, 333–341.

Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., and Murphy, S. (2012b). Ser7 Phosphorylation of the CTD Recruits the RPAP2 Ser5 Phosphatase to snRNA Genes. Molecular Cell *45*, 111–122.

El Hage, A., Koper, M., Kufel, J., and Tollervey, D. (2008). Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. Genes & Development *22*, 1069–1081.

Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.-K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., et al. (2004). Human Spt6 Stimulates Transcription Elongation by RNA Polymerase II In Vitro. Molecular and Cellular Biology *24*, 3324–3336.

Ertel, F., Dirac-Svejstrup, A.B., Hertel, C.B., Blaschke, D., Svejstrup, J.Q., and Korber, P. (2010). In Vitro Reconstitution of PHO5 Promoter Chromatin Remodeling Points to a Role for Activator-Nucleosome Competition In Vivo. Molecular and Cellular Biology *30*, 4060–4076.

Fasken, M.B., Laribee, R.N., and Corbett, A.H. (2015). Nab3 Facilitates the Function of the TRAMP Complex in RNA Processing via Recruitment of Rrp6 Independent of Nrd1. PLOS Genetics *11*, e1005044.

Feng, J., Gan, H., Eaton, M.L., Zhou, H., Li, S., Belsky, J.A., MacAlpine, D.M., Zhang, Z., and Li, Q. (2016). Noncoding Transcription Is a Driving Force for Nucleosome Instability in spt16 Mutant Cells. Molecular and Cellular Biology *36*, 12.

Fermi, B., Bosio, M.C., and Dieci, G. (2016). Promoter architecture and transcriptional regulation of Abf1-dependent ribosomal protein genes in Saccharomyces cerevisiae. Nucleic Acids Res *44*, 6113–6126.

Fermi, B., Bosio, M.C., and Dieci, G. (2017). Multiple roles of the general regulatory factor Abf1 in yeast ribosome biogenesis. Curr. Genet. *63*, 65–68.

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I.K., Sharon, E., Lubling, Y., Widom, J., and Segal, E. (2008). Distinct Modes of Regulation by Chromatin Encoded through Nucleosome Positioning Signals. PLoS Computational Biology *4*, e1000216.

Fishburn, J., Tomko, E., Galburt, E., and Hahn, S. (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. PNAS *112*, 3961–3966.

Fishburn, J., Galburt, E., and Hahn, S. (2016). Transcription Start Site Scanning and the Requirement for ATP during Transcription Initiation by RNA Polymerase II. Journal of Biological Chemistry *291*, 13040–13047.

Flaus, A., and Owen-Hughes, T. (2003). Dynamic Properties of Nucleosomes during Thermal and ATP-Driven Mobilization. Molecular and Cellular Biology *23*, 7767–7779.

Fleming, A.B., Kao, C.-F., Hillyer, C., Pikaart, M., and Osley, M.A. (2008). H2B Ubiquitylation Plays a Role in Nucleosome Dynamics during Transcription Elongation. Molecular Cell *31*, 57–66.

Floer, M., Wang, X., Prabhu, V., Berrozpe, G., Narayan, S., Spagna, D., Alvarez, D., Kendall, J., Krasnitz, A., Stepansky, A., et al. (2010). A RSC/Nucleosome Complex Determines Chromatin Architecture and Facilitates Activator Binding. Cell *141*, 407–418.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. Molecular Cell *60*, 256–267.

Fox, M.J., and Mosley, A.L. (2016). Rrp6: Integrated roles in nuclear RNA metabolism and transcription termination: Rrp6 in RNA Metabolism and Termination. Wiley Interdisciplinary Reviews: RNA *7*, 91–104.

Freitag, M. (2017). Histone Methylation by SET Domain Proteins in Fungi. Annual Review of Microbiology *71*, 413–439.

Gagnon, J., Lavoie, M., Catala, M., Malenfant, F., and Elela, S.A. (2015). Transcriptome Wide Annotation of Eukaryotic RNase III Reactivity and Degradation Signals. PLOS Genetics *11*, e1005000.

Ganapathi, M., Palumbo, M.J., Ansari, S.A., He, Q., Tsui, K., Nislow, C., and Morse, R.H. (2011). Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. Nucleic Acids Research *39*, 2032–2044.

Ganguli, D., Chereji, R.V., Iben, J.R., Cole, H.A., and Clark, D.J. (2014). RSC-dependent constructive and destructive interference between opposing arrays of phased nucleosomes in yeast. Genome Research *24*, 1637–1649.

Garbett, K.A., Tripathi, M.K., Cencki, B., Layer, J.H., and Weil, P.A. (2007). Yeast TFIID Serves as a Coactivator for Rap1p by Direct Protein-Protein Interaction. Molecular and Cellular Biology *27*, 297–311.

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., Ge, D., Gervais-Bird, J., Gagnon, J., and Abou Elela, S. (2005). Genome-Wide Prediction and Analysis of Yeast RNase III-Dependent snoRNA Processing Signals. Molecular and Cellular Biology *25*, 2981–2994.

Ghazal, G., Gagnon, J., Jacques, P.-É., Landry, J.-R., Robert, F., and Abou Elela, S. (2009). Yeast RNase III Triggers Polyadenylation-Independent Transcription Termination. Molecular Cell *36*, 99–109.

Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M., and Ponticelli, A.S. (2004). Amino Acid Substitutions in Yeast TFIIF Confer Upstream Shifts in Transcription Initiation and Altered Interaction with RNA Polymerase II. Molecular and Cellular Biology *24*, 10975–10985.

Gilbert, C., Kristjuhan, A., Winkler, G.S., and Svejstrup, J.Q. (2004). Elongator Interactions with Nascent mRNA Revealed by RNA Immunoprecipitation. Molecular Cell *14*, 457–464.

Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J.L., Barton, G.J., and Owen-Hughes, T. (2011). A Role for Snf2-Related Nucleosome-Spacing Enzymes in Genome-Wide Nucleosome Organization. Science *333*, 1758–1760.

Govind, C.K., Zhang, F., Qiu, H., Hofmeyer, K., and Hinnebusch, A.G. (2007). Gcn5 Promotes Acetylation, Eviction, and Methylation of Nucleosomes in Transcribed Coding Regions. Molecular Cell *25*, 31–42.

Graber, J.H., Cantor, C.R., Mohr, S.C., and Smith, T.F. (1999). Genomic detection of new yeast pre-mRNA 3'-end-processing signals. Nucleic Acids Research *27*, 888–894.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proceedings of the National Academy of Sciences *106*, 9613–9618.

Green, D.M., Marfatia, K.A., Crafton, E.B., Zhang, X., Cheng, X., and Corbett, A.H. (2002). Nab2p Is Required for Poly(A) RNA Export in *Saccharomyces cerevisiae* and Is Regulated by Arginine Methylation via Hmt1p. Journal of Biological Chemistry *277*, 7752–7760.

Gross, S., and Moore, C.L. (2001). Rna15 Interaction with the A-Rich Yeast Polyadenylation Signal Is an Essential Step in mRNA 3'-End Formation. Molecular and Cellular Biology *21*, 8045–8055.

Gu, B., Eick, D., and Bensaude, O. (2013). CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. Nucleic Acids Research *41*, 1591–1603.

Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nature Structural & Molecular Biology *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. Molecular Cell *48*, 409–421.

Guo, Z., Russo, P., Yun, D.F., Butler, J.S., and Sherman, F. (1995). Redundant 3' end-forming signals for the yeast CYC1 mRNA. Proceedings of the National Academy of Sciences *92*, 4211–4214.

Hahn, S., and Young, E.T. (2011). Transcriptional Regulation in Saccharomyces cerevisiae: Transcription Factor Regulation and Function, Mechanisms of Initiation, and Roles of Activators and Coactivators. Genetics *189*, 705–736.

Hainer, S.J., Pruneski, J.A., Mitchell, R.D., Monteverde, R.M., and Martens, J.A. (2011). Intergenic transcription causes repression by directing nucleosome assembly. Genes Dev. *25*, 29–40.

Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche, N.E.H., Gostan, T., Vandermoere, F., Robert, M.-C., Cusack, S., et al. (2013). CBC–ARS2 stimulates

3′-end maturation of multiple RNA families and favors cap-proximal processing. Nature Structural & Molecular Biology *20*, 1358–1366.

Han, M., and Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters in vivo. Cell *55*, 1137–1145.

Han, Z., Libri, D., and Porrua, O. (2017). Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. Nucleic Acids Research *45*, 1355–1370.

Hardy, C.F., Sussel, L., and Shore, D. (1992). A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. Genes Dev. *6*, 801–814.

Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nature Reviews Molecular Cell Biology *18*, 263–273.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Reports *15*, 2147–2158.

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. Molecular Cell *49*, 55–66.

Hector, R.E., Nykamp, K.R., Dheur, S., Anderson, J.T., Non, P.J., Urbinati, C.R., Wilson, S.M., Minvielle-Sebastia, L., and Swanson, M.S. (2002). Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. The EMBO Journal *21*, 1800–1810.

Henry, K.W., Wyce, A., Lo, W.-S., Duggan, L.J., Tolga Emre, N.C., Kao, C.-F., Pillus, L., Shilatifard, A., Osley, M.A., and Berger, S.L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes & Development *17*, 2648–2663.

Heo, D., Yoo, I., Kong, J., Lidschreiber, M., Mayer, A., Choi, B.-Y., Hahn, Y., Cramer, P., Buratowski, S., and Kim, M. (2013). The RNA polymerase II C-terminal domain-interacting domain of yeast Nrd1 contributes to the choice of termination pathway and couples to RNA processing by the nuclear exosome. J. Biol. Chem. *288*, 36676–36690.

Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S., and Stefl, R. (2011). Recognition of Transcription Termination Signal by the Nuclear Polyadenylated RNA-binding (NAB) 3 Protein. Journal of Biological Chemistry *286*, 3645–3657.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and Bradbury, E.M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 “tail” to DNA. J. Biol. Chem. *268*, 305–314.

Hong, S.W., Hong, S.M., Yoo, J.W., Lee, Y.C., Kim, S., Lis, J.T., and Lee, D. -k. (2009). Phosphorylation of the RNA polymerase II C-terminal domain by TFIIH kinase is not essential for transcription of Saccharomyces cerevisiae genome. Proceedings of the National Academy of Sciences *106*, 14276–14280.

Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C., and Tollervey, D. (2006). Microarray detection of novel nuclear RNA substrates for the exosome. Yeast *23*, 439–454.

Hsin, J.-P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD Phosphorylated on Threonine-4 Is Required for Histone mRNA 3' End Processing. Science *334*, 683–686.

Hughes, A.L., Jin, Y., Rando, O.J., and Struhl, K. (2012). A Functional Evolutionary Approach to Identify Determinants of Nucleosome Positioning: A Unifying Model for Establishing the Genome-wide Pattern. Molecular Cell *48*, 5–15.

Huisinga, K.L., and Pugh, B.F. (2004). A Genome-Wide Housekeeping Role for TFIID and a Highly Regulated Stress-Related Role for SAGA in Saccharomyces cerevisiae. Molecular Cell *13*, 573–585.

Hull, M.W., McKune, K., and Woychik, N.A. (1995). RNA polymerase II subunit RPB9 is required for accurate start site selection. Genes & Development *9*, 481–490.

Hunter, G.O., Fox, M.J., Smith-Kinnaman, W.R., Gogol, M., Fleharty, B., and Mosley, A.L. (2016). Phosphatase Rtr1 Regulates Global Levels of Serine 5 RNA Polymerase II C-Terminal Domain Phosphorylation and Cotranscriptional Histone Methylation. Molecular and Cellular Biology *36*, 2236–2245.

Iasillo, C., Schmid, M., Yahia, Y., Maqbool, M.A., Descostes, N., Karadoulama, E., Bertrand, E., Andrau, J.-C., and Jensen, T.H. (2017). ARS2 is a general suppressor of pervasive transcription. Nucleic Acids Research *45*, 10229–10241.

Iyer, V., and Struhl, K. (1995). Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO J *14*, 2570–2579.

Izban, M.G., and Luse, D.S. (1991). Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. Genes & Development *5*, 683–696.

Izban, M.G., and Luse, D.S. (1993). SII-facilitatedTranscript Cleavage inRNA Polymerase I1 Complexes Stalled Early after Initiation Occurs in Primarily Dinucleotide Increments. 10.

Jansen, A., and Verstrepen, K.J. (2011). Nucleosome Positioning in Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews *75*, 301–320.

Jasnovidova, O., Krejcikova, M., Kubicek, K., and Stefl, R. (2017a). Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. EMBO Reports *18*, 906–913.

Jasnovidova, O., Klumpler, T., Kubicek, K., Kalynych, S., Plevka, P., and Stefl, R. (2017b). Structure and dynamics of the RNAPII CTDsome with Rtt103. PNAS *114*, 11133–11138.

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with Pervasive Transcription. Molecular Cell *52*, 473–484.

Jia, H., Wang, X., Liu, F., Guenther, U.-P., Srinivasan, S., Anderson, J.T., and Jankowsky, E. (2011). The RNA Helicase Mtr4p Modulates Polyadenylation in the TRAMP Complex. Cell *145*, 890–901.

Jiang, C., and Pugh, B.F. (2009). A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. Genome Biology *10*, R109.

Jin, Y., Eser, U., Struhl, K., and Churchman, L.S. (2017). The Ground State and Evolution of Promoter Region Directionality. Cell *170*, 889-898.e10.

Johnson, A.N., and Weil, P.A. (2017). Identification of a transcriptional activation domain in yeast repressor activator protein 1 (Rap1) using an altered DNA-binding specificity variant. J. Biol. Chem. *292*, 5705–5723.

Ju, Q.D., Morrow, B.E., and Warner, J.R. (1990). REB1, a yeast DNA-binding protein with many targets, is essential for growth and bears some resemblance to the oncogene myb. Molecular and Cellular Biology *10*, 5226–5234.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental Biology *339*, 225–229.

Kadaba, S., Krueger, A., Trice, T., Krecic, A.M., Hinnebusch, A.G., and Anderson, J.T. (2004). Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae. Genes & Development *18*, 1227–1240.

Kao, C.-F., Hillyer, C., Tsukuda, T., Henry, K., Berger, S., and Osley, M.A. (2004). Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. Genes & Development *18*, 184–195.

Kaplan, C.D., Moris, J.R., Wu, C.T., and Winston, F. (2000). Spt5 and Spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. Genes & Development *14*, 2623–2634.

Kaplan, C.D., laprade, L., and Winston, F. (2003). Transcription Elongation Factors Repress Transcription Initiation from Cryptic Sites. Science *301*, 1096–1099.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L., et al. (2007). RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. Science *316*, 1484–1488.

Kasahara, K., Ohyama, Y., and Kokubo, T. (2011). Hmo1 directs pre-initiation complex assembly to an appropriate site on its target gene promoters by masking a nucleosome-free region. Nucleic Acids Research *39*, 4136–4150.

Kassabov, S.R., Zhang, B., Persinger, J., and Bartholomew, B. (2003). SWI/SNF Unwraps, Slides, and Rewraps the Nucleosome. Molecular Cell *11*, 391–403.

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 & Development *22*, 1082–1092.

Keogh, M.-C., Podolny, V., and Buratowski, S. (2003). Bur1 Kinase Is Required for Efficient Transcription Elongation by RNA Polymerase II. Molecular and Cellular Biology *23*, 7005–7018.

Keogh, M.-C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., et al. (2005). Cotranscriptional Set2 Methylation of Histone H3 Lysine 36 Recruits a Repressive Rpd3 Complex. Cell *123*, 593–605.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nature Reviews Molecular Cell Biology *13*, 700–712.

Kim, T., and Buratowski, S. (2009). Dimethylation of H3K4 by Set1 Recruits the Set3 Histone Deacetylase Complex to 5′ Transcribed Regions. Cell *137*, 259–272.

Kim, J.H., Lee, B.B., Oh, Y.M., Zhu, C., Steinmetz, L.M., Lee, Y., Kim, W.K., Lee, S.B., Buratowski, S., and Kim, T. (2016). Modulation of mRNA and lncRNA expression dynamics by the Set2–Rpd3S pathway. Nature Communications *7*.

Kim, M., Ahn, S.-H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004a). Transitions in RNA polymerase II elongation complexes at the 3′ ends of genes. The EMBO Journal *23*, 354–364.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004b). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517–522.

Kim, M., Suh, H., Cho, E.-J., and Buratowski, S. (2009). Phosphorylation of the Yeast Rpb1 C-terminal Domain at Serines 2, 5, and 7. Journal of Biological Chemistry *284*, 26421–26426.

Kim, T., Xu, Z., Clauder-Münster, S., Steinmetz, L.M., and Buratowski, S. (2012). Set3 HDAC Mediates Effects of Overlapping Noncoding Transcription on Gene Induction Kinetics. Cell *150*, 1158–1169.

Kimmerly, W., Buchman, A., Kornberg, R., and Rine, J. (1988). Roles of two DNA-binding factors in replication, segregation and transcriptional repression mediated by a yeast silencer. EMBO J. *7*, 2241–2253.

Klein-Brill, A., Joseph-Strauss, D., Appleboim, A., and Friedman, N. (2019). Dynamics of Chromatin and Transcription during Transient Depletion of the RSC Chromatin Remodeling Complex. Cell Reports *26*, 279-292.e5.

Knight, B., Kubik, S., Ghosh, B., Bruzzone, M.J., Geertz, M., Martin, V., Dénervaud, N., Jacquet, P., Ozkan, B., Rougemont, J., et al. (2014). Two distinct promoter architectures centered on dynamic nucleosomes control ribosomal protein gene transcription. Genes Dev. *28*, 1695–1709.

Kobor, M.S., Venkatasubrahmanyam, S., Meneghini, M.D., Gin, J.W., Jennings, J.L., Link, A.J., Madhani, H.D., and Rine, J. (2004). A Protein Complex Containing the Conserved Swi2/Snf2-Related ATPase Swr1p Deposits Histone Variant H2A.Z into Euchromatin. PLOS Biology *2*, e131.

Koering, C.E., Fourel, G., Binet-Brasselet, E., Laroche, T., Klein, F., and Gilson, E. (2000). Identification of high affinity Tbf1p-binding sites within the budding yeast genome. Nucleic Acids Res *28*, 2519–2526.

Komarnitsky, P., Cho, E.-J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & Development *14*, 2452–2460.

Korde, A., Rosselot, J.M., and Donze, D. (2014). Intergenic Transcriptional Interference Is Blocked by RNA Polymerase III Transcription Factor TFIIIB in *Saccharomyces cerevisiae*. Genetics *196*, 427–438.

Kornberg, R.D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. Science *184*, 868–871.

Krajewski, W.A. (2013). Comparison of the Isw1a, Isw1b, and Isw2 Nucleosome Disrupting Activities. Biochemistry *52*, 6940–6949.

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., and Korber, P. (2016). Genomic Nucleosome Organization Reconstituted with Pure Proteins. Cell *167*, 709-721.e12.

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004). Ssu72 Is an RNA Polymerase II CTD Phosphatase. Molecular Cell *14*, 387–394.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., et al. (2003a). The Paf1 Complex Is Required for Histone H3 Methylation by COMPASS and Dot1p: Linking Transcriptional Elongation to Histone Methylation. Molecular Cell *11*, 721–729.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., et al. (2003b). Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Molecular and Cellular Biology *23*, 4207–4218.

Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. Genes & Development *26*, 1891–1896.

Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.-L., Rougemont, J., and Shore, D. (2015). Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. Molecular Cell *60*, 422–434.

Kubik, S., Bruzzone, M.J., and Shore, D. (2017). TFIID or not TFIID, a continuing transcriptional SAGA. The EMBO Journal *36*, 248–249.

Kubik, S., O'Duibhir, E., de Jonge, W.J., Mattarocci, S., Albert, B., Falcone, J.-L., Bruzzone, M.J., Holstege, F.C.P., and Shore, D. (2018). Sequence-Directed Action of RSC Remodeler and General Regulatory Factors Modulates +1 Nucleosome Position to Facilitate Transcription. Mol. Cell *71*, 89-102.e5.

Kubik, S., Challal, D., Bruzzone, M.J., Dreos, R., Mattarocci, S., Bucher, P., Libri, D., and Shore, D. (2019). Opposing chromatin remodelers control transcription initiation frequency and start site selection. BioRxiv.

Kuehner, J.N., and Brow, D.A. (2006). Quantitative Analysis of in Vivo Initiator Selection by Yeast RNA Polymerase II Supports a Scanning Model. J. Biol. Chem. *281*, 14119–14128.

Kuehner, J.N., and Brow, D.A. (2008). Regulation of a Eukaryotic Gene by GTP-Dependent Start Site Selection and Transcription Attenuation. Molecular Cell *31*, 201–211.

Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. Nature Reviews Molecular Cell Biology *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. The EMBO Journal *11*, 4665–4674.

Kurtz, S., and Shore, D. (1991). RAP1 protein activates and silences transcription of mating-type genes m yeast. 14.

Kwak, H., and Lis, J.T. (2013). Control of Transcriptional Elongation. Annual Review of Genetics *47*, 483–508.

Kwapisz, M., Wery, M., Després, D., Ghavi-Helm, Y., Soutourina, J., Thuriaux, P., and Lacroute, F. (2008). Mutations of RNA polymerase II activate key genes of the nucleoside triphosphate biosynthetic pathways. The EMBO Journal *27*, 2411–2421.

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

Kyrion, G., Boakye, K.A., and Lustig, A.J. (1992). C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in Saccharomyces cerevisiae. Mol. Cell. Biol. *12*, 5159–5173.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005). RNA Degradation by the Exosome Is Promoted by a Nuclear Polyadenylation Complex. Cell *121*, 713–724.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nature Reviews Molecular Cell Biology *18*, 548–562.

Lam, K.C., Chung, H.-R., Semplicio, G., Iyer, S.S., Gaub, A., Bhardwaj, V., Holz, H., Georgiev, P., and Akhtar, A. (2019). The NSL complex-mediated nucleosome landscape is required to maintain transcription fidelity and suppression of transcription noise. Genes & Development *33*, 452–465.

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. Molecular and Cellular Biology *13*, 649–658.

Lang, W.H., and Reeder, R.H. (1995). Transcription termination of RNA polymerase I due to a T-rich element interacting with Reblp. Proc. Natl. Acad. Sci. USA 5.

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.

Larochelle, M., Robert, M.-A., Hébert, J.-N., Liu, X., Matteau, D., Rodrigue, S., Tian, B., Jacques, P.-É., and Bachand, F. (2018). Common mechanism of transcription termination at coding and noncoding RNA genes in fission yeast. Nature Communications *9*.

Lebreton, A., Tomecki, R., Dziembowski, A., and Séraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature *456*, 993–996.

Lee, S.D., and Moore, C.L. (2014). Efficient mRNA Polyadenylation Requires a Ubiquitin-Like Domain, a Zinc Knuckle, and a RING Finger Domain, All Contained in the Mpe1 Protein. Molecular and Cellular Biology *34*, 3955–3967.

Lee, C.-K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nature Genetics *36*, 900–905.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A Posttive Role for Histone in lrzmcription Factor A. 12.

Lee, J.-S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R., and Shilatifard, A. (2007a). Histone Crosstalk between H2B Monoubiquitination and H3 Methylation Mediated by COMPASS. Cell *131*, 1084–1096.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007b). A high-resolution atlas of nucleosome occupancy in yeast. Nature Genetics *39*, 1235–1244.

Lenstra, T.L., Benschop, J.J., Kim, T., Schulze, J.M., Brabers, N.A.C.H., Margaritis, T., van de Pasch, L.A.L., van Heesch, S.A.A.C., Brok, M.O., Groot Koerkamp, M.J.A., et al. (2011). The Specificity and Topology of Chromatin Interaction Pathways in Yeast. Molecular Cell *42*, 536–549.

Lenstra, T.L., Tudek, A., Clauder, S., Xu, Z., Pachis, S.T., van Leenen, D., Kemmeren, P., Steinmetz, L.M., Libri, D., and Holstege, F.C.P. (2013). The Role of Ctk1 Kinase in Termination of Small Non-Coding RNAs. PLoS ONE *8*, e80495.

Leuther, K.K., Bushnell, D.A., and Kornberg, R.D. (1996). Two-Dimensional Crystallography of TFIIB– and IIE–RNA Polymerase II Complexes: Implications for Start Site Selection and Initiation Complex Formation. Cell *85*, 773–779.

Li, B., Howe, L., Anderson, S., Yates, J.R., and Workman, J.L. (2003). The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. Journal of Biological Chemistry *278*, 8897–8903.

Li, B., Jackson, J., Simon, M.D., Fleharty, B., Gogol, M., Seidel, C., Workman, J.L., and Shilatifard, A. (2009). Histone H3 Lysine 36 Dimethylation (H3K36me2) Is Sufficient to Recruit the Rpd3s Histone Deacetylase Complex and to Repress Spurious Transcription. Journal of Biological Chemistry *284*, 7970–7976.

Lidschreiber, M., Easter, A.D., Battaglia, S., Rodríguez-Molina, J.B., Casañal, A., Carminati, M., Baejen, C., Grzechnik, P., Maier, K.C., Cramer, P., et al. (2018). The APT complex is involved in non-coding RNA transcription and is distinct from CPF. Nucleic Acids Research.

Lieb, J.D., Liu, X., Botstein, D., and Brown, P.O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat. Genet. *28*, 327–334.

Lieleg, C., Ketterer, P., Nuebler, J., Ludwigsen, J., Gerland, U., Dietz, H., Mueller-Planitz, F., and Korber, P. (2015). Nucleosome Spacing Generated by ISWI and CHD1 Remodelers Is Constant Regardless of Nucleosome Density. Molecular and Cellular Biology *35*, 1588–1605.

Liu, C., and Lustig, A.J. (1996). Genetic Analysis of Rap1p/Sir3p Interactions in Telomeric and HML Silencing in Saccharomyces cerevisiae. Genetics *143*, 81–93.

Liu, C.L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S.L., Friedman, N., and Rando, O.J. (2005). Single-Nucleosome Mapping of Histone Modifications in S. cerevisiae. PLoS Biology *3*, e328.

Liu, X., Bushnell, D.A., Silva, D.-A., Huang, X., and Kornberg, R.D. (2011). Initiation complex structure and promoter proofreading. Science *333*, 633–637.

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M., and Hahn, S. (2009). Phosphorylation of the Transcription Elongation Factor Spt5 by Yeast Bur1 Kinase Stimulates Recruitment of the PAF Complex. Molecular and Cellular Biology *29*, 4852–4863.

Lorch, Y., Maier-Davis, B., and Kornberg, R.D. (2006). Chromatin remodeling by nucleosome disassembly in vitro. Proc Natl Acad Sci U S A *103*, 3090–3093.

Lorch, Y., Griesenbeck, J., Boeger, H., Maier-Davis, B., and Kornberg, R.D. (2011). Selective removal of promoter nucleosomes by the RSC chromatin-remodeling complex. Nature Structural & Molecular Biology *18*, 881–885.

Loya, T.J., O'Rourke, T.W., and Reines, D. (2012). A genetic screen for terminator function in yeast identifies a role for a new functional domain in termination factor Nab3. Nucleic Acids Research *40*, 7476–7491.

Loya, T.J., O'Rourke, T.W., and Reines, D. (2013). Yeast Nab3 Protein Contains a Self-assembly Domain Found in Human Heterogeneous Nuclear Ribonucleoprotein-C (hnRNP-C) That Is Necessary for Transcription Termination. Journal of Biological Chemistry *288*, 2111–2117.

Lu, H., Flores, O., Weinmann, R., and Reinberg, D. (1991). The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. Proceedings of the National Academy of Sciences *88*, 10004–10008.

Luger, K. (1997). Crystal structure of the nucleosome core particle at 2.8 A˚ resolution. *389*, 10.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nature Structural & Molecular Biology *17*, 1195–1201.

Luo, K., Vega-Palas, M.A., and Grunstein, M. (2002). Rap1–Sir4 binding independent of other Sir, yKu, or histone interactions initiates the assembly of telomeric heterochromatin in yeast. Genes Dev. *16*, 1528–1539.

Luo, W., Johnson, A.W., and Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes & Development *20*, 954–965.

Luse, D.S. (2013). Promoter clearance by RNA polymerase II. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1829*, 63–68.

Lustig, A.J., Kurtz, S., and Shore, D. (1990). Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. Science *250*, 549–553.

MacKellar, A.L., and Greenleaf, A.L. (2011). Cotranscriptional Association of mRNA Export Factor Yra1 with C-terminal Domain of RNA Polymerase II. Journal of Biological Chemistry *286*, 36385–36395.

Makino, D.L., Schuch, B., Stegmann, E., Baumgärtner, M., Basquin, C., and Conti, E. (2015). RNA degradation paths in a 12-subunit nuclear exosome complex. Nature *524*, 54–58.

Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mRNA decay. ELife *4*.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3′-end processing. Cellular and Molecular Life Sciences *65*, 1099–1122.

Marquardt, S., Hazelbaker, D.Z., and Buratowski, S. (2011). Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. Transcription *2*, 145–154.

Marquardt, S., Escalante-Chong, R., Pho, N., Wang, J., Churchman, L.S., Springer, M., and Buratowski, S. (2014). A Chromatin-Based Mechanism for Limiting Divergent Noncoding Transcription. Cell *158*, 462.

Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature *429*, 571.

Martens, J.A., Wu, P.-Y.J., and Winston, F. (2005). Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. Genes Dev. *19*, 2695–2704.

Martin-Tumasz, S., and Brow, D.A. (2015). *Saccharomyces cerevisiae* Sen1 Helicase Domain Exhibits 5′- to 3′-Helicase Activity with a Preference for Translocation on DNA Rather than RNA. Journal of Biological Chemistry *290*, 22880–22889.

Mas, G., de Nadal, E., Dechant, R., de la Concepción, M.L.R., Logie, C., Jimeno-González, S., Chávez, S., Ammerer, G., and Posas, F. (2009). Recruitment of a chromatin remodelling complex by the Hog1 MAP kinase to stress genes. The EMBO Journal *28*, 326–336.

Mason, P.B., and Struhl, K. (2003). The FACT Complex Travels with Elongating RNA Polymerase II and Is Important for the Fidelity of Transcriptional Initiation In Vivo. Molecular and Cellular Biology *23*, 8323–8333.

Mattarocci, S., Hafner, L., Lezaja, A., Shyian, M., and Shore, D. (2016). Rif1: A Conserved Regulator of DNA Replication and Repair Hijacked by Telomeres in Yeasts. Frontiers in Genetics *7*.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008a). Nucleosome organization in the Drosophila genome. Nature *453*, 358–362.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008b). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Research *18*, 1073–1083.

Max, T., Søgaard, M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal Repeat Domain of RNA Polymerase II Facilitates Dissociation of Its Complex with Mediator. Journal of Biological Chemistry *282*, 14113–14120.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nature Structural & Molecular Biology *17*, 1272–1278.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. Science *336*, 1723–1725.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Program, A.E., Shuman, S., et al. (1997a). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes & Development *11*, 3306–3318.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997b). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357–361.

McKinlay, A., Podicheti, R., Wendte, J.M., Cocklin, R., and Rusch, D.B. (2018). RNA polymerases IV and V influence the 3′ boundaries of Polymerase II transcription units in *Arabidopsis*. RNA Biology *15*, 269–279.

Meinel, D.M., Burkert-Kautzsch, C., Kieser, A., O'Duibhir, E., Siebert, M., Mayer, A., Cramer, P., Söding, J., Holstege, F.C.P., and Sträßer, K. (2013). Recruitment of TREX to the Transcription Machinery by Its Direct Binding to the Phospho-CTD of RNA Polymerase II. PLOS Genetics *9*, e1003914.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature *430*, 223–226.

Mellor, J., Jiang, W., Funk, M., Rathjen, J., Barnes, C.A., Hinz, T., Hegemann, J.H., and Philippsen, P. (1990). CPF1, a yeast protein which functions in centromeres and promoters. EMBO J. *9*, 4017–4026.

Mencía, M., Moqtaderi, Z., Geisberg, J.V., Kuras, L., and Struhl, K. (2002). Activator-Specific Recruitment of TFIID and Regulation of Ribosomal Protein Genes in Yeast. Molecular Cell *9*, 823–833.

Merkl, P., Perez-Fernandez, J., Pilsl, M., Reiter, A., Williams, L., Gerber, J., Bohm, M., Deutzmann, R., Griesenbeck, J., Milkereit, P., et al. (2014). Binding of the Termination Factor Nsi1 to Its Cognate DNA Site Is Sufficient To Terminate RNA Polymerase I Transcription In Vitro and To Induce Termination In Vivo. Molecular and Cellular Biology *34*, 3817–3827.

Milligan, L., Huynh-Thu, V.A., Delan-Forino, C., Tuck, A., Petfalski, E., Lombraña, R., Sanguinetti, G., Kudla, G., and Tollervey, D. (2016). Strand-specific, high-resolution mapping of modified RNA polymerase II. Molecular Systems Biology *12*, 874.

Minsky, N., Shema, E., Field, Y., Schuster, M., Segal, E., and Oren, M. (2008). Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nature Cell Biology *10*, 483–488.

Minvielle-Sebastia, L., Beyer, K., Krecic, A.M., Hector, R.E., Swanson, M.S., and Keller, W. (1998). Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. The EMBO Journal *17*, 7454–7468.

Mischo, H.E., and Proudfoot, N.J. (2013). Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1829*, 174–185.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., and Wu, C. (2004). ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. Science *303*, 343–348.

Mizuno, T., Kishimoto, T., Shinzato, T., Haw, R., Chambers, A., Wood, J., Sinclair, D., and Uemura, H. (2004). Role of the N-terminal region of Rap1p in the transcriptional activation of glycolytic genes inSaccharomyces cerevisiae. Yeast *21*, 851–866.

Morris, S.A., Baek, S., Sung, M.-H., John, S., Wiench, M., Johnson, T.A., Schiltz, R.L., and Hager, G.L. (2014). Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. Nature Structural & Molecular Biology *21*, 73–81.

Morrow, B.E., Johnson, S.P., and Warner, J.R. (1989). Proteins That Bind to the Yeast rDNA Enhancer. 8.

Morrow, B.E., Ju, Q., and Warner, J.R. (1990). Purification and characterization of the yeast rDNA binding protein REB1. J. Biol. Chem. *265*, 20778–20783.

Morrow, B.E., Ju, Q., and Warner, J.R. (1993). A bipartite DNA-binding domain in yeast Reb1p. Molecular and Cellular Biology *13*, 1173–1182.

Morse, R.H. (2000). RAP, RAP, open up! New wrinkles for RAP1 in yeast. Trends in Genetics *16*, 51–53.

Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., and Washburn, M.P. (2009). Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation. Molecular Cell *34*, 168–178.

Murray, S.C., Serra Barros, A., Brown, D.A., Dudek, P., Ayling, J., and Mellor, J. (2012). A pre-initiation complex at the 3′-end of genes drives antisense transcription independent of divergent sense transcription. Nucleic Acids Research *40*, 2432–2444.

Nadal-Ribelles, M., Solé, C., Xu, Z., Steinmetz, L.M., de Nadal, E., and Posas, F. (2014). Control of Cdc28 CDK1 by a Stress-Induced lncRNA. Molecular Cell *53*, 549–561.

Nagawa, F., and Fink, G.R. (1985). The relationship between the “TATA” sequence and transcription initiation sites at the HIS4 gene of Saccharomyces cerevisiae. PNAS *82*, 8557–8561.

Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C.L., and Greenblatt, J. (2003). Organization and Function of APT, a Subcomplex of the Yeast Cleavage and Polyadenylation Factor Involved in the Formation of mRNA and Small Nucleolar RNA 3′-Ends. Journal of Biological Chemistry *278*, 33000–33010.

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.

Nevers, A., Doyen, A., Malabat, C., Néron, B., Kergrohen, T., Jacquier, A., and Badis, G. (2018). Antisense transcriptional interference mediates condition-specific gene repression in budding yeast. Nucleic Acids Res. *46*, 6009–6025.

Ng, H.H., Xu, R.-M., Zhang, Y., and Struhl, K. (2002). Ubiquitination of Histone H2B by Rad6 Is Required for Efficient Dot1-mediated Methylation of Histone H3 Lysine 79. Journal of Biological Chemistry *277*, 34655–34657.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity. Molecular Cell *11*, 709–719.

Niedenthal, R., Stoll, R., and Hegemann, J.H. (1991). In vivo characterization of the Saccharomyces cerevisiae centromere DNA element I, a binding site for the helix-loop-helix protein CPF1. Mol. Cell. Biol. *11*, 3545–3553.

Nielsen, M., Ard, R., Leng, X., Ivanov, M., Kindgren, P., Pelechano, V., and Marquardt, S. (2019). Transcription-driven chromatin repression of Intragenic transcription start sites. PLOS Genetics *15*, e1007969.

Ocampo, J., Chereji, R.V., Eriksson, P.R., and Clark, D.J. (2016). The ISW1 and CHD1 ATP-dependent chromatin remodelers compete to set nucleosome spacing *in vivo*. Nucleic Acids Research *44*, 4625–4635.

Oliveira, A.P., Ludwig, C., Zampieri, M., Weisser, H., Aebersold, R., and Sauer, U. (2015). Dynamic phosphoproteomics reveals TORC1-dependent regulation of yeast nucleotide and amino acid biosynthesis. Sci Signal *8*, rs4.

Orozco, I.J., Kim, S.J., and Martinson, H.G. (2002). The Poly(A) Signal, without the Assistance of Any Downstream Element, Directs RNA Polymerase II to Pause *in Vivo* and Then to Release Stochastically from the Template. Journal of Biological Chemistry *277*, 42899–42911.

Orphanides, G., LeRoy, G., Chang, C.-H., Luse, D.S., and Reinberg, D. (1998). FACT, a Factor that Facilitates Transcript Elongation through Nucleosomes. Cell *92*, 105–116.

Orphanides, G., Wu, W.-H., Lane, W.S., Hampsey, M., and Reinberg, D. (1999). The chromatin-speci®c transcription elongation factor FACTcomprises human SPT16 and SSRP1 proteins. *400*, 5.

Osheim, Y.N., Slikes, M.L., and Beyer, K. (2002). EM visualization of Pol II genes in Drosophila : most genes terminate without prior 3′ end cleavage of nascent transcripts. Chromosoma *111*, 1–12.

Ozonov, E.A., and van Nimwegen, E. (2013). Nucleosome Free Regions in Yeast Promoters Result from Competitive Binding of Transcription Factors That Interact with Chromatin Modifiers. PLoS Computational Biology *9*, e1003181.

Papai, G., Tripathi, M.K., Ruhlmann, C., Layer, J.H., Weil, P.A., and Schultz, P. (2010). TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. Nature *465*, 956–960.

Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global Regulation of H2A.Z Localization by the INO80 Chromatin-Remodeling Enzyme Is Essential for Genome Integrity. Cell *144*, 200–213.

Park, J., Kang, M., and Kim, M. (2015). Unraveling the mechanistic features of RNA polymerase II termination by the 5'-3' exoribonuclease Rat1. Nucleic Acids Res. *43*, 2625–2637.

Parnell, T.J., Huff, J.T., and Cairns, B.R. (2008). RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. The EMBO Journal *27*, 100–110.

Parnell, T.J., Schlichter, A., Wilson, B.G., and Cairns, B.R. (2015). The chromatin remodelers RSC and ISW1 display functional and chromatin-based promoter antagonism. ELife *4*.

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B Monoubiquitination Functions Cooperatively with FACT to Regulate Elongation by RNA Polymerase II. Cell *125*, 703–717.

Pearson, E.L., and Moore, C.L. (2013). Dismantling promoter-driven RNA polymerase II transcription complexes in vitro by the termination factor Rat1. J. Biol. Chem. *288*, 19750–19759.

Pinto, I., Ware, D.E., and Hampsey, M. (1992). The yeast SUA7 gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection in vivo. Cell *68*, 977–988.

Pinto, I., Wu, W.H., Na, J.G., and Hampsey, M. (1994). Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast. J. Biol. Chem. *269*, 30569–30573.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast. Cell *122*, 517–527.

Porrua, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nature Structural & Molecular Biology *20*, 884–891.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nature Reviews Molecular Cell Biology *16*, 190–202.

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: Novel determinants of degradative termination. The EMBO Journal *31*, 3935–3948.

Porrua, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes. Trends in Genetics *32*, 508–522.

Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription regulation. Critical Reviews in Biochemistry and Molecular Biology *48*, 575–608.

Powers, T., and Walter, P. (1999). Regulation of Ribosome Biogenesis by the Rapamycin-sensitive TOR-signaling Pathway in Saccharomyces cerevisiae. MBoC *10*, 987–1000.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. Science *322*, 1851–1854.

Preti, M., Ribeyre, C., Pascali, C., Bosio, M.C., Cortelazzi, B., Rougemont, J., Guarnera, E., Naef, F., Shore, D., and Dieci, G. (2010). The Telomere-Binding Protein Tbf1 Demarcates snoRNA Gene Promoters in Saccharomyces cerevisiae. Molecular Cell *38*, 614–620.

Qiu, H., Hu, C., and Hinnebusch, A.G. (2009). Phosphorylation of the Pol II CTD by KIN28 Enhances BUR1/BUR2 Recruitment and Ser2 CTD Phosphorylation Near Promoters. Molecular Cell *33*, 752–762.

Ranjan, A., Mizuguchi, G., FitzGerald, P.C., Wei, D., Wang, F., Huang, Y., Luk, E., Woodcock, C.L., and Wu, C. (2013). Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A.Z replacement. Cell *154*, 1232–1245.

Rawal, Y., Chereji, R.V., Qiu, H., Ananthakrishnan, S., Govind, C.K., Clark, D.J., and Hinnebusch, A.G. (2018). SWI/SNF and RSC cooperate to reposition and evict promoter nucleosomes at highly expressed genes in yeast. Genes Dev. *32*, 695–710.

Ray-Soni, A., Bellecourt, M.J., and Landick, R. (2016). Mechanisms of Bacterial Transcription Termination: All Good Things Must End. Annual Review of Biochemistry *85*, 319–347.

Reed, S.H., Akiyama, M., Stillman, B., and Friedberg, E.C. (1999). Yeast autonomously replicating sequence binding factor is involved in nucleotide excision repair. Genes Dev. *13*, 3052–3058.

Reeder, R.H., Guevara, P., and Roan, J.G. (1999). *Saccharomyces cerevisiae* RNA Polymerase I Terminates Transcription at the Reb1 Terminator In Vivo. Molecular and Cellular Biology *19*, 7369–7376.

Reines, D. (1992). Elongation Factor-dependent Transcript Shortening by Template-engaged RNA Polymerase 11". 6.

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: RNA polymerase I-transcription termination *in vivo*. The EMBO Journal *31*, 3480–3493.

Reja, R., Vinayachandran, V., Ghosh, S., and Pugh, B.F. (2015). Molecular mechanisms of ribosomal protein gene coregulation. Genes & Development *29*, 1942–1954.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution. Cell *147*, 1408–1419.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.

Ribar, B., Prakash, L., and Prakash, S. (2006). Requirement of ELC1 for RNA Polymerase II Polyubiquitylation and Degradation in Response to DNA Damage in Saccharomyces cerevisiae. Molecular and Cellular Biology *26*, 3999–4005.

Ribar, B., Prakash, L., and Prakash, S. (2007). ELA1 and CUL3 Are Required Along with ELC1 for RNA Polymerase II Polyubiquitylation and Degradation in DNA-Damaged Yeast Cells. Molecular and Cellular Biology *27*, 3211–3216.

Ribaud, V., Ribeyre, C., Damay, P., and Shore, D. (2012). DNA-end capping by the budding yeast transcription factor and subtelomeric binding protein Tbf1. The EMBO Journal *31*, 138–149.

Roguev, A. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. The EMBO Journal *20*, 7137–7148.

Rondón, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-Safe Transcriptional Termination for Protein-Coding Genes in S. cerevisiae. Molecular Cell *36*, 88–98.

Roy, K., Gabunilas, J., Gillespie, A., Ngo, D., and Chanfreau, G.F. (2016). Common genomic elements promote transcriptional and DNA replication roadblocks. Genome Research *26*, 1363–1375.

Sadowski, M., Dichtl, B., Hubner, W., and Keller, W. (2003). Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. The EMBO Journal *22*, 2167–2177.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nature Reviews Molecular Cell Biology *16*, 129–143.

Schaft, D., Roguev, A., kotovic, K.M., Shevvhenko, A., Sarov, M., Shevchenko, A., Neugebauer, K.M., and Stewart, A.F. (2003). The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. Nucleic Acids Research *31*, 2475–2482.

Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-Wide Mapping of Yeast RNA Polymerase II Termination. PLoS Genetics *10*, e1004632.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell *132*, 887–898.

Schreieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L.A. (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. Nature Structural & Molecular Biology *21*, 175–179.

Schwer, B., and Shuman, S. (2011). Deciphering the RNA Polymerase II CTD Code in Fission Yeast. Molecular Cell *43*, 311–318.

Sdano, M.A., Fulcher, J.M., Palani, S., Chandrasekharan, M.B., Parnell, T.J., Whitby, F.G., Formosa, T., and Hill, C.P. (2017). A novel SH2 recognition mechanism recruits Spt6 to the doubly phosphorylated RNA polymerase II linker at sites of transcription. ELife *6*.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent Transcription from Active Promoters. Science *322*, 1849–1851.

Selth, L.A., Sigurdsson, S., and Svejstrup, J.Q. (2010). Transcript Elongation by RNA Polymerase II. Annual Review of Biochemistry *79*, 271–293.

Serdar, L.D., Whiteside, D.L., and Baker, K.E. (2016). ATP hydrolysis by UPF1 is required for efficient translation termination at premature stop codons. Nature Communications *7*.

Sheridan, R.M., Fong, N., D'Alessandro, A., and Bentley, D.L. (2019). Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5′ Pause Release, Termination, and Transcription Elongation Rate. Molecular Cell *73*, 107-118.e4.

Shi, Y.G., and Tsukada, Y. -i. (2013). The Discovery of Histone Demethylases. Cold Spring Harbor Perspectives in Biology *5*, a017947–a017947.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. Cell *119*, 941–953.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic Remodeling of Individual Nucleosomes Across a Eukaryotic Genome in Response to Transcriptional Perturbation. PLoS Biology *6*, e65.

Shore, D., Stillman, D.J., Brand, A.H., and Nasmyth, K.A. (1987). Identification of silencer binding proteins from yeast: possible roles in SIR control and DNA replication. EMBO J. *6*, 461–467.

Shukla, A., and Bhaumik, S.R. (2007). H2B-K123 ubiquitination stimulates RNAPII elongation independent of H3-K4 methylation. Biochemical and Biophysical Research Communications *359*, 214–220.

Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C., Sharp, P.A., Giallourakis, C.C., et al. (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proceedings of the National Academy of Sciences *110*, 2876–2881.

Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. Current Opinion in Cell Biology *21*, 344–351.

Silve, S., Rhode, P.R., Coll, B., Campbell, J., and Poyton, R.O. (1992). ABF1 is a phosphoprotein and plays a role in carbon source control of COX6 transcription in Saccharomyces cerevisiae. Molecular and Cellular Biology *12*, 4197–4208.

Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A., and Arndt, K.M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. The EMBO Journal *22*, 1846–1856.

Smith, J.E., Alvarez-Dominguez, J.R., Kline, N., Huynh, N.J., Geisler, S., Hu, W., Coller, J., and Baker, K.E. (2014). Translation of small open reading frames within unannotated RNA transcripts in Saccharomyces cerevisiae. Cell Rep *7*, 1858–1866.

Smolle, M., Venkatesh, S., Gogol, M.M., Li, H., Zhang, Y., Florens, L., Washburn, M.P., and Workman, J.L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. Nature Structural & Molecular Biology *19*, 884–892.

Somesh, B.P., Reid, J., Liu, W.-F., Søgaard, 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.

Somesh, B.P., Sigurdsson, S., Saeki, H., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2007). Communication between Distant Sites in RNA Polymerase II through Ubiquitylation Factors and the Polymerase CTD. Cell *129*, 57–68.

Spain, M.M., Ansari, S.A., Pathak, R., Palumbo, M.J., Morse, R.H., and Govind, C.K. (2014). The RSC Complex Localizes to Coding Sequences to Regulate Pol II and Histone Occupancy. Molecular Cell *56*, 653–666.

Steinmetz, E.J., and Brow, D.A. (1998). 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. Proceedings of the National Academy of Sciences *95*, 6699–6704.

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.

Stockdale, C., Flaus, A., Ferreira, H., and Owen-Hughes, T. (2006). Analysis of Nucleosome Repositioning by Yeast ISWI and Chd1 Chromatin Remodeling Complexes. J. Biol. Chem. *281*, 16279–16288.

Sun, Z.-W., and Allis, C.D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature *418*, 104–108.

Sun, Z., Tessmer, A., and Hampsey, M. (1996). Functional interaction between TFIIB and the Rpb9 (Ssu73) subunit of RNA polymerase II in Saccharomyces cerevisiae. Nucleic Acids Research *24*, 2560–2566.

Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R.R., Grimmond, S.M., Schroder, K., et al. (2009). Tiny RNAs associated with transcription start sites in animals. Nature Genetics *41*, 572–578.

Tanny, J.C., Erdjument-Bromage, H., Tempst, P., and Allis, C.D. (2007). Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. Genes & Development *21*, 835–847.

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. Molecular Cell *23*, 853–864.

Thiebaut, M., Colin, J., Neil, H., Jacquier, A., Séraphin, B., Lacroute, F., and Libri, D. (2008). Futile Cycle of Transcription Initiation and Termination Modulates the Response to Nucleotide Shortage in S. cerevisiae. Molecular Cell *31*, 671–682.

Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome Research *18*, 1084–1091.

Tirosh, I., Sigal, N., and Barkai, N. (2010). Widespread remodeling of mid-coding sequence nucleosomes by Isw1. Genome Biology *11*, R49.

Tomar, R.S., Zheng, S., Brunke-Reese, D., Wolcott, H.N., and Reese, J.C. (2008). Yeast Rap1 contributes to genomic integrity by activating DNA damage repair genes. The EMBO Journal *27*, 1575–1584.

Tomar, R.S., Psathas, J.N., Zhang, H., Zhang, Z., and Reese, J.C. (2009). A Novel Mechanism of Antagonism between ATP-Dependent Chromatin Remodeling Complexes Regulates RNR3 Expression. Molecular and Cellular Biology *29*, 3255–3265.

Tsankov, A., Yanagisawa, Y., Rhind, N., Regev, A., and Rando, O.J. (2011). Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. Genome Res. *21*, 1851–1862.

Tsankov, A.M., Thompson, D.A., Socha, A., Regev, A., and Rando, O.J. (2010). The Role of Nucleosome Positioning in the Evolution of Gene Regulation. PLOS Biology *8*, e1000414.

Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., et al. (2014). Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. Molecular Cell *55*, 467–481.

Tudek, A., Candelli, T., and Libri, D. (2015). Non-coding transcription by RNA polymerase II in yeast: Hasard or nécessité? Biochimie *117*, 28–36.

Tudek, A., Schmid, M., Makaras, M., Barrass, J.D., Beggs, J.D., and Jensen, T.H. (2018). A Nuclear Export Block Triggers the Decay of Newly Synthesized Polyadenylated RNA. Cell Reports *24*, 2457-2467.e7.

Udugama, M., Sabri, A., and Bartholomew, B. (2011). The INO80 ATP-Dependent Chromatin Remodeling Complex Is a Nucleosome Spacing Factor. Molecular and Cellular Biology *31*, 662–673.

Uhler, J.P., Hertel, C., and Svejstrup, J.Q. (2007). A role for noncoding transcription in activation of the yeast PHO5 gene. PNAS *104*, 8011–8016.

Ursic, D., Chinchilla, K., Finkel, J.S., and Culbertson, M.R. (2004). Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. Nucleic Acids Research *32*, 2441–2452.

Uwimana, N., Collin, P., Jeronimo, C., Haibe-Kains, B., and Robert, F. (2017). Bidirectional terminators in Saccharomyces cerevisiae prevent cryptic transcription from invading neighboring genes. Nucleic Acids Research *45*, 6417–6426.

Vakirlis, N., Hebert, A.S., Opulente, D.A., Achaz, G., Hittinger, C.T., Fischer, G., Coon, J.J., and Lafontaine, I. (2018). A Molecular Portrait of De Novo Genes in Yeasts. Mol. Biol. Evol. *35*, 631–645.

Valerius, O., Brendel, C., Düvel, K., and Braus, G.H. (2002). Multiple Factors Prevent Transcriptional Interference at the Yeast *ARO4-HIS7* Locus. Journal of Biological Chemistry *277*, 21440–21445.

Vanácová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol. *3*, e189.

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.

Vary, J.C., Gangaraju, V.K., Qin, J., Landel, C.C., Kooperberg, C., Bartholomew, B., and Tsukiyama, T. (2003). Yeast Isw1p Forms Two Separable Complexes In Vivo. Molecular and Cellular Biology *23*, 80–91.

Vasiljeva, L., and Buratowski, S. (2006). Nrd1 Interacts with the Nuclear Exosome for 3′ Processing of RNA Polymerase II Transcripts. Molecular Cell *21*, 239–248.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1–Nab3–Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nature Structural & Molecular Biology *15*, 795–804.

Venkatesh, S., Li, H., Gogol, M.M., and Workman, J.L. (2016). Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. Nat Commun *7*, 13610.

Viphakone, N., Voisinet-Hakil, F., and Minvielle-Sebastia, L. (2008). Molecular dissection of mRNA poly(A) tail length control in yeast. Nucleic Acids Research *36*, 2418–2433.

Wade, J.T., Hall, D.B., and Struhl, K. (2004). The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes. Nature *432*, 1054–1058.

Wang, Q., and Donze, D. (2016). Transcription factor Reb1 is required for proper transcriptional start site usage at the divergently transcribed TFC6-ESC2 locus in Saccharomyces cerevisiae. Gene *594*, 108–116.

Wang, A., Kurdistani, S.K., and Grunstein, M. (2002). Requirement of Hos2 Histone Deacetylase for Gene Activity in Yeast | Science.

Wang, F., Ranjan, A., Wei, D., and Wu, C. (2016). Comment on “A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme.” Science *353*, 358–358.

Wang, H., Nicholson, P.R., and Stillman, D.J. (1990). Identification of a Saccharomyces cerevisiae DNA-binding protein involved in transcriptional regulation. Molecular and Cellular Biology *10*, 1743–1753.

Warfield, L., Ramachandran, S., Baptista, T., Devys, D., Tora, L., and Hahn, S. (2017). Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. Molecular Cell *68*, 118-129.e5.

Watanabe, S., and Peterson, C.L. (2016). Response to Comment on "A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme." Science *353*, 358–358.

Watanabe, S., Radman-Livaja, M., Rando, O.J., and Peterson, C.L. (2013). A Histone Acetylation Switch Regulates H2A.Z Deposition by the SWR-C Remodeling Enzyme. Science *340*, 195–199.

Weake, V.M., and Workman, J.L. (2008). Histone Ubiquitination: Triggering Gene Activity. Molecular Cell *29*, 653–663.

Wei, W., Pelechano, V., Järvelin, A.I., and Steinmetz, L.M. (2011). Functional consequences of bidirectional promoters. Trends in Genetics *27*, 267–276.

Wellinger, R.J., and Zakian, V.A. (2012). Everything you ever wanted to know about Saccharomyces cerevisiae telomeres: beginning to end. Genetics *191*, 1073–1105.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522–525.

Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature *450*, 1031–1035.

Wilson, B.A., and Masel, J. (2011). Putatively Noncoding Transcripts Show Extensive Association with Ribosomes. Genome Biol Evol *3*, 1245–1252.

Wilson, M.D., Harreman, M., and Svejstrup, J.Q. (2013). Ubiquitylation and degradation of elongating RNA polymerase II: The last resort. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1829*, 151–157.

Wiltshire, S., Raychaudhuri, S., and Eisenberg, S. (1997). An Abf1p C-terminal region lacking transcriptional activation potential stimulates a yeast origin of replication. Nucleic Acids Res. *25*, 4250–4256.

Wlotzka, W., Kudla, G., Granneman, S., and Tollervey, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts: Targets for nuclear surveillance. The EMBO Journal *30*, 1790–1803.

Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. Molecular Cell *54*, 601–612.

Woo, H., Dam Ha, S., Lee, S.B., Buratowski, S., and Kim, T. (2017). Modulation of gene expression dynamics by co-transcriptional histone methylations. Experimental & Molecular Medicine *49*, e326–e326.

Wotton, D., and Shore, D. (1997). A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in Saccharomyces cerevisiae. Genes Dev. *11*, 748–760.

Wu, A.C.K., Patel, H., Chia, M., Moretto, F., Frith, D., Snijders, A.P., and van Werven, F.J. (2018). Repression of Divergent Noncoding Transcription by a Sequence-Specific Transcription Factor. Molecular Cell *72*, 942-954.e7.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, 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.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes & Development *17*, 654–663.

Xiao, T., Kao, C.-F., Krogan, N.J., Sun, Z.-W., Greenblatt, J.F., Osley, M.A., and Strahl, B.D. (2005). Histone H2B Ubiquitylation Is Associated with Elongating RNA Polymerase II. Molecular and Cellular Biology *25*, 637–651.

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.

Xue, Y., Van, C., Pradhan, S.K., Su, T., Gehrke, J., Kuryan, B.G., Kitada, T., Vashisht, A., Tran, N., Wohlschlegel, J., et al. (2015). The Ino80 complex prevents invasion of euchromatin into silent chromatin. Genes & Development *29*, 350–355.

Xue, Y., Pradhan, S.K., Sun, F., Chronis, C., Tran, N., Su, T., Van, C., Vashisht, A., Wohlschlegel, J., Peterson, C.L., et al. (2017). Mot1, Ino80C, and NC2 Function Coordinately to Regulate Pervasive Transcription in Yeast and Mammals. Molecular Cell *67*, 594-607.e4.

Yadon, A.N., Mark, D.V. de, Basom, R., Delrow, J., Whitehouse, I., and Tsukiyama, T. (2010). Chromatin Remodeling around Nucleosome-Free Regions Leads to Repression of Noncoding RNA Transcription. Molecular and Cellular Biology *30*, 5110–5122.

Yan, C., Chen, H., and Bai, L. (2018). Systematic Study of Nucleosome-Displacing Factors in Budding Yeast. Molecular Cell *71*, 294-305.e4.

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene *389*, 52–65.

Yarragudi, A., Miyake, T., Li, R., and Morse, R.H. (2004). Comparison of ABF1 and RAP1 in Chromatin Opening and Transactivator Potentiation in the Budding Yeast Saccharomyces cerevisiae. Molecular and Cellular Biology *24*, 9152–9164.

Yarragudi, A., Parfrey, L.W., and Morse, R.H. (2007). Genome-wide analysis of transcriptional dependence and probable target sites for Abf1 and Rap1 in Saccharomyces cerevisiae. Nucleic Acids Res *35*, 193–202.

Yarrington, R.M., Richardson, S.M., Lisa Huang, C.R., and Boeke, J.D. (2012). Novel Transcript Truncating Function of Rap1p Revealed by Synthetic Codon-Optimized Ty1 Retrotransposon. Genetics *190*, 523–535.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T., and Pugh, B.F. (2012). Genome-wide Nucleosome Specificity and Directionality of Chromatin Remodelers. Cell *149*, 1461–1473.

Yen, K., Vinayachandran, V., and Pugh, B.F. (2013). SWR-C and INO80 Chromatin Remodelers Recognize Nucleosome-free Regions Near +1 Nucleosomes. Cell *154*, 1246–1256.

Yu, L., and Morse, R.H. (1999). Chromatin Opening and Transactivator Potentiation by RAP1 in Saccharomyces cerevisiae. Molecular and Cellular Biology *19*, 5279–5288.

Yu, L., Sabet, N., Chambers, A., and Morse, R.H. (2001). The N-terminal and C-terminal Domains of RAP1 Are Dispensable for Chromatin Opening and GCN4-mediated HIS4 Activation in Budding Yeast. J. Biol. Chem. *276*, 33257–33264.

Yu, Q., Qiu, R., Foland, T.B., Griesen, D., Galloway, C.S., Chiu, Y.-H., Sandmeier, J., Broach, J.R., and Bi, X. (2003). Rap1p and other transcriptional regulators can function in defining distinct domains of gene expression. Nucleic Acids Res *31*, 1224–1233.

Yuan, G.-C. (2005). Genome-Scale Identification of Nucleosome Positions in S. cerevisiae. Science *309*, 626–630.

Zhang, Z., and Dietrich, F.S. (2005). Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE. Nucleic Acids Research *33*, 2838–2851.

Zhang, Z., and Gilmour, D.S. (2006). Pcf11 Is a Termination Factor in Drosophila that Dismantles the Elongation Complex by Bridging the CTD of RNA Polymerase II to the Nascent Transcript. Molecular Cell *21*, 65–74.

Zhang, H., Rigo, F., and Martinson, H.G. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Molecular Cell *59*, 437–448.

Zhang, Z., Fu, J., and Gilmour, D.S. (2005). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. Genes & Development *19*, 1572–1580.

APPENDICES

OPEN

ARTICLE

Citation: Cell Discovery (2017) **3**, 17040; doi:10.1038/celldisc.2017.40
www.nature.com/celldisc

Binding to RNA regulates Set1 function

Pierre Luciano[1,5], Jongcheol Jeon[2,5], Abdessamad El-kaoutari[1,5], Drice Challal[3,5], Amandine Bonnet[4], Mara Barucco[3], Tito Candelli[3], Frederic Jourquin[1], Pascale Lesage[4], Jaehoon Kim[2,*], Domenico Libri[3,*], Vincent Géli[1,*]

[1]Marseille Cancer Research Center (CRCM), Aix Marseille University, Institut Paoli-Calmettes. Equipe labellisée Ligue, Marseille, France; [2]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon, South Korea; [3]Institut Jacques Monod, Univ Paris Diderot, Sorbonne Paris Cité, CNRS, Bâtiment Buffon, Paris Cedex, France; [4]Université Paris Diderot, Sorbonne Paris Cité, INSERM U944, CNRS UMR 7212, Institut Universitaire d'Hématologie, Hôpital St. Louis, Paris, France

The Set1 family of histone H3 lysine 4 (H3K4) methyltransferases is highly conserved from yeast to human. Here we show that the Set1 complex (Set1C) directly binds RNA *in vitro* through the regions that comprise the double RNA recognition motifs (dRRM) and N-SET domain within Set1 and its subunit Spp1. To investigate the functional relevance of RNA binding, we performed UV RNA crosslinking (CRAC) for Set1 and RNA polymerase II in parallel with ChIP-seq experiments. Set1 binds nascent transcripts through its dRRM. RNA binding is important to define the appropriate topology of Set1C distribution along transcription units and correlates with the efficient deposition of the H3K4me3 mark. In addition, we uncovered that Set1 binds to different classes of RNAs to levels that largely exceed the levels of binding to the general population of transcripts, suggesting the Set1 persists on these RNAs after transcription. This class includes RNAs derived from *SET1*, Ty1 retrotransposons, specific transcription factors genes and snRNAs (small nuclear RNAs). We propose that Set1 modulates adaptive responses, as exemplified by the post-transcriptional inhibition of Ty1 retrotransposition.

***Keywords*:** Set1; transcription; RNA binding; H3K4 methylation
Cell Discovery (2017) **3,** 17040; doi:10.1038/celldisc.2017.40; published online 24 October 2017

Introduction

Highly conserved histone proteins undergo several types of covalent modifications including acetylation, methylation, phosphorylation, ubiquitylation, SUMOylation, citrullination and ADP-ribosylation [1]. These modifications that are deposited and removed by specific chromatin-modifying enzymes can either directly alter the chromatin architecture or create docking sites that facilitate the binding of specific domains present in chromatin readers [2]. These readers in turn recruit chromatin remodeling enzymes or additional chromatin modifiers to shape chromatin landscapes that regulate DNA accessibility [3]. Among these marks, methylation of lysine 4 on histone H3 (H3K4) has aroused considerable interest [4]. In mammals, this modification is catalyzed by at least six different complexes that differ by their catalytic SET domain subunit (Set1a, Set1b, Mll1, Mll2, Mll3 and Mll4) [5] but share a protein module comprises WDR5, RbBP5, ASH2L and DPY-30, which binds to the catalytic SET domain and stimulate H3K4 methyltransferase activity [6]. Each complex contains additional factors specifying their recruitment to chromatin and their biological effect [7].

In *Saccharomyces cerevisiae*, all H3K4 methylation is carried out by a complex called COMPASS (for complex of proteins associated with Set1) or Set1C (for Set1 Complex) [8]. The catalytic subunit Set1 acts as a

[5]These authors contributed equally to this work.
*Correspondence: Jaehoon Kim
Tel: +82 42 350 2632;
Fax: 82 42 350 2610;
E-mail: kimjaehoon@kaist.edu
or Domenico Libri
Tel: +33 1 57 27 80 65;
E-mail: domenico.libri@ijm.fr
or Vincent Géli
Tel: +33 4 86 97 74 01;
E-mail: vincent.geli@inserm.fr
Received 24 July 2017; accepted 22 September 2017

scaffold for seven other components (Swd1 (mammalian homolog RbBP5), Swd2 (WDR82), Swd3 (WDR5), Bre2 (ASHL2), Sdc1 (DPY-30), Spp1 (CFP1) and Shg1 (BOD1)) [9]. Swd1, Swd3, Bre2 and Sdc1 associate with the SET domain of Set1 to form the SETc that is minimally sufficient to methylate free H3 *in vitro* [9], whereas Spp1 and Shg1 directly associate to Set1 by binding to the N-SET domain and the Set1 central region, respectively [10]. The loss of individual Set1C subunits differentially affects Set1 stability, complex integrity, the pattern of global H3K4 methylation and the distribution of H3K4 methylation marks along active genes [11]. The WD40 repeat protein Swd2 is the only essential subunit of Set1C and its depletion strongly affects Set1 stability and H3K4 methylation [9]. Swd2 also belongs to the APT complex (for '*associated with Pta1*'), which is part of the cleavage polyadenylation factor [12]. Several studies suggested a functional link between Set1C and 3′-end formation/termination [13] but it remains unknown how the binding of Swd2 to either of the two complexes (Set1C and APT complex) is regulated. Other regions outside of the SET domain have been reported to regulate Set1 catalytic activity, including the N-SET domain, the double RNA recognition motif (dRRM) and a centrally located auto-inhibitory domain, but the mechanism underlying such regulation is still elusive [14].

Genome-wide studies in yeast indicate that active transcription is characteristically accompanied by histone H3K4 trimethylation (H3K4me3) at the 5′-end of genes and by H3K4 di- and monomethylation (H3K4me2 and H3K4me1) at downstream nucleosomes [15]. H3K4me3 can also be found at the 3′-end of a number of genes most likely reflecting the presence of antisense ncRNAs (non-coding RNAs) [16]. These H3K4 methylation patterns correlate with Set1 occupancy that is higher at the 5′-end of coding regions of highly transcribed RNA polymerase II (RNAPII) genes and decreases at more distal nucleosomes [17]. Set1 has been reported to associate with the elongation complex in the early stages of the transcription cycle, which is thought to contribute to the prevalence of H3K4me3 at the 5′-end of active genes. Recruitment occurs when the carboxyl terminal domain of RNAPII is preferentially phosphorylated at the serine in the fifth position (Ser5) of its heptad repeats, which has been reported to depend on the Paf1 complex [18]. However, direct interactions that underpin the recruitment of Set1C to actively transcribed genes remain to be characterized. Although interaction of Set1C with chromatin was proposed to be mediated by the interaction of Swd2 with ubiquitylated H2B (H2Bub) [19], this model has been challenged by *in vitro* reconstitution experiments showing that the Swd2-deficient Set1C can methylate chromatinized H3K4 in an H2B ubiquitylation-dependent manner [20]. Thus, current models to explain Set1 recruitment and the establishment of H3K4 methylation along genes *in vivo* still need to be improved.

Set1 contains two tandem RRMs, RRM1 and RMM2 (dRRM) [14]. We previously reported that Set1 RRM1 contains the canonical RRM-fold but lacks some typical RNA-binding features. Consistently, RRM1 is necessary but not sufficient for Set1 to bind RNA *in vitro* and RRM2 was also shown to be required [21]. Deletion or mutation of RRM1 has been shown to lead to decreased H3K4me3 in the 5′ regions of active genes along with an increase in H3K4me2 [14], opening the possibility that a potential RNA-binding activity of Set1 could regulate Set1 occupancy and/or the distribution of H3K4 methylation [22].

Here, we show that Set1 binds directly RNA and that its dRRM and N-SET, as well as Spp1, contribute to Set1 RNA binding *in vitro* in the context of a reconstituted Set1C. By combining ChIP-seq and CRAC experiments of Set1 and Set1 mutants that have lost the ability to bind RNA, we show that Set1 RNA-binding activity mediated by its dRRM does not affect Set1 recruitment to chromatin per se but maintains Set1 in the 5′ region of genes. We propose that RNA binding to Set1 increases the time of residency of Set1C in the proximity of chromatin allowing additional time for H3K4 trimethylation in the 5′-end of genes. Our results also indicate that Set1 strongly associates, presumably post-transcriptionally to transcripts produced by specific classes of genes, including snRNAs small nuclear RNAs, Ty1 and adaptive response genes. In particular, we show that Ty1 retrotransposition is negatively regulated by Set1 at a post-transcriptional level.

Results

Binding of RNA in vitro by reconstituted Set1C involves the dRRM and N-SET domains of Set1

Our previous results suggested that purified Set1 RRM1-RRM2 (dRRM) binds RNA *in vitro* [21]. We reconstituted the whole Set1C [20] to analyze Set1 RNA binding in the context of the complex form of Set1 with its associated subunits. Set1C was purified from insect cells expressing FLAG-fused full-length or truncated Set1 together with the seven other subunits (Figure 1a and b). Purified complexes lacking specific

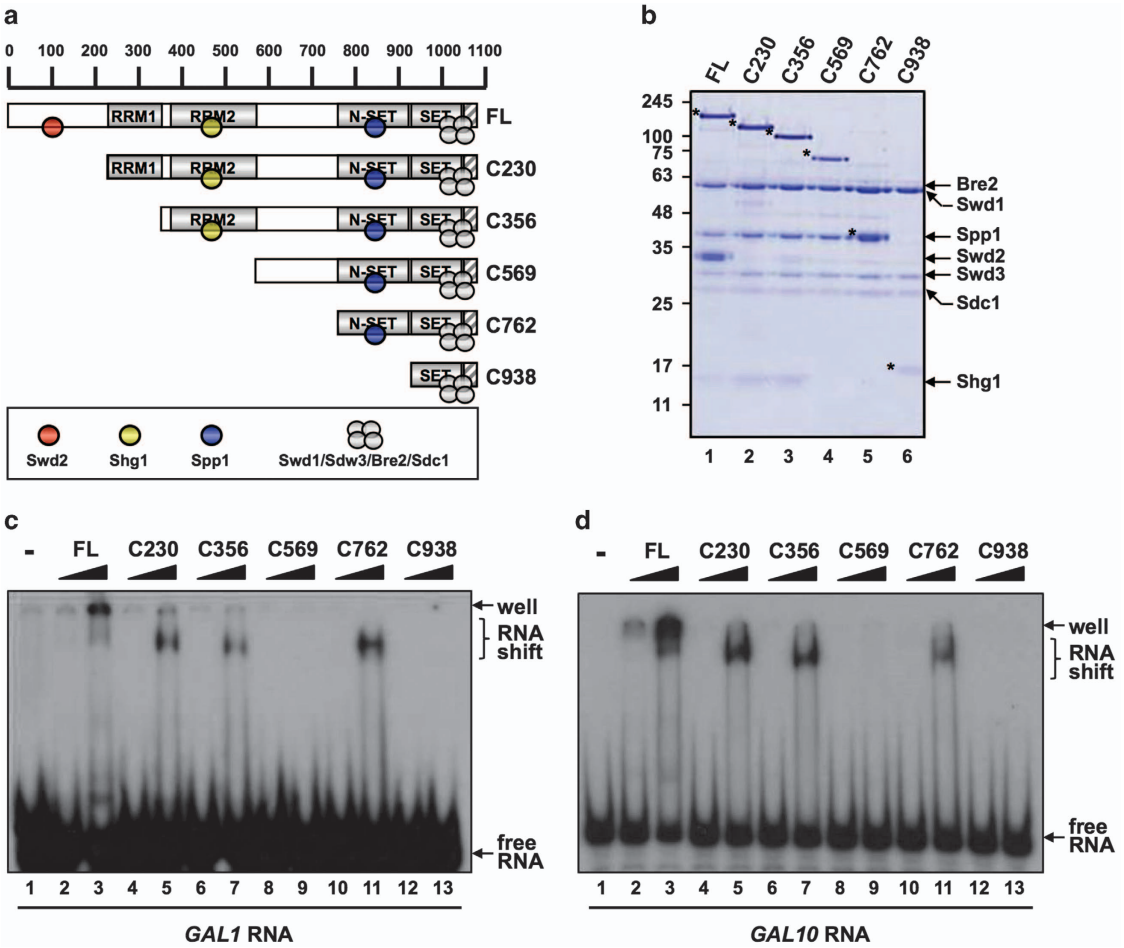


Figure 1 Purified Set1C binds directly to RNA *in vitro.* (**a**) A schematic diagram of Set1 and derived fragments with predicted RRM1, RRM2, N-SET and post-SET (hatched box) domains and associated subunits. FL indicates full-length. (**b**) Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and Coomassie blue staining of purified Set1Cs reconstituted with baculoviruses expressing FLAG-Set1 or FLAG-Set1 fragments and untagged subunits. FLAG-Set1 polypeptides are marked by asterisks. (**c, d**) Radiolabeled GAL1 (**c**) and GAL10 (**d**) transcripts were subjected to *in vitro* RNA electrophoretic mobility shift assays with 0.5 (lanes 2, 4, 6, 8, 10 and 12) or 2.5 (lanes 3, 5, 7, 9, 11 and 13) pmoles of indicated Set1Cs.

domains of Set1 were incubated with *in vitro* transcribed and purified *GAL1* and *GAL10* mRNAs. Interaction was probed by electrophoretic mobility shift assay (Figure 1c and d). We found that reconstituted Set1C was able to directly bind *GAL1* and *GAL10* mRNAs *in vitro*. In agreement with our previous results, a truncated Set1 lacking the two RRM motifs (C569) was unable to bind RNA. Surprisingly, we found that the C762 fragment encompassing the N-SET and the SET domains (Figure 1a) was able to bind RNA in contrast to C569 and C938 that could not. These results suggest that N-SET contributes to Set1 RNA binding, which might be inhibited by the region between residues 569 and 762 (Figure 1c and d).

To further confirm these results, we introduced in the full-length Set1 the $Y_{271}F_{272}$/AA mutation previously shown to decrease the RNA-binding activity of dRRM *in vitro* [21], as well as a deletion of the dRRM, or the N-SET domain, and a combination of these mutations [21] (Figure 2a and b). Consistent with Figure 1, either the $Y_{271}F_{272}$/AA mutation or dRRM deletion strongly affected Set1C RNA-binding *in vitro* (Figure 2c and d). Deleting only the N-SET domain also significantly decreased RNA binding suggesting that the N-SET domain is important for RNA binding (Figure 2c and d). However, we did not detect interaction of the N-SET domain alone with RNA (data not shown) indicating that the N-SET domain is not sufficient to bind RNA. As expected, combining alterations of the dRRM with the N-SET deletion abolished Set1 RNA-binding activity *in vitro*. We next assessed whether Set1C subunits contribute by themselves to the

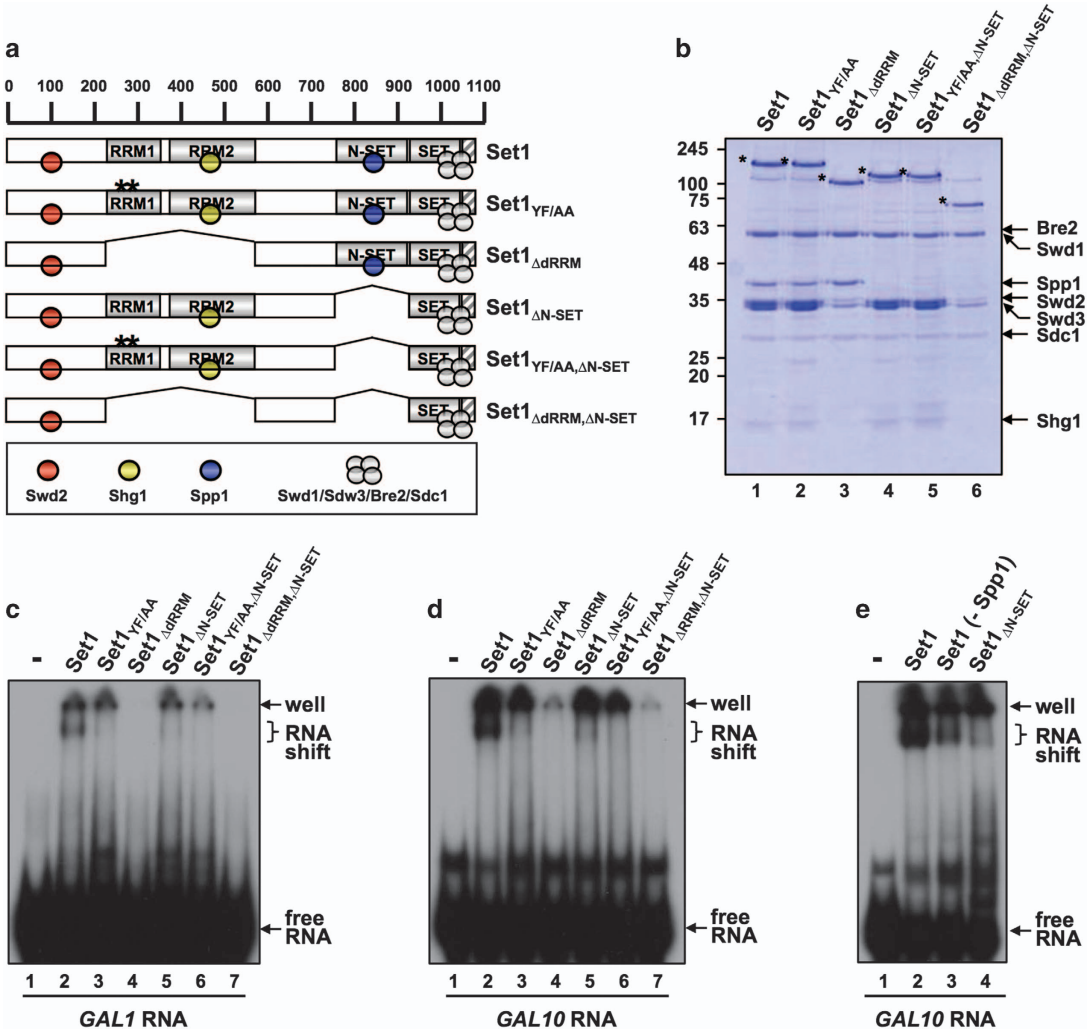


Figure 2 Set1C RNA binding requires dRRM, the N-SET domain, and Spp1. (**a**) A schematic representation of Set1, Set1$_{YF/AA}$ and Set1 deletion mutants with predicted RRM1, RRM2, N-SET and post-SET (hatched box) domains and associated subunits. (**b**) Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and Coomassie blue staining of purified Set1Cs reconstituted with baculoviruses expressing FLAG-Set1 or FLAG-Set1 fragments and untagged subunits. (**c**, **d**) Radiolabeled *GAL1* (**c**) and *GAL10* (**d**) transcripts were subjected to *in vitro* RNA electrophoretic mobility shift assays with indicated Set1Cs. (**e**) Binding of Set1C and Set1C lacking Spp1 to radiolabeled GAL10. Set1C containing Set1$_{\Delta N\text{-}SET}$ is also shown.

RNA-binding activity of Set1C by monitoring the RNA-binding activity of each subunit. None of the Set1C subunits were found to bind *GAL10* mRNA (Supplementary Figure S1A and B). Finally, as the N-SET domain binds Spp1, we asked whether only omitting Spp1 in Set1C reconstitution also affected RNA-binding activity. The results shown in Figure 2e indicated that Spp1 was required for Set1C to bind RNA despite the fact that Spp1 by itself does not bind RNA.

Collectively, these results show that the fully reconstituted Set1C has the ability to bind mRNAs *in vitro*. Unexpectedly, not only the dRRM motif but also the N-SET domain and Spp1 contributed to the ability of Set1 to bind RNA. Therefore, Set1C RNA binding requires the presence of multiple protein surfaces comprising the dRRM, as well as N-SET domain and Spp1.

Altering Set1 dRRM affects Set1 distribution along genes

Before addressing Set1 binding to RNA *in vivo,* we performed ChIP-seq analysis to determine the genome-wide occupancy of Set1 and Set1 mutants. ChIP-seq

experiments were carried out from *set1*Δ cells expressing N-terminal tagged (Z-tag-Tev-6HIS) version of Set1 (PTH-Set1) [23]. Because in cells grown in SC-LEU-TRP medium, the amount of PTH-Set1 was similar to that of endogenous Set1 (Supplementary Figure S2A) we used these conditions for ChIP-seq and all subsequent experiments.

To address the importance of the Set1 domains involved in RNA-binding *in vitro* for chromatin binding, we performed ChIP-seq experiments of PTH-Set1 and of its mutant forms (set1$_{YF/AA}$, set1$_{\Delta dRRM}$, set1$_{\Delta N\text{-}SET}$ and set1$_{\Delta dRRM,\Delta N\text{-}SET}$) with an anti-Set1 mouse monoclonal antibody (anti-Set1 mAb) [24] that recognizes a Set1 epitope between residues 700 and 761 (Supplementary Figure S2B and C). In parallel, we also performed ChIP-seq of a Myc-Set1 strain with an anti-Myc antibody (9E10). The occupancy profiles for PTH-Set1 and Myc-Set1 at selected genes were overall similar and both datasets were highly correlated (Supplementary Figure S3A, B and C). Set1 occupancy was maximum beyond the H3K4me3 peak and upstream of the H3K4me2 peak and was slightly 3′-shifted relative to RNAPII occupancy (Supplementary Figure S3D), in agreement with previous results indicating that Set1 and its subunits are recruited at the 5′ region of active genes transcribed by RNAPII [17].

Set1 protein amount was controlled in the different mutant strains by western blot (Figure 3a). We reproducibly observed a reduction of the PTH-Set1$_{YF/AA}$ amount by about 1.3–1.5-fold in all experiments [21], whereas deleting dRRM increased the stability of Set1. PTH-Set1 was functional, as it supported wild-type levels of H3K4me3 when expressed in a *set1*Δ background. In contrast, H3K4me3 was globally abolished when the dRRM and N-SET functions were compromised (Figure 3b). To evaluate the importance of the Set1 domains characterized *in vitro*, we next assessed the occupancy of PTH-Set1 and mutant forms by ChIP-seq experiments.

Surprisingly, deleting dRRM or mutating it in PTH-Set1$_{YF/AA}$ did not affect the distribution of the individual ChIP signals calculated *per* each mRNA-coding gene (Supplementary Figure S4 for PTH-Set1$_{YF/AA}$, and data not shown). To analyze the chromatin distribution of Set1 and its mutant derivatives in more details, we generated normalized occupancy profiles over large genes, which allows a better spatial resolution of recruitment regions. As shown in Figure 3c, the aggregate signal for Set1$_{YF/AA}$ and Set1$_{\Delta dRRM}$ shows reduced occupancy in the 5′ region of genes relative to wild-type (WT) Set1, which is compensated by an

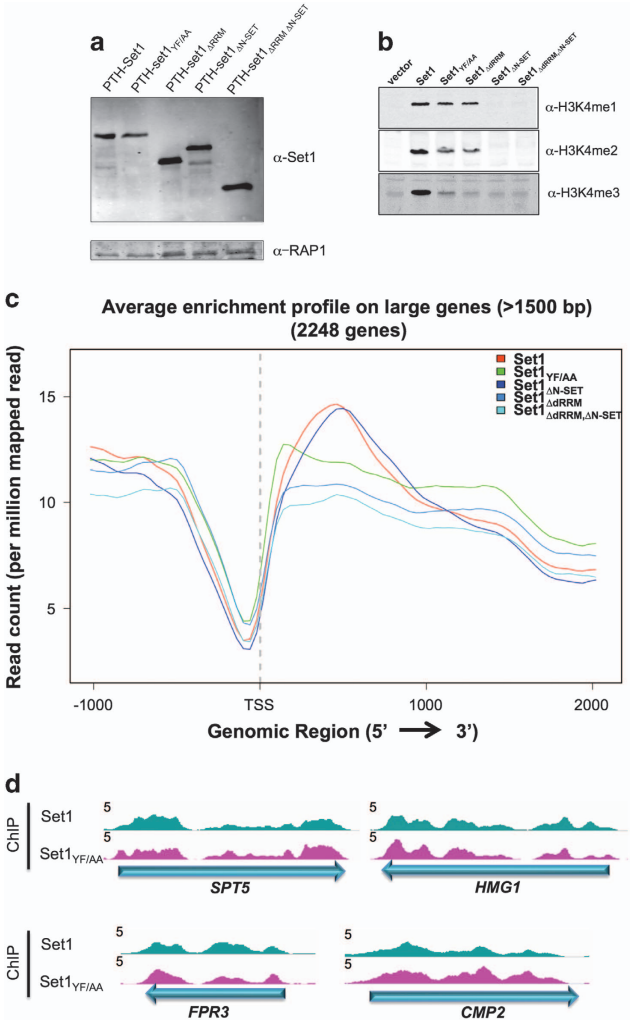


Figure 3 RRM but not N-SET regulates the genome-wide occupancy of Set1. (**a**, **b**) W303 *set1*Δ*::TRP1* pRS415-nHTP-*SET1* (and *SET1* mutant forms) cells were grown in SC -TRP-LEU. Protein levels of Set1 (**a**) and methylated histone H3 (**b**) were verified by western blots using anti-Set1 mAb and anti-H3K4me1, me2, and me3 antibodies, respectively. A Rap1 loading control is shown. (**c**) Average enrichment profiles of PTH-Set1 and PTH-Set1 mutants in genes > 1500 bp. Read counts were normalized to read counts per million of mapped reads. ChIP-seq experiments were performed with the anti-Set1 mAb from *set1*Δ*::TRP1* cells expressing PTH-Set1 and PTH-Set1 mutants from the pRS415-nHTP (grown in SD -TRP -LEU). (**d**) Normalized occupancy profiles of Set1 at the indicated genes. Graphs were normalized to 10 million mapped reads for each mutant.

average increase in the 3′-end of genes to generate the observed unchanged overall signal on a *per gene* basis. This trend is illustrated in Figure 3d for individual genes. Deletion of the N-SET domain had no detectable impact in the distribution of signals (data not shown) or the average profile of the signal (Figure 3c).

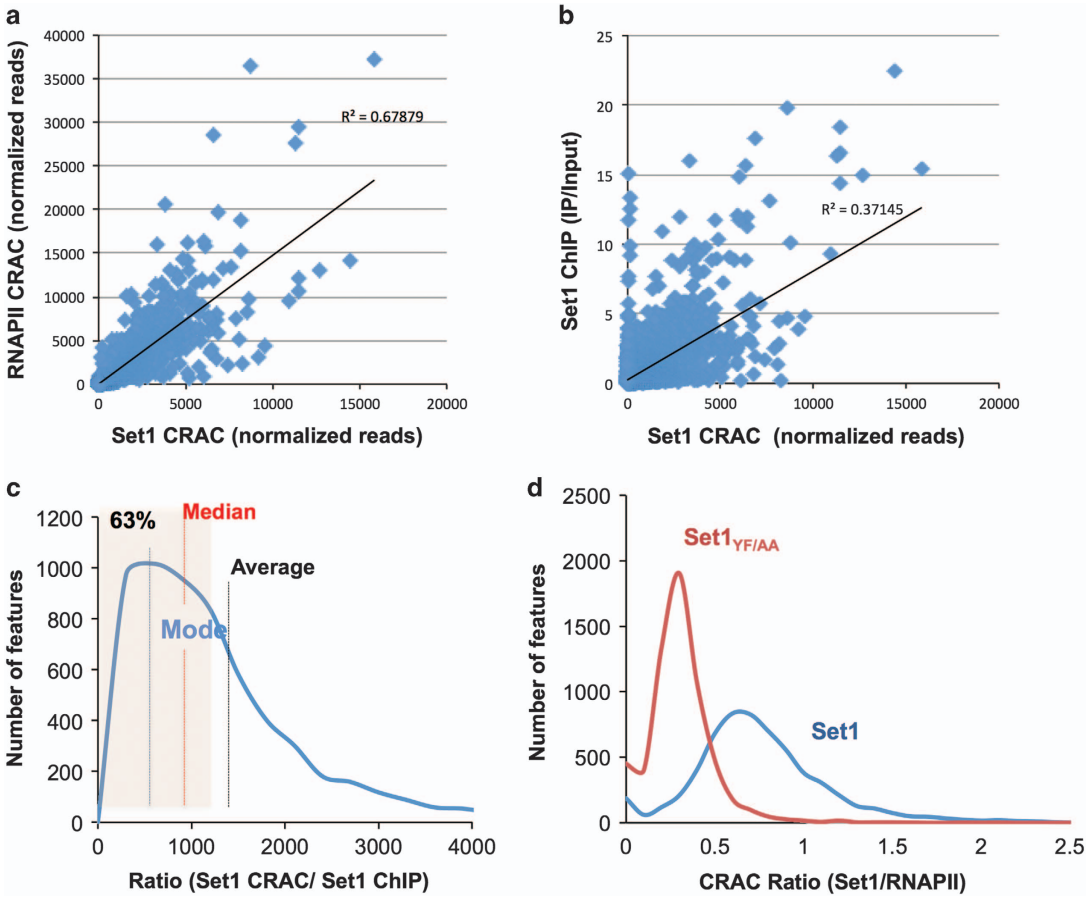


Figure 4 Set1 binding to RNA occurs via two different modes and requires its dRRM. (**a**) Dispersion plot showing high correlation between the Set1 and RNAPII CRAC signals. The determination coefficient (r^2) of the linear, least squares regression is indicated. (**b**) Dispersion plot of Set1 CRAC signals versus Set1 ChIP-Seq (IP/input). (**c**) Distribution of the Set1 CRAC/Set1 ChIP ratios for all mRNAs coding genes. The average, mode and median are indicated to highlight the asymmetry of the distribution. The shaded area corresponds to the percentage of features whose Set1 CRAC/ChIP ratios are included in a range symmetrically positioned around the mode. (**d**) Distribution of the Set1/RNAPII CRAC ratios in cells expressing PTH-Set1 and PTH-Set1$_{YF/AA}$.

These results, together with the strong impact of the Set1$_{YF/AA}$ mutation on *in vivo* RNA binding described below, indicate that the dRRM domain of Set1 is not required for the overall recruitment of Set1 to chromatin, but is essential for its normal distribution along genes. Importantly, the N-SET domain that is essential for Set1C catalytic activity (Figure 3b) and is required to bind RNA *in vitro,* is not involved in the recruitment and positioning of Set1.

Set1 binding to RNA in vivo is determined by both co-transcriptional and post-transcriptional components

We next used the CRAC procedure to analyze *in vivo* the genome-wide RNA binding of Set1 and Set1$_{YF/AA}$ whose RNA-binding activity is compromised *in vitro* to detect *in vivo* RNA-protein interactions [25]. Briefly, tagged RNA-binding proteins are UV crosslinked to their targets *in vivo* and purified by three sequential steps of affinity selection, two of which are under denaturing conditions. The associated RNA is isolated and sequenced. We used the same PTH-Set1 and PTH-*set1* $_{YF/AA}$ constructs and growth conditions used for the ChIP-seq. A non-crosslinked sample was processed in parallel as a control for specificity. A spike in control was generated by adding to the *S. cerevisiae* cultures 0.5% of *S. pombe* cells expressing a non-relevant HTP-tagged protein that binds RNA and that was purified with *S. cerevisiae* Set1. The number of reads mapping to the *S. pombe* genome was used for normalization. We also monitored RNAPII distribution by the same CRAC technique, which provides high-resolution information about the level of transcription.

We identified 2543 mRNAs displaying high-confidence Set1 RNA crosslinking sites for which the number of reads obtained for the crosslinked sample

(Set1 CL) was >5-fold over the no-crosslinked sample (Set1 No-CL) (Supplementary Table S1). Set1 was found to bind mRNAs but also several other transcript classes (Supplementary Figure S5). Set1 binds at least partially during transcription, as witnessed by the significant representation of intronic RNAs in the crosslinked material (see below). For assessing to what extent Set1 binds the RNA during transcription, we sought correlations between the Set1 CRAC signal and the levels of RNAPII occupancy as determined by RNAPII CRAC (Figure 4a) for all mRNA-coding genes. We also compared the Set1 CRAC signal with Set1 recruitment to chromatin as determined by ChIP-Seq (Figure 4b). We reasoned that if binding to the RNA were co-transcriptional, theses datasets should be highly correlated. Consistent with this notion, binding of Set1 to the RNA correlated remarkably well both with RNAPII CRAC (Figure 4a, $r^2 = 0.68$; $P = 3E^{-22}$) and recruitment of Set1 to chromatin as measured by ChIP (Figure 4b, $r^2 = 0.37$; $P = 9E^{-10}$).

The distribution of Set1 CRAC/ChIP ratios was clearly not symmetric as could have been expected for a homogeneous population with random variability (Figure 4c). Rather, it was markedly skewed toward high values (compare the difference between the mode, the median and the average in Figure 4c) with only 63% of the population symmetrically distributed around the mode (shaded area) and the remaining values tailing over a wide range of higher ratios. This suggests the existence of at least two classes of genes: one major, for which the levels of RNA binding relative to chromatin-associated Set1 are relatively homogeneous; the second displaying levels of Set1 binding to RNA that are generally higher than expected based on the sole co-transcriptional interaction. Overall, these analyses strongly suggest that the levels of Set1 binding to the RNA detected by CRAC are generally dominated by a co-transcriptional component but also contain a post-transcriptional component that might predominate for some genes (Supplementary Table S1). Snapshots of the second class of mRNA are shown in Supplementary Figure S6A and B. Interestingly, the feature with the highest Set1 crosslinking signal was the *SET1* mRNA, which is fully consistent with the notion that it associates with the Set1C containing Set1, Swd1, Spp1 and Shg1 during its co-translational assembly [26]. We further sought to determine whether genes encoding this specific class of transcripts (Supplementary Table S1) have physical and/or functional associations. Evidence that many of the proteins encoded by these genes are linked in reliable networks stemmed from computational analysis [27]. Gene ontology analysis revealed that genes whose transcripts were strongly bound by Set1 included some involved in chromosome segregation and many transcription factors (DNA-binding proteins) involved in adaptive responses (Supplementary Figure S6C).

Mutation of the RNA-binding domain in Set1$_{YF/AA}$ led to a marked decrease in the Set1 CRAC signal, which affected uniformly the whole population of Set1 targets as indicated by a general shift in the distribution of Set1/RNAPII CRAC ratios in the *set1*$_{YF/AA}$ mutant relative to WT (Figure 4d). These data demonstrate that the YF/AA mutation in Set1 dramatically affect the interaction of Set1 with mRNAs *in vivo*, although this interaction was not totally abolished but partially maintained with a different topology (see below). Set1 was found to bind with similar *set1*$_{YF/AA}$ dependency stable unannotated transcripts, cryptic unstable transcripts and sno/snRNAs (Figure 5a–c). The first two classes interact with Set1 to a somewhat lower extent even when normalized to the RNAPII CRAC signal, maybe because these RNAs are unstable and the post-transcriptional component might contribute to less to the CRAC signal. Many sno- and snRNAs appear to be bound by Set1 post-transcriptionally, as suggested by the large distribution of Set1/RNAPII CRAC values (Figure 5c) and indicated by the general lack of signal in the regions of the precursor (Figure 5d for the U1, U2 and U4 snRNAs and data not shown). Interestingly, spliceosomal snRNAs were among the strongest binders (Figure 5d), possibly suggesting a role of Set1 in splicing.

Aggregated distribution of Set1 binding on RNAs

We profiled the distribution of RNA-associated Set1 for different features aligned on the transcription start site. As co-transcriptional RNA binding at any given position is likely to be strongly dependent on the level of transcription, we plotted in parallel RNAPII occupancy as defined by the CRAC signal. As shown in Figure 6a, binding of Set1 to the RNA was slightly delayed relative to the appearance of the RNAPII signal (see inset in Figure 6a), which resulted in a relative Set1/RNAPII signal building up in the first 100–250 nt of transcription. This is consistent with the notion that Set1 is recruited co-transcriptionally to the RNA and suggests that it binds the nascent transcript after interacting directly with the polymerase. In the *set1*$_{YF/AA}$ mutant, the CRAC signal was markedly reduced over most of the length of the transcription unit, particularly in the 5′ region.

To assess the distribution of the Set1 signal in the 3′ region of genes we first calculated a positionally

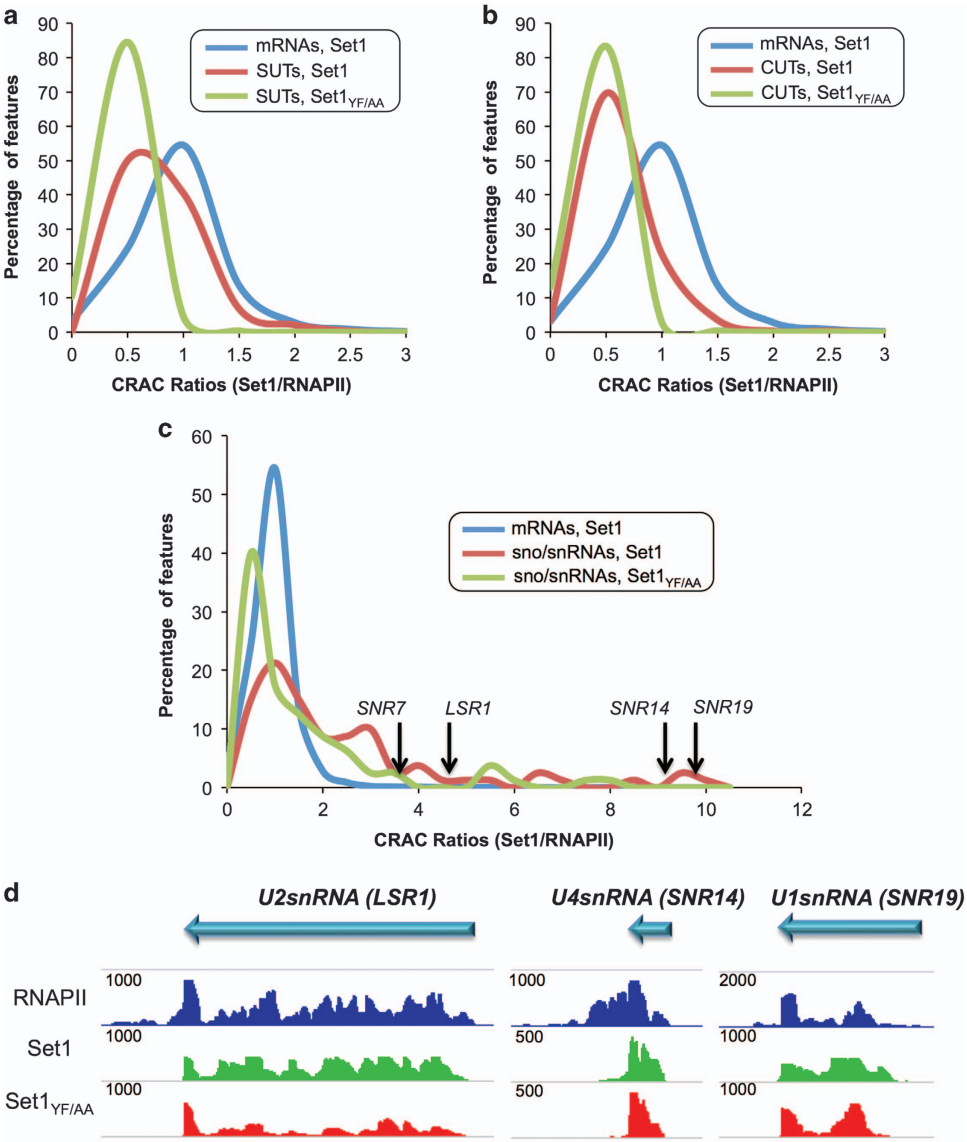


Figure 5 Set1 binds to different classes of RNA. (**a**, **b**, **c**) Normalized distribution of Set1/RNAPII CRAC ratios in cells expressing PTH-Set1 and PTH-Set1$_{YF/AA}$ for the indicated classes of RNA. The position in the distribution of each RNAPII-transcribed snRNAs is indicated in C. The distribution of ratios for mRNAs (Figure 4d) is shown in each graph. (**d**) Snapshots for Set1, Set1$_{YF/AA}$ and PolI CRAC normalized signals for the indicated spliceosomal RNAs.

weighted average p(A) site (wPAS) for every gene (see Materials and Methods section). This was necessary to improve the quality of the alignment in the 3′-end of genes as most genes have multiple polyadenylation sites. To this end, we used T-fill data [28] and assigned to every p(A) addition site a weight depending on the intensity of the signal at that position. This was used to generate a positionally weighted average p(A) site (wPAS). As shown in Figure 6b, the polymerase signal declines in this region, either because of multiple sites of termination or to increased speed. Interestingly, the Set1 signal is maintained and actually slightly increases immediately before the wPAS. Intriguingly, this increase is maintained in the YF/AA mutant, to the point that the signals for the mutant and WT Set1 are identical in this region. This surprising observation indicates that the YF/AA mutation does not affect the binding to the RNA in this region of the transcripts. Whether this 3′ peak is mainly because of the co-transcriptional or post-transcriptional binding of Set1 to the RNA cannot be determined from these experiments. Note that if the 3′ peak were formed co-transcriptionally, its intensity relative to the polymerase signal, which decreases in this region (Figure 6b),

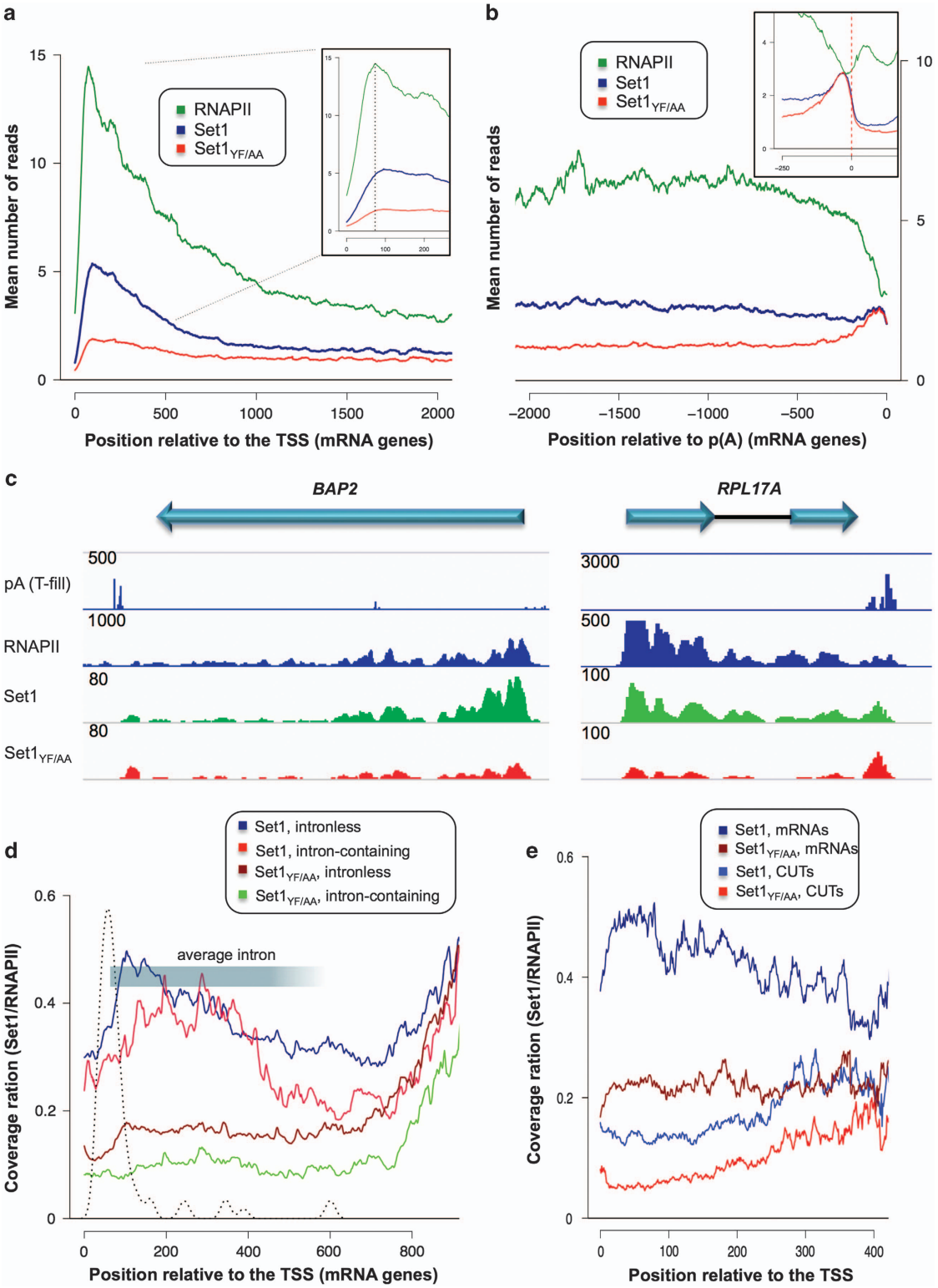


Figure 6 Distribution of Set1-binding sites on RNA. (**a**, **b**) Metagene analysis of Set1 and Set1$_{YF/AA}$ RNA-binding signals on mRNAs as observed by CRAC compared with polymerase occupancy (**a**) features aligned on transcription start site; (**b**) features aligned on wPAS. Insets contain zooms of relevant regions. (**c**) Snapshots for Set1, Set1$_{YF/AA}$, and PolI CRAC normalized signals at the 5′ region of *BAP2* and *RPL17-A* mRNAs. (**d, e**) Metagene analysis of Set1 and Set1$_{YF/AA}$ RNA-binding signals, (**d**) on intronless and intron-containing genes, the distribution of intron position is indicated as well as the average position of introns; (**e**) to cryptic unstable transcripts (CUTs) compared with size matched open reading frame (ORF).

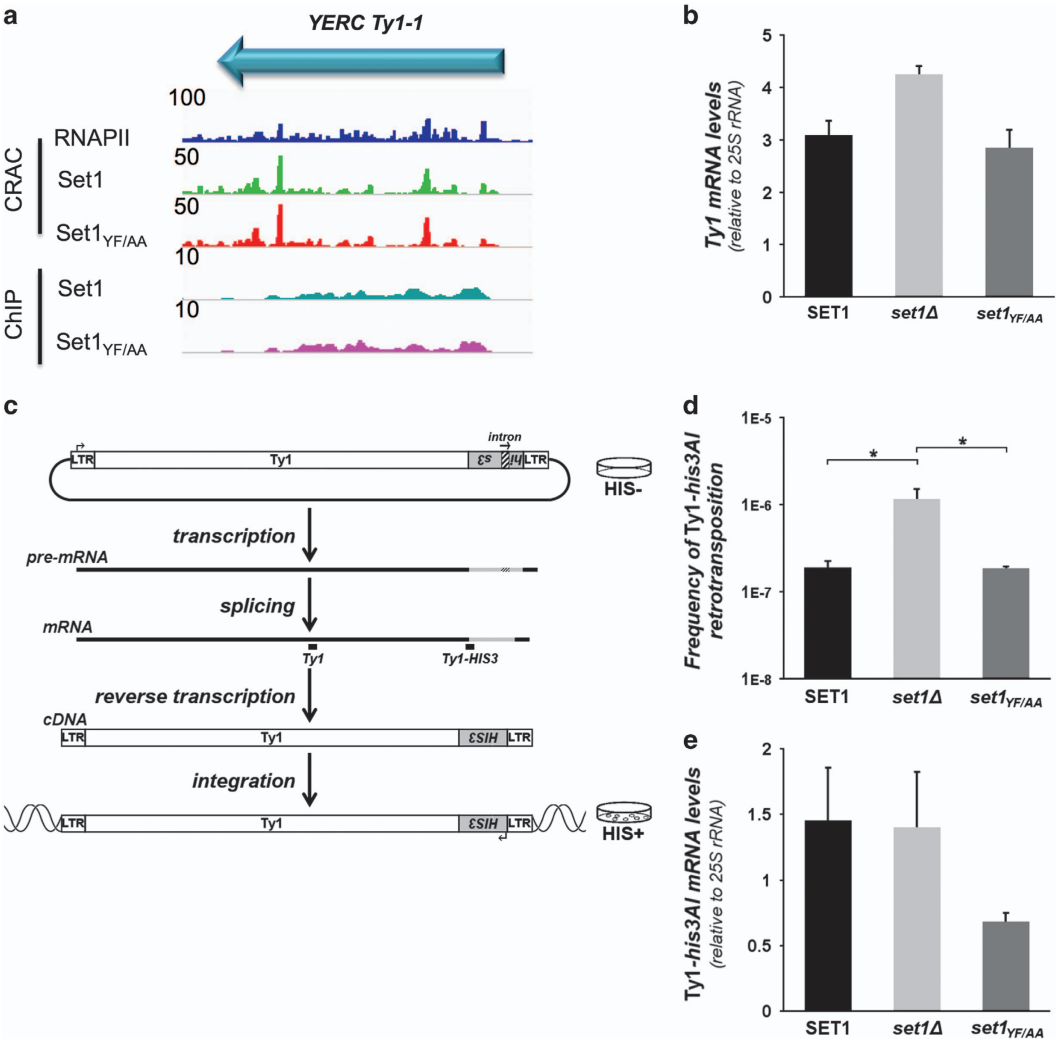


Figure 7 Set1 represses Ty1 retrotransposition post-transcriptionally. (**a**) Example of the Set1 and Set1$_{YF-AA}$ binding on Ty1-1 mRNA. The Set1 CRAC and ChIP profiles are shown. (**b**) Global expression of endogenous Ty1 retrotransposons in *SET1, set1Δ* or *set1-YF/AA* yeast cells, monitored by quantifying Ty1 mRNA levels by RT-qPCR (normalized to 25 S rRNA values; mean ± s.d.; $n = 3$). (**c**) Ty1 retrotransposition assay from a plasmid expressing a Ty1 element tagged with the *his3AI* reporter gene [33]. An intron is inserted in the *HIS3* gene in an antisense orientation in a spliceable orientation in the Ty1 transcript resulting in a Ty1 complementary DNA (cDNA) bearing a functional *HIS3* gene. The cDNA can then be integrated into the host genome. Cells that sustain a Ty1-*HIS3* retrotransposition event give rise to His+ colonies [32]. The position of the qPCR amplicons used to amplify all the Ty1 mRNAs in (**b**) and the Ty1 reporter mRNA specifically expressed from the plasmid in (**e**) are indicated (*Ty1* and *Ty1-HIS3* amplicons, respectively). (**d**) Frequency of Ty1-*his3AI* retrotransposition in *SET1, set1Δ* or *set1-YF/AA* yeast cells (number of His+ prototrophs divided by the total number of cells; mean ± s.d.; $n = 3$). *$p \leq 0.05$ (Welch's *t*-test). (**e**) Plasmid Ty1-*his3AI* expression in *SET1, set1Δ* or *set1-YF/AA* yeast cells monitored by quantifying Ty1-HIS3 mRNA levels by RT-qPCR (normalized to 25S rRNA values; mean ± s.d.; $n = 3$).

would be higher than in other regions of the RNA (see Discussion). After the wPAS, both the WT and mutant Set1 signals decrease steadily (Figure 6b, inset), indicating that they are significantly above background within the range of the transcription unit. The presence of a Set1 3′ peak can be readily observed at individual genes, most prominently in the mutant for which the signal before the peak is generally lower (Figure 6c; see also snapshots for the *SET1, SLK19,* and *SWI1* loci in Supplementary Figure S6), indicating that this behavior is not limited to a small set of genes. These data are compatible both with increased co-transcriptional recruitment of Set1 in the region of termination and with a post-transcriptional binding to the mRNA in the

immediate vicinity of the poly(A) site. Importantly and surprisingly, in both cases the interaction with the RNA is not dependent on dRRM.

Prompted by the strong binding of Set1 to snRNAs, we assessed the profile of Set1 binding to intron-containing RNAs by comparing it with size matched mRNA-coding genes. As shown in Figure 6d, Set1 bound intronic transcripts with similar or even better efficiency than non-intronic RNAs, causing a slight downstream shift of the 5′-peak of Set1 binding. Binding of the $Set1_{YF/AA}$ was similarly affected at intron-containing genes, as well as to the general population. We also assessed binding to cryptic unstable transcripts, a class of transcripts that are unstable in WT yeast because they are rapidly degraded in the nucleus [29], which we compared with matched size small open reading frames. Set1 binding to these features was lower than at small open reading frames, even when normalization to RNAPII was applied (Figure 6e) to account for the generally different levels of transcription. This could be due either to specificities residing in the sequence of the cryptic unstable transcript, or to nuclear degradation of these RNAs.

Set1 represses Ty1 retrotransposition post-transcriptionally

Among the mRNAs that were strongly bound by Set1, presumably post-transcriptionally, we also found Ty1 retrotransposon (Figure 7a). Binding of Set1 to Ty1 mRNA was not affected by the YF/AA mutation suggesting that Set1 binding to Ty mRNA does not involve its dRRM (Figure 7a). The $Set1_{YF/AA}$ mutation had no major effect on steady-state Ty1 mRNA levels (Figure 7b) as previously reported for the *set1*Δ mutant [30,31]. This indicates that Set1 binding does not affect Ty1 mRNA expression or stability. To assess whether Set1 affects Ty1 retrotransposition, we performed a typical retrotransposition assay based on a Ty1 element marked with a *his3AI* reporter gene on a plasmid, which confers His^+ prototrophy to cells upon retrotransposition (Figure 7c). In the absence of Set1, the frequency of Ty1 retrotransposition significantly increased (Figure 7d), whereas no change in Ty1*HIS3* mRNA levels was observed (Figure 7e). This indicates that Set1 can repress Ty1 mobility at a post-transcriptional stage. In contrast, Set1-YF/AA, which retains the ability to bind Ty1 mRNAs, repressed Ty1 retrotransposition as efficiently as WT Set1. These results suggest that Set1 binding to Ty1 mRNA could impair Ty1 mRNA export, translation or encapsidation, all essential steps to Ty1 retrotransposition efficiency. Of note, the less than twofold decrease in Ty1-*his3AI* mRNA levels observed in the *set1-YF/AA* mutant may not affect Ty1 retrotransposition (Figure 7e), as much more Ty1 mRNAs are produced than effective transposition events occurring in cells [32]. However, we cannot exclude that the slight defect in Ty1-*his3AI* mRNA levels may mask a slight increase in Ty1 cDNA integration that could be facilitated by the modification of the histone methylation status of the yeast genome in the *set1-YF/AA* mutant.

Reduced H3K4me3 levels are due to defective recruitment or positioning of Set1 during transcription

Although we showed that the $Set1_{YF/AA}$ mutation only marginally affects the recruitment of Set1 to chromatin on a genome-wide scale, at the gene level a variegated range of cases exists. In some instances, a strong RNA-binding defect translates into a marginal effect on recruitment (for example, *MOT3*, Figure 8), in other cases (for example, *PMA1* and *ENO1*) recruitment to chromatin is affected in spite of a moderate effect on *in vivo* crosslinking to the RNA as revealed by the CRAC signal. Although it is unclear why in these latter particular cases, the $Set1_{YF/AA}$ mutation affects Set1 occupancy, we exploited these individual differences to address the role of the nascent RNA and Set1 recruitment in H3K4 methylation. As shown in Figure 8b, in all these three cases H3K4me3 was found to be strongly reduced, indicating that neither the recruitment to chromatin (*MOT3*) nor the crosslinking to the RNA alone (*PMA1* and *ENO1*, Figure 8) are sufficient to promote methylation.

The general strong decrease in methylation when dRRM is mutated might be due to the defective positioning of the protein along transcription units, to an allosteric requirement for RNA interaction or to a general inactivation of the methylation function of Set1 by the Set1-YF/AA mutation. To distinguish between these possibilities, we analyzed *in vitro* the histone methyltransferase (HMT) activity of the Set1C containing the $Set1_{YF/AA}$ mutation in the presence or absence of RNA. As shown in Figure 8c, Set1C YF/AA reproducibly displayed a higher HMT activity compared with WT on a recombinant chromatin template containing ubiquitylated H2B. This indicates that the $Set1_{YF/AA}$ not only retained full HMT activity but its *in vitro* activity was even enhanced. Addition of purified *GAL1* RNA did not improve the activity of Set1C, and actually inhibited its function in a concentration-dependent manner. As expected, it had no effect when added to Set1C YF/AA (Figure 8c). This strongly suggests that the interaction with RNA is

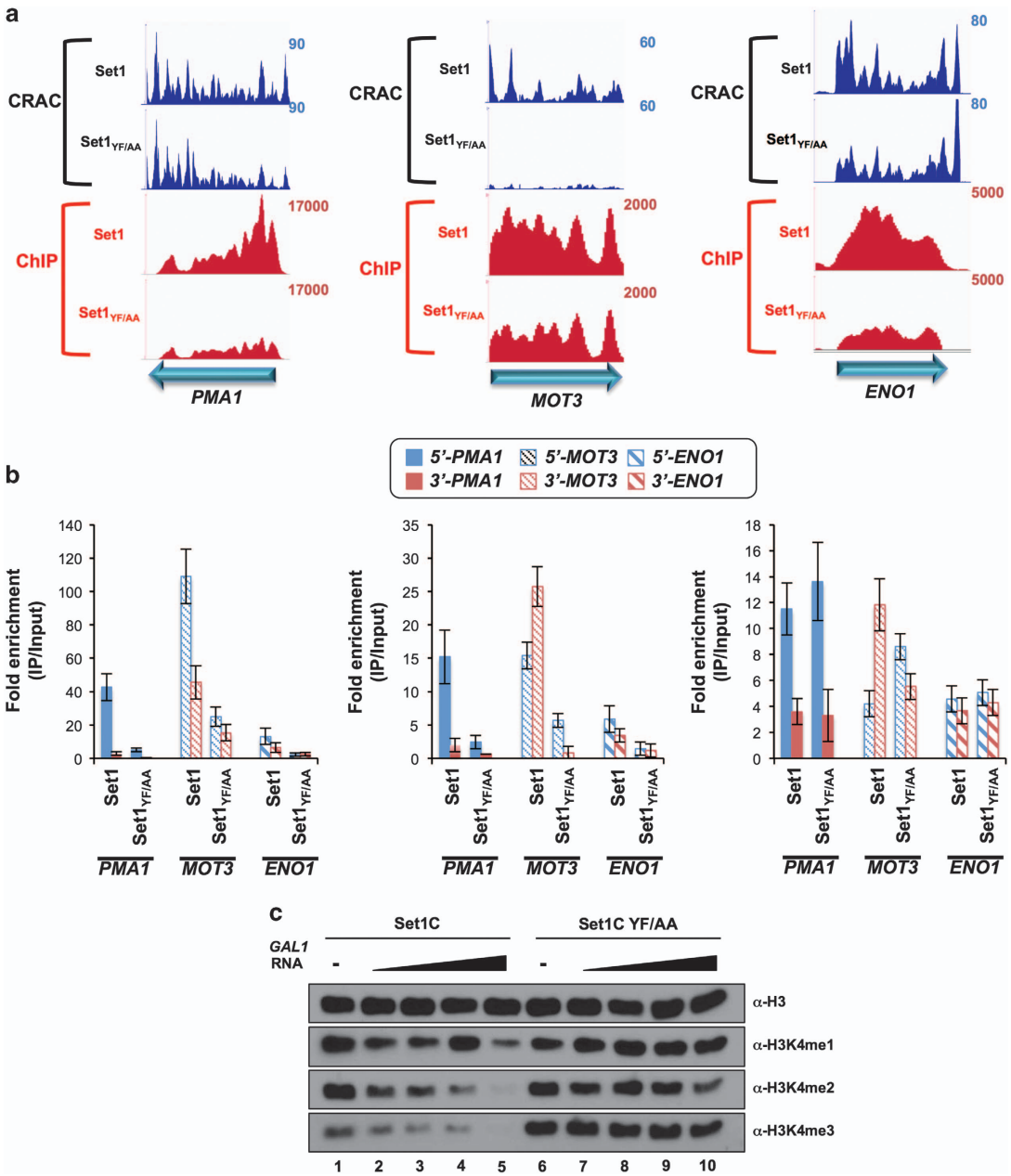


Figure 8 Reduction of Set1 occupancy and RNA-binding activity in Set1$_{YF/AA}$ correlates with reduced H3K4me3 levels. (**a**) ChIP-seq and CRAC signals of PTH-Set1 (Set1) and PTH-Set1$_{YF/AA}$ (Set1 $_{YF/AA}$) at representative genes. ChIP-seq bedGraphs were generated by normalization to 10 million mapped reads for each sample. CRAC signals were normalized as described above. (**b**) H3K4me3, me2, me1 occupancies at the indicated genes in W303 *set1*Δ::*TRP1* pRS415-nHTP-*SET1* and *SET1*$_{YF/AA}$ strains. Levels of H3K4me3, me2 and me1 are normalized to total H3. Positions of the primers used for the ChIP-qPCR for each representative gene are indicated. Errors bars represent the s.d. from three independent experiments. (**c**) Recombinant chromatin template containing fully ubiquitylated H2B (H2Bub) was subjected to *in vitro* HMT assays with purified Set1C and Set1C YF/AA in the absence and presence of purified *GAL10* RNA [0.1 (lanes 2 and 7), 0.2 (lanes 3 and 8), 0.5 (lanes 4 and 9) and 2 (lanes 5 and 10) molar ratio to Set1C. H3 methylated status was monitored by western blots with indicated antibodies.

not required for activating the HMT function of Set1, and is consistent with the possibility that RNA might negatively regulate its activity.

Together, these results strongly suggest that binding to the RNA is important to define the appropriate topology of Set1C distribution along transcription

units, which is important for the deposition of the H3K4me3 mark.

Discussion

In this study, we first showed with *in vitro* studies that the interaction between Set1C and RNA is direct. Both the dRRM and N-SET domains of Set1 contribute to Set1 RNA-binding activity *in vitro* in the context of the reconstituted complex. Although structural data indicated that dRRM has the canonical structure to bind RNA [14], binding to RNA of the C762 fragment alone was unexpected because none of the C762 constituents (N-SET, SET and post-SET domains and Spp1, Bre2, Sdc1, Swd3 and Swd1) has a canonical RNA-binding motif. In addition, no direct interaction of Set1C subunits and N-SET domain alone (data not shown) with RNA was observed, strongly suggesting that interaction with RNA is mediated by a composite surface potentially involving all or some components associated with the C762 fragment. Interestingly, addition of residues 569–762 to the C762 fragment, which are known to have an inhibitory effect on Set1 methyltransferase activity [10] inhibited the RNA-binding activity of the C762 fragment, suggesting that Set1 binding to RNA could regulate the methyltranferase activity of Set1.

The involvement of the N-SET domain in RNA binding is of particular interest as this domain acts as central regulatory region of Set1C by its ability to bind Spp1 [34] and Swd1 [20]. The N-SET domain also mediates an Spp1-dependent interaction with the SET domain and its associated subunits, an interaction that likely regulates Set1 methyltransferase activity [35]. Other studies have indicated a cross-talk between Swd2 and Spp1 suggesting a complex regulation mediated by Set1C subunits for interaction between the C- and the N-terminal regions of Set1C [21]. In this work, we found that omitting Spp1 in the reconstitution assay strongly decreases Set1C RNA-binding activity. Therefore, in the context of the full-length Set1, it is possible that dRRM, Spp1 and the N-SET cooperate to stabilize interaction with RNA. *In vivo*, inhibition of the C762 RNA-binding activity by the region of Set1 encompassing amino acids 569–762 remains an open question.

We performed ChIP-seq experiments with the same Set1 mutants studied *in vitro* using an anti-Set1 mAb [24]. We provide a high-resolution map of Set1 occupancy indicating that the peak of Set1 occupancy is shifted 3′ to the peak of H3K4me3 and slightly shifted with respect to RNAPII average occupancy. Analysis of ChIP-seq signals revealed that $set1_{YF/AA}$ mutation did not affect Set1 recruitment to chromatin per se but rather regulated Set1C distribution by maintaining Set1 in the 5′ region of genes. In addition, we show that the N-SET domain that is required to bind RNA *in vitro* is not involved in recruitment of Set1 to chromatin, whereas it is essential for Set1C catalytic activity. Assessing the functional importance of the RNA-binding activity of the N-SET remains a challenge for future studies.

To assess whether Set1 binds RNA *in vivo*, we performed CRAC experiments using Set1 and $Set1_{YF/AA}$. Our high-resolution and strand-specific CRAC analysis shows that Set1 binds to RNA *in vivo* thereby extending and providing a physiological facet to our *in vitro* analysis with the reconstituted complex. CRAC analysis is expected to detect binding to the RNA both during and after transcription. In our experiments, the occurrence of co-transcriptional binding is demonstrated by the observation that Set1 binds to intronic regions and is also strongly suggested by the highly significant correlation with Set1 and RNAPII occupancy (as determined by ChIP and CRAC, respectively) for most features. Although it is formally possible that the levels of transcription and/or Set1 chromatin occupancy also impact to some extent the post-transcriptional binding to the RNA, we favor the hypothesis that a co-transcriptional component dominates in directing Set1 binding to the RNA for the largest fraction of the population. Binding of Set1 to the nascent transcript is not sequence specific and occurs preferentially at the 5′-end of RNAs, with a peak that is slightly shifted downstream relative to the maximum of RNAPII occupancy as detected by CRAC.

Our results indicate that the substitution of residues Y271 and F272, which are part of the hydrophobic core of RRM1 and predicted to be important for maintaining the structure of whole dRRM [36], strongly decreased Set1 RNA binding, particularly in the 5′ region of RNAs *in vivo*. This result combined with our ChIP-seq data supports the notion that Set1 is recruited via protein–protein interactions and subsequently contacts the nascent RNA 5′-region via its dRRM. Transfer of Set1C to the nascent RNA once the latter emerges from the elongation complex would contribute to position Set1 predominantly to the 5′ regions of genes.

The $set1_{YF/AA}$ mutation markedly affects the deposition of the H3K4me3 mark at the 5′-end of genes for all genes tested consistent with the fact that global H3K4me3 is strongly reduced in the $set1_{YF/AA}$ mutant.

Importantly, the *set1*$_{YF/AA}$ mutation in the context of reconstituted Set1C enhanced the methyltransferase activity of Set1C assayed on recombinant chromatin containing ubiquitylated H2B.

There might be several mechanisms by which the *set1*$_{YF/AA}$ mutation affects H3K4me3. Interaction with the nascent RNA might be important to increase the persistence of Set1C in the proximity of chromatin to allow additional time for H3K4 trimethylation in the 5′-end of genes, whereas transcription proceeds at its normal speed. It is possible that binding to the RNA activates allosterically the HMT activity of the protein. However, we showed that *in vitro* addition of the RNA does not activate Set1 but actually inhibits H3K4 methylation, in a manner that depends on the RNA-binding activity of Set1. It is not clear whether this inhibition detected *in vitro* is physiologically relevant, but the possibility exists that binding to the RNA could regulate Set1 activity in some phases of the H3 methylation process. The simultaneous or sequential interaction of Set1C with the polymerase and the nascent RNA might constitute a quality control strategy to ensure deposition of the H3K4m3 mark only to regions of active transcription. Whether RNA binding to N-SET could contribute to such a process remains to be determined.

Metagene analysis experiments revealed that Set1 was also crosslinked to mRNA at the 3′-end of the molecules, showing a 3′ peak immediately before the poly(A) addition site. Importantly, formation of this 3′ RNA-binding peak was fully insensitive to the *set1*$_{YF/AA}$ mutation suggesting that the binding of Set1 at poly(A) sites is independent of its dRRM. Interestingly, although the level of this peak was low relative to the levels of Set1 at the 5′-end of genes, its intensity relative to the polymerase signal (that is low in this region) is the highest over the whole transcription unit, suggesting that it might have functional significance.

The dRRM-independent formation of the 3′ Set1 peak may occur via a 3′ recruitment of Set1 by Swd2, an essential subunit of Set1C that also belongs to the APT complex, a subcomplex of the cleavage and polyadenylation factor that is involved in mRNA and snoRNA 3'-end formation [36,37]. These observations might suggest a functional link between Set1C and 3′-end formation/termination but it remains unknown how the binding of Swd2 to either of the two complexes (Set1C and APT) is regulated. Swd2 was shown to directly interact with the N-terminus of Set1 [20]. Consistent with this, Swd2 recruitment to the 5′ region of genes is reduced when *SET1* is deleted [38] but whether Swd2 contributes to recruit Set1C at the vicinity of the poly(A) site to signal cleavage and polyadenylation remains to be determined.

We also observed a striking enrichment of Set1 to snRNA, suggesting that Set1 could fulfill a function in signaling splicing events. Consistent with this notion we observed that Set1 is enriched within introns, even when the signal is evaluated relative to the RNA polymerase (that also increases in introns).

Interestingly, cryptic unstable transcripts were under-represented when compared with matched size small mRNAs, opening the possibility that at early stages of ncRNA (non-coding RNA) transcription, dRRM may compete with Nab3 RRM (of the Nrd1 complex) for the recognition of sequences in the nascent transcript. This might explain the more efficient termination of ncRNA observed in cells lacking Set1 [39,40].

Finally, we show that Set1 binds to a class of transcripts to an extent that cannot be justified by a co-transcriptional component alone, or, at least, not by the same co-transcriptional component that holds for the majority of the population. We therefore suggest that Set1 can bind RNAs after transcription or that binding occurs during transcription but additionally persists in this class of genes. The very high level of Set1 bound to *SET1* mRNA was previously uncovered [26] and probably reflects the co-translational assembly mode of Set1C. In this model, *SET1* mRNA is bound by the nascent Set1 protein that emerges from the ribosome through indirect interactions with the translation machinery.

Set1 also binds post-transcriptionally to other transcripts. Interestingly, many of these factors are transcription factors some of which are functionally related. For instance, Mot3 and Rox1, which are transcriptional repressors of genes encoding cell wall proteins [36] and of hypoxic genes [37] are functionally linked with Msn2-Msn4 in the osmostress response [36] and with Sok2 in the ergosterol biosynthetic pathway [41]. Interestingly, Set1 was previously described to be required for the expression of genes in the ergosterol biosynthetic pathway [42]. Strikingly, Set1 binds also to mRNA of several genes functionally related to chromosome segregation, in line with recent results linking Set1 to mitotic spindle assembly [43,44]. However, the steady-state levels of these transcripts were not affected by deletion of *SET1*, at least in rich medium (YPD) [39], suggesting that Set1 might affect the expression of these genes at levels that do not involve mRNA synthesis or degradation rates, or that the impact of Set1 is revealed only under defined growth or stress conditions.

Our results also show that Set1 binds to Ty1 mRNA and repress Ty1 mobility at a post-transcriptional stage. Our data uncover a new function of Set1C as a repressor of Ty1 mobility and add another layer of regulation by Set1 to previous studies showing that Set1 had a synergistic role with the histone H4 methyltransferase Set5 in repressing transcription of Ty transposable elements [23]. Although Ty1 retrotransposition can alter yeast genome integrity and is consequently a highly controlled process, release of Ty1 repression is supposed to contribute to genome evolution and cell adaptation to stress [45]. Therefore, it would be interesting to determine whether Set1C repression could be alleviated under stress conditions that are known to stimulate Ty1 retrotransposition [46,47].

Deciphering the role of the post-transcriptional binding of Set1 to RNAs will reveal unexpected function of Set1C that might explain the incredibly complex genetic interaction map of Set1 [48].

Materials and Methods

Purification of recombinant Set1C and subunits, in vitro RNA electrophoretic mobility shift assay, and in vitro HMT assay

Preparation of FLAG-tagged recombinant Set1C and subunits were as described [20]. For radiolabeled RNA probe preparation, DNA duplexes containing T7 promoter sequence (5′-TAATACGACTCACTATAGGG-3′) followed by 100-nucleotide sequence encoding *GAL1* or *GAL10* mRNA starting from the transcription start site were generated by PCR. *GAL1* and *GAL10* RNA were transcribed by T7 RNA polymerase according to the manufacturer's instructions (Promega, Fitchburg, WI, USA) and then purified by gel elution method. After removing 5′ phosphate by Antarctic phosphatase, the 5′-end of RNA was radiolabeled by T4 polynucleotide kinase using [γ-32P]ATP and then purified by Sephadex column (iNtRon, Seongnam, South Korea) and gel elution method. For *in vitro* RNA electrophoretic mobility shift assay, reactions containing purified Set1C or individual subunits and 0.25 pmole of radiolabeled RNA in 20 µl reaction buffer (10 mM Tris-Cl (pH 7.5), 1 mM EDTA, 5 mM $MgCl_2$, 50 mM K-glutamate, 5% glycerol, 1 mM DTT and two units of RNasin) were incubated at room temperature for 30 min. The samples were resolved by electrophoresis at 4 °C on 5% polyacrylamide gels in 1 × TBE buffer and subjected to autoradiography. For *in vitro* HMT assay, 40 µl reactions containing 350 ng (histone amount) recombinant chromatin assembled as described in Kim *et al.* (2013) with H2Bub-containing histone octamer [49], purified Set1C and 100 µM S-adenosylmethionine were incubated at 30 °C for 2 h. Proteins were resolved by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and subjected to western blots.

Strains, constructs and growth conditions

For reconstitution of Set1Cs, *SET1, set1* mutants and Set1C subunits genes were subcloned in pFASTBAC1 with or without a FLAG tag [20]. Baculoviruses were generated according to the manufacturer's instruction (Gibco-Invitrogen, Waltham, MA, USA). Sf9 cells were infected with combinations of baculoviruses and proteins/complexes were affinity purified on M2 agarose (Sigma, St Louis, MO, USA) as described [20].

Yeast strains and primers used in this study are described in Supplementary Tables S2 and S3, respectively (see Supplementary Information). Full-length *SET1* and *SET1* mutants were cloned into in pRS415-nHTP [23]. Expression of the resulting constructs (*Z-tag—TEV cleavage site—His6—SET1*) is under the control of *MET25* promoter. The pRS415-nHTP-*SET1* (or *SET1* mutants) were transformed into W303 *set1*Δ::*TRP1* strain. Plasmid pRS415-nHTP-*SET1* complements all the tested *set1*Δ-associated phenotypes of the *set1*Δ*::TRP1* strain. For ChIP-seq and CRAC experiments, W303 *set1*Δ::*TRP1* pRS415-nHTP-*SET1* (or *SET1* mutants) cells were grown in SC-TRP-LEU. RNAPII CRAC experiments were performed from W303 cells expressing Rpb1-HTP and grown in SC-TRP. Construction of fully functional chromosomally encoded Myc9-tagged Set1 is described in Dehe *et al*. [10].

CRAC analyses

Cells in exponential phase were crosslinked with a Megatron for 100 s (Set1) and 50 s (Rpb1), harvested by centrifugation, 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, 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 five cycles of 3 min at 20 Hz. A non-crosslinked sample was treated in parallel as a background control.

Powders were thawed and the resulting extracts were treated for one hour at 25 °C with DNase I (165 U/g of cells) to solubilize chromatin and then clarified by centrifugation for 20′ at 20 000 *g* at 4 °C. Subsequent purifications steps were performed essentially as described with minor modifications from Granneman *et al.* [50]. For both nPTH-Set1 and Rpb1-HTP strains, adaptors were modified in order to sequence RNA molecules from the 3′-end.

The RNA was recovered after proteinase K treatment and reverse transcribed using specific primers. The resulting complementary DNA was used to perform multiple PCR reactions in a final volume of 25 µl using the following conditions: 0.4 µM of each primers 0.2 mM dNTP, 2.5 U LA Taq DNA polymerase from Takara, 1X LA PCR Buffer II and 2 µl of complementary DNA per reaction with the programme: 2′ at 95 °C, (30″ at 95 °C, 45″ at 58 °C, 1′ at 72 °C) × 13 cycles, 5′ at 72 °C. PCR were pooled and treated with 200 U of Exonuclease I (NEB) per milliliter of PCR reaction for 1 h 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, Düren, Germany) and sequenced using Illumina

technology (San Diego, CA, USA). Primers are indicated in Supplementary Table S2.

Samples were demultiplexed using the pyBarcodeFilter script from the pyCRAC utility suite. Subsequently, the 3′ adaptor is clipped with Cutadapt and the resulting insert is quality trimmed from the 3′-end using Trimmomatic rolling mean clipping (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 six-nucleotides random tag within the 5′ adaptor. The resulting sequences are reverse complemented with Fastx_reverse_-complement (part of the fastx toolkit [51]), and mapped to the R64 genome (sgd) with bowtie2 (-N 1 –f).

Read counts were normalized relative to reads derived from an *S. pombe* spike that was added to *S. cerevisiae* cells before the crosslinking step. The *S. pombe* spike cells contain a non-relevant protein tagged with the same HTP tag that was co-purified with the *S. cerevisiae* material. The positionally weighted average poly(A) addition site (wPAS) for every gene was calculated by weighting the position of each poly(A) site using its intensity and calculating an average position.

ChIP-seq, data processing and ChIP-qPCR

ChIP of Myc-Set1 and PTH-Set1 were performed as previously described [52] with 9E10 (anti-MYC, Santa Cruz Biotechnology, Dallas, TX, USA) and anti-Set1 monoclonal antibodies (P Nagy, University of Toronto, Toronto, Canada). Libraries were prepared from fragmented DNA using the Chip-seq MicroPlex Library Preparation Kit v2 samples preparation (Diagenode, Seraing, Belgium) according to the manufacturer's instructions. In all, 2 ng from IP samples were used as the starting material. Each library was barcoded using MicroPlex Single Index (Diagenode): iPCRtagT5, T6, T7 and T8 and amplified for 10 and 6 cycles for IP and input samples, respectively. Each library was quantified on Qubit with Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA) and then, size distribution was examined on the Bioanalyser with High Sensitivity DNA chip (Agilent, Santa Clara, CA, USA), to ensure that the samples have the proper size, no adaptor contamination and to estimate sample molarity. Each library was diluted to 4 nM and then pulled together at equimolar ratio. Libraries were denatured according to the manufacturer's instruction and sequenced on a mid-output flow cell (130 M clusters) using the NextSeq 500/550 High Output v2 150 cycles kit (Illumina), in paired-end 75/7 nt mode, according to the manufacturer's instructions. In all, 148 million (M) paired-end reads were generated (34–39 M per sample) with 93% > = Q30.

ChIP-Seq data quality was assessed using FastQC. FasQC: a quality control tool for high-throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Sequencing reads (FastQ format) were mapped to the *Saccharomyces cerevisiae* genome (sacCer3) using BFAST alignment tool with default parameters [53] (PMID 19907642) to obtain a Binary Alignment Mapped (BAM) file. The sorted BAM files were used to determine average profiles of ChIP-Seq read density using *ngs.plot* software [54], (PMID 24735413) around the transcription start site. Read counts were normalized to the total number of million uniquely mapped reads or to read count per million of mapped reads (RPM). The RPM values allow samples to be compared regardless of differences in sequencing depth. To generate BedGraphs for visualization on genome browsers, ChIP-Seq BAM files were processed using HOMER package. The tag directory for each sample was then created using the makeTagDirectory tool and the corresponding BedGraph was generated using makeUCSCfile tool with default options. Only uniquely mappable reads (non-secondary alignment) were considered to create BedGraphs with a normalization to 10 million mapped reads for each sample. To compare Set1 binding positions on RNA with Set1 occupancy on genes of interest, the Multicov command from Bedtools [55] was used to obtain read counts within each gene.

For ChIP-qPCR, samples were prepared as previously described [52]. DNA was analyzed by real-time qPCR using SYBR Green Premix Ex Taq (Takara, Mountain View, CA, USA) in a Rotor Gene 6000 (Corbett Research, Labgene, Archamps, France). Primers are listed in Supplementary Table S2. The following antibodies were used: anti-H3 (Abcam1791, Cambridge, UK), anti-H3K4me2 (Abcam-ab7766, Cambridge, UK), anti-H3K4me3 (Abcam-ab8580), anti-Myc 9E10 (Santa Cruz Biotechnology-sc-40) and anti-Rap1 (V. Géli's laboratory, Marseille, France).

Data access

ChIP-seq data sets (PTH-Set1, PTH-Set1$_{YF/AA}$, PTH-Set1$_{\Delta dRRM}$, PTH-Set1$_{\Delta NSET}$, PTH-Set1$_{\Delta RRM}$ ΔNSET, PTH-vector, Input PTH-Set1 and input PTH-Set1$_{YF/AA}$), as well as CRAC sequences generated during this work are deposited to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE104486 (GSE104484 for chiPseq datasets and GSE104485 for CRAC datasets).

Ty1 transposition assays

The pOY1 *URA3*, centromeric vector carrying a Ty1-*his3AI* reporter element expressed from its own promoter was previously described [33]. Total RNAs were extracted from yeast cultures after 4 h or 8 h at 20 °C, for Ty1 mRNAs or Ty1-*his3AI* mRNAs respectively, using the Nucleospin RNA II kit (Macherey-Nagel) and reverse transcribed with Superscript-II reverse transcriptase (Invitrogen, Waltham, MA, USA). Complementary DNA quantification was achieved by real-time PCR with a LightCycler 480 system (Roche, Basel, Switzerland) using SYBR Green incorporation according to the manufacturer's instructions. The amounts of the mRNAs of interest were normalized relative to 25S ribosomal RNA values. Primers used are described in Supplementary Table S2.

To estimate the frequency of Ty1*his3AI* mobility [32], overnight liquid cultures were grown at 30 °C from an individual clone in HC medium (Hartwell's synthetic complete) [56], lacking uracil and supplemented with 2% glucose. Each culture was diluted to OD 0.01 in HC-URA medium and grown to saturation at 20 °C, which is permissive for Ty1 transposition. In all, 3 ml of each culture were plated on two HC agar plates lacking histidine. Cell titer was determined by plating 10 000-fold diluted cultures on YEPD rich medium. Plates were

incubated for 3 days at 30 °C and counted to determine the fraction of [HIS$^+$] prototrophs. Ty1 retrotransposition frequencies were defined as the mean of 3 experiments, each one performed with four independent clones.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We deeply thank Peter Nagy and Bichtien Rouse for providing the Set1 mAb and Markus Bohnsack and Khaterine Sloan for pRS415-nHTP. We thank Luis Soares and Steve Buratowski for Set1 constructs, Tom Muir for ubiquiytlated H2B, and Nir Friedman and Alon Appleboim for communicating H3K4 Exo-ChIP profiles. We thank Julie Drogat, Ghislain Bidaud and Nicolas Fernandez for their involvement in the ChIP-seq procedures and Lorant Székvögyi for discussions. We also thank Judith Recht for critical reading of the manuscript. Work in VG laboratory is supported by the 'Ligue contre le Cancer' (Equipe Labéllisée). Bioinformatic support and computing sources at CRCM were provided by the CIBI and DISC platforms (CRCM Integrative Bioinformatics and Datacentre IT and Scientific Computing). High-throughput sequencing was performed at the TGML Platform, supported by grants from Inserm, GIS IBiSA, Aix Marseille Université, and ANR-10-INBS-0009-10. AE is supported by a fellowship from the Canceropole PACA. Work in the DL laboratory was supported by the CNRS, the ANR ('TermCut' grant) and the FRM. TC and DC are supported by French MRT fellowships. Work in the JK laboratory was supported by grants from the National Research Foundation of Korea (2012M3A9C6049938, 2015K1A3A1A21000296 and 2015R1A1A1A05001593). Collaboration between VG and JK laboratories is supported by the PHC STAR exchange programme (34325UF).

Author contributions

JJ performed all the *in vitro* reconstitution experiments. PL and AE-k performed all the ChIP-seq and ChIP-qPCR experiments and analysis. DC, MB and TC performed the CRAC experiments and analyzed the data. FJ constructed and characterized strains. AB and PL performed Ty1 retrotransposition experiments. VG, JK and DL analyzed the data, wrote the manuscript and directed the work.

References

1 Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. *BBA Gene Regul Mech* 2014; **1839**: 627–643.
2 Suganuma T, Workman JL. Histone modification as a reflection of metabolism. *Cell Cycle* 2016; **15**: 481–482.
3 Tessarz P, Kouzarides T. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol* 2014; **15**: 703–708.
4 Shilatifard A. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu Rev Biochem* 2006; **75**: 243–269.
5 Ruthenburg AJ, Allis CD, Wysocka J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell* 2007; **25**: 15–30.
6 Ernst P, Vakoc CR. WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Brief Funct Genomics* 2012; **11**: 217–226.
7 Smith E, Shilatifard A. The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. *Mol Cell* 2010; **40**: 689–701.
8 Briggs SD, Bryk M, Strahl BD *et al.* Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev* 2001; **15**: 3286–3295.
9 Roguev A, Schaft D, Shevchenko A *et al.* The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J* 2001; **20**: 7137–7148.
10 Dehe PM, Dichtl B, Schaft D *et al.* Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem* 2006; **281**: 35404–35412.
11 Dehé P-M, Géli V. The multiple faces of Set1This paper is one of a selection of papers published in this Special Issue, entitled 27th International West Coast Chromatin and Chromosome Conference, and has undergone the Journal's usual peer review process. *Biochem Cell Biol* 2006; **84**: 536–548.
12 Dichtl B, Blank D, Sadowski M, Hübner W, Weiser S, Keller W. Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO J* 2002; **21**: 4125–4135.
13 Terzi N, Churchman LS, Vasiljeva L, Weissman J, Buratowski S. H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. *Mol Cell Biol* 2011; **31**: 3569–3583.
14 Schlichter A, Cairns BR. Histone trimethylation by Set1 is coordinated by the RRM, autoinhibitory, and catalytic domains. *EMBO J* 2005; **24**: 1222–1231.
15 Pokholok DK, Harbison CT, Levine S *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 2005; **122**: 517–527.
16 Kirmizis A, Santos-Rosa H, Penkett CJ *et al.* Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* 2007; **449**: 928–932.
17 Ng HH, Robert F, Young RA, Struhl K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* 2003; **11**: 709–719.
18 Mueller CL, Jaehning JA. Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol Cell Biol* 2002; **22**: 1971–1980.

19 Lee J-S, Shukla A, Schneider J *et al.* Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell* 2007; **131**: 1084–1096.
20 Kim J, Kim J-A, McGinty RK *et al.* The n-SET domain of Set1 regulates H2B ubiquitylation-dependent H3K4 methylation. *Mol Cell* 2013; **49**: 1121–1133.
21 Trésaugues L, Dehé P-M, Guérois R *et al.* Structural characterization of Set1 RNA recognition motifs and their role in histone H3 lysine 4 methylation. *J Mol Biol* 2006; **359**: 1170–1181.
22 Hampsey M, Reinberg D. Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* 2003; **113**: 429–432.
23 Martín GM, King DA, Green EM *et al.* Set5 and Set1 cooperate to repress gene expression at telomeres and retrotransposons. *Epigenetics* 2014; **9**: 513–522.
24 Nagy PL, Griesenbeck J, Kornberg RD, Cleary ML. A trithorax-group complex purified from *Saccharomyces cerevisiae* is required for methylation of histone H3. *Proc Natl Acad Sci USA* 2002; **99**: 90–94.
25 Bohnsack MT, Tollervey D, Granneman S. Identification of RNA helicase target sites by UV cross-linking and analysis of cDNA. *Meth Enzymol* 2012; **511**: 275–288.
26 Halbach A, Zhang H, Wengi A *et al.* Cotranslational assembly of the yeast SET1C histone methyltransferase complex. *EMBO J* 2009; **28**: 2959–2970.
27 Szklarczyk D, Franceschini A, Kuhn M *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011; **39**: D561–D568.
28 Wyers F, Rougemaille M, Badis G *et al.* Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 2005; **121**: 725–737.
29 Wilkening S, Pelechano V, Järvelin AI *et al.* An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Rese* 2013; **41**: e65–e65.
30 Berretta J, Pinskaya M, Morillon A. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev* 2008; **22**: 615–626.
31 Martín GM, King DA, Garcia-Nieto PE, Morrison AJ. Transcriptome profiling of Set5 and Set1 methyltransferases: tools for visualization of gene expression. *GDATA* 2014; **2**: 216–218.
32 Curcio MJ, Garfinkel DJ. Regulation of retrotransposition in *Saccharomyces cerevisiae*. *Mol Microbiol* 1991; **5**: 1823–1829.
33 Lee BS, Lichtenstein CP, Faiola B *et al.* Posttranslational inhibition of Ty1 retrotransposition by nucleotide excision repair/transcription factor TFIIH subunits Ssl2p and Rad3p. *Genetics* 1998; **148**: 1743–1761.
34 Mersman DP, Du H-N, Fingerman IM, South PF, Briggs SD. Charge-based interaction conserved within histone H3 lysine 4 (H3K4) methyltransferase complexes is needed for protein stability, histone methylation, and gene expression. *J Biol Chem* 2012; **287**: 2652–2665.
35 Vitaliano-Prunier A, Menant A, Hobeika M, Géli V, Gwizdek C, Dargemont C. Ubiquitylation of the COMPASS component Swd2 links H2B ubiquitylation to H3K4 trimethylation. *Nat Cell Biol* 2008; **10**: 1365–1371.
36 Martínez-Montañés F, Rienzo A, Poveda-Huertes D, Pascual-Ahuir A, Proft M. Activator and repressor functions of the Mot3 transcription factor in the osmostress response of *Saccharomyces cerevisiae*. *Eukaryotic Cell* 2013; **12**: 636–647.
37 Klinkenberg LG, Mennella TA, Luetkenhaus K, Zitomer RS. Combinatorial repression of the hypoxic genes of *Saccharomyces cerevisiae* by DNA binding proteins Rox1 and Mot3. *Eukaryotic Cell* 2005; **4**: 649–660.
38 Soares LM, Buratowski S. Yeast Swd2 is essential because of antagonism between Set1 histone methyltransferase complex and APT (associated with Pta1) termination factor. *J Biol Chem* 2012; **287**: 15219–15231.
39 Margaritis T, Oreal V, Brabers N *et al.* Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription. Madhani HD, ed. *PLoS Genet* 2012; **8**: e1002952.
40 Castelnuovo M, Zaugg JB, Guffanti E *et al.* Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Res* 2014; **42**: 4348–4362.
41 Hongay C, Jia N, Bard M, Winston F. Mot3 is a transcriptional repressor of ergosterol biosynthetic genes and is required for normal vacuolar function in *Saccharomyces cerevisiae*. *EMBO J* 2002; **21**: 4114–4124.
42 South PF, Harmeyer KM, Serratore ND, Briggs SD. H3K4 methyltransferase Set1 is involved in maintenance of ergosterol homeostasis and resistance to Brefeldin A. *Proc Natl Acad Sci USA* 2013; **110**: E1016–E1025.
43 Schibler A, Koutelou E, Tomida J *et al.* Histone H3K4 methylation regulates deactivation of the spindle assembly checkpoint through direct binding of Mad2. *Genes Dev* 2016; **30**: 1187–1197.
44 Beilharz TH, Harrison PF, Miles DM *et al.* Coordination of cell cycle progression and mitotic spindle assembly involves histone H3 lysine 4 methylation by Set1/COMPASS. *Genetics* 2017; **205**: 185–199.
45 Curcio MJ, Lutz S, Lesage P. The Ty1 LTR-retrotransposon of budding yeast, *Saccharomyces cerevisiae*. *Microbiol Spectr* 2015; **3**: 1–35.
46 Morillon A, Springer M, Lesage P. Activation of the Kss1 invasive-filamentous growth pathway induces Ty1 transcription and retrotransposition in *Saccharomyces cerevisiae*. *Mol Cell Biol* 2000; **20**: 5766–5776.
47 Todeschini A-L, Morillon A, Springer M, Lesage P. Severe adenine starvation activates Ty1 transcription and retrotransposition in *Saccharomyces cerevisiae*. *Mol Cell Biol* 2005; **25**: 7459–7472.
48 Costanzo M, VanderSluis B, Koch EN *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* 2016; **353**: aaf1420–aaf1420.
49 McGinty RK, Köhn M, Chatterjee C, Chiang KP, Pratt MR, Muir TW. Structure-activity analysis of semisynthetic

nucleosomes: mechanistic insights into the stimulation of Dot1L by ubiquitylated histone H2B. *ACS Chem Biol* 2009; **4**: 958–968.
50 Granneman S, Kudla G, Petfalski E, 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 USA* 2009; **106**: 9613–9618.
51 Weiner A, Hsieh T-HS, Appleboim A *et al.* High-resolution chromatin dynamics during a yeast stress response. *Mol Cell* 2015; **58**: 371–386.
52 Ostrow AZ, Viggiani CJ, Aparicio JG, Aparicio OMChIP-Seq to analyze the binding of replication proteins to chromatinIn:Vengrova S, Dalgaard J (eds). *DNA Replication* Vol 1300 Methods in Molecular Biology New York, NY, USA: Springer New York. 2015, 155–168.
53 Homer N, Merriman B, Nelson SF. BFAST: an alignment tool for large scale genome resequencing. Creighton C, ed. *PLoS ONE* 2009; **4**: e7767.
54 Shen L, Shao N, Liu X, Nestler E. ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 2014; **15**: 284.
55 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010; **26**: 841–842.
56 Servant G, Pennetier C, Lesage P. Remodeling yeast gene transcription by activating the Ty1 long terminal repeat retrotransposon under severe adenine deficiency. *Mol Cell Biol* 2008; **28**: 5543–5554.

(**Supplementary information** is linked to the online version of the paper on the *Cell Discovery* website.)

Binding to RNA regulates Set1 function

Pierre Luciano[1,5], Jongcheol Jeon[2,5], Abdessamad El-kaoutari[1,5], Drice Challal[3,5], Amandine Bonnet[4], Mara Barucco[3], Tito Candelli[3], Frederic Jourquin[1], Pascale Lesage[4], Jaehoon Kim[2,*], Domenico Libri[3,*], Vincent Géli[1,*].

Supplementary information

Supplementary Figure S1. *In vitro* interaction of RNA with individual subunits of Set1C

(**A**) SDS-PAGE and Coomassie blue staining of purified Set1C subunits.

(**B**) Radiolabeled *GAL10* transcripts were subjected to *in vitro* RNA electrophoretic mobility shift assay with purified Set1C subunits. 0.5 pmoles (lanes 4, 6, 8, 12, 14, 16) or 2.5 pmoles (lanes 5, 7, 9, 11, 13, 15, 17) of each subunit were added. Set1C is shown as a positive control at the same concentrations.

Supplementary Figure S2. Expression of the PTH-Set1 and characterization of the anti-Set1 mAb

(**A**) Set1 amount in W303 and in W303 *set1*Δ*::TRP1* pRS415-Z-tag-Tev-6His-*SET1* (PTH-Set1) cells grown in SC -TRP-LEU-MET versus SC-TRP-LEU. Set1 and PTH-Set1 are detected with anti-Set1 mAb.

(**B and C**) The anti-Set1 mAb recognized an epitope comprised in a Set1 region lying between residue 700 and 761. **(B)** Reconstituted Set1C- containing either Flag-Set1 (Full length, FL) or the indicated Flag-Set1 truncations- were analysed by Western blot either with anti-Flag (top) or anti-Set1 mAb (bottom). **(C)** Yeast strains expressing the indicated Set1 deletion mutants (Soares et al. 2014) were analysed by Western blot with anti-Set1 mAb.

Only the Set1 mutant lacking the region from 700 to 761 is not recognized by the anti-Set1 mAb.

Supplementary Figure S3. Comparison of Myc-Set1 and PTH-Set1 occupancy profiles

(A) Enrichment profiles of Myc-Set1 and PTH-Set1 at the indicated region of Chr VII. Graphs are normalized to 10 million mapped reads for each ChIP-seq.

(B) Occupancy snapshots of Myc-Set1 and PTH-Set1 on representative genes.

(C) The correlation plot between Myc-Set1 and PTH-Set1 datasets. Myc-Set1 and PTH-Set1 read coverages were compared by Pearson correlation. Read coverages were computed using deepTools utility multiBamSummary version 2.5.3 after binning in 100 bp intervals and calculating the count per regions.

(**D**) Metagene analysis of PTH-Set1 occupancy on big genes (> 1500 bp). Enrichment profiles were compared to those of H3K4me1, me2, and me3 [50] and RNAPII [57].

Supplementary Figure S4. Distribution of ChIP signals for PTH-Set1 and PTH-Set1$_{YF/AA}$ in the whole population of mRNA coding genes

Supplementary Figure S5. Distribution of Set1 CRAC reads across transcript classes. Reads for rRNA have not been included as they are strongly represented in the sequencing reaction from the non-crosslinked sample and it cannot be established if these represent artefacts or bona fide signals.

Supplementary Figure S6. Specific classes of mRNA are highly bound by Set1 relative to RNAPII

(A) Examples of transcripts with high Set1/RNAPII CRAC signals. The RNAPII CRAC (Blue) and Set1 CRAC (green) signals are shown. The scale is indicated for each snapshot.

(B) BAP2 illustrates a transcript that is co-transcriptionally bound by Set1.

(C) Integration of protein-protein interactions among genes whose mRNA are highly bound by Set1, including direct (physical) as well as indirect (functional) associations. The graph was performed using the STRING database.

Supplementary Table Legends.

Supplementary Table S1. CRAC and ChIP-seq datasets.

Reads are indicated. Set1CL: Set1 cross-linked; Set1noCL:Set1 not Cross-linked.

Supplementary Table S2. Strains used in this study

Name	genotype	Ref.
W303-1A	Mat a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15	Rothstein RJ (1983)
W303-1A Myc-Set1	Mat a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 Myc-Set1::TRP1	Dehé et al. 2016
W303-1A *set1::TRP1*	Mat a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 *set1::TRP1*	This study
W303-1A *set1::TRP1* pRS415-nHTP	Mat a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 *set1::TRP1* pRS415-nHTP	This study
W303-1A *set1::TRP1* pRS415-nHTP-*SET1*	Mat a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 set1Δ::TRP1 pRS415-nHTP-SET1	This study
W303-1A *set1::TRP1* *pRS415-nHTP-set1*$_{YF/AA}$	Mat a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 set1::TRP1 *pRS415-nHTP- set1*$_{YF/AA}$	This study
W303-1A *set1::TRP1* *pRS415-nHTP- set1*$_{\Delta RRM}$	Mat a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 set1::TRP1 *pRS415-nHTP- set1*$_{\Delta RRM}$	This study
W303-1A *set1::TRP1* *pRS415-nHTP- set1*$_{\Delta N\text{-}SET}$	Mat a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 *set1::TRP1 pRS415-nHTP- set1*$_{\Delta N\text{-}SET}$	This study
W303-1A *set1::TRP1* *pRS415-nHTP- set1*$_{\Delta RRM, \Delta N\text{-}SET}$	Mat a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 *set1::TRP1 pRS415-nHTP- set1*$_{\Delta RRM, \Delta N\text{-}SET}$	This study
W303-1A *set1YF/AA::TRP1*	Mat a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 *set1YF/AA::TRP1*	This study

Supplementary Table S3. Primers used in this study

Name	Sequence 5' to 3'	
5'-PMA1-F	TCAGCTCATCAGCCAACTCAAG	qPCR
5'-PMA1-R	CGTCGACACCGTGATTAGATTG	
3'-PMA1-F	TACTGTCGTCCGTGTCTGGATCT	
3'-PMA1-R	CCTTCATTGGCTTACCGTTCA	
5'-MOT3-F	AACACGACTACTGTTTCCTCT	
5'-MOT3-R	AAGGGTATATATACTGCTGCT	
3'-MOT3-F	GTTACGATACAAACATCAAGA	
3'-MOT3-R	CTATTTGTTGTGACTAACAAT	
5'-ENO1-F	CGATGACTTCTTGATTTCTTT	
5'-ENO1-R	GTGCTTGTATAATGGGACATT	
3'-ENO1-F	ACTTTCATTGCTGACTTGGTC	
3'-ENO1-R	AACAGCGTTGTCACCTAATTC	
5'-Ty1	CATTGCGTCAAATGAGATCCAA	
3'-Ty1	GGTGTGGAATCGGTTGGACTC	
5'-Ty1-HIS3	TGTGATGACAAAACCTCTTCCG	
3'-Ty1-HIS3	ACGATGTTCCCTCCACCAAA	
5'-25S rRNA	AACGTCTATGCGAGTGTTTGG	
3'-25S rRNA	TTCCTCTGGCTTCACCCTATT	
L3-6N-GA	/5rApp/GCTtcNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAG AGTGT/3ddC/	3' adapter
L3-6N-GU	/5rApp/GCTacNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAG AGTGT/3ddC/	
L3-6N-AC	/5rApp/GCTgtNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAG AGTGT/3ddC/	
L3-6N-UC	/5rApp/GCTgaNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAG AGTGT/3ddC/	
L5miRCat	5-/5InvddT/CTTGrGrCrArCrCrCrGrArGrArArUrUrCrCrA-3	5' adapter
RT L3-2	ACACTCTTTCCCTACACGACGCTCTTCCG-3	RT
P5_3prime	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA CGCTCTTCCGATCT	PCR
miRCat_PCR2	CAAGCAGAAGACGGCATACGAgatcCTTGGCACCCGAGAAT	PCR

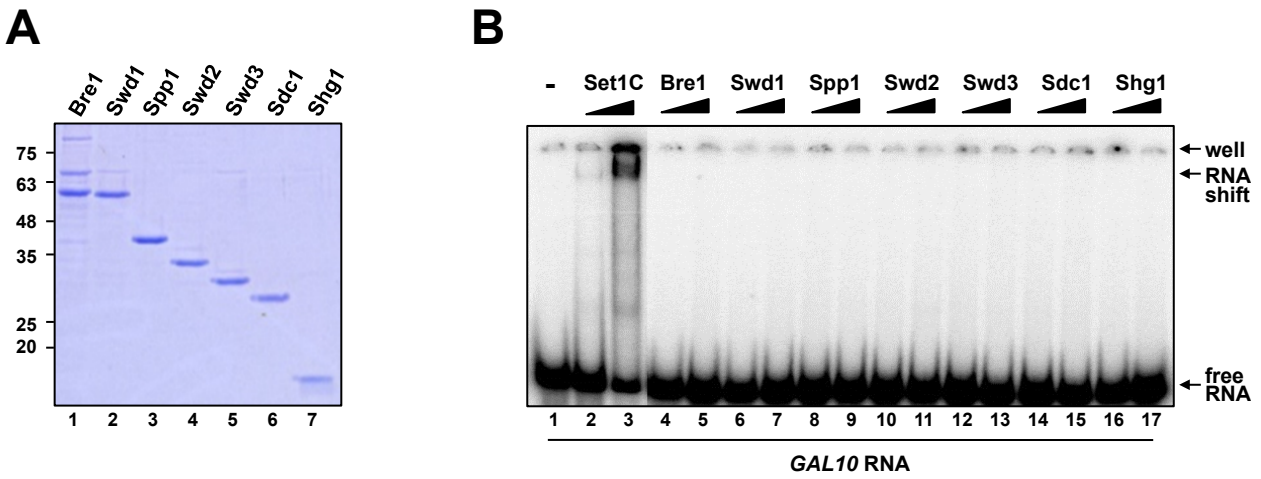
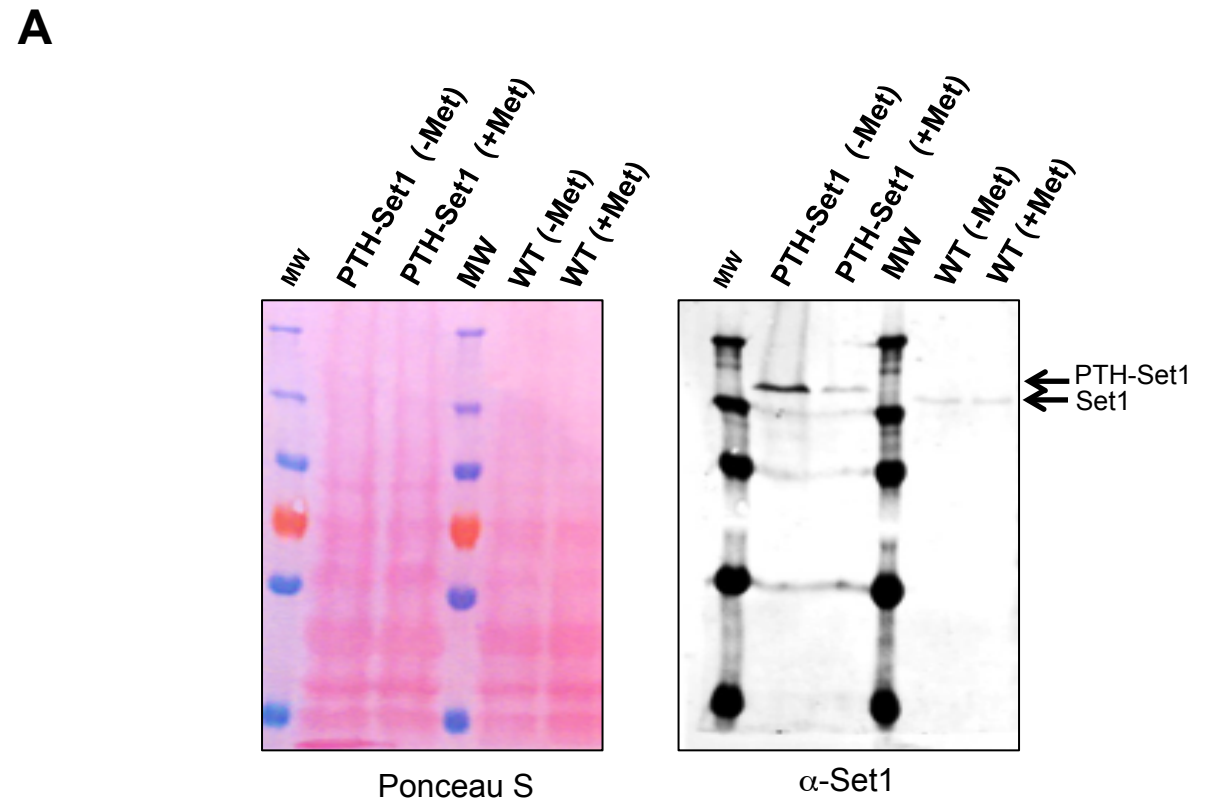
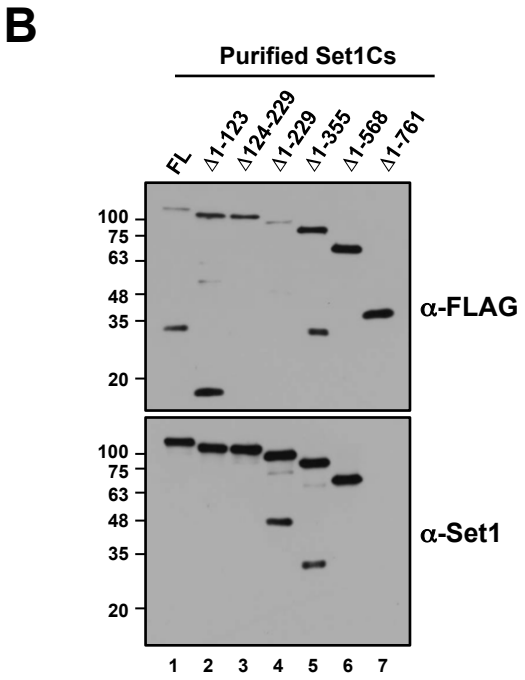

A
Bre1
Swd1
Spp1
Swd2
Swd3
Sdc1
Shg1
75
63
48
35
25
20
1 2 3 4 5 6 7
B
-
Set1C
Bre1
Swd1
Spp1
Swd2
Swd3
Sdc1
Shg1
well
RNA shift
free RNA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
GAL10 RNA

Figure S1


A
MW
PTH-Set1 (-Met)
PTH-Set1 (+Met)
MW
WT (-Met)
WT (+Met)
Ponceau S
α-Set1
PTH-Set1
Set1


B
Purified Set1Cs
FL
Δ1-123
Δ124-229
Δ1-229
Δ1-355
Δ1-568
Δ1-761
100
75
63
48
35
20
α-FLAG
α-Set1
1 2 3 4 5 6 7

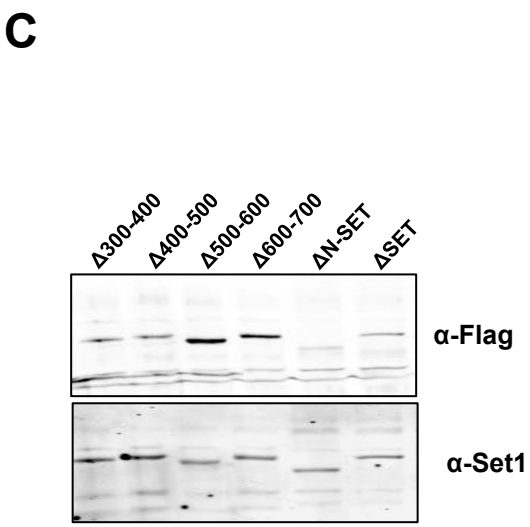

C
Δ300-400
Δ400-500
Δ500-600
Δ600-700
ΔN-SET
ΔSET
α-Flag
α-Set1

Figure S2

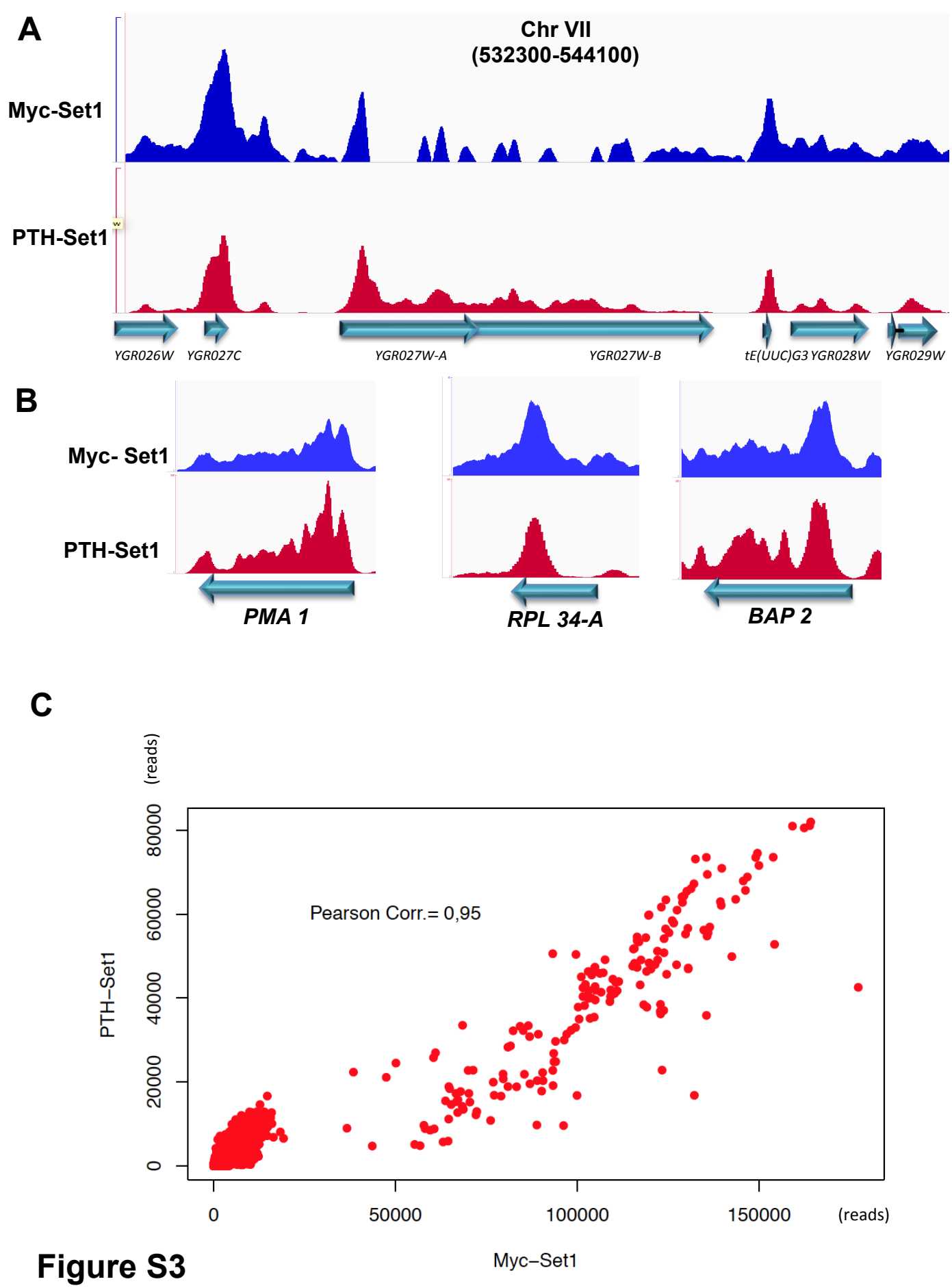

A
Chr VII
(532300-544100)
Myc-Set1
PTH-Set1
YGR026W
YGR027C
YGR027W-A
YGR027W-B
tE(UUC)G3
YGR028W
YGR029W
B
Myc- Set1
PTH-Set1
PMA 1
RPL 34-A
BAP 2
C
Pearson Corr.= 0,95
PTH–Set1
Myc–Set1
(reads)
0
20000
40000
60000
80000
0
50000
100000
150000
(reads)

Figure S3

D

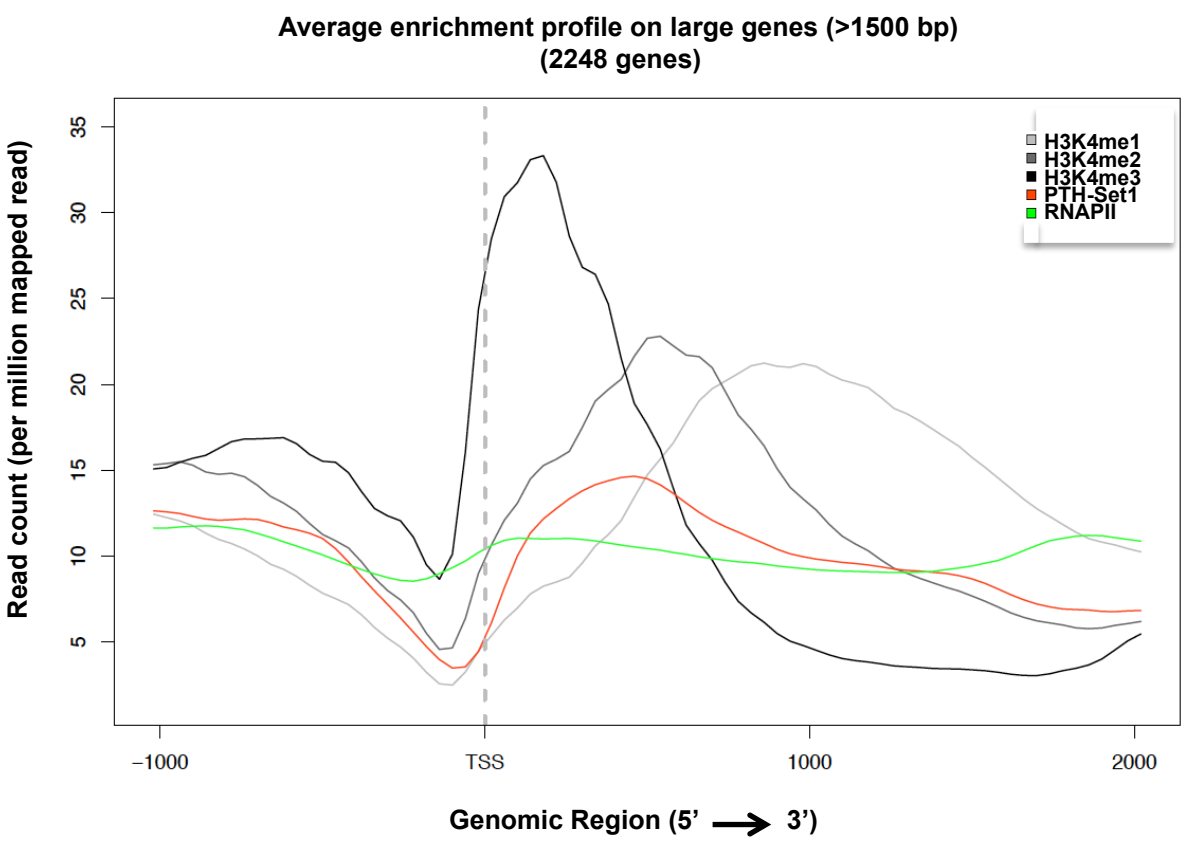

Average enrichment profile on large genes (>1500 bp)
(2248 genes)
Read count (per million mapped read)
5
10
15
20
25
30
35
H3K4me1
H3K4me2
H3K4me3
PTH-Set1
RNAPII
−1000
TSS
1000
2000
Genomic Region (5' → 3')

Figure S3

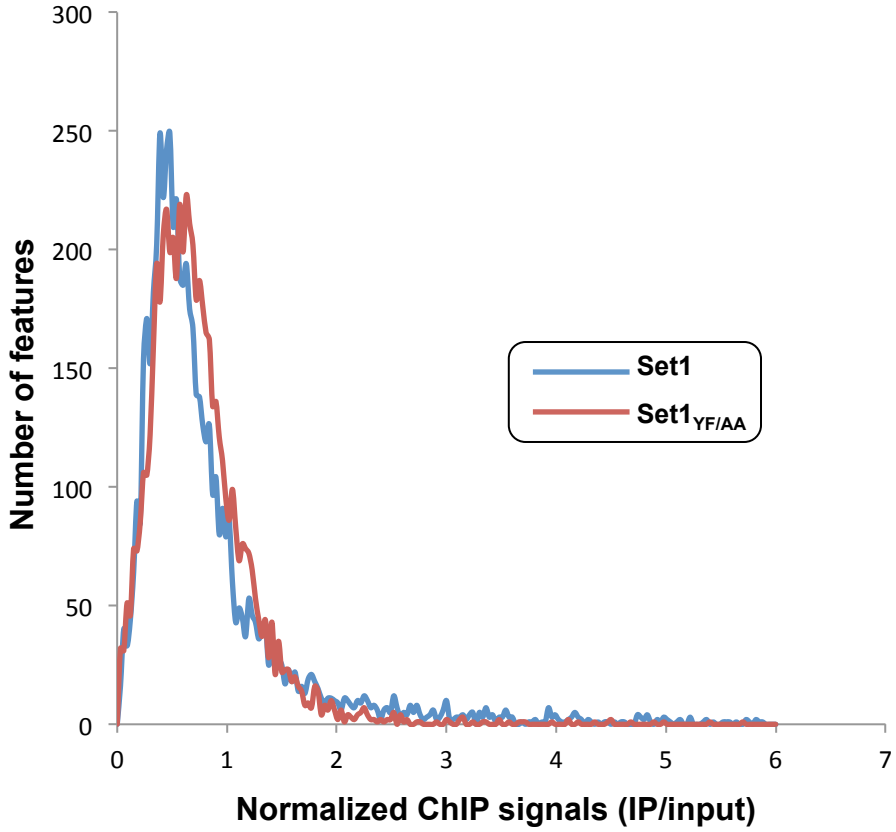

300
250
200
150
100
50
0
Number of features
0
1
2
3
4
5
6
7
Normalized ChIP signals (IP/input)
Set1
Set1YF/AA

Figure S4

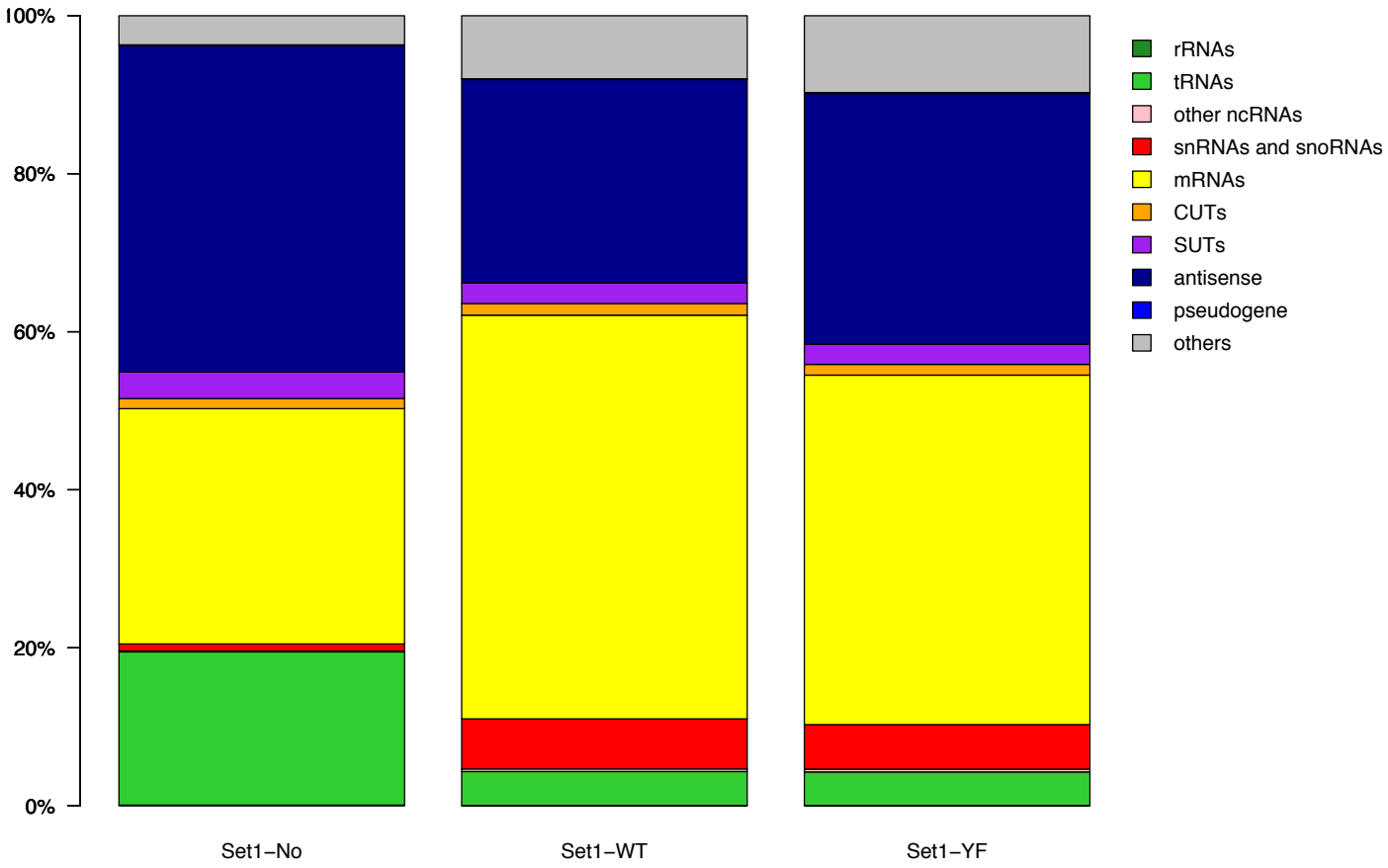

100%
80%
60%
40%
20%
0%
Set1−No
Set1−WT
Set1−YF
rRNAs
tRNAs
other ncRNAs
snRNAs and snoRNAs
mRNAs
CUTs
SUTs
antisense
pseudogene
others

Figure S5

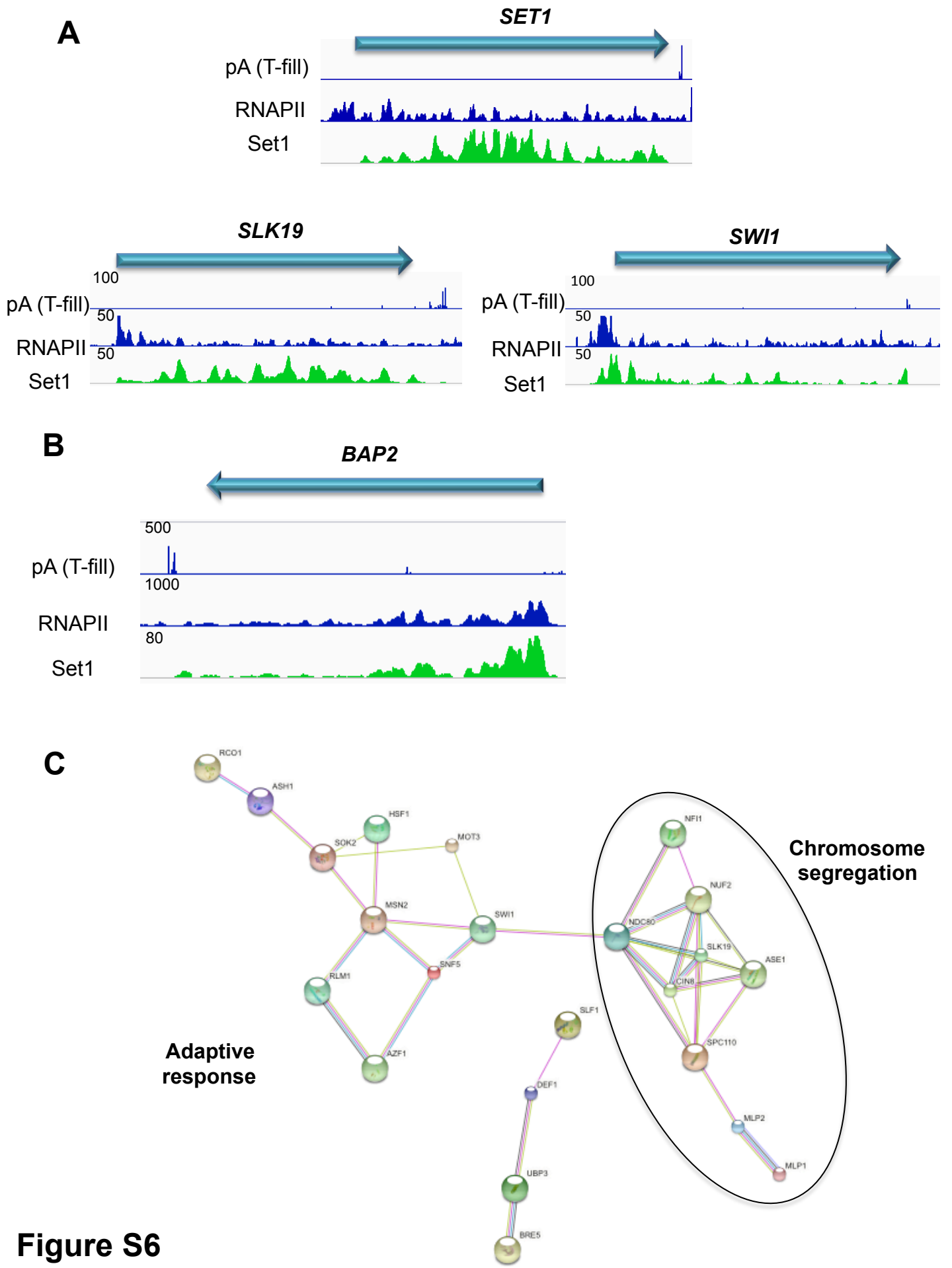

A
SET1
pA (T-fill)
RNAPII
Set1
SLK19
100
50
50
SWI1
100
50
50
B
BAP2
500
1000
80
C
RCO1
ASH1
HSF1
SOK2
MOT3
MSN2
SWI1
SNF5
RLM1
AZF1
SLF1
DEF1
UBP3
BRE5
NFI1
NUF2
NDC80
SLK19
ASE1
CIN8
SPC110
MLP2
MLP1
Chromosome segregation
Adaptive response

Figure S6

RESUME DE THESE

INTRODUCTION

Mes travaux de thèse ont porté sur le rôle d'une classe de facteurs de transcription appelée GRF (General Regulatory Factors) dans le contrôle de l'expression et de la fidélité de la transcription des gènes de classe II. Chez les eucaryotes, l'ARN polymérase II (ARNPII) est responsable de la synthèse des ARNs messagers (ARNm). Le processus de transcription de ces ARNs par la polymérase est communément divisé en trois grandes étapes : l'initiation, l'élongation et la terminaison de la transcription.

L'initiation de la transcription nécessite la formation du complexe de pré-initiation (PIC) composé de l'ARNPII et des facteurs généraux de la transcription. Au niveau des promoteurs, l'absence de nucléosomes et la présence de séquences spécifiques favorisent la reconnaissance du PIC et son assemblage. Une fois le complexe assemblé, la polymérase démarre la synthèse de la molécule d'ARN au cours du processus d'élongation. Cette étape fait intervenir de nombreux facteurs contribuant à l'avancer de la machinerie de transcription à travers la chromatine. Enfin, à l'extrémité 3' des gènes, des signaux de terminaisons présents le long de la molécule d'ADN sont transcrits et aident au recrutement de facteurs capables de promouvoir le démantèlement de l'ARNPII et la libération de l'ARN naissant. Chez la levure, le complexe CPF-CF (cleavage and polyadenylated factor – cleavage factor) est le principal acteur de la terminaison des ARNm. Il est notamment impliqué dans le clivage du transcrit en court de synthèse et sa polyadénylation. Une fois la maturation de l'ARNm achevée, il est exporté vers le cytoplasme pour être traduit.

En plus des ARNm, la polymérase II est impliquée dans la transcription de nombreux ARN non-codants (ARNnc) issus de la transcription dite « cachée » ou « pervasive ». Les CUTs

(cryptic unstable transcripts) sont parmi les premiers ARNnc de classe II découverts chez la levure *S. cerevisiae* et constituent la majeure partie de la transcription pervasive. D'autres ARNnc ont par la suite été décrits chez la levure (SUTs, XUTs, MUTs…) et diffèrent des CUTs de par leur métabolisme. De façon similaire aux ARNm, la transcription des ARNnc initie au niveau de régions dépourvues en nucléosomes (NDRs, nucleosome depleted regions). Une grande partie de la transcription cachée prend d'ailleurs naissance au niveau des promoteurs de gènes codant pour des ARNm. Ils peuvent chevaucher le gène voisin (même orientation) ou au contraire, être transcrits de façon divergente. La notion de promoteurs « bidirectionnels », c'est-à-dire capable de générer des évènements de transcription dans les deux orientations possibles, est particulièrement répandue et conservée de la levure à l'homme et est notamment considérée comme étant la source majeure de ces transcrits pervasifs.

Chez *S. cerevisiae*, les ARNnc sont hautement instables et dégradés dans le noyau des cellules. La dégradation rapide de ces transcrits est principalement associée avec une voie de terminaison spécifique nommée NNS (Nrd1, Nab3, Sen1). Tout comme la terminaison CPF-CF, le complexe NNS reconnait des signaux de terminaison présents sur l'ARN naissant conduisant finalement à la libération du transcrit. Le complexe NNS est couplé avec l'exosome nucléaire à travers de multiples interactions directes et indirectes aboutissant finalement à la polyadénylation et la dégradation (ou la maturation dans le cas des petits ARN nucléaires ou nucléolaires) des molécules d'ARNnc. Certains ARNnc peuvent être terminés par le complexe CPF-CF et sont alors dégradés dans le cytoplasme par la voie NMD (nonsense-mediated decay) principalement dédiée à l'élimination de transcrits avec un décalage du cadre de lecture ou présentant une région 3' UTR (utranslanted) anormalement longue.

De façon intéressante, le rôle des ARNnc chez *S. cerevisiae*, si tant est qu'il existe, reste énigmatique. En effet, contrairement à beaucoup d'eucaryotes et de levures proches, la levure de boulanger ne possède pas de machinerie permettant la formation de siRNA, miRNA ou

autre petit ARNs impliqués dans le phénomène d'interférence par l'ARN. En revanche, de nombreuses études ont démontré un rôle important de la transcription pervasive dans le contrôle de l'expression des gènes codant des protéines par un mécanisme d'interférence transcriptionnelle. En effet, lorsqu'une ARN polymérase transcrit au travers du promoteur d'un gène voisin, cela conduit à une réorganisation de la chromatine et une diminution de la transcription de ce dernier. De ce fait, il est important que la transcription pervasive soit contrôlée afin d'éviter la survenue de conflits entre des évènements de transcription se chevauchant, ou entre la transcription et d'autres machineries associées à l'ADN (réplication, réparation, …).

RESULTATS

I. Analyse de la terminaison « roadblock » à l'échelle du génome entier

Au cours de ma thèse, j'ai participé à l'étude d'une nouvelle voie de terminaison de la transcription par l'ARNPII appelée « roadblock ». La terminaison roadblock repose sur la capacité de certains activateurs transcriptionnels tels que Reb1 ou Rap1, à induire la terminaison en se liant à leur site de fixation sur l'ADN et en bloquant physiquement la progression du complexe d'élongation (Colin et al., 2014; Candelli et al., 2018). La collision entre l'ARNPII et Reb1 ou Rap1 induit un arrêt transcriptionnel qui est résolu par l'ubiquitinylation et la dégradation de la polymérase. Les ARNs ainsi terminés sont instables et dégradés dans le noyau par l'exosome nucléaire.

Afin de déterminer l'importance de la terminaison roadblock à l'échelle du génome, nous avons dans un premier temps travaillé sur l'optimisation d'une technique appelée « CRAC » (UV crosslinking and analysis of cDNA). Le CRAC consiste à réaliser un pontage aux UVs afin de lier de façon covalente les protéines et ARNs en contact dans la cellule (Granneman et al., 2009). Par la suite, un complexe d'intérêt est purifié via plusieurs étapes successives afin de récupérer et séquencer les ARNs. Dans notre cas, cette méthode a été utilisée dans le but

d'isoler les ARNs naissants associés à l'ARNPII afin d'évaluer la distribution de la polymérase le long du génome de *S. cerevisiae* et plus particulièrement au niveau des sites éventuels de roadblock.

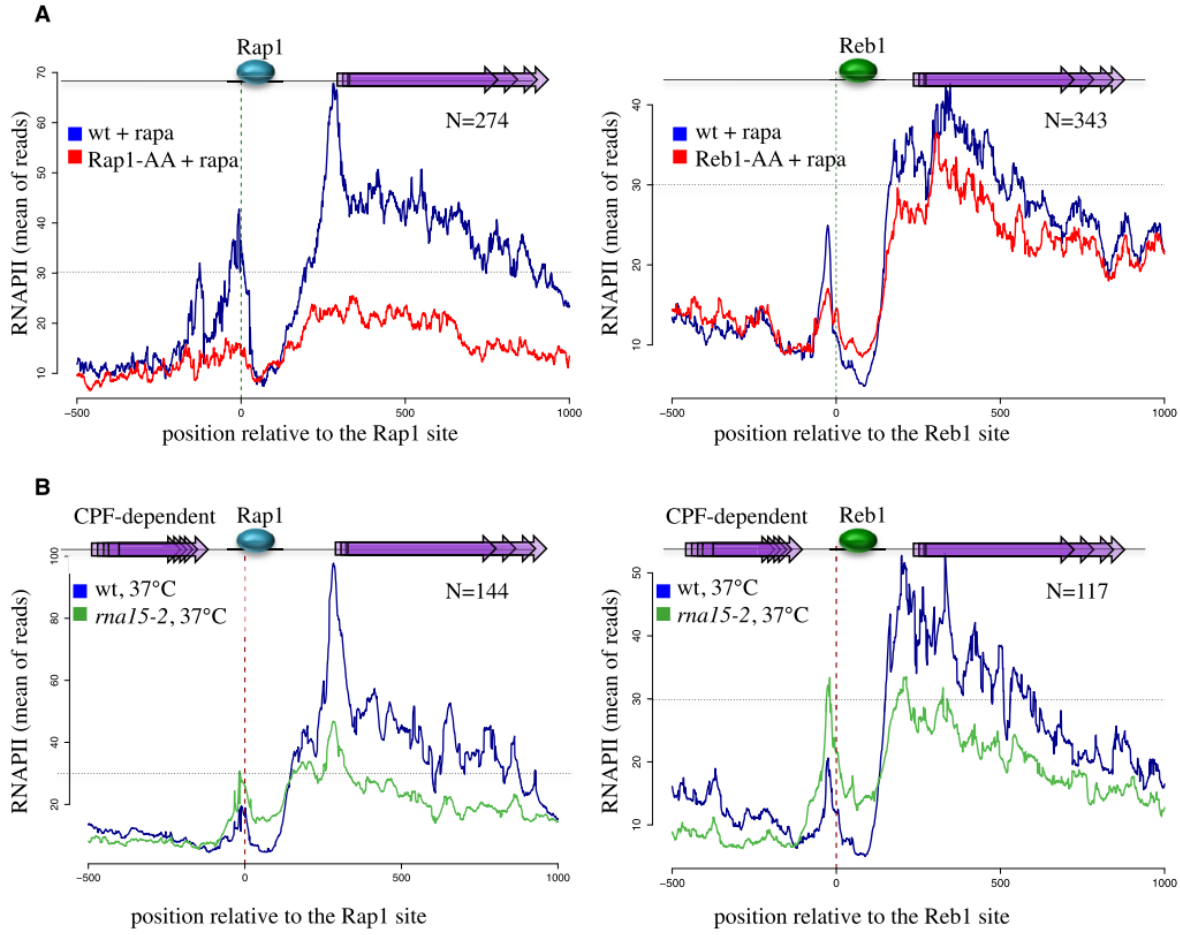


Figure 1. Analyse de la terminaison roadblock au niveau des sites Rap1 et Reb1.
A. Distribution de l'ARN polymérase 2 déterminée par CRAC au niveau de régions génomiques alignées aux sites Rap1 (gauche) ou Reb1 (droite) en présence (bleu) ou après déplétion nucléaire (rouge) des facteurs de roadblock. Les flèches en violet représentent des unités de transcription transcrites par l'ARNPII en aval des sites de fixation de Reb1 et Rap1. **B.** Comme (A) mais en considérant uniquement les sites Reb1 et Rap1 localisés dans une région de 300 pb en aval de sites de terminaison CPF-CF. La distribution de la polymérase a été déterminée dans une condition sauvage (bleu) ou dans la mutant thermosensible *rna15-2* (vert) a température non-permissive (37°C) pour le mutant. Figure issue de Candelli et al., 2018.

L'étude des données de CRAC a permis de détecter l'accumulation de l'ARNPII en amont de nombreux sites de fixations de Reb1 et Rap1 (Figure 1A). En accord avec le rôle de ces

facteurs dans la terminaison, leur déplétion nucléaire est associée à la disparition de la pause transcriptionnelle au profit d'évènements de translecture en aval des sites Reb1 et Rap1 (Candelli et al., 2018) (Figure 1A).

De nombreux sites de fixation pour Reb1 ou Rap1 sont localisés au niveau de régions intergéniques, en aval d'autres gènes. Dans une précédente étude il a été montré que dans ce contexte, la terminaison roadblock pourrait fonctionner comme mécanisme de secours afin de limiter la progression d'ARNPII n'ayant pas terminé de façon efficace au niveau des sites de terminaisons primaire (Colin et al., 2014). Néanmoins, seul un nombre de cas limité a été analysé. En étudiant la distribution de la polymérase au niveau des sites Reb1 et Rap1 localisés en aval des sites de terminaison canoniques (CPF-CF notamment), nous avons pu démontrer la généralité de ce mécanisme. En effet, dans ce contexte, une accumulation de polymérase peut être observée au niveau des sites de roadblock (Figure 1B). De plus, nous avons montré que l'augmentation des fuites transcriptionnelles (utilisation de mutants CPF-CF) s'accompagne d'un accroissement du signal de la polymérase en amont des sites Reb1 et Rap1. Cet afflux de polymérases, en augmentant localement la transcription en amont des sites de roadblock, a permis de faciliter la détection des pics de pause associés (Figure 1B). Cette étude nous a permis d'avoir une vision plus claire de l'importance de la terminaison par roadblock comme mécanisme de secours.

De façon intéressante, nous avons découvert que l'ARNPII s'accumulait en amont de nombreux autres facteurs de liaison à l'ADN tel que les activateurs transcriptionnels Abf1, Ume6 et Ste12. De plus, Cbf1 et TFIIB, impliqués dans la structure des centromères et l'activation des gènes de classe III (dépendants de l'ARNPIII) respectivement, sont également capable de limiter la progression de l'ARNPII *via* un mécanisme de roadblock. Ensemble, ces données suggèrent que la terminaison roablock pourrait avoir un rôle important dans la protection et le maintien de l'intégrité des processus associés à l'ADN.

II. Rôle des GRFs dans le contrôle de la fidélité de la transcription

Au cours de la précédente étude visant à étudier le rôle de Rap1 dans la terminaison roadblock, nous avons observé un phénomène intéressant se produisant en absence de Rap1. En effet, en plus de l'effet de répression ou d'activation de l'expression des gènes, nous avons découvert que la déplétion de Rap1 s'accompagne d'une augmentation du signal de l'ARNPII au niveau des région promotrices de nombreux gènes (Challal et al., 2018). Une analyse approfondie a révélé que ces évènements de modifications de l'initiation de la transcription se produisent pour environ 30% des cibles de Rap1 (Figure 2).

Afin d'analyser l'effet de la modification des sites d'initiation sur l'expression des gènes, nous avons déterminé la position des TSSs en présence ou absence de Rap1. En réalisant cette expérience dans des cellules délétées pour *UPF1* (impliqué dans la voie de dégradation NMD) nous avons montré que la majeure partie des ARNs produits à partir de sites ectopiques sont instables et dégradés dans le cytoplasme. Cette forte sensibilité de ces transcrits à la voie NMD est la conséquence de la présence de codons initiateurs situés en amont et entrainant ainsi un décalage du cadre de lecture ou la traduction de petites ORFs localisées en amont (uORF, upstream open reading frame). Cependant, un certain nombre de transcrits émanant de sites d'initiation ectopique s'avèrent être partiellement stables et probablement traduits. De façon générale, cette expérience révèle que le facteur de transcription Rap1 est capable de contrôler l'expression des gènes en partie en favorisant la transcription à partir de sites d'initiations appropriés.

De façon intéressante, dans un contexte d'initiation ectopique (en absence de Rap1), la quantité de polymérase (CRAC de l'ARNPII) ou la quantité d'ARN totale (RNA-Seq) ne semble pas toujours un bon indicateur de l'expression des gènes. En effet, des analyses du niveau de protéines (western blot) révèlent que beaucoup de gènes sont réprimés en absence de Rap1 malgré l'augmentation apparente de la quantité d'ARN et/ou d'ARNPII. Cette différence

observée s'explique notamment par l'existence de ces sites d'initiations ectopiques favorisant la production d'ARNs dont le potentiel codant est différent des ARNs issus des sites canoniques.

Comme précédemment indiqué, une grande partie des promoteurs des gènes sont bidirectionnels favorisant ainsi la production d'une multitude d'ARNs non-codants. De façon inattendue, la déplétion de Rap1 s'accompagne d'une forte augmentation de la transcription au niveau des régions intergéniques, favorisant la production d'ARNnc. Cette observation suggère que Rap1 joue un rôle important dans le contrôle de la bidirectionnalité des promoteurs en régulant le nombre d'ARN polymérases au niveau de ces régions. Cette fonction de Rap1 comme répresseur de la transcription cachée pourrait notamment être importante afin de limiter les évènements d'interférence transcriptionnelle pouvant impacter l'expression de gènes codants.

Afin de comprendre le mécanisme de régulation de l'initiation de la transcription par Rap1, nous avons cartographié la position des nucléosomes en présence ou absence de Rap1. En effet, Rap1 a été montré dans de nombreuses études comme étant important pour la formation des NDRs qui, elles même, favorisent l'initiation de la transcription. En comparant la position des nucléosomes avec celle des TSSs, nous avons montré qu'il existe une forte corrélation entre la réorganisation des nucléosomes en absence de Rap1 et l'émergence de nouveaux sites d'initiations ectopiques. Après déplétion de Rap1, la taille des NDRs liées par Rap1 est réduite. Cependant, la persistance de petites NDRs semble suffisante pour permettre à l'ARNPII d'initier de façon efficace. Ceci suggère que Rap1 n'est pas nécessaire au recrutement du PIC au niveau de ses gènes cibles et que le contrôle de l'initiation par Rap1 est étroitement lié à sa capacité à influencer le positionnement des nucléosomes.

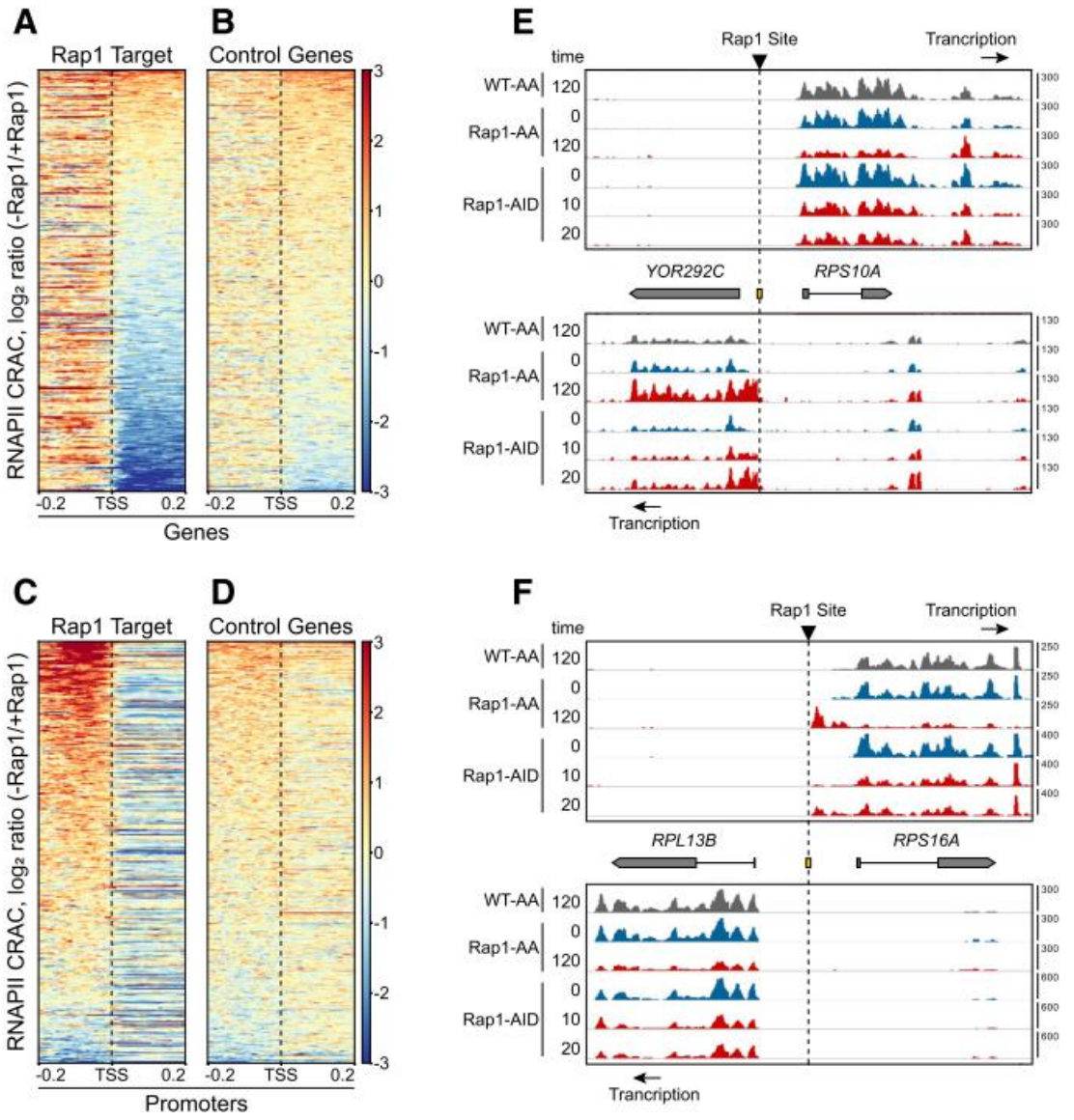


Figure 2. L'analyse de la distribution de l'ARNPII en absence de Rap1 révèle l'apparition d'évènements de transcription ectopique.

A-D. Heatmaps montrant la distribution de la variation du signal (ratio log_2) de l'ARNPII déterminée par CRAC au niveau de gènes contrôlés par Rap1 (n = 334) ou de gènes contrôles (n = 424) après 2h de déplétion de Rap1. Le signal est aligné sur les TSS (Transcription Start Sites) des gènes et est trié par ordre décroissant selon la valeur déterminée dans les corps du gènes (TSS à +200 pb) (A et B) ou dans la région promotrice (-200 pb au TSS). **E-F.** Distribution de l'ARNPII au niveau des locus *RPS10A-YOR292C* (en E) et *RPL13B-RPS16A* (en F). Le temps de déplétion de Rap1 est indiqué en minute pour le système anchor-away (Rap1-AA) ou le système dégron (Rap1-AID). Les flèches noires indiquent le sens de la transcription. Figure issue de Challal et al., 2018.

Comme beaucoup de régulateurs transcriptionnels, Rap1 possède un domaine de liaison à l'ADN (DBD) séparé du domaine de régulation (ou d'activation). Le domaine carboxy-terminal de Rap1 joue un rôle important dans le contrôle de l'expression des gènes. Il a notamment été montré comme étant important pour l'interaction et le recrutement de différents complexes incluant certains composants du PIC. Afin de déterminer si le domaine de régulation joue un rôle dans la répression des TSS ectopiques, nous avons exprimé une version tronquée de Rap1 ne comportant que le domaine de liaison à l'ADN. Dans ce contexte, nous avons analysé la position des nucléosomes et l'expression des gènes. De façon très surprenante, nous avons découvert que l'expression du DBD de Rap1 est suffisante pour rétablir presque complétement la position normale des nucléosomes ainsi que pour empêcher, dans de nombreux cas, la transcription à partir des sites ectopiques observée en absence de Rap1. De plus, l'expression des gènes pour lesquels la position des nucléosomes est rétablie en présence du DBD est également restaurée. Ces résultats démontrent qu'une grande partie de la fonction de Rap1 dans le contrôle de la position des nucléosomes, la régulation de l'expression des gènes et l'inhibition des TSSs ectopiques est portée par le domaine de liaison de l'ADN. A partir de ces résultats, nous avons proposé que Rap1 agit probablement par un mécanisme d'encombrement stérique en empêchant les nucléosomes d'envahir les régions promotrices.

La capacité de Rap1 à promouvoir la formation des NDRs a longtemps été attribuée à son aptitude à recruter les facteurs de remodelage de la chromatine au niveau des promoteurs. Parmi les différents remodeleurs, les complexes RSC et SWI/SNF ont été particulièrement étudiés. En effet, ces facteurs agissent au niveau de nombreuses NDRs en expulsant les nucléosomes des promoteurs. Afin de tester le lien entre Rap1, les complexes de remodelage de la chromatine et le choix du site d'initiation de la transcription, nous avons cartographié et comparé la position des TSSs en présence / absence de Rap1, de RSC (déplétion de Sth1) ou de SWI/SNF (déplétion de Snf2). De façon surprenante et en accord avec une récente

étude (Kubik et al., 2018), la déplétion de RSC et SWI/SNF a un effet moindre sur le déplacement des nucléosomes en comparaison avec la déplétion de Rap1. De plus, les évènements d'initiation ectopique observés en absence de Rap1 ne se retrouvent pas présents au sein des cellules pour lesquelles RSC et SWI/SNF ont été déplétés. L'ensemble de ces données indique que les facteurs de remodelage de la chromatine et Rap1 agissent de façon indépendante mais coordonnée afin de permettre le positionnement correct des nucléosomes situés de part et d'autre de la région promotrice. Ceci renforce également l'idée que Rap1 régulerait l'expression des gènes en partie par un mécanisme d'encombrement stérique.

Enfin, en analysant la distribution de l'ARNPII en absence d'autres facteurs de transcription (Reb1 et Abf1), nous avons montré que le rôle de Rap1 dans le contrôle de l'initiation de la transcription semble être une caractéristique partagée par d'autres GRFs. De plus, d'autres facteurs de liaison à l'ADN ont été récemment décrits comme étant important pour la régulation du choix du TSS et le maintien de la fidélité de la transcription chez différents organismes (Homme et *D. melanogaster* par exemple) (Oldfield et al., 2018 bioRxiv ; Lam et al., 2019).

CONCLUSION

Mes travaux de thèse ont permis de mettre en évidence de nouvelles fonctions de certains facteurs de transcription en lien avec l'expression des gènes et le maintien de la fidélité de la transcription. Nous avons notamment montré que la fixation des GRFs au niveau de leur site de liaison permet de prévenir de la progression de complexes d'élongation issus d'évènements de translecture en amont du GRFs. Les ARNs ainsi terminés sont instables et peuvent être considérés comme produits de la transcription cachée. En plus de contrôler la transcription cachée au niveau de la terminaison de la transcription, les GRFs ont aussi un rôle important dans la régulation de l'initiation. En effet, ils sont capables de limiter la quantité de polymérase impliquée dans la production d'ARNs non-codants issus des promoteurs bidirectionnels. En

agissant à la fois sur l'initiation et la terminaison de la transcription cachée les GRFs permettent très probablement de restreindre les effets potentiellement délétères de cette dernière en limitant les évènements de conflit entre la transcription RNAPII dépendante et d'autres évènements associés à l'ADN.

Finalement, nous avons également montré que les GRFs ont une fonction très importante dans le contrôle de la transcription des gènes codants et ce, non seulement sur le plan quantitatif en agissant sur le nombre de polymérases initiant à un endroit donné, mais également d'un point de vu qualitatif en favorisant l'utilisation de TSSs permettant la production de transcrits ayant un fort potentiel codant. De ce fait, les GRFs s'avèrent essentiels pour assurer une expression fidèle des gènes chez la levure *S. cerevisiae*.

Titre : Rôle des Facteurs de Transcription dans le Contrôle de l'Expression des Gènes et de la Fidélité de la Transcription

Mots clés : Facteurs de transcription, transcription pervasive, initiation, terminaison, contrôle qualité des ARNs

Ces dernières décennies ont été marquées par la découverte de la transcription dite « cachée » ou « pervasive ». Il a été en effet montré que la majeure partie du génome des eucaryotes est transcrite, donnant naissance à la formation de nombreux ARNs non-codants. La délimitation des unités de transcription apparait essentielle dans le contrôle de l'expression des gènes mais également dans le maintien de l'intégrité des processus associés à l'ADN en limitant notamment l'apparition de conflits avec la transcription. Dans ce contexte, l'initiation et la terminaison de la transcription représentent des étapes clés dans le partitionnement du génome et le métabolisme des ARNs. Nous avons montré que certains facteurs de transcription, appelés GRFs (General Regulatory Factors) chez la levure *S. cerevisiae*, jouent un rôle important dans le contrôle de la transcription pervasive à la fois au niveau de l'initiation mais également de la terminaison de la transcription et sont également requis pour assurer la fidélité de la transcription des gènes codant les ARN messagers.

Nous avons prouvé que les GRFs liés au niveau des régions promotrices sont capables d'induire la terminaison de la transcription en bloquant physiquement la progression d'ARN polymérases issues de la translecture des terminateurs situés en amont. D'après nos études, cette voie de terminaison appelée « roadblock » est très répandue à l'échelle du génome et joue un rôle important dans la protection des promoteurs contre l'interférence transcriptionnelle. Nous avons également découvert que les GRFs limitent la transcription pervasive en obstruant les sites d'initiations ectopiques situés à proximité de leur site de fixation sur l'ADN. Ces facteurs sont aussi impliqués dans le contrôle de l'expression des gènes codants en favorisant l'utilisation de sites d'initiations les plus appropriés, c'est-à-dire, permettant la synthèse d'ARNs ayant un fort potentiel codant. Le rôle des GRFs dans le contrôle de l'initiation apparait intimement lié à leur capacité à correctement positionner les nucléosomes au niveau des promoteurs en collaboration avec les facteurs de remodelage de la chromatine.

Title : Role of General Regulatory Factors in the Control of Gene Expression and Transcription Fidelity

Keywords : General regulatory factors, pervasive transcription, initiation, termination, RNA quality control

The last decades have been marked by the discovery of pervasive transcription. Indeed, many studies have shown that transcription by RNA polymerase II is not restricted to annotated regions but is widespread in eukaryotic genomes, leading to the production of a plethora of non-coding RNAs. Precise delimitation of transcriptional units appears to be essential to ensure robust fidelity of gene expression and to maintain the integrity of DNA-associated events by preventing the occurrence of conflicts with transcription. In this respect, accurate transcription initiation and termination represent crucial mechanisms to partition the genome and define the correct processing of RNA molecules. Here, we show that yeast general regulatory factors (GRFs), a class of highly expressed transcription regulators, control pervasive transcription at the level of initiation and termination and are also involved in the fidelity of initiation of mRNA-coding genes. We demonstrate that GRFs bound at promoter regions can elicit transcription termination by physically impeding the progression of polymerases mainly deriving from readthrough transcription at upstream canonical termination sites. We provide evidence that this termination pathway named roadblock is widespread throughout the yeast genome and protects promoter regions from transcriptional interference. Furthermore, we establish that the presence of general regulatory factors limits pervasive transcription at the level of initiation, notably by occluding spurious transcription start sites present in the vicinity of their binding sites. We also unveil the importance of these factors in promoting correct transcription start site selection at mRNA-coding genes thus favouring the synthesis of transcripts with an appropriate coding potential. Finally, we determine that the role of GRFs in controlling proper initiation is intimately linked to their ability to correctly position nucleosomes in promoters, a role that occurs independently from but in cooperation with chromatin remodelers.