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

Genome-wide occupancy profiling of the RNA polymerase II transcription machinery in *S. cerevisiae*

Michael Maximilian Lidschreiber
aus
München, Deutschland

2013

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 12.08.2013

Michael Maximilian Lidschreiber

Dissertation eingereicht am 12.08.2013

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

Mündliche Prüfung am 03.09.2013

Acknowledgments

Finding the right PhD position was not an easy task. It was about finding the perfect balance between bench work and bioinformatics, and having the freedom to work on my own yet getting helpful feedback and support whenever needed. On top of that, I was looking for interesting and challenging projects in an excellent scientific environment.

I would like to express my deep gratitude to you, Patrick Cramer, for giving me the opportunity to work in your lab and for giving me everything that I was looking for. You were an outstanding supervisor and mentor, and I am especially grateful for the trust you placed in me and my capabilities.

I would also like to thank my second research supervisor, Dietmar Martin, for your ongoing interest in my research work and your enlightening feedback after my presentations.

I wish to express my gratitude to Johannes Söding, Ulrike Gaul, Dirk Eick, Klaus Förstemann and Tobias Straub for their time and support as part of my advisory and examination committees.

I especially thank Andreas Mayer for all the fruitful collaborations (I'm sure there will be more to come!). I couldn't have done it without you. We were a perfect team, like ChIP and chip, or to quote Patrick: "a win-win team". But the team wouldn't be complete without Matthias Siebert. I'm deeply grateful for all your help in establishing our ChIP-chip analysis pipe-line and for the many nights we spent dancing away to the finest drum and bass tunes. It was a great pleasure working with both of you!

I am particularly grateful for the excellent technical assistance given by Kristin Leike. You supported me from the beginning to the end. Thank you for the great work and also for plenty of non-science related chats during our daily coffee breaks.

I also would like to thank the computational biology team, especially Johannes Söding for supervising ChIP-chip data evaluation, and Achim Tresch, Julien Gagneur and Bene Zacher for their efforts to establish novel tools for genome-wide ChIP data analysis based on our experiments. Thanks also go to Björn Schwalb for all his advices on R coding, as well as Philipp Eser for sharing our cosy office, tolerating the mess ... and the many discussions we had about systems biology and the life beyond science.

Likewise, I would like to thank Amelie Schreick and Stefanie Etzold for their considerable help, discussions, and collaborations. Being part of your quest to identify the CTD tyrosine phosphatase (together with Lori Passmore and her group) is very exciting.

I want to thank Dirk Eick and his team for collaborating with us and for their constant efforts (together with Elisabeth Kremmer) to generate the best antibodies for studying the Pol II CTD.

I especially would like to thank Alan Cheung for constant scientific advice, proofreading parts of this thesis, and sharing the pink tent at THE festival (an experience, I'm sure, even Patrick is jealous of).

I also thank Kerstin Maier. You took care of me when I started my PhD studies and introduced me to the yeast world. I truly appreciate all your help over the years, including several incidents of microarray troubleshooting and proofreading parts of this thesis. I loved our "special" dog walking excursion.

I am thankful to Mai Sun for his help on cDTA and plenty of discussions, as well as great fun during the SFB646 retreat in Bad Wiessee.

I would like to extend my thanks to Claudia Buchen and Stefan Benkert for taking good care of the lab and keeping everything running.

Advice given by Heidi Feldmann has been a tremendous help especially during the revision of my last project.

Thanks go also to all the other present and former members of the Cramer lab for providing an extraordinary working atmosphere. It was always a great pleasure working with all of you!

I would also like to thank Dominik Meinel from Katja Sträßer's group for helpful discussions and for providing yeast strains and antibodies.

Additionally, many thanks go to the coordinators of the International Max Planck Research School for Molecular and Cellular Life Sciences, the International Doctorate Program NanoBioTechnology, and the SFB TR5 PhD program. I was very lucky to profit from a plethora of interesting workshops and seminars offered by these programs. Moreover, they provided financial support for external workshops and scientific training.

I am very thankful to all my friends, who supported me or influenced me along the way.

My eternal gratitude goes to my beloved girlfriend Katja Frühauf, who proofread this entire thesis and who supports me in everything that I do.

Above all, I want to thank my family for their everlasting appreciation, love, and support during my studies and my PhD work.

Summary

Transcription of protein-coding genes by RNA polymerase (Pol) II begins with the assembly of an initiation complex at the promoter. As the RNA is synthesized, initiation factors dissociate from Pol II, and elongation factors are recruited. Whereas Pol II is unphosphorylated during initiation, it is phosphorylated at its C-terminal repeat domain (CTD) during elongation. The CTD forms a flexible extension of Pol II that consists of 26 (yeast) or 52 (human) heptapeptide repeats of the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. Changes in CTD phosphorylation coordinate the recruitment of mRNA transcription and processing factors to Pol II. Ser5 phosphorylation recruits the mRNA 5′ capping machinery near the promoter, whereas Ser2 phosphorylation recruits mRNA 3′ processing and termination factors near the polyadenylation (pA) site.

This work presents genome-wide chromatin-immunoprecipitation (ChIP) occupancy profiles for Pol II, its phosphorylated forms and transcription factors in proliferating budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) cells. Pol II exchanges initiation factors for elongation factors during a 5′ transition that is completed about 150 nucleotides (nt) downstream of the transcription start site (TSS). The resulting elongation complex is composed of all elongation factors at all genes and shows high levels of Ser7 and Ser5 phosphorylation on the CTD. Ser7 and Ser5 phosphorylation levels then decrease over the transcribed region, whereas Tyr1 and Ser2 phosphorylation levels increase and saturate at 600–1,000 nt downstream of the TSS. Elongation complexes disassemble in a two-step 3′ transition. Bur1, Ctk1, Paf1 and Spt16 exit upstream of the pA site, whereas Elf1, Spt4, Spt5, Spt6 and Spn1 exit downstream. Peak levels for mRNA 3′ processing and termination factors are observed downstream of the pA site.

After defining the transitions in Pol II complex composition along genes, we further elucidated the underlying molecular mechanisms that are responsible for transcription factor exchange. First, we show that during the 5′ initiation-elongation transition the early Pol II elongation factor Spt5 contributes to gene recruitment of the mRNA capping enzymes Cet1, Ceg1 and Abd1. Genome-wide occupancy for Cet1 and Ceg1 is restricted to the TSS, whereas occupancy for Abd1 peaks ∼110 nt downstream, and occupancy for the cap-binding complex (CBC) rises subsequently. Abd1 and CBC are shown to be important for recruitment of the kinases Ctk1 and Bur1, which promote elongation and capping enzyme release. Second, we show that Tyr1 phosphorylation of the CTD impairs recruitment of termination factors during elongation. About 180 nt upstream of the pA site, Tyr1 phosphorylation levels drop, whereas Ser2 phosphorylation levels remain high. This enables recruitment of termination factors Rtt103 and Pcf11, resulting in mRNA 3′ processing and transcription termination.

Taken together, our results point to a general Pol II transcription complex that mediates transcription and mRNA processing at all actively transcribed Pol II genes in yeast. At both ends of genes this complex undergoes uniform transitions that are independent of gene length, gene type or expression level. Moreover, our results indicate that mRNA 5′ cap completion underlies the initiation-elongation transition and triggers productive Pol II elongation and that CTD Tyr1 dephosphorylation underlies the elongation-termination transition and triggers mRNA 3′ processing and termination.

Publications

Parts of this work have been published or are in the process of publication :

2013 **Cap completion and CTD kinase recruitment underlie the initiation-elongation transition of RNA polymerase II**
M. Lidschreiber, K. Leike, and P. Cramer
Molecular and Cellular Biology 13 (Epub ahead of print, July 22).

Author contributions: ML and KL carried out experiments, ML analyzed ChIP-chip data, ML and PC prepared the manuscript.

2013 **RNA polymerase II termination involves CTD tyrosine dephosphorylation by CPF subunit Glc7**
A. Schreieck*, A. Easter*, S. Etzold, K. Wiederhold, **M. Lidschreiber**, P. Cramer, and L. Passmore (* contributed equally)
Manuscript submitted.

Author contributions: AS, AE, SE, and KW performed experiments. ML analyzed data. AS, LP, and PC wrote the manuscript, with help from all authors.

2012 **CTD tyrosine phosphorylation impairs termination factor recruitment to transcribing RNA polymerase II**
A. Mayer*, M. Heidemann*, **M. Lidschreiber**, A. Schreieck, M. Sun, C. Hintermair, E. Kremmer, D. Eick, and P. Cramer (* contributed equally)
Science 336, 1723.

Author contributions: AM carried out ChIP-chip and fluorescence anisotropy experiments. MH validated antibodies. ML analyzed ChIP-chip data. MS carried out modelling. AS carried out additional ChIP assays. EK generated the 3D12 antibody. AM and PC prepared the manuscript, with help from all authors.

2012 **The Spt5 C-terminal region recruits yeast 3´-RNA cleavage factor I**
A. Mayer*, A. Schreieck*, **M. Lidschreiber**, K. Leike, D. Martin, and P. Cramer (* contributed equally)
Molecular and Cellular Biology 32, 1321.

Author contributions: AM and AS carried out experiments. ML analyzed ChIP-chip data. AM, AS and PC prepared the manuscript, with help from all authors.

2010 **Uniform transitions of the general RNA polymerase II transcription complex**
A. Mayer*, **M. Lidschreiber***, M. Siebert*, K. Leike, S. Söding, and P. Cramer (* contributed equally)
Nature Structural and Molecular Biology 17, 1272.

Author contributions: AM established experimental protocols. AM, ML, and KL performed experiments. ML and MS evaluated data sets. JS designed and supervised data evaluation. PC designed and supervised research and wrote the manuscript, with help from all authors.

Contents

Part I
Introduction

Transcription represents one of the most fundamental biological processes, during which DNA is transcribed into RNA by DNA-dependent RNA polymerases [229]. In eukaryotes exist three nuclear RNA polymerases: RNA polymerase (Pol) I, Pol II, and Pol III [36]. Pol I synthesizes most of the ribosomal RNAs (rRNAs), whereas Pol III synthesizes transfer RNAs (tRNAs), 5S rRNA, and other small RNAs. Pol II produces messenger RNAs (mRNAs), as well as several non-coding RNAs, including small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs). Recent genome-wide transcriptome analyses revealed that Pol II transcription is pervasive and not limited to protein-coding and sno/snRNA genes [89]. These studies revealed that up to 80% of the yeast [44, 159] and human [12, 102] genomes are transcribed, identifying hundreds of novel non-coding RNAs. In yeast, stable unannotated transcripts (SUTs, [242]) exist stably in cells, while cryptic unstable transcripts (CUTs, [242]) and Xrn1-sensitive unstable transcripts (XUTs, [223]) are rapidly degraded after synthesis. To what extent these novel transcripts or their synthesis are biologically functional [228], or whether the majority simply accounts for transcriptional noise [211], is still unclear.
Pol I, II and III are multi-subunit enzymes comprised of a conserved ten-subunit core and up to seven addtional subunits [36]. Pol II is composed of 12 subunits, termed Rpb1 – 12, with Rpb1 being the largest. Rpb1 has evolved a highly repetitive C-terminal domain (CTD) which is unique to Pol II and plays a central role in transcription mediated by its dynamic phosphorylation [77].

1 mRNA synthesis by RNA polymerase II

The first step of Francis Crick's central dogma of molecular biology [39], the transcription of protein-coding genes into mRNA, is catalyzed by RNA polymerase (Pol) II with the help of many auxiliary factors in a repetitive, cyclic process, named the transcription cycle (Section 1.1). In eukaryotes, the genetic information contained in the DNA is first transferred to pre-mRNA that must be further co-transcriptionally processed by RNA processing factors (Section 1.2). Correct processing and export to the cytoplasm enable translation of the mature mRNA by the ribosomes.

1.1 The RNA polymerase II transcription cycle

The main phases of the Pol II transcription cycle consist of initiation and promoter clearance (Section 1.1.1), elongation (Section 1.1.2), as well as termination and re-initiation (Section 1.1.3) of transcription.

1.1.1 Initiation and promoter clearance

Transcription of protein-coding genes by Pol II begins with the assembly of a pre-initiation complex (PIC) at the promoter. The PIC consists of general transcription factors (GTFs) that recognize core promoters and recruit Pol II [71, 205]. Moreover, GTFs interact with coactivators like Mediator or SAGA to modulate transcription. The GTFs TBP, TFIIA and TFIIB form a complex with promoter DNA. TFIIB, along with TFIIF, is critical for subsequent Pol II recruitment and positioning of the promoter DNA above the Pol II cleft. The remaining factors, TFIIE and TFIIH are required for separation and stabilization of promoter DNA strands, leading to formation of the open complex that allows the single-stranded DNA template to enter the active site of

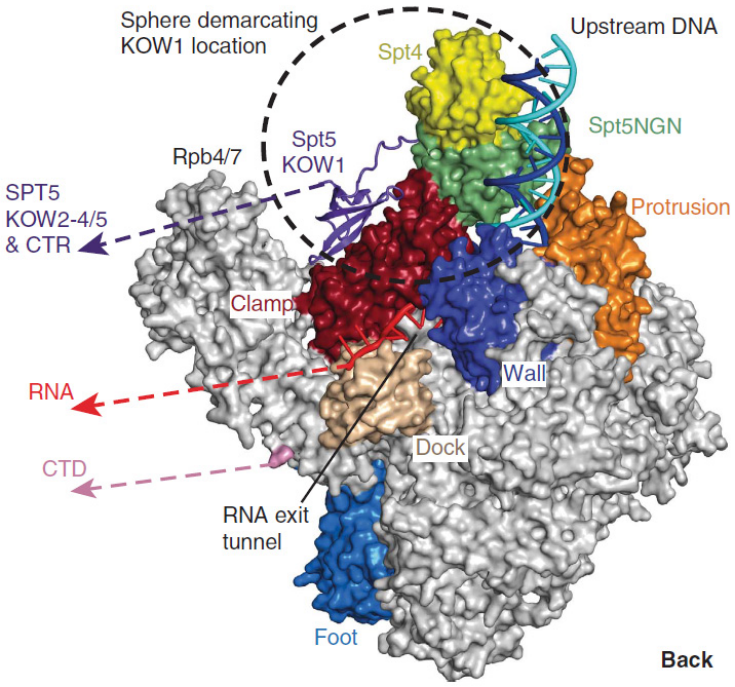

Figure 1: Structure of yeast Spt4-5 bound to Pol II. The complete yeast Pol II elongation complex with bound Spt4-5 is viewed from the back [141]. Pol II and Spt4-5 are shown as molecular surfaces with key domains highlighted in color. Exiting RNA, the C-terminal KOW domains and C-terminal repeat region of Spt5 (CTR), and the Pol II largest subunit C-terminal domain (CTD) extend towards the same side around the Rpb4/7 subcomplex.

Pol II [105]. The subunit of TFIIH that is required for the structural remodeling of the PIC during this process is Rad25 (human XPB), an ATP-dependent DNA helicase. The complex then scans for the transcription start site (TSS) downstream and initiates transcription, leading to the initially transcribing complex (ITC) [26]. Recent structural data propose additional roles for TFIIB in DNA opening and start site recognition via the 'B-linker' and 'B-reader' domains [115]. The ITC is unstable and the transition into a stable elongation complex is prone to abortive initiation and release of short RNAs [137]. As the nascent RNA grows, the complex stabilizes, disengages from promoter DNA, and initiation factors dissociate from Pol II. This process is referred to as promoter clearance or promoter escape [137]. Subsequently, elongation factors are recruited [167, 168, 176]. The initiation-elongation transition begins when the nascent RNA grows to about 12 nucleotides (nt), which releases initiation factor TFIIB [169, 194]. Release of initiation factors frees the Pol II clamp domain, which can then bind the conserved elongation factor Spt5 [65, 141] (Figure 1).

1.1.2 Elongation

After the transition into a productively transcribing elongation complex (TEC), Pol II is assisted by a number of proteins (elongation factors, EFs) to facilitate efficient gene transcription [195, 206]. Some of these EFs can influence the rate and processivity of Pol II directly (e.g. TFIIS [104]), whereas others are needed to transcribe efficiently through the chromatin environment (see below). Furthermore, some EFs are needed to facilitate co-transcriptional recruitment of RNA processing factors (see below and Section 1.2). The major EFs in *S. cerevisiae* which are of particular interest for this work are listed in Table 1 and are further described in the following paragraphs:

The conserved elongation factor Spt5 The essential elongation factor Spt5 [236], which directly binds the Pol II clamp domain [65, 141] (Figure 1) and forms the Spt4-Spt5 complex [75] (human DSIF, [227]), is of particular importance as it is the only known Pol II-associated factor that is conserved in all three kingdoms of life [231]. The core function of Spt4-5 is to stimulate transcription elongation and Pol II processivity [74]. Spt5 copurifies with over 90 yeast proteins that are involved in transcription elongation, RNA processing, transcription termination, and mRNA export [130]. Eukaryotic Spt5 possesses a repetitive C-terminal region (CTR) [215, 254] that has

Factor	Function
Bur1	Cyclin (Bur2)-dependent kinase; phosphorylates Pol II CTD and Spt5 CTR
Ctk1	Catalytic subunit of CTD kinase I (CTDK-I); phosphorylates serine 2 on the Pol II CTD
Elf1	Implicated in maintaining proper chromatin structure during elongation
PafC	Modulates H2B monoubiquitination and H3K4 methylation; recruits Set1 and Set2
Spn1	Associates with Spt6
Spt4-5	Stimulates elongation and capping; suppresses early transcript termination
Spt6	Stimulates elongation; histone chaperone activity; modulates chromatin structure
Spt16	Facilitates elongation through chromatin; histone chaperone activity

Table 1: Transcription elongation factors of RNA polymerase II in *S. cerevisiae*.
See text for references and details. This list contains only elongation factors that are further investigated in this work (Section 6). For a complete list of elongation factors and their functions refer to [195, 206].

been shown to act as a platform for recruiting accessory factors to elongating Pol II. Spt5 interacts with capping enzymes [130, 173, 198, 230] and the CTR recruits the Paf complex [132, 254] (see below). Spt4-5 also interacts with the elongation factor and histone chaperone Spt6 to modulate chromatin structure [75, 216] (see below).

Elongation through chromatin requires the Paf complex, FACT, and Spt6 Chromatin is a nucleoprotein structure built from nucleosomes that contain 147 bp of DNA wrapped around an octamer of histone proteins [126]. The histone octamer consists of two copies of each histone protein H2A, H2B, H3, and H4. The chromatin architecture represents a barrier to Pol II transcription, and thus needs to be overcome during transcription elongation with the help of auxiliary factors.
The Paf complex (PafC), consisting of Paf1, Ctr9, Cdc73, Rtf1 and Leo1 in yeast, is a conserved elongation factor [90, 221]. Together with Spt4-Spt5, PafC is involved in recruitment of chromatin factors during early transcription elongation. These factors include the histone chaperones Spt6 and FACT (facilitates chromatin transcription), as well as the histone H3K4 methyltransferase Set1 complex (COMPASS). The FACT complex (consisiting of Spt16 and Pob3 in yeast) [10, 166, 197, 237] and Spt6 [16, 101], help Pol II transcribe through nucleosomes by displacing histones and rapidly reassembling them after Pol II passage. Maintaining proper chromatin structure during elongation is important, as mutations in FACT or Spt6 lead to transcription initiation from cryptic start sites, which are exposed within gene bodies when redeposition of histones fails [101, 142]. Spt6 was identified in the same genetic screen as Spt4-5 and Spt16 [236], and physically interacts with Pol II, as well as elongation factors Spn1, Spt5, and Spt16 [117, 130, 216].
Methylations of histones H3K4 and H3K36 by the methyltransferases Set1 and Set2, respectively, represent elongation-related histone modifications observed at actively transcribed genes [177, 207]. During early elongation Set1 is recruited to the serine 5 phosphorylated CTD of Pol II, whereas Set2 is recruited later during elongation via binding also to the serine 2 phosphorylated CTD (see Section 1.2.1 for details). Proper H3K4 and H3K36 methylation patterns are important to recruit further histone acetylation and chromatin remodeling complexes during elongation [126, 207].

Pol II CTD kinases Bur1 and Ctk1 Differential phosphorylation of the Pol II CTD during the transcription cycle plays a key role in co-transcriptional coupling of pre-mRNA processing [77, 86, 250] (Section 3). Responsible for the changing phosphorylation patterns of the CTD are CTD kinases and phosphatases, many of which act during transcription elongation. Serine 2 phosphorylation of the CTD represents the hallmark of productively elongating Pol II and is catalized by the Bur1 [182] and Ctk1 [28] kinases. Dynamic phosphorylation of the CTD as well as detailed information on CTD function are discussed in Section 1.2.1.

1.1.3 Termination and re-initiation

Transcription termination of protein-coding genes requires several protein complexes to bind to Pol II as well as to specific sequences in the nascent mRNA when the transcription complex reaches the polyadenylation (pA) site [138, 155, 189]. The pA site marks the point of RNA 3´-cleavage and polyadenylation (see Section 1.2.4 for details), but transcription continues beyond this site until termination where Pol II is released from the DNA template over a range of about 200 bp (1500 bp in humans) downstream. Initially, two models have been proposed how Pol II iss dismantled from the DNA during termination. In the 'anti-terminator' model, transcription of the pA site triggers a change in the Pol II machinery that allows for termination [193]. In the 'torpedo' model, pA-dependent RNA cleavage results in a new RNA 5'-end that is recognized by the Rat1/Rai1/Rtt103 exonuclease complex that degrades nascent RNA and triggers termination [109, 233]. However, recent data indicate that transcription termination may be a combination of both models [234].
Transcription is a cyclic process that can be initiated again after termination either by recruitment of the complete transcription machinery to the gene promoter or by re-initiation. Re-initiation is facilitated by a subset of initiation factors, including TFIIA, -D, -E, -H and Mediator, which remain bound to the promoter after the first round of trancription initiation, forming a so-called scaffold complex [247]. This scaffold complex then enables rapid PIC formation during successive rounds of transcription.

1.2 Coordination of transcription and pre-mRNA processing

Processing of the newly transcribed pre-mRNA occurs co-transcriptionally and starts as soon as the nascent transcript emerges from the RNA exit tunnel of Pol II [37, 139]. The CTD of Pol II plays a pivotal role in this process (Section 1.2.1, Figure 2). It couples the three main pre-mRNA processing events, which are 5'-capping (Section 1.2.2), splicing (Section 1.2.3) and 3'-end formation (Section 1.2.4) to the three main phases of the transcription cycle (Section 1.1), initiation, elongation and termination, respectively (Figure 3) [77, 86, 250]. This spatio-temporal coupling allows pre-mRNA processing events to occur in an ordered manner. Moreover, pre-mRNA processing events are interdependent, as capping is required for splicing [202], and splicing enhances 3'-end formation [135, 164] and vice versa [222].

1.2.1 The CTD code

A unique feature that distinguishes Pol II from the other two eukaryotic polymerases is the presence of an extended carboxy-terminal domain (CTD) at Rpb1, the largest of the 12 subunits of Pol II [36]. The CTD consists of 26 (yeast) or 52 (human) highly conserved heptad repeats of the consensus Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. In an extended conformation the CTD has a length of over four times the diameter of the core polymerase (Figure 2A). The CTD becomes differentially phosphorylated and dephosphorylated at five of the seven residues (Tyr1 (Y1), Ser2 (S2), Thr4 (T4), Ser5 (S5), Ser7 (S7)) in coordination with the transcription cycle (Figure 2B) [77, 86, 250]. Furthermore, threonine and the serines can be glycosylated, and the prolines can undergo cis-trans isomerization.
The writers and erasers of the CTD phosphorylation marks are cyclin-dependent kinases and phosphatases, respectively (Figure 2B) [92]. Whereas Pol II is unphosphorylated during PIC assembly [134], it is then phosphorylated at S5 and S7 during initiation, primarily by the Kin28 (human Cdk7) subunit of TFIIH [110]. Srb10 (human Cdk8) also contributes to basal S5 phosphorylation. During elongation, S2 gets phosphorylated by Bur1 [182] and Ctk1 [28] which enables efficient RNA elongation and 3'-processing [3, 103]. Bur1 has also been shown to be responsible for additional S7 phosphorylation during elongation [218]. Bur1 and Ctk1 share homology with mammalian Cdk9, the kinase of the P-TEFb complex that triggers the transition

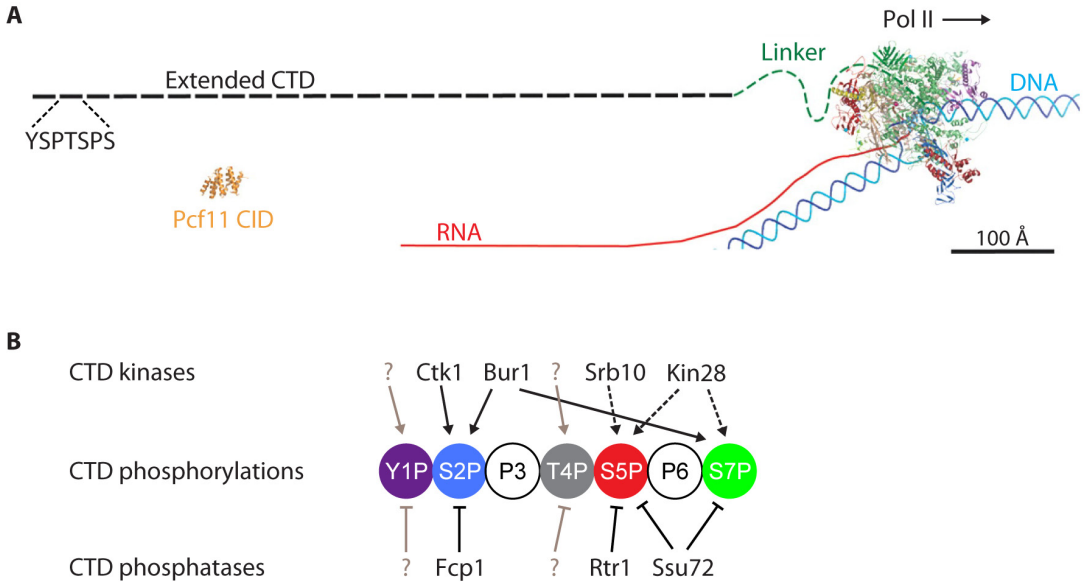

Figure 2: RNA polymerase II C-terminal domain (CTD) structure and phosphorylation marks. (A) Illustration of the relative length between the CTD in fully extended conformation and the core Pol II adapted from [151]. Each black bar represents one of the 26 YSPTSPS heptad repeats of the *S. cerevisiae* CTD. For size comparison, the structure of the CTD-interacting domain (CID) of termination factor Pcf11 is depicted in gold [150]. (B) The CTD can be phosphorylated at five of the seven repeat residues (Y1, S2, T4, S5, S7). Respective kinases and phosphatases identified in *S. cerevisiae* are displayed above and below the YSPTSPS heptad repeat, respectively. Kinases which act during initiation and elongation are depicted using dashed and solid arrows, respectively. For CTD phosphorylation marks the same color-code is used throughout this work.

into productive elongation [25, 140, 163]. For a long time it was believed that Cdk9 is the major contributor to S2 phosphorylation in higher eukaryotes, but recent evidence suggests that the fly and human Cdk12 proteins also contribute to S2 phosphorylation, and that Bur1 and Ctk1 might be the orthologs of Cdk9 and Cdk12, respectively [7]. Dephosphorylation of the CTD by specific phosphatases is less understood, but the current knowledge is that Fcp1 preferentially dephosphorylates S2 [29, 113], while Rtr1 and Ssu72 are required for S5 dephosphorylation during elongation and at termination, respectively [151, 156]. It has to be noted that the role of Rtr1 in S5 dephosphorylation is rather controversial [240]. Moreover, it was recently shown that Ssu72 also dephosphorylates S7 at the 3'-ends of genes [9]. Complete dephosphorylation of the Pol II CTD after termination is also critical for Pol II recycling [29]. The yeast kinases and phosphatases required for Y1/T4 phosphorylation and dephosphorylation remain to be identified.

The readers of this proposed "CTD code" are transcription and mRNA-processing factors with binding preferences for different phosphorylation patterns (Figure 3) [92]. During early transcription S5 phosphorylation recruits the mRNA capping enzymes, early termination factor Nrd1, and the Set1 histone methyltransferase [92, 114, 199, 224]. S2 phosphorylation occurs during transcription elongation and functions, for example, in recruitment of chromatin modifying factors like Spt6 [213], as well as RNA 3´-processing and termination factors Pcf11 and Rtt103 [3, 150]. Interaction between the CTD and Spt6 is mediated by the C-terminal tandem SH2 domain of Spt6 [213], whereas termination factors Nrd1, Pcf11, and Rtt103 contact the CTD via a conserved CTD-interacting domain (CID) [150, 224]. Phosphorylations at S7 [22, 52, 110] and T4 [87] have roles in processing of specific RNAs. Y1 phosphorylation was described for human Pol II almost two decades ago [8], but whether this has a functional role and whether it exists in other species is unknown.

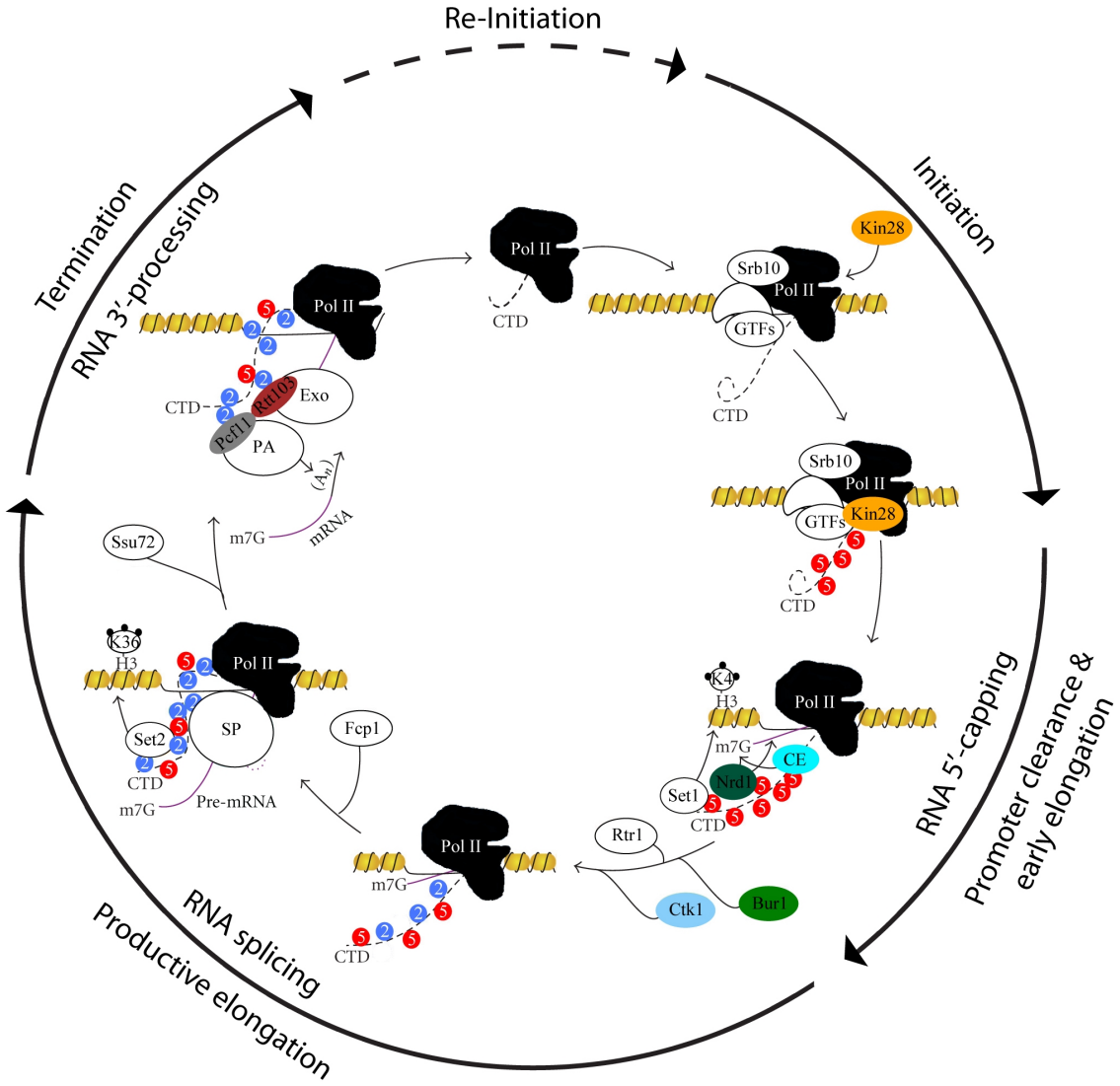

Figure 3: Coordination of co-transcriptional pre-mRNA processing with the transcription cycle.
Figure adapted from [250]. The main phases of the transcription cycle and the coupled pre-mRNA processing steps are indicated outside and inside the black arrows, respectively. For factors and Pol II CTD phosphorylation marks that are further investigated in this work the same color-code is used throughout. For details and references see text. Briefly, non-phosphorylated Pol II assembles at the preinitiation complex (PIC) together with Mediator and general transcription factors (GTFs). TFIIH associates last and its associated kinase Kin28 phosphorylates Ser5 (shown in red) and Ser7 (not shown) on the CTD enabling promoter clearance. Mediator-associated kinase Srb10 also contributes to basal levels of Ser5 phosphorylation. The Ser5P mark is important for recruitment of 5'-capping enzymes (CE), CTD-interacting domain (CID)-containing termination factor Nrd1, and Set1 histone methyltransferase. Ser5P also contributes to initial recruitment of the Bur1 kinase. Bur1 places promoter-proximal Ser2P marks (shown in blue), which contributes to recruitment of the Ctk1 kinase. Ctk1 is the primary Ser2 kinase in yeast, and its phosphorylation is required for efficient transcription elongation and further pre-mRNA processing. Ser2P recruits the Set2 histone methyltransferase and elongation factor Spt6. Moreover, Ser2P is required for recruitment of the splicing machinery (SP), as well as subunits of the cleavage and polyadenylation (PA) machinery. 3'-RNA processing and termination factors Pcf11 and Rtt103 both contain CIDs, which bind the Ser2P CTD. The exonuclease complex (Exo) is also recruited through interaction between CTD and Rtt103 and through cooperative interaction between Rtt103 and Pcf11. Finally, complete dephosphorylation of the Pol II CTD after termination is critical for Pol II recycling. Ser2P dephosphorylation is mediated by the phosphatase Fcp1, whereas Ser5P dephosphorylation requires two phosphatases, Rtr1 and Ssu72.

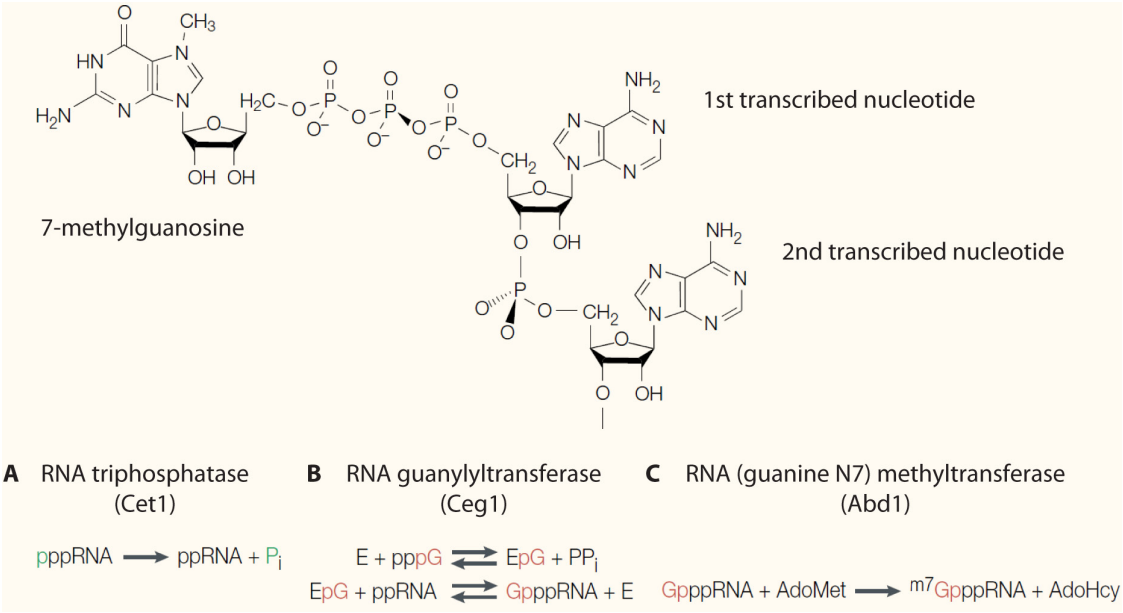

Figure 4: Structure and enzymatic synthesis of the m7G cap. Figure adapted from [204]. The cap consists of 7-methylguanosine linked to the 5' nucleoside of the nascent RNA through an inverted 5'-5' triphosphate bridge. The cap is formed by the sequential action of three enzymes (*S. cerevisiae* enzymes are given in brackets). (A) RNA triphosphatase hydrolyses the γ-phosphate (green) of the 5´-triphosphate end of the nascent RNA. (B) The truncated RNA 5´-end is then linked to an inverted guanylyl group by the guanylyltransferase. First, a lysine side chain of the enzyme (E) reacts with the alpha phosphorus of GTP (red) to form a covalent enzyme-guanylate intermediate (EpG). Second, the enzyme transfer of the GMP to the 5' diphosphate RNA end forms a G cap and regenerates the apoenzyme. (C) RNA methyltransferase transfers the methyl group from S-adenosylmethionine (AdoMet) to the G cap (red) to form a m7G cap and releases S-adenosylhomocysteine (AdoHyc).

1.2.2 5'-capping

During initiation, the CTD of Pol II is phosphorylated at residue serine 5 (S5), which is important for recruitment of RNA 5'-capping enzymes [114, 149, 199]. Capping starts with removal of the terminal γ-phosphate from the 5´-triphosphate end of the nascent RNA, which is catalysed by the RNA triphosphatase [35, 59, 203] (Figure 4). The truncated RNA 5´-end is then linked to an inverted guanylyl group by the guanylyltransferase. Finally, the cap methyltransferase methylates position N7 of the newly added terminal guanine. In *S. cerevisiae* the three catalytic activities are encoded by Cet1, Ceg1 and Abd1. In metazoans the first two steps of the capping reaction are catalyzed by a single capping enzyme consisting of an N-terminal triphosphatase and a C-terminal guanylyltransferase domain. The resulting 7-methyl-guanosine (m7G) cap protects the transcript from degradation and promotes translation initiation of the mRNA [201, 208]. The complete cap associates with the cap-binding complex (CBC) that functions in pre-mRNA splicing and mRNA export [125, 202]. The CBC is a hetero-dimer consisiting of the cap-binding subunit CBP20 and an auxiliary subunit CBP80 [21, 148], both of which are required for binding of the m7G cap structure [239].

1.2.3 Splicing

During pre-mRNA splicing the coding segments of a gene (exons) are joined together, and the interrupting, non-coding fragments (introns) are removed [235]. Splicing is mediated by the spliceosome, a macromolecular ribonucleoprotein (RNP) complex comprised of five snRNPs and numerous proteins. The spliceosome recognizes introns by short, conserved sequences in the nascent

pre-mRNA (5´-splice site, branch point and 3´-splice site [20]) and helps to position reactive groups of the nascent pre-mRNA for catalysis. Introns are then removed by two consecutive transesterification reactions [98]. Alternative splicing (e.g skipping of a particular exon) represents a way to increase the complexity of transcriptomes and proteomes by producing several variable mRNAs from a single pre-mRNA species. Alternative splicing is prevalent in higher eukaryotes and is beyond the scope of this thesis (for a detailed review see [24]). Although *S. cerevisiae* contains only few introns (∼95% of genes are intronless), almost one third of all mRNAs are produced from the intron-containing genes – two thirds of which are highly expressed ribosomal protein genes (RPGs) [162].

1.2.4 3'-end processing

The polyadenylation (pA) site in humans consists primarily of a highly conserved sequence element with the consensus AAUAAA [179], whereas it is more degenerate in yeast. In *S. cerevisiae* it is organized in an AU-rich efficiency element, an A-rich positioning element, and U-rich regions surrounding the cleavage position and the site of poly(A) addition [155, 180]. When Pol II transcribes past these signals, the transcript is processed by endonucleolytic cleavage and by subsequent addition of a protective poly(A) sequence of about 70 nt (200 nt in humans).
The 3'-end RNA processing machinery consists of two multimeric complexes, CPF (cleavage and polyadenylation factors) and CF (cleavage factor IA and B) [155]. Recruitment of these complexes occurs via binding of subunits to the pA elements in the nascent transcript and/or the phosphorylated CTD of Pol II (Section 1.2.1) [114, 150, 154]. Focused recruitment of RNA 3'-end processing factors to the pA region via binding to the CTD of Pol II greatly enhances the cleavage and polyadenlytion reactions [127].
A requirement of RNA 3'-end processing for efficient transcription termination was already established over 25 years ago [133]. Pcf11, an essential subunit of CF IA that interacts with nascent RNA, also contains a CTD-interacting domain (CID) for binding the S2 phosphorylated CTD of Pol II, promoting its recruitment to the 3'-end of genes [114, 150, 154]. Pcf11 has been shown to possess an ability to dismantle elongation complexes *in vitro* [251, 252], dependent on its function in RNA cleavage [111], thus playing a key role in coupling of transcription termination (Section 1.1.3) to transcript cleavage and polyadenylation [13]. After cleavage and polyadenylation mature mRNAs are exported to the cytoplasm for translation by the ribosomes.

1.3 Transcription regulation and RNA checkpoints

Regulation of gene expression is essential for normal cellular development and often occurs at the level of transcription [55]. Although it was initially assumed that all regulation of transcription takes place at the initiation stage via regulated recruitment of Pol II to gene promoters [168], studies over the last decade have revealed that Pol II is subject to regulatory control at all stages of the transcription cycle, especially during the initiation-elongation transition [1, 45].
In higher eukaryotes the transition from intitiation to elongation is often regulated by promoter-proximal pausing (PPP) of Pol II [1, 243]. Although it was already observed in the 1980s that transcription elongation could be a rate-limiting step in gene expression [192], only recent advances in genome-wide studies (Section 2.2) enabled the identification of PPP as a widespread phenomenom in metazoans [34, 61, 158, 249]. To establish PPP, Pol II has first to be recruited to promoters, which depends on the intrinsic strength of the core promoter [99] and on specific transcription factors that recruit chromatin-remodelling proteins and the transcription machinery [71, 205]. Pol II then initiates transcription (Section 1.1.1) and starts to produce mRNA. When the nascent RNA reaches 20-60 nt in length, DSIF (yeast Spt4-5) and NELF (negative elongation factor) associate with the early elongation complex, inhibiting further synthesis. P-TEFb can release Pol II from promoter-proximal pause sites by phosphorylating the CTD,

DSIF and NELF, leading to dissociation of NELF and transformation of DSIF into a positive elongation factor [244, 243]. P-TEFb recruitment, and thus pause release, can be triggered by multiple mechanisms, including transcription factors, chromatin modifications, and co-regulatory complexes [1]. This allows the duration of pausing to be differentially regulated from gene to gene. While most of the key players controlling the initiation-elongation transition are conserved between *S. cerevisiae* and higher eukaryotes, yeast and *C. elegans* apparently lack a NELF ortholog and promoter-proximal pausing.

Despite regulation of gene expression via controlled recruitment of Pol II to gene promoters or release of Pol II from promoter-proximal pause sites, co-transcriptional RNA processing also presents a potential layer to modulate transcription [45, 168]. Like cell cycle checkpoints [238], RNA checkpoints could ensure that the processing steps at each phase of the transcription cycle have been accurately completed before progression into the next phase. Thus faulty transcripts, which can lead to a large number of human diseases [32, 42], are degraded and only correctly processed transcripts are exported for translation. Although a list of publications suggests the existence of such RNA checkpoints (see below), the molecular mechanisms involved remain largely elusive.
The first checkpoint may operate shortly after initiation at the level of 5' pre-mRNA capping. Since promoter-proximal pausing co-occurs with capping of the nascent transcript and interactions between the mammalian capping enzyme and DSIF have been reported [230], it has been proposed that this pausing serves as a quality control checkpoint to ensure that only correctly capped transcripts pursue the productive elongation phase of transcription [139, 168]. Although PPP has not been observed in yeast under normal growth conditions so far, several lines of evidence suggest that such a capping checkpoint could also exist in yeast. First, Spt5 interacts with the *S. cerevisiae* and *S. pombe* capping enzymes [130, 173, 198], and second, the *S. pombe* P-TEFb homolog interacts with the cap methyltransferase [68, 172]. Moreover, a quality control mechanism for mRNA 5'-end capping was recently identified in *S. cerevisiae* [95], during which nascent RNAs without or with aberrant cap structures are co-transcriptionally removed by the Rai1-Rat1 decay pathway.
During subsequent RNA processing steps control points may also operate. Splicing factors recruited to nascent transcripts have been shown to stimulate productive elongation [54, 129]. Furthermore, transcription termination depends on correct processing of RNA 3'-ends and defects in 3'-end formation lead to accumulation of unprocessed transcripts near the sites of transcription [40, 80]. Hence, aberrant RNAs are not released and exported to the cytoplasm.
Finally, the recent discovery that most of the DNA of eukaryotic genomes is transcribed [89], giving rise to a wealth of novel non-coding and potentially non-functional RNA species, further implies a requirement for massive RNA quality surveillance in the nucleus. The future challenge will be to unravel the underlying molecular mechanisms that are employed to distinguish functional from non-functional RNAs and degrade the latter.

2 Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) is a widely used method to measure the occupancy level of proteins bound to DNA *in vivo* [5]. During the ChIP protocol, proteins are first cross-linked to DNA. In early ChIP studies, UV light was used for irreversible cross-linking of proteins to DNA [62]. Nowadays the most used cross-linking agent is formaldehyde, which forms heat-reversible protein-DNA and protein-protein cross-links. Thus, not only direct protein-DNA contacts are observed, but also indirect protein-DNA interactions mediated by bridging proteins. After cell lysis, the cross-linked chromatin is sheared into 100 – 800 bp fragments, either by sonication or restriction enzyme digestion. It has been shown that extensive fragmentation, leading to smaller fragment sizes between 75 and 300 bp, improves enrichment of protein binding sites as well as resolution [53]. Sheared DNA fragments that are bound by the protein of interest (or are bound by a protein carrying a specific protein modification of interest) are then selectively enriched via immunoprecipitation (IP) with protein/modification-specific antibodies. Alternatively, epitope-tagged proteins can be immunoprecipitated using tag-specific antibodies. Typical control samples include input chromatin (sheared chromatin before IP), as well as mock IPs using non-specific antibodies or untagged controls (see Section 2.1.2). The immunoprecipitated protein-DNA complexes are then washed, cross-links reversed by heat, and proteins removed by proteinase K digestion. Finally, the originally bound DNA fragments are identified either via quantitative real-time PCR (qPCR) amplification of a region of interest or via genome-wide detection approaches (see below).
The ChIP method was first described by Gilmour and Lis almost 30 years ago [62, 63]. During the years 2000 – 2002, the first genome-wide ChIP approaches were published [88, 122, 186]. Since then, ChIP coupled to microarray hybridization (ChIP-chip, Section 2.1) has emerged as the standard method for obtaining global DNA binding profiles of proteins. Furthermore, many bioinformatics tools have been developed for the analysis of ChIP-chip data (Section 2.1.2). Over the past few years, rapid advances in high-throughput sequencing technologies have facilitated ChIP coupled to massively parallel sequencing (ChIP-seq, Section 2.2) to be commonly used as an alternative to ChIP-chip. In accordance with that, many complementary genome-wide methods have been established, some of which are also discussed in Section 2.2.

2.1 ChIP-chip

ChIP-chip essentially consists of ChIP (see above) followed by microarray (chip) hybridization, allowing global detection of protein-DNA interactions. The capability of performing genome-wide ChIP-chip studies of transcription factors as well as chromatin states and modifications, has greatly aided our understanding of global gene regulation. Major breakthroughs include the dissection of transcription regulatory networks [72, 122], genome-wide mapping of nucleosomes [123, 246] and histone modifications [67, 78, 177], and the identification of widespread post-initiation gene regulation mechanisms [158, 249]. The ChIP-chip method consists of two parts (Figure 5), an experimental "wet lab" part (Section 2.1.1) including ChIP, DNA amplification, and array hybridization, and a bioinformatics "dry lab" data analysis part (Section 2.1.2).

2.1.1 Experimental procedure

Since a single ChIP experiment often does not provide enough DNA suitable for microarray hybridization, immunoprecipitated DNA fragments have to be amplified either by whole genome amplificiation (WGA) or ligation-mediated PCR (LM-PCR), two methods that have been shown to introduce no (or only minimal) amplification bias [96, 165]. After amplification the DNA is labeled, for instance by incorporation of a biotinylated nucleotide analog, and eventually loaded onto the

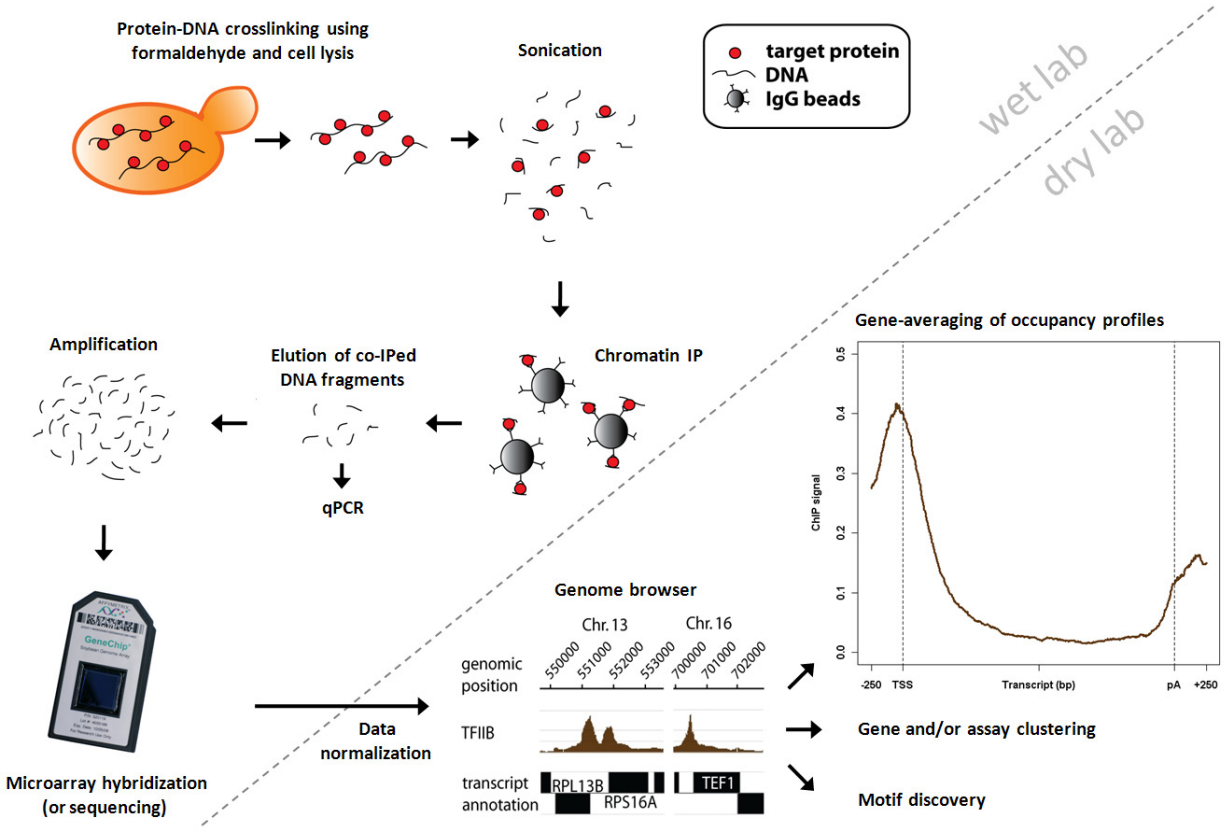

Figure 5: ChIP-Chip is a powerful technique for genome-wide detection of *in vivo* protein-DNA interactions. This method combines chromatin immunoprecipitation (ChIP) with DNA tiling microarray analysis (chip). The figure illustrates the ChIP-chip protocol for *S. cerevisiae* cells containing a TAP-tagged version of the protein of interest (POI). Briefly, cells are grown to mid-exponential phase, followed by formaldehyde cross-linking of proteins to other proteins and DNA. After cell lysis the chromatin is sheared by sonication, and DNA fragments crosslinked to the POI are enriched by immunoprecipitation (IP) with IgG beads directed against the TAP tag. The IP efficiency as well as the fold-enrichment over unoccupied genomic regions is controlled by quantitative PCR (qPCR). Since the amount of DNA is usually too low for subsequent genome-wide detection approaches such as ChIP-chip or ChIP-seq, IPed DNA fragments have to be amplified. In case of ChIP-chip, amplified DNA is fragmented, labeled and eventually hybridized to an Affymetrix *S. cerevisiae* high-resolution tiling array. After hybridization the array is washed, stained and scanned. The scanned array image provides intensity data for each probe indicating a relative level of hybridization with the labeled target. Downstream bioinformatics analysis starts with data preprocessing and normalization. Biological interpretation of the results involves inspection of the normalized ChIP enrichments in a genome browser, followed by further analyses tailored to answer specific biological questions. These analyses include gene-averaging of occupancy profiles, gene and/or assay clustering, and motif discovery.

microarray. Labeling, hybridization, washing, and array scanning are performed using standard protocols and kits provided by the manufacturer of the array platform used.
The major differences between state-of-the-art microarray platforms are single-color (Affymetrix, Santa Clara, CA; Illumina, San Diego, CA) versus two-color (Agilent Technologies, Palo Alto, CA; NimbleGen Systems[1], Madison, WI) systems. In the case of single-color arrays the IP and reference (either input or mock IP) samples are hybridized on different arrays, whereas the two-color system allows the IP and reference samples to be labeled with different dyes and to be hybridized on the same array. Both systems provide comparable quality of results [96]. Whole-genome tiling arrays represent the most robust array design since the entire genome, including the non-coding regions are present. Tiling arrays consist of millions of short (25 to 60 nucleotides long) probes that cover the genome at a constant spacing (4 to 100s of nucleotides). Compact genomes of model organisms like yeast can be covered on a single microarray, whereas larger genomes need several microarrays for full coverage. A state-of-the-art *S. cerevisiae* tiling microarray from Affymetrix contains ∼3.2 million 25-mer oligonucleotide probes, covering the entire yeast genome at 4 bp resolution [44].

2.1.2 Bioinformatics data analysis

The scanned array images provide intensity data for each probe indicating a relative level of hybridization with the labeled target. Since the genomic location of each probe is known, construction of a genome-wide map of *in vivo* protein-DNA interactions is possible. The resolution of the resulting ChIP-chip data depends mainly on two factors. First, the lengths of the sheared DNA fragments enriched by the IP, and second the spacing and coverage of the probes on the microarray.

Preprocessing of ChIP-chip data starts with quality controls of the raw data, followed by intra-array and inter-array normalization (for a detailed protocol see online[2] and Methods 5.7.1). In any ChIP-chip experiment, it is important to correct for sequence- and genomic region-specific biases in the efficiency of the various biochemical and biophysical steps of the protocol. Cross-linking, chromatin fragmentation, IP, amplification, labeling, and hybridization to the array all produce biases that must be corrected for. Although reference-free normalization procedures have been suggested that can reduce these biases [97], the cleanest and most efficient method is to measure a reference sample and to divide the true signals obtained from the IP by the reference intensities. In our experience, this is the most important step in data normalization when using single-color arrays with short probes such as the Affymetrix arrays used here. These arrays exhibit a strong probe sequence bias, meaning that the GC-content of a probe influences its measured intensity. Usually input chromatin (i.e. sheared chromatin before IP) is used as reference. Mock IP (i.e. using an unrelated antibody for which there is no corresponding epitope) represents a further control experiment. The best kind of mock IP, which also allows to control for non-specific antibody binding, is to use a strain that lacks the IP epitope, thus preventing specific binding of the antibody used. In experiments using an epitope-tagged factor, this is easily achieved by using the wild-type strain lacking the tag. Whereas published ChIP-chip studies usually use either input DNA [200] or mock IP [225] as reference control, we decided to incorporate both controls for our data normalization model which should correct for all sources of bias (see Methods 5.7.1).
It is advisable to measure at least two biological replicates for each factor or condition, as this allows corrupted measurements to be easily identified. Furthermore, using N replicates reduces the standard deviation of unsystematic noise by a factor of $\sqrt{N}$.

[1] NimbleGen was taken over by Roche in 2007. In June 2012 Roche exited the DNA microarray business.
[2] http://www.epigenesys.eu/images/stories/protocols/pdf/20111025112610_p47.pdf

After quality control and proper normalization of the data, normalized probe intensity ratios or occupancies can be viewed in open source genome browsers, e.g. IGV, IGB, or UCSC genome browser, given that the data is in a suitable format, e.g. Gene Feature Format (GFF), or Browser Extensible Data (BED). Further in-depth analysis largely depends on the factors/modifications which are analysed and the biological questions that one wants to address.
Binding regions of site-specific transcription factors (TFs) are identified using peak detection algorithms [97, 253], followed by motif discovery [73]. Many software packages for ChIP-chip analysis have been published, which are tailored to the analysis of site-specific TFs [11, 93, 97]. Data normalization and peak detection using these methods rely heavily on the assumption that most of the array probes are not bound by the factor of interest and can be used for modeling non-specific hybridization background.
Since most of the yeast genome is transcribed by Pol II [44, 159], a large percentage of array probes represent binding targets for Pol II and different normalization and analysis techniques are needed (see below and Methods 5.7.1). The same holds true for other frequent binders like factors widely associated with Pol II, or certain histone modifications. ChIP-chip data of such frequent binders can be represented by plotting gene-averaged metagene profiles covering entire genes or specific gene regions (e.g. around the transcription start or polyadenylation sites), allowing the occupancy distributions between gene groups to be compared. Gene groups can be either predefined, depending on ChIP enrichment, expression level, gene ontology, etc., or groups of genes with similar occupancy identified using clustering methods [48]. When comparing experiments (e.g. time courses) clustering can also be performed on the assay level. Features used for comparing gene groups or clustering of genes, include peak levels or the shapes of the binding profiles of the genes. Again, these features can be extracted for entire genes or specific gene regions. Of course all these analysis methods can be applied not only to genes but also to other transcripts or any predefined targets. Finally, so far unannotated, novel targets that are bound be the protein of interest can be identfied as well.

Preprocessing and analysis of ChIP-chip data can be carried out within the R environment for statistical data analysis [183], using packages from the Bioconductor project [58]. Bioconductor packages are available for data normalization as well as peak calling [219], and provide additional functions, often specifically designed for a particular microarray platform (one-color [248] vs. two-color [220]) or a particular sample type (e.g. site-specific TFs [50]). Additional user-defined methods for downstream analysis can be easily implemented in R.

2.2 ChIP-seq and other novel genome-wide approaches

Recently, ChIP-seq has emerged as a complementary or alternative to ChIP-chip for mapping global protein-DNA interactions [171]. The ChIP part of the protocol is similar, but instead of labeling and microarray hybridization, the immunoprecipitated material is used to construct a library of DNA fragments which are amplified and then sequenced in parallel. Since the high-throughput sequencing technology was still in its infancy and expensive when this work was initiated, we decided to use the ChIP-chip method as the underlying technique for our studies. Recent comparison of ChIP-chip and ChIP-seq data revealed that both methodologies produce comparable results and generate highly reproducible profiles within each platform [84]. Overall ChIP-seq shows a better signal-to-noise ratio and higher spatial resolution (i.e. narrower peaks). While this might to some extend affect the precision of peak identification of site-specific TFs, ChIP profiles of frequent binders like Pol II are very similar when microarray and sequencing data are compared [60]. Furthermore, it has to be noted that general principles for ChIP-seq data normalization and the use of suitable control (input) measurements are still not as mature as those implemented for ChIP-chip [84]. As the costs for high-throughput sequencing will further decrease, and general normalization principles are being

established, ChIP-seq will become the method of choice in future research.
Advances in high-throughput sequencing technology gave rise to many novel genome-wide approaches over the past years. In 2012, Frank Pugh's lab published a new method that further improves the resolution of ChIP-seq by applying lambda exonuclease to chromatin immunoprecipitates to digest DNA that is not directly contacted by the protein of interest. This allows mapping of transcription factor binding sites at single nucleotide resolution [188]. Moreover, genome-wide densities of Pol II can also be obtained using global run-on sequencing (GRO-seq, [34, 43, 196]) or native elongating transcript sequencing (NET-seq, [30]). Whereas ChIP is not strand-specific and cannot determine whether Pol II complexes are active, GRO-seq and NET-seq allow identification of transcriptionally engaged Pol II in a strand-specific manner. Compared to ChIP and GRO-seq, NET-seq provides nucleotide resolution but is so far only established for *S. cerevisiae*.
In addition to genomic profiling of the transcription machinery and transcription associated chromatin states, gene expression analysis and more recently transcriptomics represent prominent complementary assays that are essential for studying global gene regulation. Whereas traditional transcriptomics methods (using gene-expression/tiling microarrays [44] or RNA-sequencing [159]) allow only to obtain steady state transcript levels, the recently developed dynamic trancriptome analysis (DTA) allows to measure mRNA synthesis and decay rates, without perturbation of the cellular system, providing higher sensitivity and temporal resolution [153, 214].

3 Aims and scope of this thesis

Transcription of protein-coding genes is a multi-step process in which Pol II synthesizes a mature mRNA in concert with various transcription initiation, elongation, termination, chromatin modifying and RNA processing factors (Section 1). Since the synthesis of mRNA depends on the interaction of proteins with DNA, it is essential to investigate genome-wide occupancy profiles of the various factors involved in transcription, to better understand global gene regulation.
The aim of this thesis was to perform high-resolution genome-wide occupancy profiling by chromatin immunoprecipitation (ChIP, Section 2) of Pol II, its phosphorylated CTD isoforms (Section 1.2.1), as well as components of the initiation (Section 1.1.1), elongation (Section 1.1.2), termination (Section 1.1.3), and RNA processing machinery (Section 1.2) in proliferating *S. cerevisiae* cells. ChIP was coupled to Affymetrix tiling microarrays that cover the entire yeast genome at 4 bp resolution [44] (ChIP-chip, Section 2.1). For data normalization (Section 2.1.2), we developed a novel procedure that corrects for non-specific antibody binding by using input reference as well as mock immunoprecipitation control measurements (Methods 5.7.1). Since previous ChIP publications focused only on a small subset of factors [225] and/or investigated only single genes [108, 176] and/or were obtained using microarrays with low resolution [177], we aimed at generating a comprehensive genome-wide binding map covering all phases of the transcription cycle at great detail.
Open questions in the field were whether all Pol II associated transcription factors assemble at all genes or if subsets of genes recruit distinct sets of proteins according to the gene-specific needs of the transcribing Pol II complex. Moreover, although it was known that the Pol II complex undergoes changes in factor composition during the course of transcription [108, 176], the precise changes in factor composition as well as the exact genomic locations of such transitions were poorly characterized. The mechanistic details underlying these transitions were also hardly understood. Furthermore, no genome-wide studies have examined whether factor recruitment correlates with specific Pol II CTD phosphorylations.
By establishing and performing ChIP-chip occupancy profiling in *S. cerevisiae* cells in combination with statistical data analysis and further biochemical and biophysical assays, we were able to obtain answers to all these questions. Statistical analysis of the genome-wide occupancy levels of the Pol II transcription machinery presented in this work point to a general Pol II transcription complex – that is, one composed of all factors – that mediates transcription and mRNA processing at all actively transcribed Pol II genes in yeast. At both ends of genes this complex undergoes uniform transitions in factor composition that are independent of gene length, gene type or expression level (Section 6). Changes in the Pol II transcription complex occur as a function of the distance from the transcription start site and polyadenylation site, respectively. Moreover, our results indicate that mRNA 5´ cap completion underlies the initiation-elongation transition and triggers productive Pol II elongation (Section 7) and that Pol II CTD tyrosine 1 dephosphorylation underlies the elongation-termination transition and triggers mRNA 3´ processing and termination (Section 8).

Parts of this work were performed in cooperation with Andreas Mayer, Matthias Siebert, and Amelie Schreieck (see page 9 and the individual Results Sections for detailed contributions).

Part II
Materials and Methods

4 Materials

4.1 Bacterial and yeast strains

Strain	Description	Source
XL-1 Blue	Rec1A; endA1; gyrA96; thi-1; hsdR17; supE44; relA1; lac[F' proAB lacIqZΔM15Tn10(Tetr)]	Stratagene

Table 2: *E. coli* strain used in this work

Strain	Description	Source
wild-type	BY4741; MATa; *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	Open Biosystems
Bur1-TAP	BY4741; *BUR1::TAP::HIS3MX6*	Open Biosystems
Cet1-TAP	BY4741; *CET1::TAP::HIS3MX6*	Open Biosystems
Ctk1-TAP	BY4741; *CTK1::TAP::HIS3MX6*	Open Biosystems
Elf1-TAP	BY4741; *ELF1::TAP::HIS3MX6*	Open Biosystems
Kin28-TAP	BY4741; *KIN28::TAP::HIS3MX6*	Open Biosystems
Paf1-TAP	BY4741; *PAF1::TAP::HIS3MX6*	Open Biosystems
Pcf11-TAP	BY4741; *PCF11::TAP::HIS3MX6*	Open Biosystems
Rpb3-TAP	BY4741; *RPB3::TAP::HIS3MX6*	Open Biosystems
Spn1-TAP	BY4741; *SPN1::TAP::HIS3MX6*	Open Biosystems
Spn1-TAP Elf1-3HA	BY4741; *SPN1::TAP::HIS3MX6; ELF1::3HA::KANMX6*	A. Mayer
Spt4-TAP	BY4741; *SPT4::TAP::HIS3MX6*	Open Biosystems
Spt5-TAP	BY4741; *SPT5::TAP::HIS3MX6*	Open Biosystems
Spt6-TAP	BY4741; *SPT6::TAP::HIS3MX6*	Open Biosystems
Spt6ΔC-TAP	FY119; Matα; *his4-912 lys2-128 leu2-1 ura3-52 trp1-63*	S. Dengl
Spt16-TAP	BY4741; *SPT16::TAP::HIS3MX6*	Open Biosystems
TFIIB-TAP	BY4741; *SUA7::TAP::HIS3MX6*	Open Biosystems
Tfg1-TAP	BY4741; *TFG1::TAP::HIS3MX6*	Open Biosystems

Table 3: *S. cerevisiae* strains used in Section 6

Strain	Description	Source
wild-type	BY4741; MATa; *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	Open Biosystems
Cet1-TAP	BY4741; *CET1::TAP::HIS3MX6*	Open Biosystems
Ceg1-TAP	RS453; MATa; *CEG1::TAP::TRP1-KL*	K. Sträßer
Abd1-TAP	BY4741; *ABD1::TAP::HIS3MX6*	This work
Cet1-TAP Spt5ΔCTR	BY4741; *CET1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Ceg1-TAP Spt5ΔCTR	RS453; *CEG1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Abd1-TAP Spt5ΔCTR	BY4741; *ABD1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
bur2Δ	BY4741; *bur2::KANMX6*	Open Biosystems
bur2Δ Cet1-TAP	BY4741; *bur2::KANMX6; CET1::TAP::HIS3MX6*	This work
bur2Δ Abd1-TAP	BY4741; *bur2::KANMX6; ABD1::TAP::HIS3MX6*	This work
CBP20-TAP	BY4741; *CBP20::TAP::HIS3MX6*	Open Biosystems
CBP80-TAP	BY4741; *CBP80::TAP::HIS3MX6*	Open Biosystems
cbp20Δ	BY4741; *cbp20::KANMX6*	Invitrogen
cbp20Δ Bur1-TAP	BY4741; *cbp20::KANMX6; BUR1::TAP::HIS3MX6*	This work
cbp20Δ Ctk1-TAP	BY4741; *cbp20::KANMX6; CTK1::TAP::HIS3MX6*	This work
cbp20Δ Elf1-TAP	BY4741; *cbp20::KANMX6; ELF1::TAP::HIS3MX6*	This work
cbp20Δ Spn1-TAP	BY4741; *cbp20::KANMX6; SPN1::TAP::HIS3MX6*	This work
cbp20Δ Spt4-TAP	BY4741; *cbp20::KANMX6; SPT4::TAP::HIS3MX6*	This work
cbp20Δ Spt5-TAP	BY4741; *cbp20::KANMX6; SPT5::TAP::HIS3MX6*	This work
cbp20Δ Spt6-TAP	BY4741; *cbp20::KANMX6; SPT6::TAP::HIS3MX6*	This work
cbp20Δ Spt16-TAP	BY4741; *cbp20::KANMX6; SPT16::TAP::HIS3MX6*	This work
Anchor-away wild-type (AA)	HHY168; MATα; *ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3 GAL psi+ tor1-1 fpr1::NAT RPL13A-2xFKBP12::TRP1*	Euroscarf
AA Bur1-TAP	HHY168; *BUR1::TAP::HIS3MX6*	This work
AA Ctk1-TAP	HHY168; *CTK1::TAP::HIS3MX6*	This work
AA Abd1-FRB	HHY168; *ABD1::FRB::KANMX6*	This work
AA Abd1-FRB-GFP	HHY168; *ABD1::FRB::GFP::KANMX6*	This work
AA Abd1-FRB Bur1-TAP	HHY168; *ABD1::FRB::KANMX6; BUR1::TAP::HIS3MX6*	This work
AA Abd1-FRB Ctk1-TAP	HHY168; *ABD1::FRB::KANMX6; CTK1::TAP::HIS3MX6*	This work
AA Abd1-FRB CBP20-TAP	HHY168; *ABD1::FRB::KANMX6; CBP20::TAP::HIS3MX6*	This work

Table 4: *S. cerevisiae* strains used in Section 7

Strain	Description	Source
Nrd1-TAP	BY4741; *NRD1::TAP::HIS3MX6*	Open Biosystems
Rna14-TAP	BY4741; *RNA14::TAP::HIS3MX6*	Open Biosystems
Rna15-TAP	BY4741; *RNA15::TAP::HIS3MX6*	Open Biosystems
Rtt103-TAP	BY4741; *RTT103::TAP::HIS3MX6*	Open Biosystems
AA Glc7-FRB	HHY168; *GLC7::FRB::KANMX6*	A. Schreieck

Table 5: *S. cerevisiae* strains used in Section 8

4.2 Primers and plasmids

All DNA primers were synthesized by Thermo Fisher Scientific GmbH Ulm, Germany.

Name	Sequence (5' → 3')
ADH1 5' Fw	TCCTTGTTTCTTTTTCTGCAC
ADH1 5' Rv	GAGATAGTTGATTGTATGCTTGG
ADH1 ORF Fw	AGCCGCTCACATTCCTCAAG
ADH1 ORF Rv	ACGGTGATACCAGCACACAAGA
ADH1 3' Fw	AAAACGAAAATTCTTATTCTTGA
ADH1 3' Rv	TACCTGAGAAAGCAACCTGA
ACT1 5' Fw	AAACCAAACTCGCCTCTCT
ACT1 5' Rv	GGAAGGAAAGGATCAAACAA
ACT1 ORF Fw	TCAGAGCCCCAGAAGCTTTG
ACT1 ORF Rv	TTGGTCAATACCGGCAGATTC
ACT1 3' Fw	TTTATCCATTGGACCGTGTA
ACT1 3' Rv	GGGCAATTGCATAAACCTAT
ILV5 5' Fw	ACCCAGTATTTTCCCTTTCC
ILV5 5' Rv	TTGTCTATATGTTTTTGTCTTGC
ILV5 ORF Fw	CTATCAAGCCATTGTTGACC
ILV5 ORF Rv	CTTGAAGACTGGGGAGAAAC
PMA1 5' Fw	TGACTGATACATCATCCTCTT
PMA1 5' Rv	TTGGCTGATGAGCTGAAACAGAA
PMA1 ORF Fw	AAATCTTGGGTGTTATGCCATGT
PMA1 ORF Rv	CCAAGTGTCTAGCTTCGCTAACAG
PMA1 3' Fw	GGTTTCTCTGGATGGTACTTT
PMA1 3' Rv	TGACTTGTGTGCGTTTCATA
YER (Control) Fw	TGCGTACAAAAAGTGTCAAGAGATT
YER (Control) Rv	ATGCGCAAGAAGGTGCCTAT

Table 6: qPCR primers used in this work

Name	Sequence (5' → 3')
Abd1 FRB(-GFP) Fw	ATTCCGTAAGGTAAAACAGTATATCGAACCGGAAAGCGTAAAGCCCAAC cggatccccgggttaattaa
Abd1 FRB(-GFP) Rv	ACGGCCGAAATACAGATGCTTTATAGTAGGGTTATTGTTTCTATTCATTTTTATT gaattcgagctcgtttaaac
Spt5△CTR Fw	AGCTGTAAATGCGCATGGAGGCTCAGGTGGTGGCGGTGTC taacggatccccgggttaattaag
Spt5△CTR Rv	TTGATTTCTTCTTGGGTGATATTGGTTCTCCTTTTGGTGA cgcataggccactagtggatc

Table 7: Primers used for C-terminal FRB(-GFP) tag integration and for deletion of the 15 C-terminal hexapeptide repeats (CTR, amino acids 931 to 1063) of Spt5.
Sequence parts which are homologous to the protein of interest and the FRB(-GFP) tag (plasmids pFA6a-FRB-KanMX6 and pFA6a-FRB-GFP-KanMX6; Table 9) or the KanMX6 casette (plasmid pFA6a-3HA-KanMX6; Table 9) are in capital and lower case letters, respectively.

Name	Sequence (5' → 3')
Abd1 TAP Fw	CATTCCGTAAGGTAAAACAGTATATCGAACCGGAAAGCGTAAAGCCCAAC
	cgtacgctgcaggtcgac
Abd1 TAP Rv	CGAAATACAGATGCTTTATAGTAGGGTTATTGTTTCTATTCATTTTTATT
	atcgatgaattcgagctcg
Bur1 TAP Fw	GATGAGTCTAAGGAGTTCCAAAATAGTGATATTGCAGATCTATAT
	cgtacgctgcaggtcgac
Bur1 TAP Rv	GTAATTAGCCACGAGGCCAGAAAGGAAGAGAGAATAGTATAACCT
	atcgatgaattcgagctcg
CBP20 TAP Fw	GACCAGGTTTCGATGAAGAAAGAGAAGATGATAACTACGTACCTCAG
	cgtacgctgcaggtcgac
CBP20 TAP Rv	TATATATCTGTGTGTAGAATCTTTCTCAGATATAAATTGATTGATTCT
	atcgatgaattcgagctcg
Cet1 TAP Fw	GTTATCGTCTTTATCATATGAAATTTTTGAAGGTTCAAAGAAAGTCATG
	cgtacgctgcaggtcgac
Cet1 TAP Rv	TCTCGCTCAAGGGCATTTGCTTATTTTTTTTTGAAATGATTCAAATA
	atcgatgaattcgagctcg
Ctk1 TAP Fw	GTAATAGTAATAATAATAATAATAATAATGACGATGATGATAAA
	cgtacgctgcaggtcgac
Ctk1 TAP Rv	TAATCTATTTTTTGTGTCTACTTATTTCAATTGGCTATATATCCTT
	atcgatgaattcgagctcg
Elf1 TAP Fw	GCCAAGTTAAAAGAGGCAGAGGCGCCTTGGTAGATAGTGACGATGAA
	cgtacgctgcaggtcgac
Elf1 TAP Rv	ATATAAAATATATATGACCTAAGTAAATATGGTTTTTTCTCAGGACCGGA
	atcgatgaattcgagctcg
Spn1 TAP Fw	GTGAGATGTACAAGAGGTTGACTTCAAGATTAAACAAGAAGCATAAA
	cgtacgctgcaggtcgac
Spn1 TAP Rv	CATATGATACATATCTCAAAGCATTACGGAATTACCTGTTTTGTTATT
	atcgatgaattcgagctcg
Spt4 TAP Fw	CTGTTGCCTCACTACAAACCGAGGGATGGCAGTCAAGTTGAG
	cgtacgctgcaggtcgac
Spt4 TAP Rv	AATTCATTACTATTATACATGTGATATCAGAACGGAAGGTTT
	atcgatgaattcgagctcg
Spt5 TAP Fw	CAAGGAAATAAGTCAAACTATGGTGGTAACAGTACATGGGGAGGTCAT
	cgtacgctgcaggtcgac
Spt5 TAP Rv	CTTTTTTATTGATTTCTTCTTGGGTGATATTGGTTCTCCTTTTGGTGATT
	atcgatgaattcgagctcg
Spt6 TAP Fw	GACGCTTCTAAAATCTAACAGTAGTAAGAATAGAATGAACAACTACCGT
	cgtacgctgcaggtcgac
Spt6 TAP Rv	GGTCAAAGTAATAATAAAATTAATAATAACAATGGACACTACATACGCAT
	atcgatgaattcgagctcg
Spt16 TAP Fw	GAGAAAAAGGCTGCTAGGGCTGATAGGGGTGCAAACTTTAGAGAT
	cgtacgctgcaggtcgac
Spt16 TAP Rv	CAGATCAAGGTCTTGCTGGTGAAACCCAGTAAGTGTTATAAGTCTA
	atcgatgaattcgagctcg

Table 8: Primers used for C-terminal TAP tag integration by homologous recombination. Sequence parts which are homologous to the protein of interest and the TAP tag (plasmid LL280; Table 9) are in capital and lower case letters, respectively.

Name	Description	Source	Application
LL280	T7; TAP tag; ColEI origin; Amp^r	L. Larivière	TAP tagging
pFA6a-FRB-KanMX6	FRB::KanMX6; pBR322 origin; Amp^r	Euroscarf	Anchor away
pFA6a-FRB-GFP-KanMX6	FRB::GFP::KanMX6; pBR322 origin; Amp^r	Euroscarf	Anchor away
pFA6a-3HA-KanMX6	3HA::KanMX6; pBR322 origin; Amp^r	H. Feldmann	Spt5 CTR deletion

Table 9: Plasmids used in this work

4.3 Antibodies

Antibody	Amount in ChIP	Source
1Y26 (Rpb3)	5 µl (lyophilized ascites dissolved in 100 µl ddH_2O)	NeoClone Biotechnology
3D12 (Y1P)	50 µl (supernatant cell culture)	Elisabeth Kremmer/Dirk Eick
3E8 (S5P)	20 µl (supernatant cell culture)	Elisabeth Kremmer/Dirk Eick
3E10 (S2P)	25 µl (supernatant cell culture)	Elisabeth Kremmer/Dirk Eick
4E12 (S7P)	50 µl (supernatant cell culture)	Elisabeth Kremmer/Dirk Eick
BL2894 (S2P)	2 µl	Bethyl Laboratories

Table 10: Antibodies used for ChIP. All antibodies but BL2894 are monoclonal.

Antibody	Usage	Host	Source
α-Pgk1	1:10000	Mouse	K. Sträßer
α-Tubulin	1:1000	Rat	Santa Cruz (sc-69971)
PAP (Peroxidase-Anti-Peroxidase)	1:2000	Rabbit	SIGMA (P1291)
α-Mouse-HRP	1:3000	Goat	Bio-Rad (170-6516)
α-Rat-HRP	1:3000	Goat	SIGMA (A9037)

Table 11: Antibodies used for western blotting

4.4 Media and supplements

Name	Description	Application
LB	1% (w/v) tryptone; 0,5% (w/v) yeast extract; 0,5% (w/v) NaCl; (+ 1,5% (w/v) agar for solid media plates)	*E. coli* culture
YPD	2% (w/v) petone; 2% (w/v) glucose; 1,5% (w/v) yeast extract; (+ 2% (w/v) agar for solid media plates)	*S. cerevisiae* culture
SC -his	0.69% (w/v) yeast nitrogen base; 0.077% (w/v) drop-out -his; 2% (w/v) glucose; (+ 2% (w/v) agar for solid media plates)	*S. cerevisiae* culture

Table 12: Growth media

Additive	Description	Working concentration
Ampicillin	Antibiotic	100 µg/ml
Geneticin (G418)	Antibiotic	400 µg/ml
Rapamycin	Immunosuppressive drug	1 µg/ml

Table 13: Growth media additives

4.5 Buffers and solutions

Name	Description	Application
1x TBE	8.9 mM Tris-HCl; 8.9 mM Boric acid; 2 mM EDTA; pH 8.0 at 25°C	Agarose gels
TFB-1 Buffer	30 mM KOAc; 50 mM $MnCl_2$; 100 mM RbCl; 10 mM $CaCl_2$; 15% (v/v) Glycerol; pH 5.8 at 25°C	Chemically competent *E. coli* cells
TFB-2 Buffer	10 mM MOPS, pH 7.0 at 25°C; 10 mM RbCl; 75 mM $CaCl_2$; 15% (v/v) Glycerol	Chemically competent *E. coli* cells
TELit	10 mM Tris-HCl, pH 8.0 at 25°C; 155 mM LiOAc; 1 mM EDTA, pH 8.0;	Chemically competent *S. cerevisiae* cells
LitSorb	10 mM Tris-HCl, pH 8.0 at 25°C; 155 mM LiOAc; 1 mM EDTA, pH 8.0; 18.2% (w/v) D-Sorbitol	Chemically competent *S. cerevisiae* cells
LitPEG	10 mM Tris-HCl, pH 8.0 at 25°C; 155 mM LiOAc; 1 mM EDTA, pH 8.0; 40% (w/v) PEG 3350	Chemically competent *S. cerevisiae* cells
LiOAc/DTT	0.1 M LiOAc; 30 mM DTT	Electroporation of *S. cerevisiae* cells
Lyticase Buffer	1 M Sorbitol; 100 mM EDTA, pH 8.0; 14.3 mM β-Mercaptoethanol;	Preparation of genomic DNA
Spheroblast wash Buffer	1 M Sorbitol; 100 mM EDTA, pH 8.0	Preparation of genomic DNA
TE 50/100	50 mM Tris-HCl, pH 7.5 at 4°C; 100 mM EDTA, pH 8.0	Preparation of genomic DNA
Lysis Buffer	50 mM Tris-HCl, pH 7.5; 10 mM EDTA, pH 8.0, 1:500 β-Mercaptoethanol; 1:1000 Lyticase (10 U/ml)	Preparation of genomic DNA
2x SDS loading Buffer	25 mM Tris-HCl, pH 7.0 at 25°C; 0.05% (w/v) Bromophenol blue; 0.5% (v/v) β-Mercaptoethanol; 7% (w/v) Dithiothreitol; 0.05% (w/v) Lauryl sulfate; 5% (v/v) Glycerol	SDS-PAGE
4x Stacking gel Buffer	0.5 M Tris-HCl; 0.4% (w/v) SDS; pH 6.8 at 25°C	SDS-PAGE
4x Separation gel Buffer	3 M Tris-HCl; 0.4% (w/v) SDS; pH 8.9 at 25°C	SDS-PAGE
SDS running Buffer	25 mM Tris-HCl; 0.1% (w/v) SDS; 250 mM Glycine	SDS-PAGE
Gel stain	50% (v/v) Ethanol; 7% (v/v) Acetic acid; 0.125% (w/v) Coomassie Brilliant Blue R-250;	SDS-PAGE
Gel destain	5% (v/v) Ethanol; 7.5% (v/v) Acetic acid	SDS-PAGE
1x PBS	2 mM KH_2PO_4; 4 mM Na_2HPO_4; 140 mM NaCl; 3 mM KCl; pH 7.4 (25°C)	Western blotting
WB transfer Buffer	25 mM Tris; 192 mM Glycine; 20% (v/v) Ethanol	Western blotting
WB blocking Buffer	2% (w/v) milk powder in 1x PBS	Western blotting

Table 14: General buffers and solutions

Name	Description
1x TBS	20 mM Tris-HCl, pH 7.5 at 4°C; 150 mM NaCl
FA lysis Buffer	50 mM HEPES-KOH, pH 7.5 at 4°C; 150 mM NaCl; 1 mM EDTA; 1% (v/v) Triton X-100; 0.1% (v/v) Na deoxycholate; 0.1% (v/v) SDS; PI; PhoI*
FA lysis Buffer 2	Identical to FA lysis Buffer, but with 500 mM NaCl instead of 150 mM NaCl
ChIP wash Buffer	10 mM Tris-HCl, pH 8.0 at 4°C; 0.25 M LiCl; 1 mM EDTA; 0.5% (v/v) NP-40; 0.5% (v/v) Na deoxycholate
TE Buffer	10 mM Tris-HCl, pH 7.4 at 4°C; 1 mM EDTA
ChIP elution Buffer	50 mM Tris-HCl, pH 7.5 at 25°C; 10 mM EDTA; 1% (v/v) SDS
RNase storage Buffer	10 mM HEPES, pH 7.5 at 25°C; 20 mM NaCl; 0.1% (v/v) Triton X-100; 1 mM EDTA; 50% (v/v) Glycerol
Protease-inhibitor mix (PI)	1 mM Leupetin; 2 mM Pepstatin A; 100 mM Phenylmethylsulfonyl fluoride; 280 mM Benzamidine
Phosphatase-inhibitor mix (PhoI)	1 mM NaN3; 1 mM NaF; 0.4 mM Na3VO4

Table 15: Buffers and inhibitor mixes used for ChIP.
* The phosphatase inhibitor mix (PhoI) was only used for ChIP experiments of Pol II phopho-isoforms.

All standard buffers and solutions were prepared according to Sambrook & Russell, 2001.

5 Methods

5.1 Molecular Cloning

5.1.1 Polymerase Chain Reaction (PCR)

A standard PCR reaction such as for the validation of TAP-tagged yeast strains was performed as follows. PCR reactions contained 3 µl (∼150 ng) genomic DNA template (Section 5.2.3), 1.25 µl of 10 µM forward and reverse primer (usual length 18 to 22 nt), 5 µl 10x Taq Buffer, 6 µl of 10 mM $MgCl_2$, 2.5 µl of 2 mM dNTP-Mix, 30 µl H_2O and 1 µl Taq DNA Polymerase. PCR was performed on a T3000 Thermocycler (Biometra) using a 2 min denaturing step at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 50°C and 1 min/kb at 72°C. Finally, the PCR reaction was incubated at 72°C for 10 min and stored at 4°C.
PCR products amplified from plasmids (Table 9) used to generate new yeast strains by homologous recombination (Section 5.2.1) were typically generated as follows. PCR reactions contained 1 µl (∼100 ng) plasmid DNA template, 2 µl of 10 µM forward and reverse primer (∼50 nt long, Tables 7 and 8), 10 µl 5x HF Buffer, 5 µl of 2 mM dNTP-Mix, 29 µl H_2O and 1 µl Phusion DNA Polymerase. PCR was performed on a T3000 Thermocycler (Biometra) using a 30 sec denaturing step at 98°C, followed by 35 cycles of 7 sec at 98°C, 20 sec at 55°C and 25 sec/kb at 72°C. Finally, the PCR reaction was incubated at 72°C for 10 min and stored at 4°C.
In cases where no or only low yield of PCR product was obtained, the standard protocols were adjusted by changing the PCR parameters, including the type of DNA polymerase, primer annealing temperature and elongation time.

5.1.2 Agarose gel electrophoresis

Electrophoretic separation of DNA was performed in horizontal 1x TBE agarose gels that contained SYBR Safe (0.01 µg/ml; Invitrogen). Depending on the size of the DNA molecules that were separated, the agarose concentration was varied between 0.8% and 2% (w/v). Agarose gel

electrophoresis was carried out in PerfectBlue Gelsystem electrophoresis devices from Peqlab (110 V, 35 min). DNA samples were mixed with 6x loading dye (Fermentas). The sizes of the separated DNA molecules were assessed with the help of the GeneRulerTM 1kb DNA ladder or the GeneRulerTM 100bp DNA ladder (Fermentas).

5.1.3 Transformation of *E. coli* cells and preparation of plasmid DNA

For **transformation of *E. coli* cells** chemically competent XL-1 Blue cells (see Table 2) were used. Approximately 2 µg of DNA were added to 50 µl of competent cells and incubated for 5 min on ice. Next, cells were heated for 30 sec at 42°C in a water bath and then put back on ice for 2 min. Transformed *E. coli* cells were recovered by incubation at 37°C for 1 h in 700 µl of LB medium. Recovered cells were centrifuged at 13,000 rpm for 30 sec (at room temperature) and 650 µl of the supernatant was removed. Cells were resuspended in the remaining volume and plated onto solid LB medium plates containing the respective antibiotic (see Table 13). Plates were incubated at 37°C overnight.
Chemically competent *E. coli* cells were prepared as follows. 200 ml LB medium (containing the respective antibiotic) were inoculated with 5 ml of an overnight culture of the desired *E. coli* strain. Cells were grown at 37°C (160 rpm) to an OD600 of 0.4 to 0.5 and incubated on ice for 10 min. The following steps were performed at 4°C with pre-cooled buffers (see Table 14). Cells were harvested by centrifugation and washed with 50 ml TFB-1 Buffer. After a second centrifugation step, cells were resuspended in 4 ml TFB-2 Buffer. Aliquots of competent *E. coli* cells were frozen in liquid nitrogen and stored at -80°C.

For **preparation of plasmid DNA**, 5 ml LB medium (containing the respective antibiotic) were inoculated with a single colony and incubated at 37°C overnight (160 rpm). Cells were harvested by centrifugation at 4,500 rpm for 10 min (4°C). Next, plasmid DNA was prepared with the QIAquick Miniprep Kit (Qiagen) according to the manufacturer's instructions.

5.2 Yeast methods

5.2.1 Generation of *S. cerevisiae* strains and epitope tagging

All *S. cerevisiae* strains used in this work are listed in Tables 3 to 5. TAP-tagged yeast strains were either obtained from Open Biosystems (Huntsville, USA; Yeast TAP-Tagged Collection) or generated by integration of the TAP tag (amplified from the LL280 plasmid) into the genome, C-terminal of the respective genes, by homologous recombination. Deletion of the 15 C-terminal hexapeptide repeats (CTR, amino acids 931 to 1063) of Spt5 was done by homologous recombination with the KanMX6 cassette, amplified from the pFA6a-3HA-KanMX6 vector. For nuclear depletion of Abd1 using the anchor-away method [76] a FKBP12-rapamycin-binding (FRB) tag was introduced at the C-terminus of Abd1. Anchor-away strains were generated as described [76] using PCR products amplified from plasmids pFA6a-FRB-KanMX6 and pFA6a-FRB-GFP-KanMX6 (P30578/P30580, Euroscarf). All plasmids used are listed in Table 9.
The *bur2*△ and *cbp20*△ deletion strains were obtained from Open Biosystems and Invitrogen, respectively.

5.2.2 Transformation of *S. cerevisiae* cells

For **transformation of chemically competent *S. cerevisiae* cells** 10 µl of DNA ($\geq$ 1 µg of linear DNA) and 360 µl LitPEG (Table 14) was added to 50 µl of competent cells. After incubation at room temperature for 30 min, 47 µl DMSO was added and in case that the yeast strain was not temperature-sensitive heated at 42°C for 15 min. Alternatively, if the yeast

strain was temperature-sensitive cells were incubated at 30°C for 10 min and at 37°C for 5 min. After recovery (30°C, 1 h, 150 rpm), cells were collected by centrifugation at 2,000 rpm for 3 min (room temperature) and the supernatant was removed. Transformed yeast cells were resuspended in 50 µl sterile ddH_2O, plated on selective media plates and incubated at 30°C for 2-5 days.
Chemically competent *S. cerevisiae* cells were prepared as follows. 50 ml YPD medium was inoculated (start OD600 of 0.2) with yeast that was grown to stationary phase over-night. Next, yeast cells were grown at 30°C (150 rpm) to an OD600 of 0.5 to 0.7 and harvested by centrifugation at 4,000 rpm for 10 min (room temperature). Yeast cells were washed with 25 ml sterile ddH_2O and finally resuspended in 360 µl LitSorb. 40 µl of heated (10 min at 100°C) salmon sperm DNA was added. Aliquots of competent yeast cells were stored at -80°C.

For **transformation by electroporation** 100 ml YPD medium was inoculated (start OD600 of 0.15) with yeast that was grown to stationary phase overnight (cells of 100 ml culture were enough for one transformation). Next, yeast cells were grown at 30°C (150 rpm) to an OD600 of 0.7 to 0.8 and harvested by centrifugation at 4,000 rpm for 5 min (4°C). Afterwards the cells were washed two times with 50 mL ice-cold sterile ddH_2O and one time with 50 mL ice-cold 1 M sorbitol (5 min, 4,000 rpm, 4°C). Next the cell pellet was resuspended in 20 mL 0.1 M LiOAc/30 mM DTT and incubated for 30 min at 150 rpm and 30°C. The cells were harvested and the pellet was washed two times with 50 mL ice-cold 1 M sorbitol and finally resuspended in 1 ml 1 M sorbitol. After transfer to an Eppendorf reaction tube and centrifugation for 1 min at 4,000 rpm (4°C) the pellet was resuspended in 100 µl ice-cold 1 M sorbitol. For transformation 5 µl of DNA ($\geq$ 1 µg of linear DNA) was added to 100 µl of cell suspension. The mixture was transferred to an electroporation cuvette which was put into a MicroPulserTM Electroporator (Bio-Rad) and a pulse of 2.5 kV (5 ms) was applied (program Sc2). Then 1 ml of pre-warmed YPD medium was added to the cells and the suspension transferred to an Eppendorf reaction tube. If needed, the transformed *S. cerevisiae* cells were recovered by incubation at 30°C and 350 rpm for 2 h. Finally, the cells were plated on selective plates (for TAP tag insertion –Histidin, Table 12) and incubated at 30°C for 2-5 days.

5.2.3 Preparation of genomic DNA

For preparation of genomic DNA 5 ml yeast were grown at 30°C (150 rpm) overnight. Cells were harvested by centrifugation at 2,500 rpm for 5 min (4°C), washed with 1 ml sterile ddH_2O and resuspended in 500 µl Lyticase Buffer. Next, 20 µl Lyticase (10 U/µl) were added. After incubation at 37°C for 45 min, yeast spheroblasts were collected by centrifugation at 5,000 rpm for 5 min (4°C). The supernatant was removed and spheroblasts were gently resuspended in 1 ml Spheroblast wash Buffer. Spheroblasts were centrifuged again and resuspended in 500 µl Spheroblast wash Buffer. Spheroblasts were collected again and resupended in 500 µl of TE 50/100 Buffer. 50 µl 10% SDS were added and incubated at 70°C for 30 min. Then, 250 µl 5 M Potassium acetate were added and incubated on ice for 15 min. The precipitated genomic DNA was collected by centrifugation at 13,000 rpm for 20 min (4°C), the supernatant was removed and 700 µl Isopropanol was added. Precipitated genomic DNA was collected by centrifugation at 13,000 rpm for 10 min (4°C). Precipitated DNA was washed with 70% Ethanol and resuspended in 500 µl TE 50/100 Buffer via heating at 42°C for 30 min. DNA quantity and quality control was performed with a ND-1000 Spectrophotometer (NanoDrop Technologies). Genomic DNA was stored at -20°C.
Alternatively, yeast genomic DNA was prepared using the QIAcube robotic workstation (Qiagen). From the overnight culture a volume corresponding to OD600 of 5.0 was centrifuged for 5 min at 3,000 rpm and resuspended in 180 µl Lysis Buffer. The QIAcube was run according to the manufacturer's instructions using the protocol "DNeasy Blood & Tissue - Bacteria (Gram+) or yeast – Enzymatic lysis". The yield of DNA was within the range of 50-100 ng/µl.

5.3 Protein methods

5.3.1 Preparation of whole cell extracts for western analysis

For preparation of whole cell extracts, yeast cultures were grown overnight in YPD and transferred into 100 ml of YPD to adjust the starting OD600 to 0.15. Cells were grown at 30°C to mid-log phase (OD600 ∼0.8). Subsequent steps were performed at 4°C with pre-cooled FA lysis Buffer containing protease inhibitors (Table 15). Cells were collected by centrifugation (5 min at 4,000 rpm), and washed with 20 ml FA lysis Buffer. Cell pellets were resuspended in 1 ml FA lysis Buffer, and disrupted by bead beating (Retsch) in the presence of 1 ml silica-zirconia beads for 30 min at 4°C. The beads were removed and the lysates clarified by centrifugation at top speed for 30 min at 4°C. Protein concentration was determined using the Bio-Rad Protein Assay (Bio-Rad), and ∼40 µg of total protein was applied to SDS-PAGE (5.3.2) followed by western analysis (Section 5.3.3).

5.3.2 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Electrophoretic separation of protein samples was performed by SDS-PAGE with 15% acrylamide gels (acrylamide:bisacrylamide ratio = 37.5:1) [119] in Bio-Rad gel systems. Before loading onto the gel, protein samples were mixed with SDS-PAGE loading Buffer and boiled at 95°C for 5 min. Protein samples requiring broader or higher resolution separation, were separated by ready-to-use NuPAGE Novex Bis-Tris minigels (Invitrogen) according to the manufacturer's instructions. MES and MOPS running buffers (Invitrogen) were used. Gels were stained with Gel staining solution at room temperature for at least 20 min and destained at room temperature overnight in Gel destaining solution. Corresponding buffers and solutions are listed in Table 14.

5.3.3 Western blotting

For western blotting the protein sample was separated by SDS-PAGE as described in Section 5.3.2. Separated proteins were then transferred to a PVDF membrane (Schleicher & Schuell; pre-equilibrated with 100% Ethanol) in the presence of the WB transfer Buffer (see Table 14). Thereby, the wet blotting system from Bio-Rad was used according to the manufacturer's instructions. Transfer was performed at 100 V for 1 h at 4°C. After transfer, the membrane was blocked for at least 30 min with WB blocking Buffer. The blot was then incubated for 1 h (room temperature) with the primary antibody in WB blocking Buffer. The blot was washed three times with 1x PBS for 10 min and then incubated with the secondary antibody in WB blocking Buffer for 1 h (room temperature). Afterwards, the blot was washed three times with 1x PBS for 10 min. Secondary antibodies were usually coupled to horseradish peroxidase. Final signals were detected with the Chemiluminescence Kit (Pierce) followed by exposure of the blot to high-sensitivity films (Invitrogen). Films were developed with the X-omat M35 developing machine (Kodak).

5.4 Microscopy

To confirm nuclear depletion of Abd1 using the anchor-away method [76], we tagged the Abd1-FRB fusion protein with GFP (strain AA Abd1-FRB-GFP, Table 4, Section 5.2.1) and monitored GFP fluorescence upon rapamycin treatment. *S. cerevisiae* cells were grown in YPD with 40 mg/l adenine hemisulfate at 30°C to an OD600 of ∼0.6. Cultures were split and incubated with equal volumes of either rapamycin (1 µg/ml f.c. in DMSO) or DMSO at 30°C for another 60 min. Cells were resolved in water and inspected under the microscope (Leica DM2500, EL 6000). Camera DFC365FX and Software LAS AF 6000 Modular Systems Version 2.6.0.7266 (Leica) were used for image analysis.

5.5 Chromatin immunoprecipitation followed by quantitative real-time PCR (ChIP-qPCR)

5.5.1 ChIP using TAP-tagged proteins

TAP-tagged yeast strains were usually obtained from Open Biosystems (Huntsville, USA; Yeast TAP-Tagged Collection) and were isogenic to BY4741 wild-type strain. All TAP-tagged strains used in this work are listed in Tables 3 to 5. Before these yeast strains were applied to ChIP analysis, the strains were validated. First, gene-specific PCR was performed to confirm that the DNA coding for the TAP-tag was at the correct genomic position (Section 5.1.1). Second, western blotting with anti-TAP antibody (PAP, Table 11) was performed to verify whether the tagged protein of interest was properly expressed. Western blot experiments were conducted as described in Section 5.3.3. Third, the growth of the various tagged yeast strains compared to non-tagged wild-type strain was monitored to rule out any influence of the epitope tag on yeast growth. This was done by serial dilutions of the various yeast strains on YPD plates at 30°C for two days. Only yeast strains which passed all quality controls were used for ChIP analysis.

For standard ChIP-qPCR experiments yeast cultures were grown in 40 ml YPD medium at 30°C to mid-log phase (OD600 ∼0.8), treated with 1% formaldehyde (Sigma) for 20 min at 20°C, and cross-linking was quenched with 5 ml 3 M glycine for 10 min. For Abd1 anchor-away ChIP experiments yeast cultures were grown in 80 ml YPD to an OD600 ∼0.6, split and incubated with equal volumes of either rapamycin (1µg/ml f.c. in DMSO) or DMSO at 30°C for another 60 min before formaldehyde cross-linking. Subsequent steps were performed at 4°C with pre-cooled buffers containing protease inhibitors (PI; Table 15). Cells were collected by centrifugation, washed twice with 1x TBS and twice with FA lysis Buffer. Cell pellets were flash-frozen in liquid nitrogen and stored at -80°C. Pellets were thawed, resuspended in 1 ml FA lysis Buffer, and disrupted by bead beating (Retsch) in the presence of 1 ml silica-zirconia beads (Roth) for 30 min at 4°C. Lysis efficiency was typically >80% as determined by spectrophotometer measurements (BioPhotometer, Eppendorf). Chromatin was solubilized and fragmented via sonication with a BioruptorTM UCD-200 (Diagenode Inc.). Sonication was performed at intensity setting "high" (cycles of 0.5 min on and 0.5 min off) for 35 min with sample cooling on ice after 5, 15 and 25 min. 30 µl and 100 µl of fragmented chromatin samples were saved as input and for control of the average chromatin fragment size (described in Section 5.5.3), respectively. 700 µl of sample (IP sample) was immunoprecipitated with 20 µl IgG Sepharose 6 Fast Flow beads (GE Healthcare) at 4°C for 1 h. The IgG beads were directed against the Protein A content of the C-terminal TAP tag. Immunoprecipitated chromatin was washed three times with FA lysis Buffer, twice with FA lysis Buffer 2, twice with ChIP wash Buffer and once with TE Buffer. Immunoprecipitated chromatin was eluted for 10 min at 65°C with ChIP elution Buffer. Eluted chromatin was digested with Proteinase K (20 µl of 20 mg/ml Proteinase K from *Engyodontium album*, Sigma) at 37°C for 2 h and the reversal of crosslinks was performed at 65°C overnight. DNA was purified with the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions. Thereby, DNA was eluted with 50 µl ddH_2O and was further analyzed by quantitative real-time PCR (Section 5.5.4). All buffers and reagents used for ChIP are listed in Table 15.

5.5.2 ChIP using antibodies against Pol II

ChIP analysis of the various Pol II phospho-isoforms was performed as described in the previous section, but with the following modifications. First, ChIP experiments were conducted in the presence of phosphatase inhibitors (PhoI; Table 15). Second, for chromatin immunoprecipitation a set of monoclonal antibodies with strong specificity and affinity for particular phosphorylated states of the Pol II CTD were applied (Table 10). These antibodies were generated and validated in

the laboratories of Dirk Eick and Elisabeth Kremmer (Helmholtz Zentrum München). The amount of antibody that was used for immunoprecipitation of chromatin was optimized as is described in [146]. 700 µl of chromatin sample was immunoprecipitated with the optimized amount of the respective antibody (see Table 10) at 4°C overnight on a rotating wheel. 25 µl of Protein A and Protein G Sepharose beads (GE Healthcare) were added and incubated at 4°C for 1.5 h on a rotating wheel. Immunoprecipitated chromatin was treated as described in the previous section.

5.5.3 Control of average chromatin fragment size

100 µl of chromatin solution was used to determine the average DNA fragment size that was reached by sonication. 92 µl of TE Buffer and 8 µl of Proteinase K (20 mg/ml) was added to the chromatin sample and incubated at 37°C for 2 h and at 65°C overnight. Next, 20 µl of LiCl (4 M), 1 µl Glycogen and 120 µl Phenol was added, mixed and centrifuged at 13,000 rpm for 10 min (20°C). The supernatant was mixed with 400 µl of pre-cooled Ethanol and incubated at -20°C for 5 h. The immunoprecipitated DNA (and RNA) was pelletized by centrifugation at 13,000 rpm for 20 min (4°C). The pellet was resuspended in 20 µl TE Buffer and RNA was removed by RNase treatment. 10 µl of RNase A/T1 mix (2 mg/ml RNase A, 5000 U/ml RNase T1, Fermentas) were added and incubated at 37°C for 1 h. The resulting DNA sample was electrophoretically separated on a 1.5% agarose gel (see Section 5.1.2). The average DNA fragment size was ∼250 bp.

5.5.4 Quantitative real-time PCR (qPCR)

For ChIP experiments, input and immunoprecipitated (IP) samples were assayed by qPCR to assess the extent of protein occupancy at different genomic regions.
Before ChIP DNA was analyzed by qPCR, primers were designed and the PCR efficiencies of the corresponding primer pairs were determined. DNA primers used for qPCR experiments were 18 to 24 nt long and were designed with the OligoPerfectTM Designer (Invitrogen). The length of the amplified qPCR product was between 60 and 70 nt. The PCR efficiency was determined by qPCR (as described below) with the same primer pair applied to at least four different dilutions of a DNA template (usually DNA that was fragmented by sonication). Based on the resulting standard curve, the PCR efficiency was calculated with the Bio-Rad CFX Manager software version 1.1 according to the manufacturer's instructions. Only primer pairs with a PCR efficiency of $\geq 90\%$ were used in ChIP-qPCR experiments. A list of all qPCR primers used in this work is given in Table 6.
PCR reactions for the analysis of ChIP DNA as well as for validation of qPCR primer pairs contained 1 µl DNA template, 2 µl of 10 µM primer pairs and 12.5 µl iTaq SYBR Green Supermix (Bio-Rad). Quantitative PCR was performed on a Bio-Rad CFX96 Real-Time System (Bio-Rad) using a 3 min denaturing step at 95°C, followed by 49 cycles of 30 sec at 95°C, 30 sec at 61°C and 15 sec at 72°C. Threshold cycle (Ct) values were determined by application of the corresponding Bio-Rad CFX Manager software version 1.1 using the Ct determination mode "Regression". The IP efficiency and fold enrichment of any given region over control regions, such as an open reading frame (ORF)-free heterochromatic region on chromosome V, was determined as described [53].

5.6 Chromatin immunoprecipitation followed by tiling microarray analysis (ChIP-chip)

5.6.1 ChIP for ChIP-chip

For ChIP-chip experiments the standard ChIP protocol as described in Sections 5.5.1 to 5.5.3 had to be adapted as follows. First, yeast cultures were grown in 600 ml YPD medium at 30°C to mid-log phase (OD600 ∼0.8). Second, cell lysis via bead beating was performed for 2 h, with cooling of the sample after 30, 60 and 90 min. Third, the chromatin pellet was washed two times

with FA lysis Buffer before sonication. Washing was performed as follows. The cell lysate was centrifuged at 13,200 rpm for 15 min (4°C), the supernatant was discarded and the chromatin pellet was resuspened with 1 ml FA lysis Buffer. Fourth, the elution from the IgG Sepharose beads was performed at 65°C for 60 min. Fifth, after purification of DNA with the QIAquick PCR Purification Kit (Qiagen), DNA was eluted with 100 µl H_2O and 5 µl RNase A (10 mg/ml, Sigma) was added, and incubated at 37°C for 20 min. DNA was purified with the QIAquick PCR Purification Kit, eluted with 50 µl H_2O, concentrated via vacuum centrifugation, and amplified as described in the following section.

5.6.2 DNA amplification

DNA samples were amplified and re-amplified with GenomePlex© Complete Whole Genome Amplification 2 (WGA2) Kit using the Farnham Lab WGA Protocol for ChIP-chip[3] [165]. Briefly, 10 µl of concentrated ChIP DNA was used to generate the PCR-amplifiable OmniPlex© Library, consisting of ChIP DNA molecules flanked by universal priming sites. Library preparation was performed essentially as described in the technical bulletin of the WGA2 Kit (Sigma). The OmniPlex© Library was then amplified by PCR within a limited number of cycles. This first whole genome amplification step was conducted according to the manufacturer's instructions. The amplified DNA was purified with the QIAquick PCR Purification Kit (Qiagen). DNA was eluted with 50 µl H_2O and the DNA quantity and quality control was performed with a ND-1000 Spectrophotometer (NanoDrop Technologies), and was usually larger than 1 µg. In addition, DNA quality was monitored by agarose gel electrophoresis. 15 ng of purified DNA was re-amplified as described in the technical bulletin of the WGA2 Kit, but with the following modification. Re-amplification was carried out in the presence of 0.4 mM dUTP. Incorporation of dUTP was a prerequisite for the enzymatic fragmentation of DNA (described in the following section). After amplification DNA was purified with the QIAquick PCR Purification Kit. DNA was eluted with 50 µl H_2O and the DNA quantity and quality control was assessed with a ND-1000 Spectrophotometer, and agarose gel electrophoresis.

5.6.3 DNA fragmentation, labeling and microarray processing

The enzymatic fragmentation, labeling, hybridization and array scanning were done according to the manufacturer's instructions[4]. Enzymatic fragmentation and terminal labeling were performed by application of the GeneChip WT Double-Stranded DNA Terminal Labeling Kit (P/N 900812, Affymetrix). Briefly, re-amplified DNA was fragmented in the presence of 1.5 µl Uracil-DNA-glycosylase (10 U/µl) and 2.25 µl APE1 (100 U/µl) at 30°C for 1 h 15 min. The average fragment size was in the range of 50-70 bp as determined by automated gel electrophoresis on an Experion system (Bio-Rad) that allows the analysis of small amounts of DNA. The fragmented DNA was then labeled at the 3´-end by adding 2 µl and 1 µl of Terminal nucleotidyl transferase (TdT, 30 U/µl) and GeneChip DNA Labeling Reagent (5 mM), respectively.
5.5 µg of fragmented and labeled DNA were hybridized to a high-density custom-made Affymetrix tiling array [44] (PN 520055) at 45°C for 16 h with constant rotational mixing at 60 rpm in a GeneChip Hybridization Oven 640 (Affymetrix). Washing and staining of the tiling arrays were performed using the FS450_0001 script of the Affymetrix GeneChip Fluidics Station 450. The arrays were scanned using an Affymetrix GeneChip Scanner 3000 7G. The resulting raw data image files (.DAT) were inspected for any impairment.

[3] http://www.genomecenter.ucdavis.edu/farnham/protocol.html
[4] Affymetrix Chromatin Immunoprecipitation Assay Protocol P/N 702238

5.7 ChIP-chip data analysis

The data generated by one ChIP-chip experiment consists of an intensity value for each DNA probe. These values measure the relative quantity of DNA at the probe's genomic position in the immunoprecipitated or input (reference) material. The fluorescence intensity for each probe is determined from the scanned array image, and when using the Affymetrix platform saved in a binary .CEL file. The .CEL files together with a mapping (.bpmap) file, which maps each probe to its genomic position along the reference genome, are the input files for further data processing and analysis.

5.7.1 Data preprocessing and normalization

The first steps in data preprocessing comprise the quality control of the obtained data as well as normalizations to render the data comparable between different arrays, to correct for saturation effects, and to obtain enrichment and occupancy values.
All data normalization procedures performed in this work were done using R [183] and Bioconductor [58]. For data import of the Affymetrix .CEL files and the conversion into the basic Bioconductor object class for microarray data *ExpressionSet*, we used the R package Starr [248]. Quality assessment of each measured array was done by inspection of raw image files, density-plots, scatter-plots, and MA-plots, in order to avoid processing of flawed arrays (for a detailed protocol see online[5]). Subsequent data normalization consisted of three steps. First, we performed quantile normalization between replicate measurements (not between non-replicate measurements). Second, for each condition (including the reference measurements) we averaged the signal for each probe by calculating the geometric average over the replicate intensities. Third, data obtained using TAP-tagged factors was normalized using a combined mock IP plus input reference normalization, whereas data obtained using specific antibodies (e.g. Pol II phospho-isoforms [Y1P, S2P, T4P, S5P, S7P]) was normalized simply by dividing through input reference intensities. The rationale for the last step is explained in detail below.

In ChIP-chip experiments, it is important to correct for sequence- and genomic region-specific biases in the efficiency of the various biochemical and biophysical steps. Cross-linking, chromatin fragmentation, IP, amplification, labeling, and hybridization to the array can produce biases. Although reference-free normalization procedures have been suggested that can reduce these biases [97], the cleanest and most efficient method is to measure a reference signal and to divide the true signal obtained from the IP by the reference intensities, allowing accurate ratio measurements to be obtained. This is in our experience the most important step in data normalization when using single-color arrays with short probes such as the Affymetrix arrays used here. These arrays exhibit a strong probe sequence bias, meaning that the GC-content of a probe influences its measured intensity.
As reference signal one can use the intensities obtained by hybridizing the input (sheared chromatin before IP) fraction to a tiling array or, alternatively, a mock IP (using an unrelated antibody for which there is no corresponding epitope). Here, we developed a simple mathematical model to describe the fragmentation, amplification, labeling, and hybridization biases, as well as the bias through unspecific antibody binding, in order to understand the effects of different normalization procedures. We found that, using either the mock IP or the matched input as reference signal, one can only correct for some of these effects. Based on our model, we derive a combined normalization using both mock IP and input signal which should correct for all these sources of bias. Crucial to our method is that our mock IP employs the same antibody as the factor IP, but using an isogenic wild-type yeast strain lacking the IP epitope. In experiments using epitope-tagged factors,

[5] http://www.epigenesys.eu/images/stories/protocols/pdf/20111025112610_p47.pdf

as the TAP-tagged factors investigated here, this is easily achieved by using the corresponding wild-type strain lacking the tag. This way, we measure unspecific binding of the TAP tag-directed antibody that can be used for subtracting the unspecific binding component in the factor IP. Since for ChIP-chip of the Pol II phospho-isoforms no mock IP measurements with the same antibody as used for the IP could be done, only input reference normalization was performed.

Combined mock IP and input reference normalization: Let x be the genomic coordinate, $B(x)$, $M(x)$, and $S(x)$ the array signals obtained from hybridizing the input, the mock IP, and the factor IP, respectively, and $p(x)$ the occupancy profile due to the specific binding of the antibody to the factor of interest. We can model these array signals by

$$B(x) = aB \times b(x),$$

$$M(x) = aM \times b(x) \times u(x),$$

$$S(x) = aS \times b(x) \times (u(x) + p(x)),$$

with aB, aM, and aS unknown scaling constants for the input, mock IP, and factor IP array measurements, respectively, $b(x)$ the input intensity profile describing the fragmentation, amplification, labeling, and hybridization biases, and $u(x)$ the profile describing the effect of unspecific binding of the antibody to other DNA-bound proteins and protein complexes. We seek to obtain the occupancy profile $p(x)$. Using normalization with the input, we would get

$$S(x)/B(x) = (aS/aB) \times (p(x) + u(x)),$$

showing that the unspecific binding of the antibody will lead to distortions of the ChIP enrichment signal which will be the more serious the less specific the antibody binds to the factor and the less sequence specificity the factor has in binding to the genomic DNA. Using normalization with the mock IP, we would obtain

$$S(x)/M(x) = (aS/aM) \times (1 + p(x)/u(x)),$$

which should work better than the previous version in cases where unspecific binding dominates the signal from the specific binding, but which introduces a bias through sequence-specific effects of the unspecific binding (e.g. through a nucleosome density-mediated GC bias of the unspecific binding signal).
We therefore introduce a combined normalization using both the input and mock IP signals:

$$(S(x) - (aS/aM) \times M(x))/B(x) = (aS/aB) \times p(x).$$

If we assume that $X\%$ of the genomic regions do not bind the factor of interest, we can estimate the factor aS/aM as the global $(X/2)\%$ quantile of $S(x)/M(x)$,

$$aS/aM = Q_{(X/2)\%}[S(x)/M(x)]$$

since with this choice $(X/2)\%$ of the values of the normalized signal $(S(x) - (aS/aM) \times M(x))/B(x)$ will be below zero. Note that, in practice, the value of aS/aM depends only weakly on the value of X and can be estimated to much better than a factor 2. Only when severely overestimating it (by more than a factor of 2) the correction of unspecific binding will be detrimental.

To be able to give absolute occupancy values on a scale between 0% and 100%, we need to estimate the factor aS/aB. For this purpose, we assumed that the highest occupancies of our measured factors correspond to 100% occupancy. We estimate the highest genome-wide occupancies using the $Y\%$ quantile instead of the genome-wide maximum probe signal to obtain an estimation that is robust against statistical noise. In this study, we chose a quantile of $Y = 99.8\%$ for all factors, which corresponds to the highest-bound ∼6000 probes on our tiling arrays. Then, aS/aB is estimated as the $Y\%$ quantile of the factor occupancy:

$$aS/aB = Q_{Y\%}[(S(x) - (aS/aM)\times M(x))/B(x)]$$

Our normalization method should improve on the normalization procedures commonly used by correcting the signal for unspecific antibody binding. Our method cannot correct for sequence dependent effects of cross-linking efficiency. These effects represent, however, an inherent limitation of ChIP-chip and ChIP-seq techniques.

We used the combined mock IP plus input normalization method to calculate occupancy profiles for all TAP-tagged factors. ChIP-chip data obtained using specific antibodies (e.g. Pol II phospho-isoforms [Y1P, S2P, T4P, S5P, S7P]) was normalized simply by dividing through genomic input intensities, since for these no mock IP measurements with the same antibody as used for the IP could be done.

Replicate measurements, reference samples, and data quality control: At least two independent biological replicates were analyzed for each factor (Pearson correlations between replicates were usually $R > 0.8$). Using N replicates reduces the standard deviation of unsystematic noise by a factor of $\sqrt{N}$. It is important that correlation calculations between replicates are performed after reference normalization. Mock IP and input measurements were used for normalization (see above). Two biological replicates were used for the mock IP ($R = 0.65$). Due to the very high reproducibility/correlation between input samples of different factors (comparable to factor replicate correlations, $R > 0.8$), we took input samples of three factors, Rpb3, Spt4, and Spt6, and used them as triplicate measurements to normalize all factors investigated in Section 6 except the Pol II phospho-isoforms and Spt6ΔC. The latter and all factors investigated in Sections 7 and 8 were normalized using their matched input data.

5.7.2 Calculation of transcript-wise occupancy profiles

In order to calculate occupancy profiles over genes or other genomic features the normalized occupancy signal at each nucleotide of the region was calculated as the median signal of all probes overlapping this position (6.5 probes on average). Individual probe intensities were further smoothed using the sliding window smoothing procedure (window half size of 75 bp) implemented in the R package Ringo [220].

5.7.3 Transcript-averaging of occupancy profiles

For metagene analysis we averaged the normalized ChIP-chip profiles over representative sets of prefiltered genes: We start with all nuclear *S. cerevisiae* S288C protein-coding genes classified as 'verified' or 'uncharacterized' by the Saccharomyces Genome Database (SGD, 5769 genes). To align gene profiles across entire transcripts, only genes with available TSS and pA assignments from RNA-seq experiments [159] were taken into account (4366 genes). Genes with TSS (pA) measurements downstream (upstream) of the annotated ATG (Stop) codon were excluded. To remove possible wrongly annotated TSSs and pAs, we only included genes with TSS (pA)

annotations showing a distance of less than 200 bp to the corresponding downstream (upstream) ATG (Stop) codon (3448 genes). As a result of the limited ChIP-chip resolution and the compactness of the yeast genome with its short intergenic regions (median inter-ORF length: 368 bp, median inter-transcript length: 259 bp), a gene's factor occupancy profile can have spurious contributions from flanking genes. To minimize these "spill-over" effects, we focused on genes exhibiting a minimal ORF and transcript distance to flanking genes of 250 bp and 200 bp, respectively (1786 genes). Furthermore, we restricted our analysis to the 50% highest expressed nuclear protein-coding genes according to [46] (1140 genes, ALL set). We grouped genes into four ORF length classes: Xtremely Short (XS) ranging from 256 to 511 bp, Short (S) 512 to 937 bp, Medium (M) 938 to 1537 bp, and Long (L) 1538 to 2895 bp, comprising 93, 266, 339, and 299 genes, respectively. Profiles within these or other (e.g. ribosomal protein genes) selected gene classes were scaled to median gene length, and gene-averaged profiles calculated by taking the median or the 5% trimmed mean at each genomic position over the ChIP-chip profiles.

To facilitate the comparison of elongation factor occupancies and in particular their slope in the region near the TSS, we shifted the traces in Figure 6E and F (Section 6) by up to 0.1 on the occupancy scale such that they overlapped in the region [TSS-250, TSS].

To avoid scaling of profiles within gene classes, ALL genes were cut around the TSS (for Figure 17A: from 250 bp upstream to 650 bp downstream; only genes longer than 680 bp were considered; 910 genes) or the pA site (for Figure 28: from 400 bp upstream to 400 bp downstream; only genes longer than 800 bp and with more than 400 bp distance to neighboring genes were considered; 619 genes) and averaged at each genomic position.

5.7.4 Calculation of factor profile peaks and transitions

To calculate peak positions of averaged ChIP-chip profiles, median profiles were calculated in a region ± 250 bp around the TSS or the pA site, and smoothed with cubic splines (R package: stats, function: smooth.spline, parameter: spar = 0.9). The maximum value of the smoothed curve was selected as peak position. To estimate the uncertainty of the peak position, we drew 1000 bootstrap samples from the set of genes, recalculated peak positions for each bootstrap sample as described, and estimated their scatter using the Median Absolute Deviation (MAD) measure.
We also used the gene-averaged ChIP-chip profiles to calculate transitions in the transcription complex, i.e. genomic positions where factors associate with or dissociate from transcribing Pol II. To determine 5´ transitions, median profiles were calculated in a region ± 150 bp around the TSS, and smoothed with cubic splines. The position at 11.5% occupancy of the smoothed curve (corresponding to the point after which the elongation factor profiles start to change) was defined as 5´ transition point. 3´ transitions were determined as the percentage of the maximum occupancy at 100 bp downstream of the pA site. Median profiles were calculated in a region from 250 bp upstream of the TSS to 250 bp downstream of the pA site, and smoothed with cubic splines. The uncertainty of the 5´ and 3´ transitions was determined as described above.

To calculate gene-wise peak distances between capping enzymes Cet1-Ceg1 and Abd1 (Section 7, Figure 18B), ChIP-chip profiles were cut around the TSS (from 150 bp upstream to 350 bp downstream) for each gene contained in the M gene length class (see Section 5.7.3), and smoothed with cubic splines. The maximum value of the smoothed curve was selected as peak position, and the distances calculated. Only genes with detectable peaks were selected.

5.7.5 Calculation of pairwise profile correlations and correlation network construction

Analyses were done using 4366 genes with available TSS and pA annotations [159]. Pairwise Pearson correlations over factor occupancy profiles were calculated between concatenated gene profiles, ranging each from TSS−250 bp to pA+250 bp, and provided as a similarity metric. The correlation-based network was calculated using the GraphViz's Neato algorithm [56] employing an edge-weighted, spring-embedded layout procedure attempting to minimize a global energy function, which is equivalent to statistical multi-dimensional scaling.

5.7.6 Singular value decomposition (SVD)

For each factor (or Pol II phospho-isoform) f and for each of the 4366 genes g for which we had TSS and pA annotations [159], we calculated 90%-quantiles of occupancies within a region [TSS−250 bp, pA+250 bp] as a robust proxy for peak occupancies. This resulted in a $f \times g$ matrix. From each matrix element, we subtracted the average over its row (i.e. over its factor). The resulting matrix X_{fg} was subjected to singular value decomposition (SVD), yielding singular values $\sigma_1 \geq ... \geq \sigma_f \geq 0$ and unit-length, orthogonal, singular vectors $u_1, .., u_f, v_1, .., v_f$, such that $X_{fg} = \Sigma_{i=1,..,f} \sigma_i \times u_{if} \times v_{ig}$. The k'th term in this sum, $\sigma_k \times u_{kf} \times v_{kg}$, can explain a fraction $\sigma_{k^2} / \Sigma_i \sigma_{i^2}$ of the data variance.

To reveal correlations contained in the fraction of the total variance that was not contributed by the first term in the SVD, we subtracted from the data matrix the values from the first term of the SVD, i.e., $X_{fg} - \sigma_1 \times u_{1f} \times v_{1g}$. This resulted in a matrix of residual correlations (see Section 6, Figure 8D). To ensure that the residual correlations were not caused by spill-over effects among neighboring genes, we used a very stringently filtered set of 97 spatially well-separated genes. First, we demanded that the distances to the nearest 'verified' or 'uncharacterized' nuclear ORF, snoRNA, snRNA, ncRNA, CUT or SUT (according to SGD and [242]) be at least 500 bp. Second, we used only genes whose neighboring genomic transcripts were both annotated to be transcribed from the same strand. This way, we made sure that TSSs and pAs of neighboring transcripts were well separated. The resulting matrix of residual correlations [146] is very similar to the one shown in Figure 8D, confirming the validity of the analysis on the full set of 4366 genes. We determined the standard errors of each of the correlation coefficients by taking bootstrap samples from the columns of matrix $X_{fg} - \sigma_1 \times u_{1f} \times v_{1g}$ and obtained errors of ± 0.047 and below. Hence, the residual correlations are not caused by statistical noise but rather mirror actual physical and functional associations.

Part III
Results and Discussion

6 Uniform transitions of the general RNA polymerase II transcription complex

All results presented in this Section were obtained in collaboration with Andreas Mayer and Matthias Siebert and are published in [146]. For detailed author contributions see page 9.

Gene transcription begins with the assembly of Pol II and its initiation factors on promoter DNA. Pol II then starts mRNA synthesis and exchanges initiation factors for elongation factors, which are required for chromatin passage and RNA processing [167, 168, 176] (for details see Section 1). Whereas Pol II is unphosphorylated during initiation, it is phosphorylated at its CTD during elongation. The CTD is phosphorylated at Ser5 residues in the 5´ region of a gene and at Ser2 residues in the 3´ region [114, 199]. The phosphorylated CTD recruits elongation factors to ensure cotranscriptional RNA processing and chromatin modification [19, 81, 151, 167, 168, 174]. A genome-wide study has shown that initiation factors are present at all active Pol II gene promoters [225], but it is unknown whether all elongation factors are recruited to all active genes, and no genome-wide studies have examined whether factor recruitment correlates with specific Pol II phosphorylations.
To address these questions, we used high-resolution genome-wide occupancy profiling by chromatin immunoprecipitation (ChIP) of Pol II, its phosphorylated forms, its elongation factors and components of the Pol II initiation and termination machinery in proliferating yeast cells. Statistical analysis provides strong evidence for a general elongation complex—that is, one composed of all elongation factors—that mediates chromatin transcription and mRNA processing at all Pol II genes. The general elongation complex is apparently established during a 5´ transition within a narrow window just downstream of the TSS, and it is disassembled in two major steps during a 3´ transition around the polyadenylation (pA) site. The results also show that CTD phosphorylation patterns previously observed at individual genes occur globally and that levels of CTD phosphorylation do not correlate with the *in vivo* occupancy of two factors that bind the phosphorylated CTD *in vitro*. General elongation complexes are active, as their gene occupancy predicts mRNA expression levels.

6.1 Genome-wide profiling reveals Pol II on a majority of genes

We determined genome-wide occupancy profiles by ChIP in exponentially growing *S. cerevisiae* strains expressing tandem affinity purification (TAP)–tagged proteins. Chromatin immunoprecipitation was performed as described in Sections 5.5 and 5.6. Enriched DNA fragments of an average size of ∼250 bp were analyzed with tiling microarrays that cover the yeast genome at 4-bp resolution [44]. For data normalization, we developed a procedure that corrects for nonspecific antibody binding by using input measurements as well as mock immunoprecipitations (Section 5.7). Data from two or three highly reproducible replicates ($R \geq 0.8$) were averaged. The ChIP-chip occupancy profile for the Pol II subunit Rpb3 (Figure 6) matched previous profiles [91] obtained with different strains, experimental protocols and array platforms, but the new profile showed more details.
Pol II was observed at genes encoding proteins, small nuclear RNA and small nucleolar RNA, and at regions producing cryptic unstable and unannotated transcripts [242], but was lacking at genes transcribed by Pol I and Pol III (Figure 6A-C and [146]). Of 4,366 yeast genes with annotated TSS and pA sites [159], 2,465 (56%) showed Pol II peak occupancies above 20%, consistent with transcription of most of the genome [44]. To average Pol II profiles over genes, we examined the 50% most highly expressed genes [46] that were at least 200 bp away from neighboring genes. These

were sorted into four main length classes, scaled to adjust for length differences and aligned at their TSS and pA sites (Section 5.7.3). The pA site marks the point of RNA 3´ cleavage and polyadenylation, but transcription continues beyond this site until termination. Consistent with this, the gene-averaged Rpb3 profile revealed Pol II occupancy through the transcribed region and into the region flanking the pA site on the 3´ side (Figure 6C).

6.2 Initiation and termination factors flank the transcribed region

Gene-averaged profiles for the initiation factors TFIIB, TFIIF and TFIIH showed a single strong peak 50–30 bp upstream of the TSS (Figure 6D and Table 16). This indicates the presence of initiation complexes at promoters and is consistent with a scanning mechanism for TSS location in yeast [115, 118]. TFIIF was found only at promoters and not within transcribed regions, indicating that its reported elongation-stimulatory activity *in vitro* [187] is restricted *in vivo* to early RNA elongation and to downstream sites of transient association. The weaker peaks for initiation factors observed downstream of the pA site are mostly due to residual spillover effects from closely spaced genes on the same strand. When we averaged only over convergently transcribed genes, the peaks were reduced two- to three-fold (Figure 7). The remaining peaks may indicate gene looping at selected genes. Occupancy of the capping enzyme subunit Cet1 peaked just downstream of the TSS, consistent with capping when the nascent RNA appears on the Pol II surface. The symmetric peaks of averaged occupancy of initiation factor and capping enzyme indicate that these factors are restricted to defined locations just upstream and downstream, respectively, of the TSS. Occupancy of the 3´ processing and termination factor Pcf11 peaked downstream of the pA site, consistent with transcription and completion of mRNA 3´-end formation downstream of the pA site. Thus, representative initiation and termination factors show peak occupancies outside the transcribed region and are apparently not present during mRNA chain elongation.

6.3 Elongation factors enter during a single 5´ transition

Elongation factor profiles did not correlate with profiles of initiation factors or termination factors (Figure 8A). Elongation factors were absent at the promoter, but their occupancies sharply increased downstream of the TSS within a narrow window of ∼50 bp, indicating coordinated elongation complex assembly during a single 5´ transition (Figure 6E,F and Table 16). Spt16 was an exception, entering ∼30 bp further upstream. Elongation factors showed characteristic distributions over the transcribed region. We observed three distinct profile shapes and used them to group the factors (Figure 8B). Group 1 includes Spt4, Spt5 and Spt6, group 2 includes Spn1 and Elf1, and group 3 includes Spt16, Paf1 and the CTD kinases Bur1 and Ctk1.

6.4 Spn1 and Elf1 interact within a Pol II complex

The similar gene-averaged profiles of the poorly characterized factors Spn1 and Elf1 suggested that these factors interact. To test this, we purified Spn1 from yeast using a TAP tag. Spn1 copurified with Elf1 and Pol II (Figure 9), consistent with an interaction between Spn1 and Elf1. To probe for a direct Spn1-Elf1 interaction, we coexpressed the two factors in bacteria. Spn1 and Elf1 did not copurify after coexpression (data not shown). These results suggest that Spn1 and Elf1 interact indirectly within a Pol II complex, and the profiling data suggest that their recruitment and functions during the transcription cycle are distinct from those of other elongation factors.

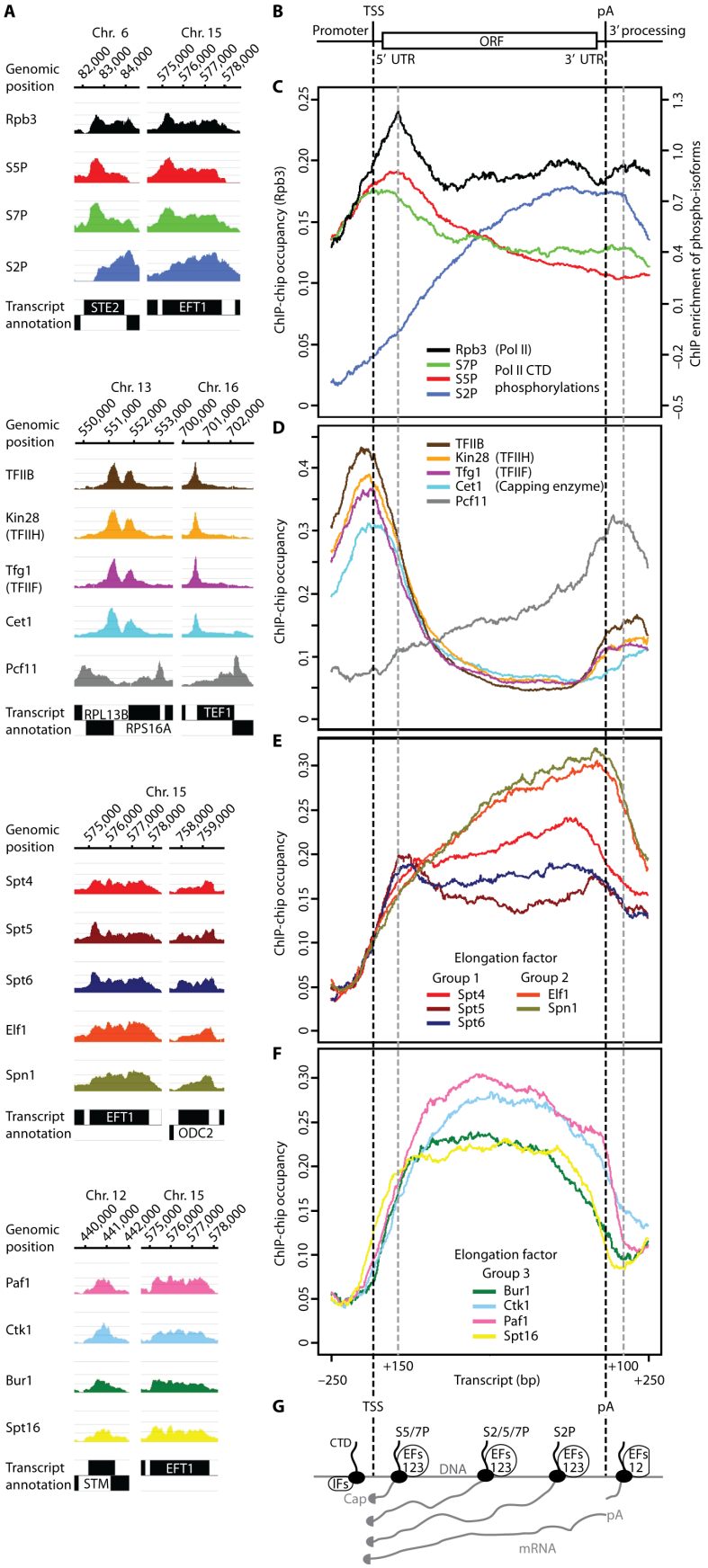

Figure 6: Genome-wide occupancy profiling of the Pol II machinery. (A) Factor occupancy on selected genes. Colored profiles represent normalized factor occupancies smoothed by a 150-bp window running median. The color code is used throughout figures. Black boxes indicate transcripts [44] on the Watson (top) and Crick strands (bottom). (B) DNA frame with promoter, 5′ untranslated region (UTR), open reading frame (ORF) and 3′ UTR. Dashed black lines indicate the TSS and pA site. Dashed gray lines mark the positions 150 bp downstream of the TSS and 100 bp downstream of the pA site. (C) Gene-averaged profiles for the median gene length class (1,238 ± 300 bp, 339 genes) of Pol II and its phosphorylated forms. Profiles of other length classes are generally similar [146]. Occupancies and signal intensities are given for Rpb3 and phosphorylated Pol II on the left and right y axes, respectively. For details on ChIP-chip profile calculations see Sections 5.7.1-5.7.3. (D) Gene-averaged profiles as in C for initiation (TFIIB, TFIIF, TFIIH), 5′ capping (Cet1) and termination (Pcf11) factors. (E,F) Gene-averaged profiles as in C for elongation factors of groups 1 (Spt4, Spt5, Spt6), 2 (Elf1, Spn1) and 3 (Spt16, Paf1, Ctk1, Bur1). (G) Cartoon representation of Pol II (black dots) and its CTD (black lines) transcribing along DNA (horizontal gray line) from left to right, to produce mRNA (gray lines). IFs, initiation factors; EFs 123, elongation factors of groups 1, 2 and 3; S2/5/7P, phosphorylation of serines 2, 5 and 7.

Feature:	Peaks around TSS[a]		5´ transition (position)[b]	Peaks around pA site		3´ transition (% of max)[c]
Genes[d]:	M	ALL	M	M	ALL	M
Pol II						
Rpb3	149 ± 10	131 ± 6	−20 ± 21			
S7P	26 ± 44	18 ± 21				
S5P	128 ± 16	117 ± 10				
S2P				−252 ± 36	−204 ± 51	
Initiation						
TFIIB	−52 ± 13	−45 ± 6				
TFIIH (Kin28)	−41 ± 10	−25 ± 6				
TFIIF (Tfg1)	−34 ± 10	−36 ± 6				
5´ capping						
Cet1	6 ± 18	14 ± 8				
Elongation						
Spt4			20 ± 13	−204 ± 16	−211 ± 12	67 ± 4
Spt5	195 ± 12	197 ± 9	15 ± 19			71 ± 6
Spt6	202 ± 12	204 ± 9	25 ± 16			75 ± 4
Elf1			26 ± 15	−73 ± 28	−74 ± 9	83 ± 3
Spn1			42 ± 15	−52 ± 28	−52 ± 9	83 ± 3
Bur1			62 ± 12			26 ± 4
Ctk1			66 ± 9			46 ± 3
Paf1			36 ± 10			29 ± 3
Spt16			−13 ± 12			21 ± 4
Termination						
Pcf11				72 ± 13	58 ± 6	

Table 16: Peaks and transitions of gene-averaged ChIP-chip profiles.
[a] Peak positions were determined separately for the ALL gene set (1140 genes) and length class M (339 genes). Median profiles were calculated in a region ± 250 bp around the TSS or the pA site, and smoothed with cubic splines (R package: stats, function: smooth.spline, parameter: spar = 0.9). The maximum value of the smoothed curve was selected as peak position. To estimate the uncertainty of the peak position, we drew 1000 bootstrap samples from the set of genes, recalculated peak positions for each bootstrap sample as described, and estimated their scatter using the Median Absolute Deviation (MAD) measure.
[b] 5´ transitions were determined for length class M. Median profiles were calculated in a region ± 150 bp around the TSS, and smoothed with cubic splines. The position at 11.5% occupancy of the smoothed curve (corresponding to the point after which the elongation factor profiles start to change) was defined as 5´ transition point. To estimate the uncertainty of the 5´ transition, we drew 1000 bootstrap samples from the set of genes, recalculated 5´ transitions for each bootstrap sample as described, and estimated their scatter using the MAD measure.
[c] 3´ transitions were determined for length class M as the percentage of the maximum occupancy at 100 bp downstream of the pA site. Median profiles were calculated in a region from 250 bp upstream of the TSS and 250 bp downstream of the pA site, and smoothed with cubic splines. To estimate the uncertainty of the 3´ transition, we drew 1000 bootstrap samples from the set of genes, recalculated 3´ transitions for each bootstrap sample as described, and estimated their scatter using the MAD measure.
[d] Gene selection to calculate gene-averaged profiles is described in Section 5.7.3.

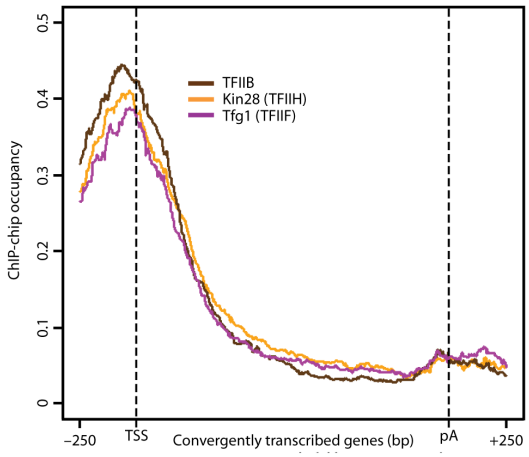

Figure 7: Gene-averaged profiles restricted to convergently transcribed genes. Gene-averaged median profiles for initiation (TFIIB, IIF, IIH) factors (genes from length class M as shown in Figure 6D) showing only convergently transcribed genes (99 out of 339 genes) in order to diminish spillover effects.

6.5 Elongation factors exit during a two-step 3´ transition

Around the pA site, two steps of a 3´ transition could be distinguished. The two-step transition was most easily seen at genes with high factor occupancies, such as ribosomal protein genes (Figure 10 and Table 16). Whereas group 3 factor occupancies sharply decreased upstream of the pA site, factors from groups 1 and 2 apparently exited further downstream, suggesting they are present during RNA 3´-end formation and possibly during transcription termination. Spn1 and Elf1 peaked just upstream of the pA site, and 100 bp downstream of this site they were still present at about 80% of their peak occupancies (Figure 6E and Table 16).

6.6 A general elongation complex for chromatin transcription

High correlations between elongation factor profiles (Figures 8A and 11) suggested that all elongation factors cooccupy active genes. To investigate this, we measured covariation in the data sets by singular value decomposition (SVD, Section 5.7.6). We calculated peak occupancies for nine elongation factors within 4,366 genes. After subtracting the row mean of the 9 × 4,366 matrix from each element, we subjected the resulting matrix to SVD. The first singular value, which describes strict covariation, explained 92% of the total variance (Figure 8C, left, and [146]). Thus, elongation factor occupancies covaried over all genes, consistent with a general composition of the elongation complex. The apparent elongation complex composition and coordinated assembly during a 5´ transition were independent of gene length, expression, function, transcript type, size of the nucleosome-depleted promoter region and the presence of introns (Figure 12 and [146]). Although differences in the composition of elongation complexes in individual cells cannot be ruled out, these results indicate a general initiation-elongation transition and a general elongation complex composition on Pol II genes.

6.7 CTD phosphorylation profiles depend on TSS location

To investigate how the observed profiles and transitions correlate with CTD phosphorylation, we determined ChIP-chip profiles for Pol II phosphorylated at CTD residues Ser7, Ser5 and Ser2 using site-specific antibodies [22] (Table 10). The averaged profiles revealed broad peaks of Ser7 and Ser5 phosphorylation at around 20 and 120 bp downstream of the TSS (Figures 6C, 13A and Table 16). Ser7 and Ser5 phosphorylation decreased over the transcribed region, whereas Ser2 phosphorylation increased, saturated at 600–1,000 bp downstream of the TSS and sharply

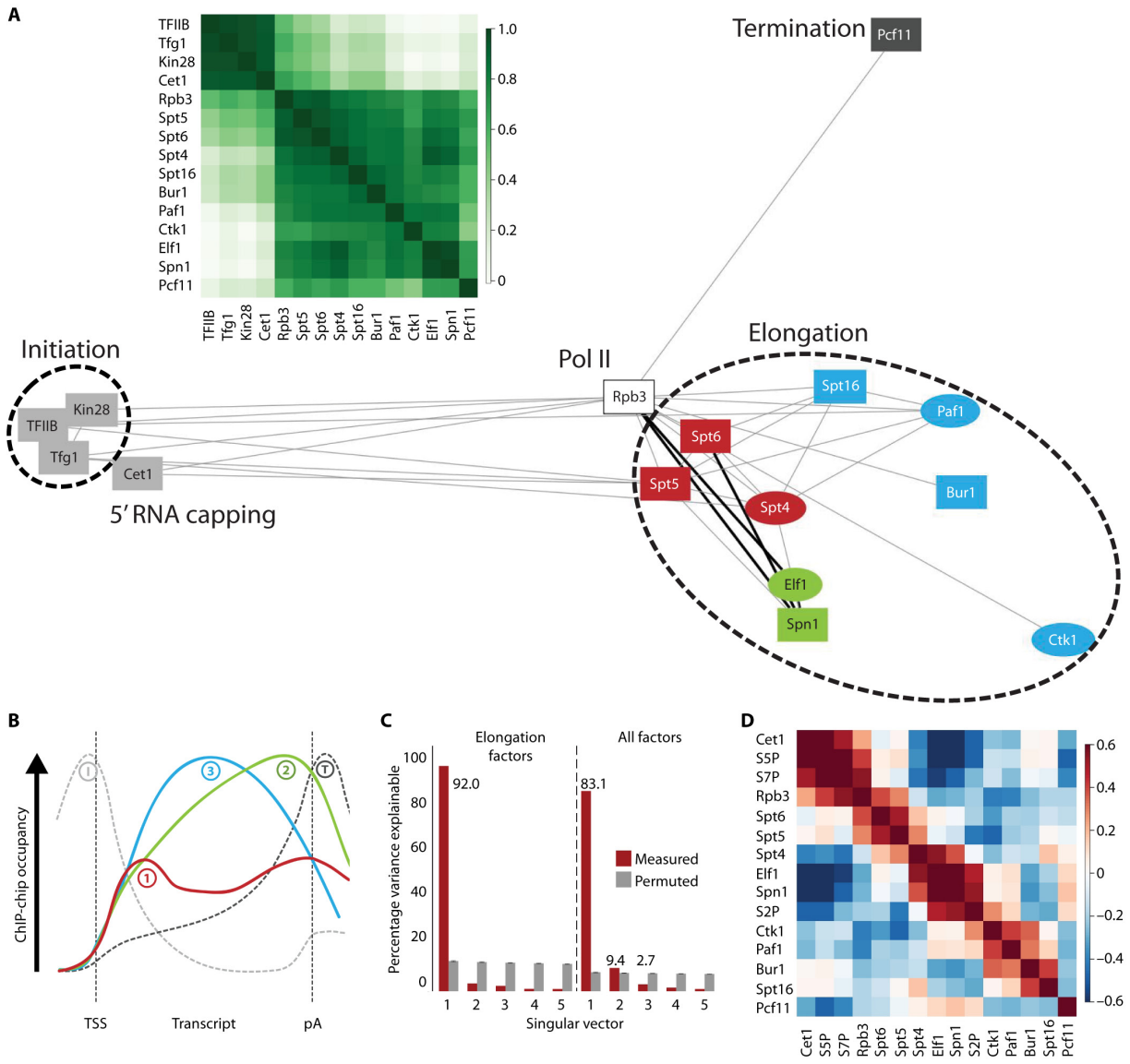

Figure 8: Statistical analysis indicates a general elongation complex. (A) Correlation analysis of genome-wide occupancy profiles. Initiation factors TFIIB, Tfg1 (TFIIF), Kin28 (TFIIH) and the capping enzyme Cet1 have similar profiles that are distinct from those of the nine elongation factors and the termination factor Pcf11. Elongation factors of groups 1, 2 and 3 (dark red, light green and blue) cluster in two dimensions when Pearson correlation coefficients between occupancy profiles are provided as similarity metric (see Section 5.7.5). Lines represent direct and functional interactions previously known (gray) or described here (black). Factors represented by ovals are not essential in yeast. (B) Pol II factors can be grouped by their gene-averaged profiles. See text for details. (C) SVD analysis. The contributions of the first five singular vectors to the variance (red) are shown in comparison to a control with randomly permuted matrix elements (gray). SVD reveals that 92% of the variance of peak occupancies of elongation factors at each gene can be explained by strictly covarying factor occupancies as a contribution from the first singular vector (left). When all factors are included, 83.1% of the variance is explained by covariation (right). (D) Residual correlations described by all but the first singular vector reveal a modular substructure among factors and phosphorylated Pol II forms, suggesting physical and functional interactions. For details on SVD analysis see Section 5.7.6.

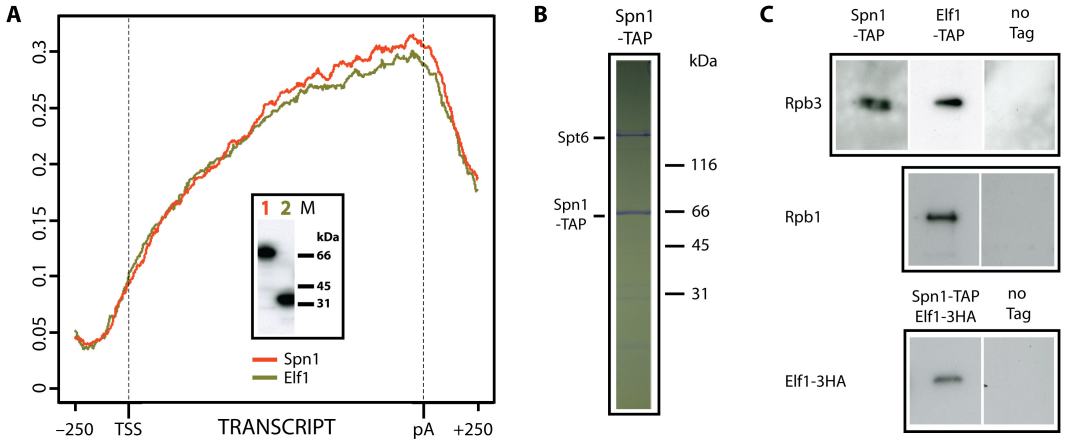

Figure 9: Elf1 copurifies with Spn1 and Pol II. (A) Gene-averaged profiles for Elf1 (orange) and Spn1 (green). Western blotting analysis was performed to validate the TAP strains (inset). The results for the Spn1- and Elf1-TAP strains are indicated on lane 1 and 2, respectively. The molecular weight marker is given on lane M. (B) Spn1 TAP: the purified protein sample was analyzed by SDS-PAGE and Coomassie staining. Spt6 could be identified by mass spectrometry. The molecular weight marker (kDA) is given on the right. Spn1 copurifies with Spt6. (C) Elf1 and Spn1 TAPs: the protein samples were analyzed by SDS-PAGE and western blotting. The TAP strains as well as the control strain (without TAP tag) are given above the panels. The proteins identified by western blotting are indicated on the left. Elf1 copurifies with two Pol II subunits Rpb1 and Rpb3. Spn1 copurifies with Rpb3 and Elf1 bearing a C-terminally 3HA tag. For the latter experiment, a yeast strain was generated that expressed both, a C-terminally 3HA-tagged version of Elf1 as well as a C-terminally TAP-tagged version of Spn1. Experiments in B and C were done by Andreas Mayer.

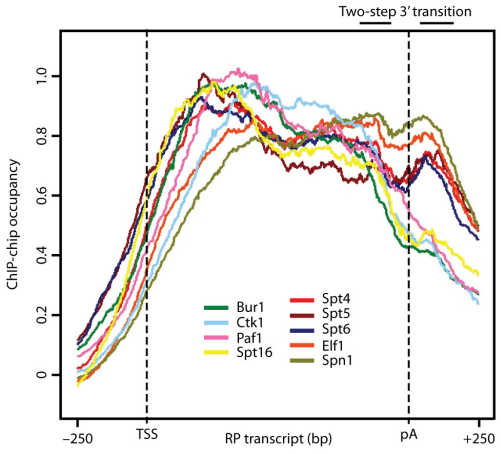

Figure 10: Two-step 3′ transition observed at ribosomal protein (RP) genes. Shown are averaged elongation factor profiles on selected ribosomal protein genes (for details on gene selection see Section 5.7.3). Dashed lines indicate the TSS and pA sites. The regions of the two-step 3′ transition are indicated. In contrast to the averaged profiles in Figure 6, the very high occupancies did not allow us to align profiles with their promoter minima along the y axis.

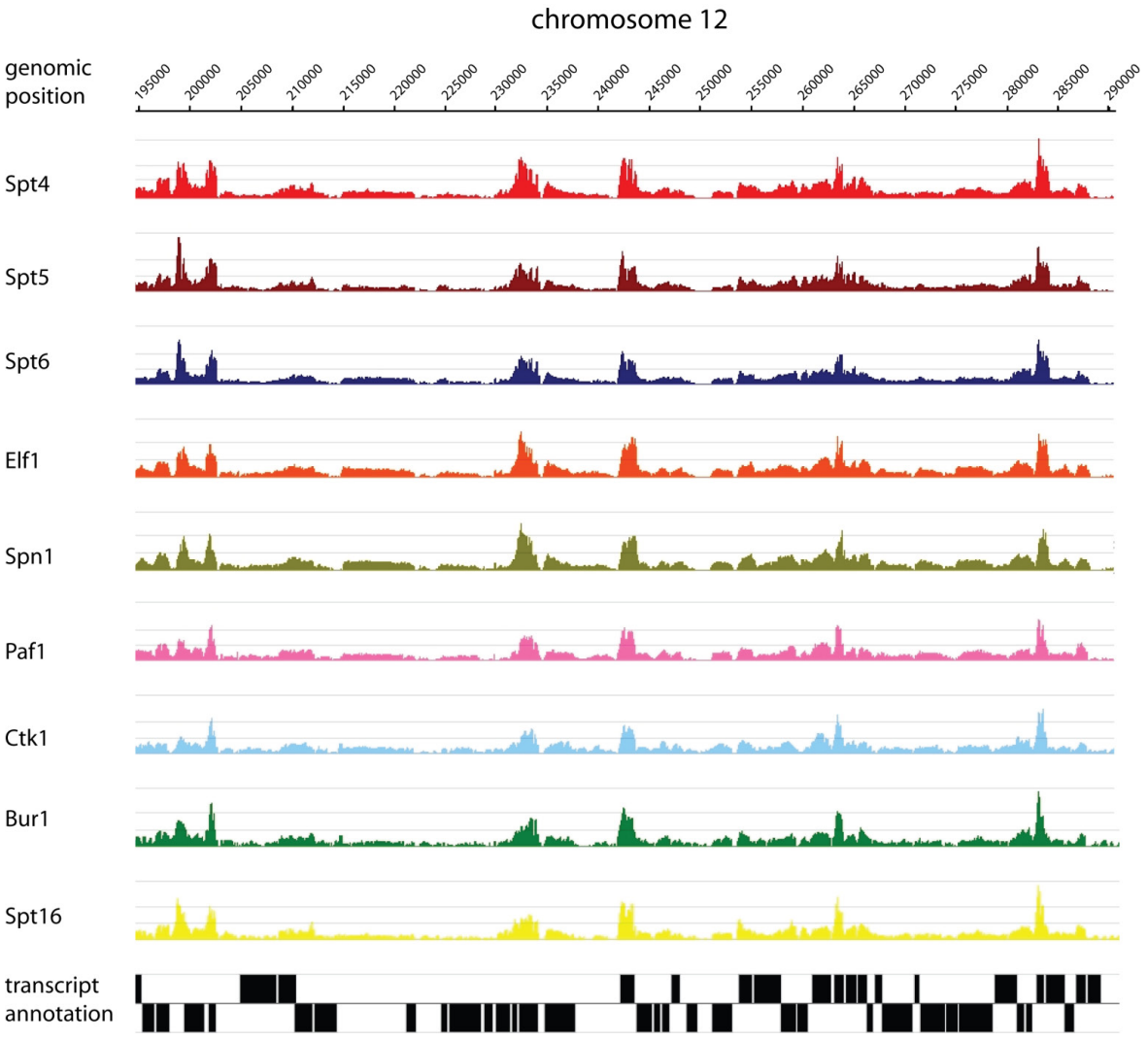

Figure 11: Elongation factor occupancy profiles along chromosome 12. Representative 100 kb genome region showing ChIP-chip occupancy profiles of elongation factors. Note the high degree of covariation among all factor occupancy profiles.

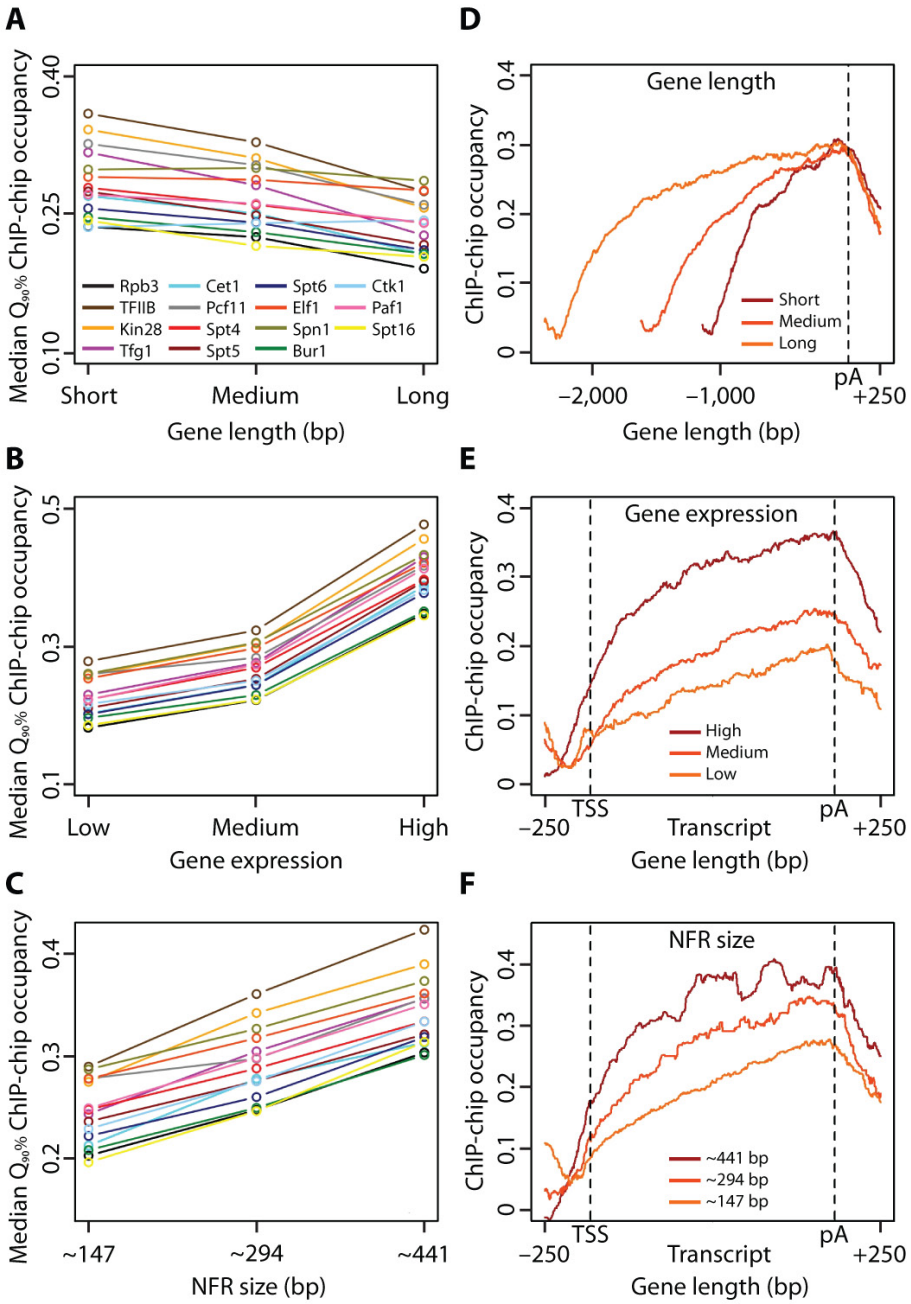

Figure 12: Transcription complex composition and transitions are independent of gene length, expression and NFR size. (A–C) Medians of the peak factor occupancies covary between different length classes (A), expression level classes (B) and nucleosome-free promoter region (NFR) size classes (C) [79, 246]. Q90%, 90% quantiles of gene occupancies, used as a proxy for peak occupancies. (D–F) Gene-averaged profiles of the representative elongation factor Elf1 have shapes and transition points that are independent of gene length (D), expression level (E) and NFR size (F). The same holds true for all other profiled factors and also for genes grouped by transcript type and functional class [146].

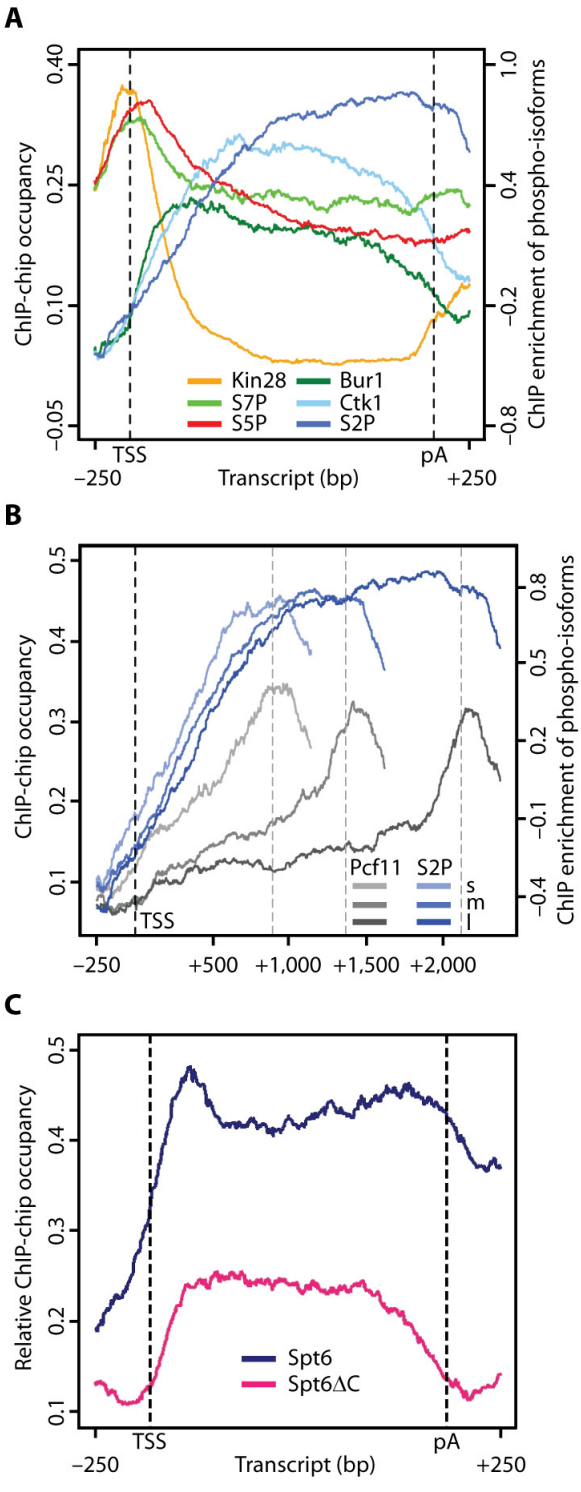

Figure 13: Pol II phosphorylation and factor occupancy. (A) Gene-averaged profiles (for length class L only) of CTD phosphorylations, CTD Ser5 kinase Kin28 and Ser2 kinases Bur1 and Ctk1. (B) Profiles for Ser2-phosphorylated Pol II and Pcf11 aligned at the TSS and pA site, respectively, and averaged for length classes short (s), medium (m) and long (l). Occupancy and signal intensity for Pcf11 and S2P are plotted on the left and right y axes, respectively. (C) Gene-averaged profiles for Spt6 and the C-terminal deletion variant Spt6ΔC (lacking the 202 C-terminal residues) [46]. As 100% occupancy levels are not expected for Spt6ΔC, the y axis shows ChIP enrichments obtained by normalization with input measurements as well as mock IPs (Section 5.7.1) without scaling to 100% occupancy.

decreased 100–200 bp downstream of the pA site (Figures 6C, 13AB and Table 16). The point where full Ser2 phosphorylation was reached did not depend on pA site location, but rather on TSS location (Figure 13B). Although we cannot rule out changes in ChIP efficiency due to the accessibility of the phosphorylated CTD to antibodies, our results overall indicate that CTD phosphorylation patterns previously observed on individual genes [114, 199] occur globally and depend on TSS location.

6.8 Recruitment of CTD kinases explains CTD phosphorylations

The Ser7 and Ser5 peaks just downstream of the TSS are consistent with the presence of the Ser7 and Ser5 kinase Kin28 [4, 110] just upstream of the TSS (Figure 13A and Table 16). Furthermore, the early peak of Ser7 phosphorylation is consistent with dependence of Ser7 phosphorylation on the coregulatory Mediator complex [15], which binds at promoters. The subsequent increase in Ser2 phosphorylation is consistent with the Ser2 kinases Bur1 and Ctk1 [28, 182] entering during the 5´ transition and remaining present in transcribed regions (Figure 13A). Thus, specific CTD phosphorylations can be explained by the recruitment of corresponding specific CTD kinases at defined distances from the TSS. Unfortunately, we were not able to obtain satisfactory profiles of CTD phosphatases to compare their recruitment with the observed decreases in CTD phosphorylation.

6.9 CTD phosphorylation and factor recruitment

To clarify the relationships between CTD phosphorylations and factor occupancies, we subjected all profiles to SVD and correlated residual profiles lacking the contributions of the first SVD term (Section 5.7.6); the eliminated first term described 85.6% of the covariation of factor occupancies. Ser7 and Ser5 phosphorylation correlated with the occupancy of the capping enzyme Cet1 (Figure 8D), as expected from binding of capping enzyme to Ser5-phosphorylated CTD *in vitro* [190]. As Ser5 phosphorylation levels peaked more than 100 bp downstream from the Cet1 peak, the capping enzyme may already be bound when the first Ser5 residues are phosphorylated. Cet1 occupancy dropped very sharply further downstream, whereas Ser5 phosphorylation levels remained high, suggesting an active mechanism to release the capping enzyme from the CTD. Ser2 phosphorylation correlated strongly with Spn1 and Elf1 occupancy (Figure 8D), suggesting these factors are stabilized within the elongation complex by Ser2 phosphorylation.

6.10 Possible CTD masking and CTD-independent recruitment

CTD Ser2 phosphorylation did not correlate with occupancy of Pcf11 and Spt6, although these factors bind to the Ser2-phosphorylated CTD *in vitro* [6, 245]. Pcf11 was recruited mainly at the pA site (Figure 13B), consistent with the known role of Pcf11 in RNA 3´ processing. This may be explained if the Ser2-phosphorylated CTD becomes accessible to Pcf11 only after pA site passage and 3´ RNA cleavage. Alternatively, Pcf11 cross-linking may be increased by cooperative interactions of factors and RNA around the pA site or by conformational changes in the elongation complex. Spt6 entered early, during the 5´ transition, suggesting a recruitment mechanism independent of CTD Ser2 phosphorylation. To investigate this, we determined the ChIP-chip profile of a variant of Spt6 lacking its CTD-binding C-terminal domain (Spt6ΔC), using a yeast strain that expresses only a truncated Spt6 lacking the last 202 residues [46]. Deletion of the Spt6 CTD-binding domain led to much less recruitment of Spt6 but did not abolish its entry during the 5´ transition (Figure 13C). Thus, Spt6 is apparently recruited in a CTD-independent manner during the 5´ transition, but full recruitment requires the CTD-binding domain. The CTD-binding domain was required for retaining Spt6 until the pA site was reached (Figure 13C), consistent with Spt6's preference for binding the Ser2-phosphorylated CTD. These results indicate that binding of a factor to the phosphorylated CTD

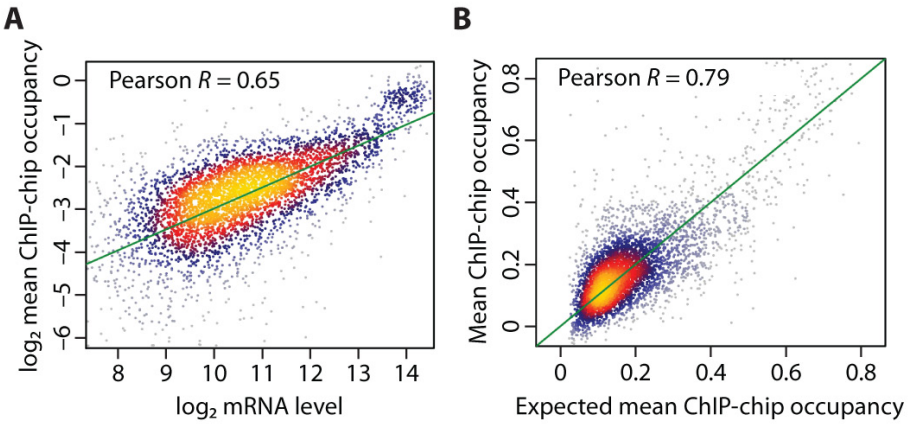

Figure 14: Elongation complex occupancy predicts mRNA expression. (A) The logarithm of the average elongation factor and Rpb3 transcript occupancy is highly correlated with the logarithm of mRNA levels. For constant mRNA synthesis rates (elongation complex speeds) v and decay rates r, we expect a linear relationship between elongation factor occupancy o and mRNA levels c, because at equilibrium the rates of mRNA synthesis and decay are equal, and thus $ov = cr$. This would result in a linear dependence between log$o$ and log$c$ with slope 1:log$o$ = log$c$+log(r/v). The actual slope of 0.49 (green) implies that the ratio of Pol II speed to decay rate increases slightly with increasing mRNA level ($v/r \propto c^{0.51}$). For ribosomal protein genes, v/r is about three-fold higher than average. (B) The averaged transcript occupancy of Rpb3 and the nine elongation factors is highly correlated with the occupancy expected on the basis of the relationship between occupancies and observed mRNA levels in A.

in vitro cannot predict factor recruitment *in vivo*. This suggests that the CTD may be transiently masked and its accessibility regulated, and that factor recruitment includes CTD-independent and CTD phosphorylation type–independent recruitment.

6.11 No evidence for promoter-proximally stalled Pol II

In higher eukaryotes, Pol II is often stalled early during elongation near the promoter [34, 160, 249] and can be released by activators [185]. Our data do not provide evidence for the presence of such promoter-proximally stalled Pol II in growing yeast. Genes with stalled Pol II would show Ser5 but not Ser2 phosphorylation, or at least more Ser5 than Ser2 phosphorylation. However, we did not find genes with a peak for Ser5 phosphorylation and no peak for Ser2 phosphorylation (data not shown). SVD analysis of initiation and elongation factor profiles revealed a high covariance of 83.1% (Figure 8C, right, and [146]), suggesting that initiation complexes are generally efficiently converted to elongation complexes. Although Rpb3 occupancy peaks around 150 bp downstream of the TSS (Figure 6C), this does not indicate polymerase stalling, as stalling generally occurs closer to the TSS. Instead, this Pol II peak may be explained by the 5´ transition being slow because of capping, phosphorylation and assembly events, leading to an apparent accumulation of Pol II. Alternatively or additionally, the peak may reflect transient pausing of Pol II between the +1 and +2 nucleosomes, which are positioned around 40 and 210 bp downstream of the TSS, respectively [94]. Our data from exponentially growing yeast also do not show evidence for polymerase peaks upstream of the TSS, as observed in yeast during stationary growth [184].

6.12 General elongation complexes are productive

To investigate whether the general elongation complexes are active on most genes, we correlated averaged Rpb3 and elongation factor occupancies with mRNA levels [46]. The mRNA level should

be proportional to the mRNA synthesis rate of a single elongation complex times its occupancy, divided by the mRNA decay rate (see legend of Figure 14). For constant mRNA synthesis and decay rates, we would therefore expect a linear dependence of occupancy on mRNA level, corresponding to a slope of 1 in a log-log plot. We found a robust correlation of 0.65 between the log occupancy and the log mRNA levels (Figure 14A); however, the slope was only ∼0.5. A correlation of 0.71 was obtained when we used the distance-filtered gene set. This shows that increased mRNA levels are due not only to greater elongation complex occupancy, but also to a higher ratio of mRNA synthesis rates over decay rates. The same dependence leads to a high correlation of 0.79 between the observed average occupancy and the expected occupancy calculated from the mRNA level (Figure 14B). These correlations indicate that most general elongation complexes are active in producing mRNA and that occupancy of genes by the general elongation complex is a good predictor for gene expression level.

6.13 Discussion & Summary

Here we have established an improved ChIP-chip protocol for obtaining high-resolution genome-wide occupancy profiles for components of the Pol II transcription machinery (Sections 5.6 and 5.7), to investigate the transcription cycle *in vivo* (Figure 6G). We demonstrate that Pol II elongation factors, which are required for chromatin passage and RNA processing on individual genes, associate with all transcribed Pol II genes in proliferating yeast cells. Elongation factors show three distinct, nonrandom patterns of distribution over genes, and these distribution patterns are independent of gene length, type, function or nucleosome structure. The underlying general elongation complex is established and disassembled during uniform transitions at the beginnings and the ends of genes. Elongation factors enter during a sharp 5´ transition just downstream of the TSS and exit in a two-step 3´ transition around the pA site. Our genome-wide Pol II phosphorylation profiles match patterns observed at individual genes and are explained by the recruitment of specific CTD kinases at defined distances from the TSS. Analysis of CTD phosphorylation profiles does not support the existence of promoter-proximally stalled Pol II in growing yeast. CTD phosphorylation is not predictive of the recruitment of factors that bind the phosphorylated CTD *in vitro*. Instead, we obtained evidence that CTD accessibility is regulated by transient CTD masking and that recruitment mechanisms are CTD independent. Finally, occupancy of genes by the general elongation complex predicts the resulting mRNA levels, suggesting that most or all elongation complexes are active.

Published biochemical and genetic data suggest that the 5´ transition corresponds to a coordinated conversion of a general initiation complex to a general elongation complex. The conversion includes initiation factor dissociation, which liberates the Pol II clamp domain [65, 141] for binding Spt5 [82]. Spt5 could coordinate entry of group 1 factors because it binds Spt4 *in vitro* [69] and associates with Spt6 *in vivo* [130]. Group 1 factors could recruit group 2 factors, as Spt6 binds Spn1 [117, 130]. Consistent with this, group 1 factors interact genetically with Elf1 [178] and, as we show here, Spn1 and Elf1 interact within a Pol II complex. Recruitment of group 3 factors may commence with CTD Ser5 phosphorylation; this recruits Bur1 [182], which in turn phosphorylates Spt5, thereby recruiting Paf1 [120, 132, 254]. Spt16 enters around 30 bp upstream from other elongation factors (Figure 6F and Table 16), perhaps because it binds to the +1 nucleosome [212]; this is consistent with its role as a histone chaperone [10]. Initiation factors are not present when the 5´ transition is completed around 150 bp downstream of the TSS, consistent with Ctk1 promoting dissociation of initiation factors [2].

The general two-step 3´ transition we observed is consistent with ChIP data obtained at individual genes. The early exit of group 3 factors Paf1, Ctk1 and Bur1, and the continued presence of Spt4, Spt5 and Spt6, have previously been observed [103, 108]. Our results, however, show that the reported Spt16 occupancy downstream of the pA site [108] does not occur globally. Also, our observation of peak levels of the *bona fide* 3´ processing factor Pcf11 downstream of the pA site

challenges the idea of an early loading of 3´ processing factors in the 5´ region of genes [64]. We observed continued presence of group 1 and group 2 factors downstream of the pA site, consistent with their reported roles in mRNA 3´ processing [100], mRNA export [245] and re-establishment of chromatin structure after Pol II passage [101].

Our results also provide insights into the role of CTD phosphorylation during transcription complex transitions and in the coordination of transcription-coupled events. First, peak levels of Ser7 and Ser5 phosphorylation were generally observed in the 5´ regions of genes, and peak Ser2 phosphorylation in the 3´ regions of genes. Second, the 5´ transition occurs before any substantial Ser2 phosphorylation, suggesting that the assembly of the general elongation complex is independent of Ser2 phosphorylation, consistent with the observation that the Ser2 kinase Ctk1 is not required for association of elongation factors with transcribing Pol II [3]. Third, peak levels of Ser2 phosphorylation are always reached 600–1,000 bp downstream of the TSS, regardless of the position of the pA site. This argues against a role of Ser2 phosphorylation in triggering the 3´ transition, although Ser2 phosphorylation is required for cotranscriptional 3´ RNA processing [3]. Fourth, the recruitment of Pcf11 and Spt6, which both bind the Ser2-phosphorylated CTD *in vitro*, cannot be explained solely by factor binding to the Ser2-phosphorylated CTD *in vivo*. Instead, late Pcf11 entry suggests CTD masking within the transcribed region and an increase in CTD accessibility upon RNA cleavage at the pA site, allowing for Pcf11 binding. Furthermore, Spt6 enters during the 5´ transition even when it lacks its CTD-binding domain, indicating that a CTD-independent recruitment mechanism exists. The CTD-binding domain seems to be more important for retaining Spt6 until the pA site is reached than for recruiting it during the 5´ transition.

In conclusion, our data support the following view of a productive transcription cycle. The initiation complex forms ∼30–50 bp upstream of the TSS and contains unphosphorylated Pol II and initiation factors. The complex then scans for the TSS downstream, begins RNA synthesis and triggers RNA 5´ capping. Next, the complex is converted into a general elongation complex during a sharp and efficient but presumably slow 5´ transition that is completed around 150 bp downstream of the TSS, where Ser5 phosphorylation levels peak. During subsequent elongation, Ser2 phosphorylation increases until it reaches peak levels 600–1,000 bp downstream of the TSS. During a two-step 3´ transition, a group of elongation factors exits upstream of the pA site, whereas another group persists downstream, where it is joined by additional factors such as Pcf11, resulting in mRNA 3´ processing and transcription termination.

7 Cap completion and CTD kinase recruitment underlie the initiation-elongation transition of RNA polymerase II

All results presented in this Section are published in [128].

Transcription of protein-coding genes by RNA polymerase (Pol) II begins with the assembly of an initiation complex at the promoter. As the RNA grows, initiation factors dissociate from Pol II, and elongation factors are recruited [167, 168, 176] (for details see Section 1). The initiation-elongation transition begins when the nascent RNA grows to about 12 nucleotides (nt), which releases initiation factor TFIIB [169, 194]. Release of initiation factors frees the Pol II clamp domain, which binds the conserved elongation factor Spt5 [65, 141]. Here we refer to the exchange of initiation by elongation factors as initiation-elongation transition. The transition starts with the dissociation of initiation factors, and ends with completed recruitment of elongation factors.
During initiation, the C-terminal repeat domain (CTD) of Pol II gets phosphorylated at residue serine 5 (S5) and S7 [19, 86, 174]. CTD S5 phosphorylation is important for recruitment of RNA 5´-capping enzymes [114, 149, 199]. Capping starts with removal of the terminal γ-phosphate from the 5´-triphosphate end of the nascent RNA, which is catalysed by the RNA triphosphatase [35, 59, 203]. The truncated RNA 5´-end is then linked to an inverted guanylyl group by the guanylyltransferase. Finally, the cap methyltransferase methylates position N7 of the newly added terminal guanine. In *S. cerevisiae* the three catalytic activities are encoded by Cet1, Ceg1 and Abd1, respectively. In metazoans the first two steps of the capping reaction are catalyzed by a single capping enzyme consisting of a N-terminal triphosphatase and a C-terminal guanylyltransferase domain. The resulting 7-methyl-guanosine (m7G) cap protects the transcript from degradation and promotes translation initiation of the mRNA [201, 208]. The complete cap associates with the cap-binding complex (CBC) that functions in pre-mRNA splicing and mRNA export [125, 202].
We have previously obtained high-resolution genome-wide occupancy profiles for Pol II initiation and elongation factors by chromatin immunoprecipitation (ChIP) in *S. cerevisiae* (Section 6). This study showed that initiation factor occupancy peaked around 50 bp upstream of the transcription start site (TSS) (Section 6.2), consistent with recent nucleotide-resolution ChIP-exo studies [188]. Exchange of Pol II initiation factors by elongation factors mainly occurred within 150 bp downstream of the TSS (Section 6.3). Further downstream, CTD phosphorylation levels decreased at residues S5 and S7, and increased at residues S2 (Section 6.7). S2 phosphorylation is achieved by the CTD kinases Bur1 [182] and Ctk1 [28] that enable efficient RNA elongation and 3´-processing [3, 103]. The mechanisms to recruit these elongation-promoting kinases are poorly understood, although recent work showed that Bur1 recruitment is stimulated by its C-terminal region that binds the S5-phosphorylated CTD [182].
Bur1 and Ctk1 share homology with mammalian Cdk9, the kinase of the P-TEFb complex which triggers the transition into productive elongation [25, 140, 163]. For a long time it was believed that Cdk9 is the major contributor to S2 phosphorylation in higher eukaryotes, but recent evidence suggests that the fly and human Cdk12 proteins also contribute to S2 phosphorylation, and that Bur1 and Ctk1 might be the orthologs of Cdk9 and Cdk12, respectively [7]. P-TEFb can release Pol II from promoter-proximal pause sites by phosphorylating the CTD, DSIF (human Spt4-Spt5) and NELF (negative elongation factor), leading to dissociation of NELF and transformation of DSIF into a positive elongation factor [1, 244]. While most of the key players controlling the initiation-elongation transition are conserved between *S. cerevisiae* and higher eukaryotes, yeast and *C. elegans* apparently lack a NELF ortholog and promoter-proximal pausing.
Human Cdk9 physically interacts with the CBC [124], and Cdk9 from the fission yeast *S. pombe* interacts with the cap methyltransferase [68, 209]. Further, Spt5 interacts with human, *S. cerevisiae* and *S. pombe* capping enzymes [130, 173, 198, 230]. These interactions have led to a capping

checkpoint model which suggests that Pol II pauses near the TSS to allow for co-transcriptional capping before entering productive elongation [139, 168, 172, 173].
The relevance of these interactions for the recruitment of the capping machinery and early elongation factors to transcribed genes has not been studied genome-wide *in vivo*. To address this we extended our ChIP-chip occupancy profiling to capping enzymes and the CBC. We found that the first two capping enzymes, Cet1 and Ceg1, are recruited near the TSS, whereas the third enzyme, the methyltransferase Abd1, is recruited about 110 bp downstream of the TSS, and CBC is recruited further downstream. We report that capping enzyme recruitment requires the C-terminal region (CTR) of Spt5 and that Bur1 helps to release capping enzymes from Spt5 once capping is completed. Moreover we show that Abd1 and CBC are involved in the recruitment of Ctk1 and Bur1 *in vivo*, thus maintaining proper levels of S2-phosphorylated Pol II and elongation factors.
A recent study showed that gene recruitment of Bur2 and Ctk2, the cyclin partners of Bur1 and Ctk1, respectively, required the CBC, and that the kinase complexes interacted with the CBC [85]. This study also demonstrated an effect of CBC deletion on S2 phosphorylation. Our data are consistent with these findings but also provide additional insights that lead to a model for factor exchange during the initiation-elongation transition of Pol II that supports the capping checkpoint model and has implications for understanding the promoter-proximal transition in higher cells.

7.1 Involvement of the Spt5 CTR in recruitment of capping enzymes

Spt5 may help to recruit capping enzymes to transcribed genes, since it copurifies with Cet1, Ceg1, and Abd1 [130]. *In vitro*, Spt5 binds and activates the human capping enzyme [230], and its C-terminal repeat region (CTR) binds *S. pombe* triphosphatase and guanylyltransferase [173, 198]. To test whether Spt5 is involved in recruiting capping enzymes *in vivo*, we carried out ChIP analysis in *S. cerevisiae* strains expressing only Spt5 lacking the CTR [132, 147]. We detected a decrease in occupancy of Cet1, Ceg1, and Abd1 to about 65%, 50%, and 35% of wild-type levels, respectively, at the 5´ region of the *ADH1* gene, whereas Pol II occupancy was relatively unaffected (Figure 15A). Similar results were obtained for *ACT1* (Figure 16A). Total protein levels of Cet1, Ceg1, and Abd1 remained unchanged in the CTR deleted cells (Figure 16B). These results show that substitution of Spt5 with a mutant lacking the CTR reduces occupancy of capping enzymes over the *ADH1* and *ACT1* genes. Since S5 phosphorylation of the CTD by the initiation factor TFIIH kinase Kin28 (human Cdk7) is also required for capping enzyme recruitment [114, 149, 199], these results suggest that the Spt5 CTR and the Pol II CTD both contribute to ensure full recruitment of capping enzymes to the 5' region of genes *in vivo*.

7.2 The Bur1 kinase complex promotes release of capping enzymes

We demonstrated that deletion of the CTR of Spt5 reduced Abd1 occupancy at the 5' region of *ADH1* and *ACT1*, but not further downstream (Figures 15A and 16A), suggesting that Abd1 is released from Spt5 during the initiation-elongation transition. Since the elongation-promoting kinase complex Bur1-Bur2 phosphorylates the Spt5 CTR during this transition [132, 254], Bur1 may help to release capping enzymes from Spt5 after cap completion. To investigate this we carried out ChIP analysis of Cet1 and Abd1 in a strain lacking Bur2, the cyclin subunit of the Bur1 complex (Figure 15B and C). In the *bur2*$\triangle$ mutant phosphorylation of Spt5 is reduced [254]. Cet1 and Abd1 occupancies were normalized against Pol II occupancy since Bur2 deletion had a negative effect on Pol II recruitment (Figure 15B). We found that deletion of Bur2 led to a significant increase ($\sim$2 fold) of Abd1/Pol II ratios at the ORF and 3' regions of *ADH1*, *ACT1*, and *ILV5* (Figures 15B and 16C). For Cet1 we observed a similar increase in Cet1/Pol II ratios, suggesting that residual amounts of Cet1 molecules remain associated with the transcription complex until the 3' end of genes when Bur2 is deleted (Figures 15C and 16D). At the *ADH1* gene the Cet1/Pol II ratio was also increased at the 5' region (Figure 15C). We do not know the reason for this

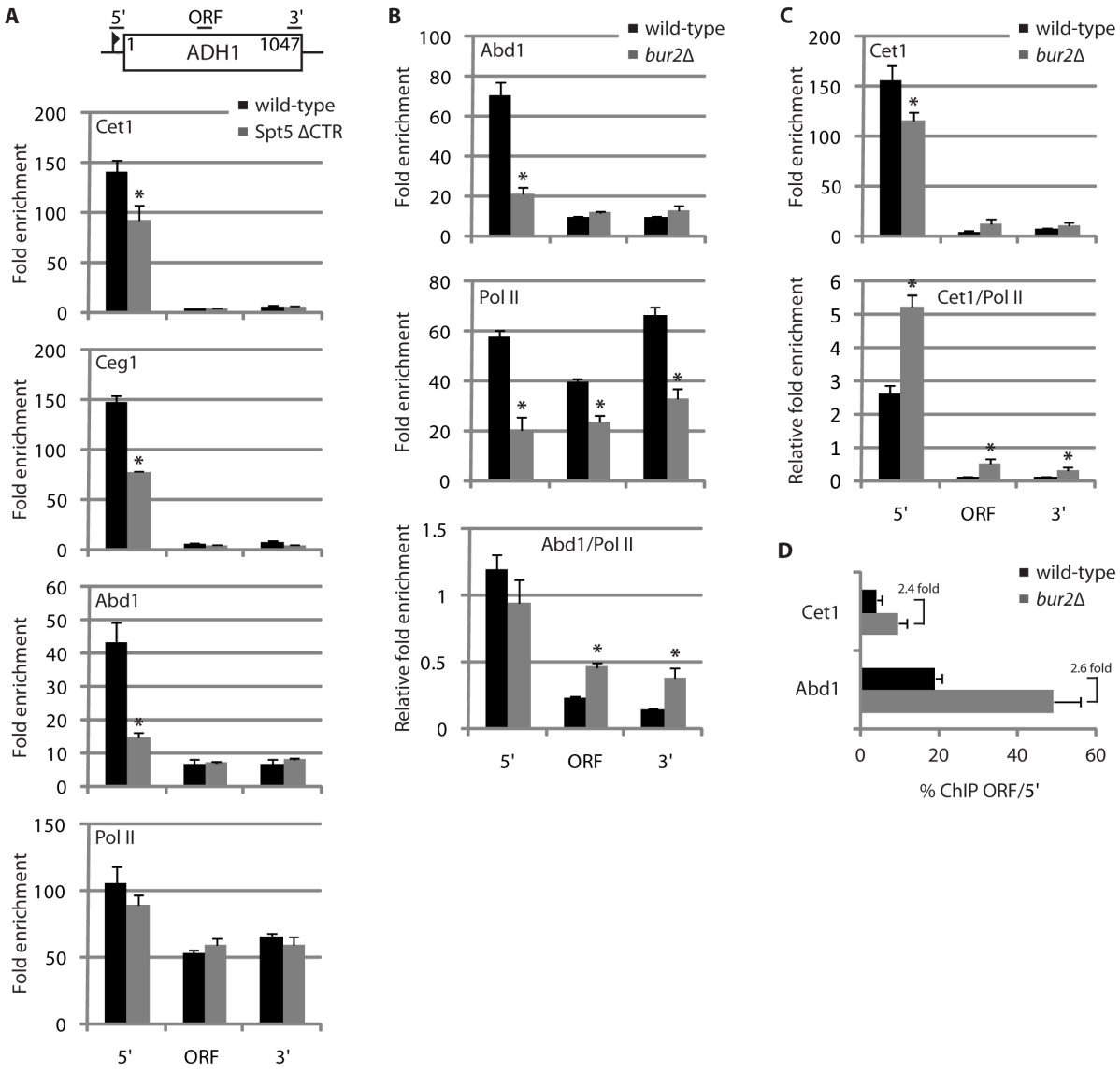

Figure 15: The Spt5 CTR is involved in recruitment of capping enzymes. ChIP-qPCR analysis was performed to monitor capping enzyme and Pol II recruitment at three different gene regions of *ADH1*: TSS (5´), coding (ORF, open reading frame) and terminator region (3´). Occupancies were calculated as fold enrichments over an ORF-free untranscribed region on chromosome V and are indicated on the y axes (see Section 5.5.4). Error bars show SD from three independent experiments of biological replicates, and the asterisk (*) indicates if the factor occupancies are significantly different (p-value < 0.05) between the wild-type and mutant condition using Student's t-test. (A) Cet1, Ceg1, Abd1 and Pol II occupancies for wild-type (black bars) compared to Spt5$\triangle$CTR (gray bars) cells are shown. (B) Abd1 and Pol II occupancies for wild-type (black bars) compared to *bur2*$\triangle$ (gray bars) cells are shown. Abd1 fold enrichments relative to Pol II (Abd1/Pol II) were calculated by dividing Abd1 occupancies by Pol II occupancies. (C) Occupancies as in B for Cet1. (D) Cet1/Pol II and Abd1/Pol II ChIP signals for the ORF region of *ADH1* relative to the 5' end.

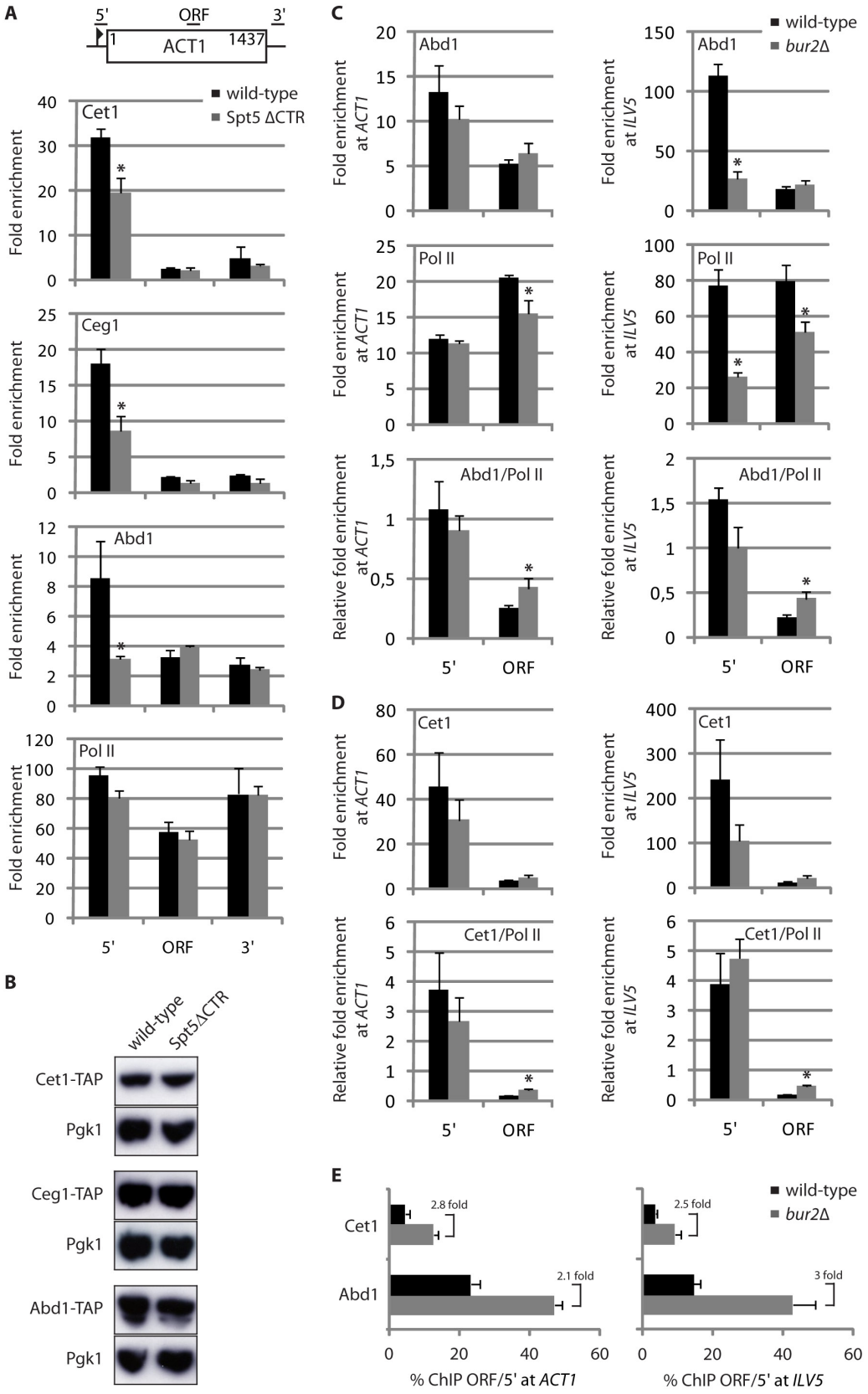

Figure 16: The Spt5 CTR is involved in recruitment of capping enzymes. (A, C-E) ChIP-qPCR analysis was performed as described in Figure 15 to monitor capping enzyme and Pol II recruitment at different gene regions *of ACT1* and *ILV5*. (B) Western blot analysis showing equal levels of Cet1-TAP, Ceg1-TAP and Abd1-TAP in wild-type and Spt5ΔCTR cells under normal growth conditions. 40 μg of total protein were loaded from each strain. Pgk1 is included as an internal loading control.

increase, but to check whether the Cet1/Pol II increase at the ORF region is an indirect effect due to higher levels at the 5' region we compared the ORF/5' ChIP ratio between wild-type and *bur2*△ experiments. In wild-type, the Cet1/Pol II ratio decreased from 100% (5') to 3.9% (ORF), while in *bur2*△ it decreased to 9.5% (ORF) (Figure 15D). Likewise decreased the Abd1/Pol II ratio from 100% (5') to 18.8% (ORF) in wild-type, and to 48.8% (ORF) in *bur2*△ cells. Similar effects were seen at other tested genes (Figure 16E). Overall these results support a model in which the Bur1-Bur2 kinase complex helps to release capping enzymes from Spt5, maybe by phosphorylating the CTR [132, 254].

7.3 Cet1-Ceg1 recruitment is restricted to the TSS

To determine the regions of capping enzyme recruitment to DNA at high resolution we collected genome-wide occupancy profiles for capping enzymes by ChIP-chip in exponentially growing *S. cerevisiae* culture. ChIP-chip and data analysis were performed as described in Sections 5.6 and 5.7, respectively. Averaging of ChIP profiles after alignment of genes at their TSS [159] revealed a sharp peak for Ceg1 (Figure 17A and B, upper panel) ∼20 bp downstream of the TSS. The peak location and shape were virtually identical to that for Cet1 (Figure 17A and B, upper panel, and Section 6.2). Both peaks were independent of gene length, gene type or expression level [128]. The highly similar profiles for Cet1 and Ceg1 ($R = 0.93$) are consistent with the formation of a stable Cet1-Ceg1 heterodimer [66], and confirm the high resolution of our occupancy profiling. Cet1-Ceg1 occupancy increases and decreases sharply, with a peak width similar to that obtained for the initiation factor TFIIB (Section 6.2) that binds a defined promoter region (Figure 17A and B, lower panel). Thus the recruitment and activity of the Cet1-Ceg1 heterodimer is apparently restricted to a narrow window near the TSS. Although ChIP data reflect only physical presence of proteins, and although cross-linking can be indirect via other proteins, these results are consistent with the model that the first two capping reactions occur on the polymerase surface when the nascent RNA 5'-end emerges from the Pol II RNA exit channel.

7.4 Abd1 recruitment occurs downstream of Cet1-Ceg1

ChIP-chip profiling of the third capping enzyme, the methyltransferase Abd1, revealed a strong peak ∼110 bp downstream of the TSS, about 90 bp further downstream from the peak for Cet1-Ceg1 (Figure 17A and B, upper panel). This separation of occupancy peaks for Cet1-Ceg1 and Abd1 is highly significant and independent of gene length, gene type or expression level (Figure 18 and [128]). A separation of occupancy peaks was also visible for the Cet1-Ceg1 dimer and the initiation factor TFIIB, which is located ∼70 bp upstream of Cet1-Ceg1 (compare Figure 17A upper and lower panel), a location that was confirmed by nucleotide-resolution ChIP-exo data [188]. The Abd1 profile is very similar to the profile for S5-phosphorylated Pol II ($R = 0.75$), and the peaks of both profiles are at the same position (Figure 17A and B, upper panel). This is consistent with Abd1 binding to the S5-phosphorylated CTD *in vitro* [149]. These results suggest that cotranscriptional pre-mRNA capping occurs in two steps, triphosphatase/guanylyltransferase action and methylation, and that completion of the cap by methylation is likely a spatially separated event that occurs around 110 bp downstream of the TSS.

7.5 Spt5 is involved in Abd1 recruitment genome-wide

We demonstrated that full Abd1 recruitment to the 5' region of *ADH1* and *ACT1* requires the Spt5 CTR (Figures 15A and 16A). To further investigate whether deletion of the Spt5 CTR has an effect on Abd1 occupancy genome-wide, we carried out ChIP-chip profiling of Abd1 in cells expressing only Spt5 lacking the CTR. We performed metagene analysis and detected a decrease in Abd1 occupancy across the 5' region of genes (Figure 17C, upper panel). Clustering analysis

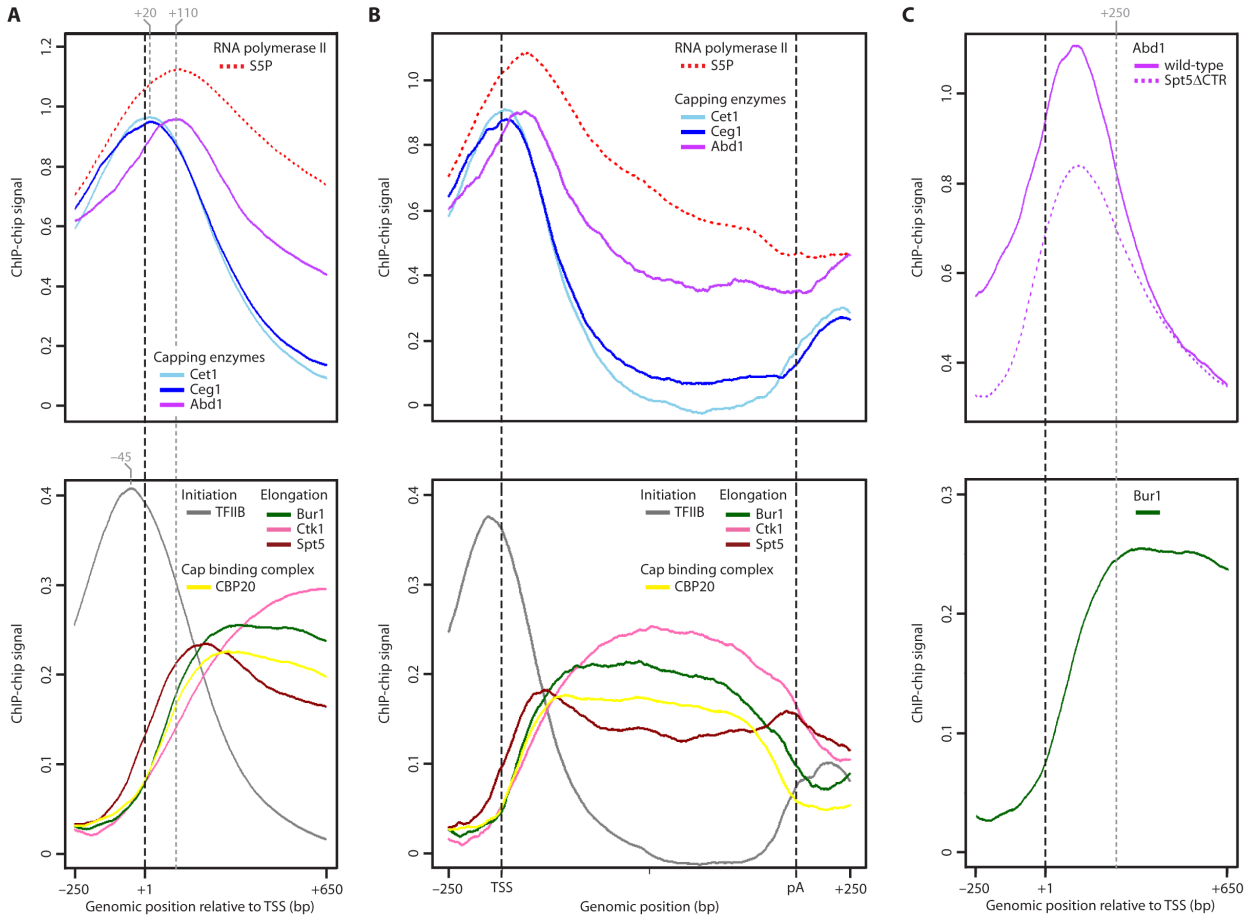

Figure 17: Genome-wide occupancy profiling of the capping machinery. (A) Gene-averaged ChIP-chip profiles for capping enzymes and Pol II phosphorylated at serine 5 (S5P) residues of the CTD (upper panel), and for initiation factor TFIIB, CBC subunit CBP20, and elongation factors Bur1, Ctk1 and Spt5 (lower panel). Profiles in the lower panel were described in Section 6, except for CBP20. The profile of the other CBC subunit, CBP80, looks similar than CBP20 but shows much lower signals [128]. Occupancy profiles taken from the quality-filtered ALL gene set (1140 genes) were cut around the TSS (from 250 bp upstream to 650 bp downstream; only genes longer than 680 bp were considered) and averaged using a 5% trimmed mean at each genomic position. Profiles for gene length classes are similar [128]. Dashed gray lines mark the peak positions of the averaged ChIP-chip profiles. For details on ChIP-chip profile calculation see Section 5.7. (B) Gene-averaged ChIP-chip profiles as in A but for medium length genes (1238 ± 300 bp, n = 339). Genes were aligned at their TSS and pA sites [159], scaled to median length, and averaged using a 5% trimmed mean at each genomic position. Profiles of other gene length classes are similar [128, 146]. (C) Gene-averaged ChIP-chip profiles as in A for Abd1 in wild-type compared to Spt5$\triangle$CTR cells (upper panel) and for Bur1 (lower panel). The dashed grey line marks the position were full recruitment of Bur1 is reached $\sim$250 bp downstream of the TSS.

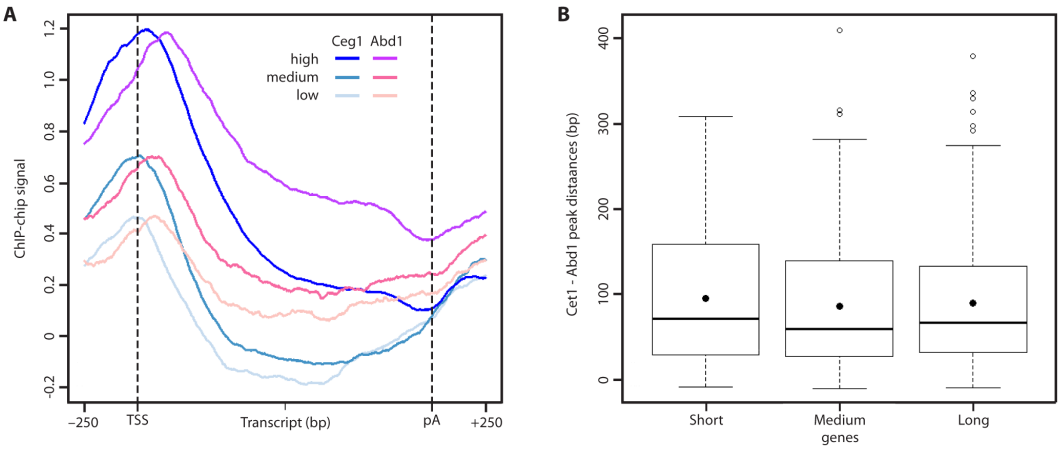

Figure 18: Capping enzyme profile peak distances are independent of expression level and gene length. (A) Gene-averaged profiles for Ceg1 and Abd1 for genes in three expression level classes. The quality-filtered set of medium length genes (Figure 17B) was partitioned into three groups: low (25%-50% quantile), medium (50%-75% quantile), and high (>75% quantile) expression levels [46]. (B) Box plots showing the gene-wise variation of peak distances between Cet1 and Abd1 for short (725 ± 213 bp, n = 266), medium (1238 ± 300 bp, n = 339), and long (2,217 ± 679 bp, n = 299) genes (Section 5.7). Mean distances are indicated by filled black circles. The distribution of peak distances between Ceg1 and Abd1 is similar (not shown).

revealed that the effect is of general nature and observed at most genes (not shown). The main differences in Abd1 occupancy between wild-type and Spt5△CTR backgrounds were restricted to a region from upstream of the TSS to ∼250 bp downstream (Figure 17C, upper panel), consistent with our ChIP data at *ADH1* and *ACT1* (Figures 15A and 16A). Bur1 occupancy is low around the TSS and increases further downstream with peak occupancy also reached ∼250 bp downstream of the TSS (Figure 17C, lower panel). This is consistent with the model that the Bur1-Bur2 kinase complex promotes release of capping enzymes from Spt5 (Figures 15B – D and 16C – E). Taken together our ChIP-chip results indicate that the requirement of Spt5 for full Abd1 recruitment occurs globally and is restricted to the 5' region of genes where Bur1 occupancy is low.

7.6 Recruitment of the cap-binding complex

The above data suggested that the cap structure can generally only be completed around 110 bp downstream of the TSS, where peak occupancies of Abd1 were observed. If true, then recruitment of the cap-binding complex (CBC), which binds the complete cap and interacts directly with the N7 methyl group *in vitro* [21, 148], should occur only from this point on downstream. Indeed, ChIP-chip profiling revealed that the two CBC subunits CBP20 and CBP80 are recruited around 110 bp downstream of the TSS and show full occupancy from 200-300 bp downstream onwards (Figure 17A and B, lower panel, and [128]). CBP20 showed higher signals than CBP80, probably since it contains the cap-binding site. CBC occupancy drops sharply just upstream of the pA site. This is consistent with release of the CBC-containing mRNA from DNA after 3´ RNA cleavage that occurs upon polymerase passage over the pA site. These results indicate that direct binding to the complete m7G cap dominates CBC recruitment to active genes *in vivo*.

7.7 Involvement of CBC in recruitment of kinases Bur1 and Ctk1

The CBC subunit occupancy profiles resembled those of the CTD kinases Bur1 and Ctk1, which are also released before the pA site (Figure 17A and B, lower panel, and Section 6.5). Since binding of these kinases is generally completed 200-300 bp downstream of the TSS, we wondered whether cap completion and CBC binding trigger Bur1 and Ctk1 recruitment *in vivo*. We carried out ChIP analysis of Bur1 and Ctk1 in knock-out strains lacking CBP20 (Figure 19), in which the CBC-cap complex is disrupted [239]. In *cbp*20△ strains, Bur1 occupancy was decreased up to 50% of wild-type levels throughout the *ADH1* and *PMA1* genes (Figure 19). The reduction in kinase occupancy was not due to a difference in Pol II occupancy (Figure 19), consistent with the finding that the CBC does not affect Pol II processivity [239]. In contrast, CBP20 deletion did not affect Ctk1 recruitment to the 5' region of *ADH1* and *PMA1*, but reduced Ctk1 occupancy in the central and the 3' region of the genes (Figure 19). Taken together, these results demonstrate that the CBC is required for normal recruitment of the CTD kinase Bur1 to active genes and for maintaining high occupancy of Ctk1 further downstream.

7.8 Role of CBC in CTD S2 phosphorylation and elongation factor recruitment

Since Bur1 and Ctk1 phosphorylate S2 residues of the Pol II CTD [28, 182], we suspected that CBP20 deletion would affect levels of S2-phosphorylated Pol II and elongation factor recruitment. To test this we performed ChIP experiments using specific antibodies that recognize the S2-phosphorylated CTD. As expected, deletion of CBP20 led to a two-fold decrease of S2 phosphorylation levels in the central and 3' region of the *ADH1* and *PMA1* genes (Figure 19). Also, CBP20 deletion affected downstream occupancy with the Pol II elongation factors Elf1, Spn1, and Spt6 (Figure 19 and [128]), the latter of which binds the S2-phosphorylated CTD *in vitro* [213]. Moreover, genome-wide occupancy of Elf1 and Spn1 correlates with S2 phosphorylation levels (Section 6.9). Elongation factor Spt4 was reduced at the 3' region of *ADH1* but not *PMA1*, and elongation factors Spt5 and Spt16 were unaffected [128]. Thus proper Bur1 and Ctk1 recruitment by CBC is crucial for maintaining high levels of S2-phosphorylated Pol II and elongation factors.

7.9 Abd1 contributes to Bur1 and Ctk1 recruitment

To further investigate a role for capping in recruitment of the CTD kinases Bur1 and Ctk1 *in vivo*, we conditionally depleted the essential Abd1 methyltransferase from the nucleus using the anchor-away (AA) method [76]. Abd1 was tagged with the FKBP12-rapamycin binding (FRB) domain and depleted from the nucleus upon rapamycin treatment by its anchoring to FKBP12 that was fused to the ribosomal protein RPL13A. Strains expressing Abd1-FRB from the endogenous *ABD1* promoter grew normally, but rapamycin addition induced a growth defect (Figure 21A). To confirm nuclear depletion of Abd1, we tagged the Abd1-FRB fusion protein with GFP and monitored fluorescence upon rapamycin treatment. Abd1-FRB-GFP was exclusively located in nuclei, and rapamycin treatment led to cytoplasmic fluorescence (Figure 21B), as expected.
Nuclear depletion of Abd1 decreased Bur1 occupancy at the *ADH1* and *PMA1* genes (Figure 20), as observed upon CBP20 deletion (Figure 7.7). This suggested that a lower level of cap completion could have led to reduced CBC binding and thus reduced Bur1 recruitment. Consistent with this, CBP20 occupancy was reduced to about 40% throughout *PMA1* and in the central and 3' region of *ADH1* upon nuclear depletion of Abd1 (Figure 20). Moreover, Abd1 depletion reduced Ctk1 recruitment to the central and 3' regions of *ADH1* and *PMA1* by ∼50% (Figure 20). In addition, Ctk1 occupancy was also decreased at the 5' region of *PMA1* upon Abd1 depletion, but was unaffected by CBP20 deletion (compare Figures 20B and 19B). Changes in ChIP signals did neither originate from differences in Pol II occupancy (Figure 20) nor from rapamycin treatment (Figure 22). Total protein levels of Bur1 and Ctk1 remained unchanged upon Abd1 depletion [128]. Taken

together, the cap methyltransferase Abd1 is important for recruitment of CBC and maintaining normal levels of the CTD kinases Bur1 and Ctk1.

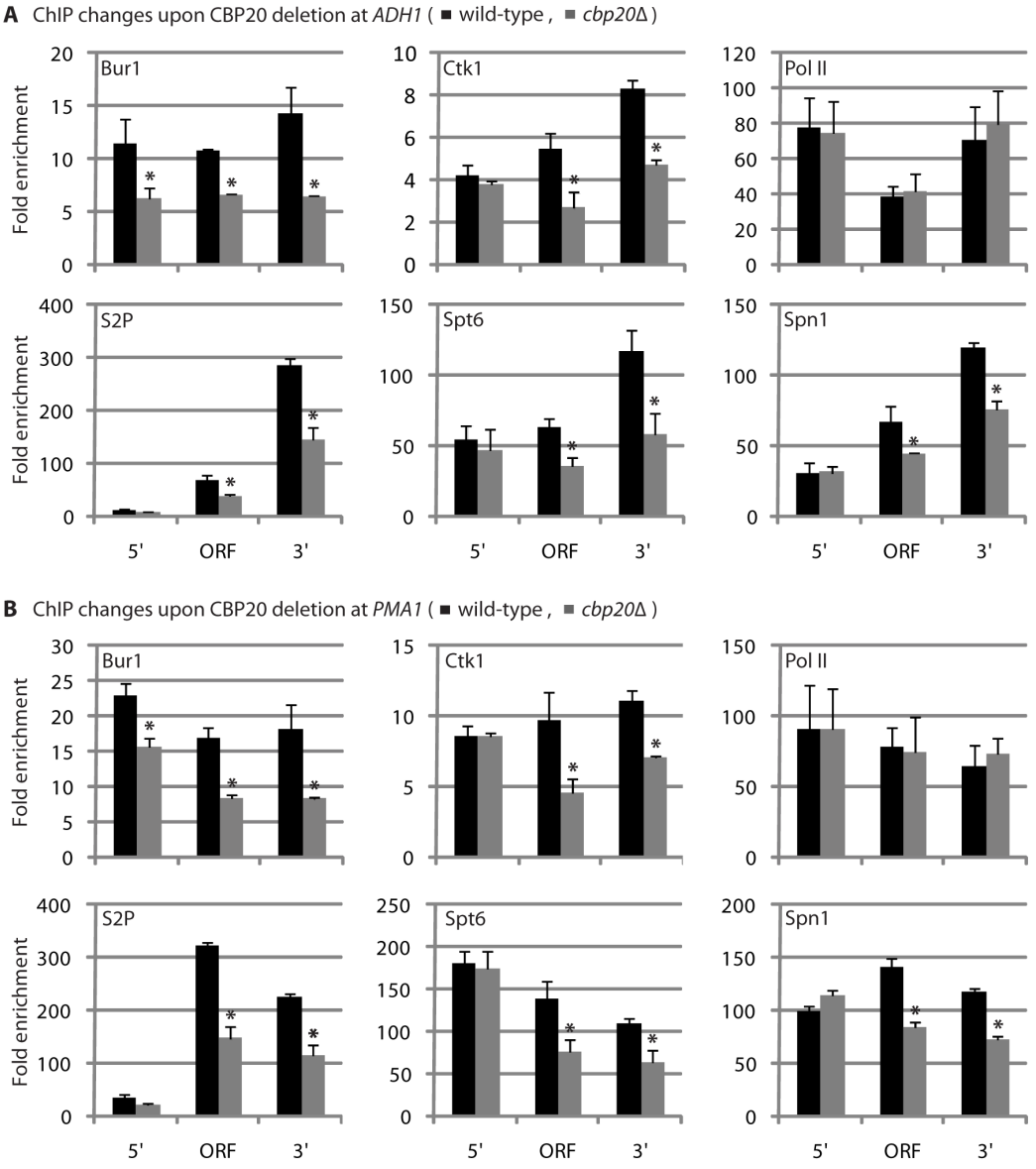

Figure 19: CBP20 is involved in recruitment of the kinases Bur1 and Ctk1. ChIP-qPCR analysis was performed to monitor changes in elongation factor and Pol II recruitment upon CBP20 deletion. Changes in levels of Ser2 phosphorylated Pol II CTD (S2P) were also analyzed. Three gene regions of (A) *ADH1* and (B) *PMA1* were investigated. Occupancies for wild-type (black bars) compared to *cbp20*$\triangle$ (gray bars) cells are shown. Occupancies were calculated as fold enrichments over an ORF-free untranscribed region on chromosome V and are indicated on the y axes (see Section 5.5.4). Error bars show SD from at least three independent experiments of biological replicates, and the asterisk (*) indicates if the factor occupancies are significantly different (p-value < 0.05) between the wild-type and mutant condition using Student's t-test.

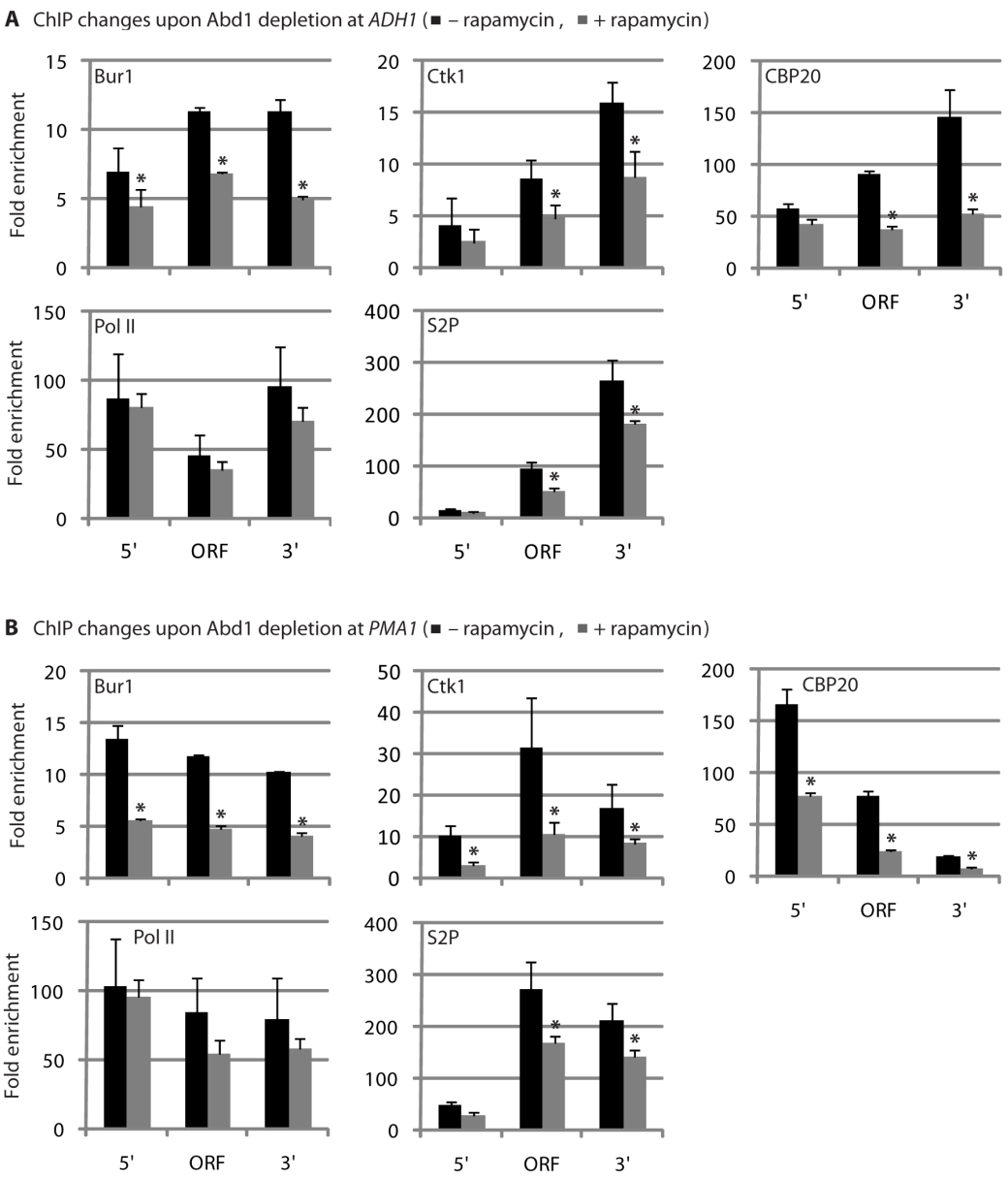

Figure 20: Abd1 contributes to recruitment of the CBC, Bur1 and Ctk1. ChIP-qPCR analysis was performed to monitor changes in Bur1, Ctk1, CBP20 and Pol II recruitment upon Abd1 nuclear depletion using the anchor-away technique. Changes in levels of Ser2 phosphorylated Pol II CTD (S2P) were also analyzed. Three gene regions of (A) *ADH1* and (B) *PMA1* were investigated. Occupancies in the Abd1-anchor-away strain that was not (black bars) or was (gray bars) treated with rapamycin for 60 min are indicated. Occupancies were calculated as fold enrichments over an ORF-free untranscribed region on chromosome V and are indicated on the y axes (see Section 5.5.4). Error bars show SD from at least three independent experiments of biological replicates, and the asterisk (*) indicates if the factor occupancies are significantly different (p-value < 0.05) between the treated and untreated condition using paired samples Student's t-test.

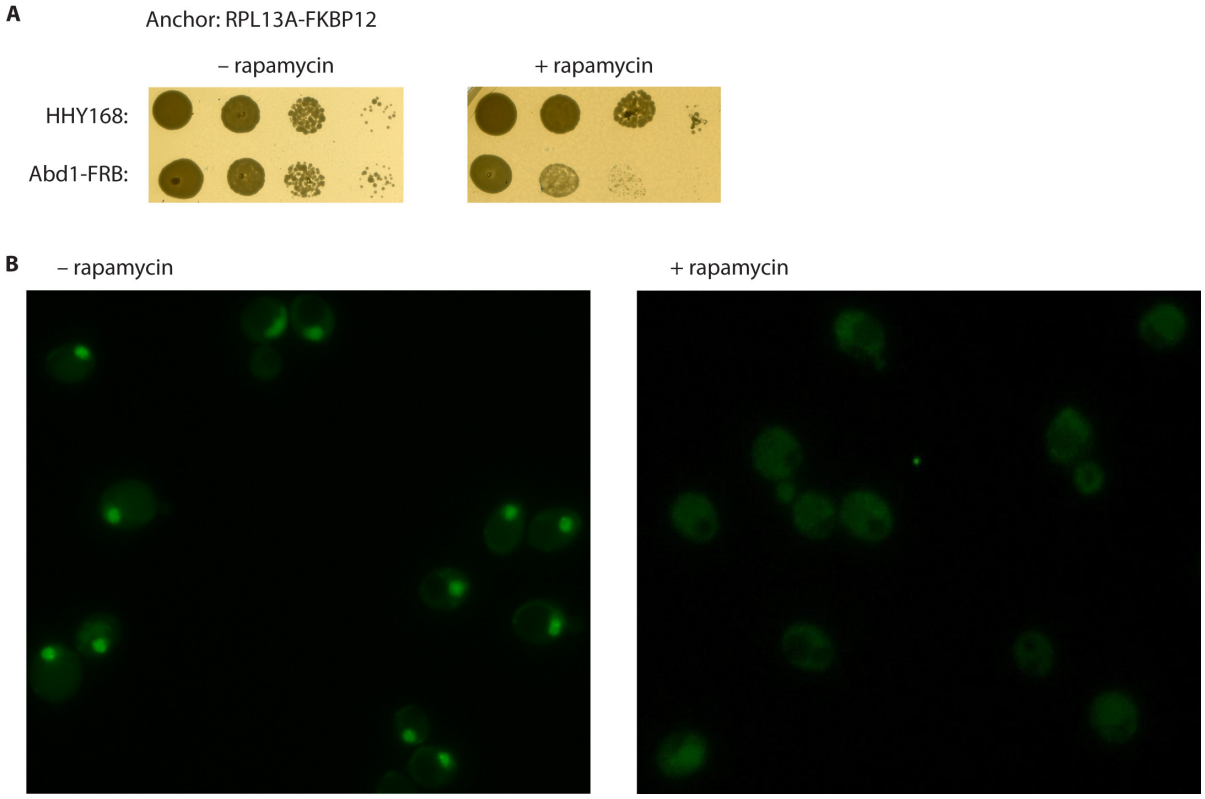

Figure 21: Rapamycin treatment leads to nuclear depletion of Abd1 and causes growth delay. (A) Spot dilution assay on YPD plates in the absence or presence of 1µg/ml rapamycin. The parental strain HHY168 was used as negative control. Dilutions were 10-fold. (B) GFP fluorescence of Abd1-FRB-GFP cells incubated without or with rapamycin (1µg/ml in DMSO) for 60 min. Addition of rapamycin causes depletion of Abd1-FRB-GFP from the nucleus.

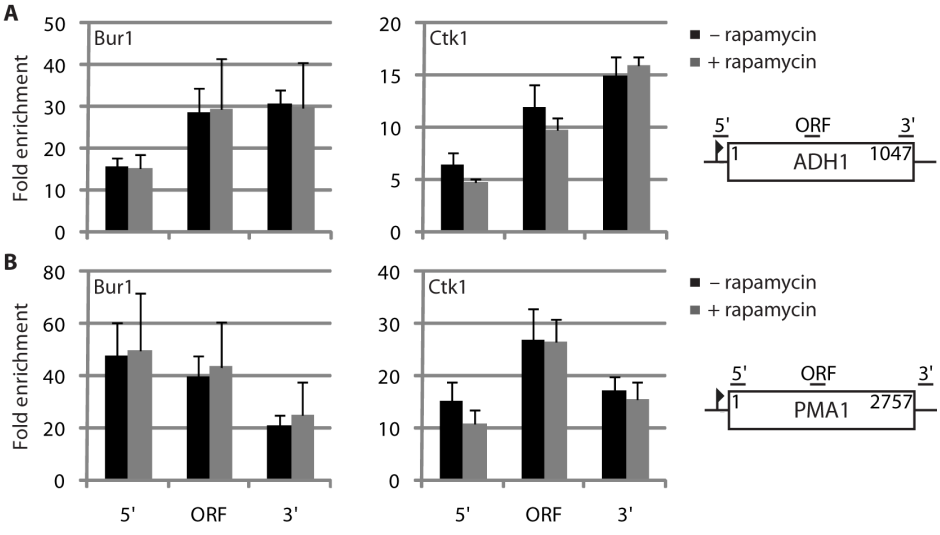

Figure 22: Rapamycin treatment does not affect Bur1 or Ctk1 occupancy at *ADH1* (A) or *PMA1* (B) using control strains expressing only untagged Abd1. ChIP-qPCR analysis was performed to monitor Bur1 and Ctk1 recruitment at the indicated gene regions. Occupancies in the anchor-away parental strain that was not (black bars) or was (gray bars) treated with rapamycin for 60 min are shown. Fold enrichments over an ORF-free untranscribed region on chromosome V are indicated on the y axes (see Section 5.5.4). Error bars show SD from three independent experiments of biological replicates.

7.10 Discussion & Summary

We previously observed that genome-wide ChIP occupancy profiles for yeast transcription initiation and elongation factors are not overlapping, consistent with exchange of factors during a general initiation-elongation transition (Section 6). Here we show that this transition and factor exchange involves stepwise recruitment of proteins that form and recognize the RNA 5´-cap structure. We first analyzed the occupancy of cap-forming and cap-binding proteins on active Pol II genes, and then investigated how formation of a complete cap triggers the recruitment of elongation-promoting kinases Bur1 and Ctk1.

In the first part of this work, we found that the RNA triphosphatase Cet1 and the guanylyltransferase Ceg1 have virtually identical occupancy profiles with a sharp peak ∼20 bp downstream of the TSS. This is consistent with the existence of a stable Cet1-Ceg1 heterodimer in fungi [66] and the presence of both activities within a single enzyme in metazoa [83]. Occupancy for the cap methyltransferase Abd1 was clearly distinct and peaked ∼110 bp downstream of the TSS. This is consistent with earlier studies, also showing a significant lag from guanylylation to methylation using a functionally coupled human *in vitro* transcription/capping system [27, 157]. Thus, whereas Cet1 and Ceg1 are obviously recruited in a single step and are present only near the TSS, Abd1 is recruited further downstream, and can remain until the 3´-end of genes, consistent with ChIP data at individual genes [199]. Abd1 recruitment apparently leads to cap completion immediately downstream, because the cap-binding complex reaches full occupancy levels about 200-300 bp downstream of the TSS, supporting the hypothesis that the methylated cap recruits CBC to nascent mRNA in yeast [239]. We further demonstrated that full capping enzyme recruitment requires the CTR of Spt5. Since our genome-wide ChIP-chip data revealed that the main peak for Spt5 occupancy lies downstream of the capping enzyme peaks, we suggest that the capping factors are loaded onto the Pol II machinery first via binding to the S5-phosphorylated CTD [114, 149, 199] and that this association is subsequently stabilized by Spt5.

In the second part of our work we found that cap completion is coupled to the recruitment of the kinases Bur1 and Ctk1. These yeast enzymes share homology with mammalian Cdk9, the kinase subunit of the P-TEFb complex that triggers the transition into productive transcription elongation [25, 140, 163]. Normal Bur1 and Ctk1 recruitment to active genes, and thus proper S2 phosphorylation levels of Pol II, required Abd1 and CBC. The CBC and its effect on CTD S2 phosphorylation were also important for maintaining high levels of elongation factors. Recruitment of Bur1 additionally depends on binding the S5-phosphorylated CTD [182], suggesting that CBC and the CTD cooperate for Bur1 recruitment. Since Bur1 also phosphorylates the Spt5 CTR [132], and since we show here that the Spt5 CTR contributes to capping enzyme recruitment, Bur1 may help to release capping enzymes from Spt5 after cap completion. This is consistent with our finding that Abd1 occupancy downstream of Bur1 recruitment was unaffected by CTR deletion. In addition, we demonstrated that Bur2 deletion leads to increased Cet1 and Abd1 occupancies along the gene bodies of *ADH1*, *ACT1*, and *ILV5*, further supporting our model.

Our findings agree with published results that showed physical interactions between the capping machinery and early elongation factors [68, 124, 130, 173, 198, 230]. In addition, a recent study showed that gene recruitment of Bur2 and Ctk2, the cyclin partners of Bur1 and Ctk1, respectively, required the CBC [85]. Furthermore this study demonstrated an effect of CBC deletion on S2 phosphorylation. Our data are consistent with these findings but also provide additional insights that lead to a more complete model for factor exchange during the initiation-elongation transition of Pol II. All available results support a previously proposed capping-dependent checkpoint during early elongation [139, 168, 173, 172].

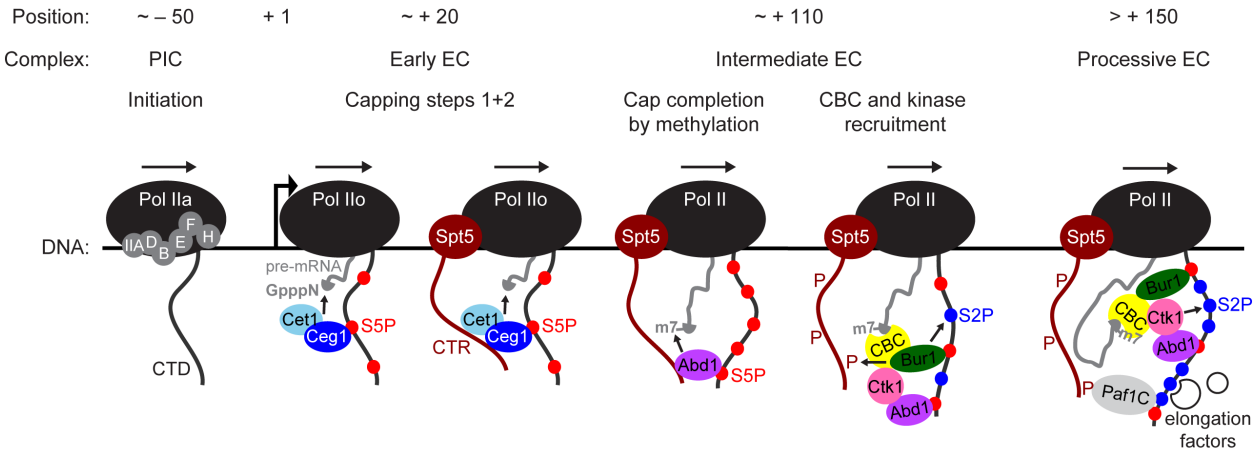

Figure 23: Model for the Pol II initiation-elongation transition. Pol II and its CTD (black) transcribe DNA (horizontal black line) from left to right to produce capped mRNA (gray). PIC, preinitiation complex; EC, elongation complex; IIA/B/D/E/F/H, initiation factors; S2/5P, phosphorylated serine 2 and 5 residues; m7, N7 methyl group of the cap; Paf1C, Paf1 complex. For details see text.

From these studies emerges the following model for the initiation-elongation transition (Figure 23). During initiation, the growing RNA transcript triggers initiation factor dissociation, which liberates the Pol II clamp domain for binding Spt5 [65, 141]. The S5-phosphorylated CTD stimulates recruitment of capping enzymes to Pol II [114, 149, 199] and this association is subsequently stabilized by the CTR of Spt5. When the nascent RNA appears on the Pol II surface, it receives an inverse GMP moiety by the action of the Cet1-Ceg1 heterodimer. Rapid dissociation of Cet1-Ceg1 gives way to the cap methyltransferase Abd1, which is recruited downstream and can remain with partially S5-phosphorylated elongating Pol II. Abd1 and the S5-phosphorylated CTD [182] initially recruit Ctk1 and Bur1, respectively. The cap is completed by methylation and binds CBC, which ensures high occupancy for both Ctk1 and Bur1 when levels of Abd1 and S5 phosphorylation drop downstream. Nascent RNAs with unmethylated cap structures are removed by the Rai1-Rat1 decay pathway [95]. Bur1 phosphorylates promoter-proximal CTD S2 residues and the Spt5 CTR, promoting release of capping enzymes. This facilitates recruitment of the Paf complex [132, 181, 182], which establishes histone modifications linked to active transcription [90]. Ctk1 phosphorylates S2 residues of the CTD further downstream [28, 182] and stimulates Pol II association with elongation, chromatin-modifying and RNA-processing factors [112, 150, 213], resulting in a productive elongation complex. Because recent evidence indicates that Bur1 is the P-TEFb (Cdk9) homologue [7], the same events likely occur during the initiation-elongation transition in higher cells.

8 CTD tyrosine dephosphorylation underlies the elongation-termination transition of RNA polymerase II

This Section presents ChIP-chip data that I analyzed during several collaborations. All results presented in Sections 8.1 to 8.3 are published in [145] and [147]. All results presented in Section 8.4 are in the process of publication. For detailed author contributions see page 9.

During the transcription cycle, CTD phosphorylation patterns coordinate the recruitment of transcription and mRNA processing factors to Pol II [19, 23, 33] (for details see Section 1.2.1). During early transcription, Ser5 phosphorylation recruits the mRNA capping enzymes [114, 199]. Ser2 phosphorylation occurs during elongation and functions in recruitment of 3´ RNA processing and termination factors [3]. Phosphorylation at Ser7 [52] and Thr4 [87] has roles in processing of specific RNAs. Tyr1 phosphorylation was described for human Pol II almost two decades ago [8], but whether this has a functional role and whether it exists in other species is unknown.
Two major transitions occur during the Pol II transcription cycle, the initiation-elongation transition at the 5´ ends of genes and the elongation-termination transition at the 3´ ends of genes (Section 6). These transitions are coupled to 5´ RNA capping and 3´ RNA processing, respectively (for details see Section 1.2). Whereas the first transition has been extensively studied by us (Section 7) and others [2, 57, 114, 121, 136, 176], less is known about the second transition. Studies of the second transition revealed a role of the Ser2-phosphorylated Pol II CTD in the recruitment of 3´-end processing and termination factors [3, 127, 136, 150, 251]. However, our genome-wide ChIP-chip data of Ser2-phosphorylated CTD (Section 6.10; Figure 13B) argues against a role of Ser2 phosphorylation in triggering the 3´ transition. On the one hand, peak levels of Ser2 phosphorylation are always reached 600–1,000 bp downstream of the TSS, regardless of the position of the pA site. And on the other hand, recruitment of Pcf11, a 3´-end processing and termination factor that binds the Ser2-phosphorylated CTD *in vitro*, is restricted to the pA site. Thus, late Pcf11 entry suggests CTD masking within the transcribed region.
Here we extended our ChIP-chip profiling using a novel monoclonal antibody against the Tyr1-phosphorylated CTD (3D12, Table 10). Moreover, we determined genome-wide occupancy profiles for termination factors Nrd1, Rtt103, and Pcf11, which contain a CTD-interacting domain (CID). We show that Tyr1 phosphorylation levels rise downstream of the TSS and decrease upstream of the pA site, coinciding with recruitment of transcription termination factors. This indicated that Tyr1 phosphorylation might impair CTD binding of termination factors before the pA site and suggested a role for Tyr1 dephosphorylation in triggering the 3´ transition. We further found that Tyr1 phosphorylation indeed impairs recruitment of termination factors Nrd1, Pcf11, and Rtt103 *in vitro*. Moreover, we show that the Glc7 phosphatase subunit of the cleavage and polyadenylation factor (CPF) is required for Tyr1 dephosphorylation at the pA site and Pol II termination *in vivo*. Taken together these results show that CTD modifications trigger and block factor recruitment and lead to an extended CTD code that explains transcription cycle coordination on the basis of differential phosphorylation of Tyr1, Ser2, and Ser5.

8.1 Genome-associated Pol II is phosphorylated at Tyr1

To investigate whether genome-associated Pol II is phosphorylated at Tyr1, we used high-resolution chromatin immunoprecipitation (ChIP) profiling in proliferating yeast (Sections 5.6 and 5.7). Data from two biological replicates ($R = 0.94$) were averaged and revealed strong signals over protein-coding and snoRNA genes. To test whether Tyr1 phosphorylation occurs on all transcribed protein-coding genes, we measured covariation in ChIP data for other CTD phosphorylations by singular value decomposition (Section 5.7.6). The first singular vector explained 83.8% of the variance (Figure 24), indicating a similar occurrence of phosphorylations at Tyr1, Ser2, and Ser5.

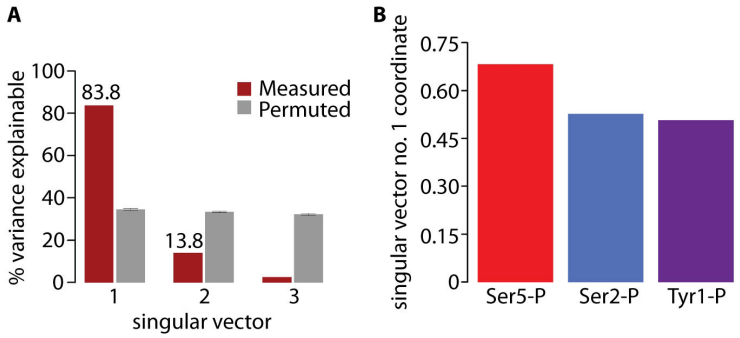

Figure 24: CTD Tyr1 phosphorylation, like Ser2 and Ser5 phosphorylation, is a general type of CTD modification. (A) Singular value decomposition (SVD) analysis (Section 5.7.6) of Tyr1-P, Ser2-P, and Ser5-P genome-wide levels. The contributions of the first three singular vectors to the variance (red) are shown in comparison to a control with randomly permuted matrix elements (gray). SVD reveals that 83.8% of the variance is explained by covariation. (B) Coefficients of first left singular vector for the SVD of the occupancy matrix for the three Pol II phospho-isoforms. The first eigenvector with its similar-sized coefficients describes the co-variation of all three Pol II phospho-isoforms.

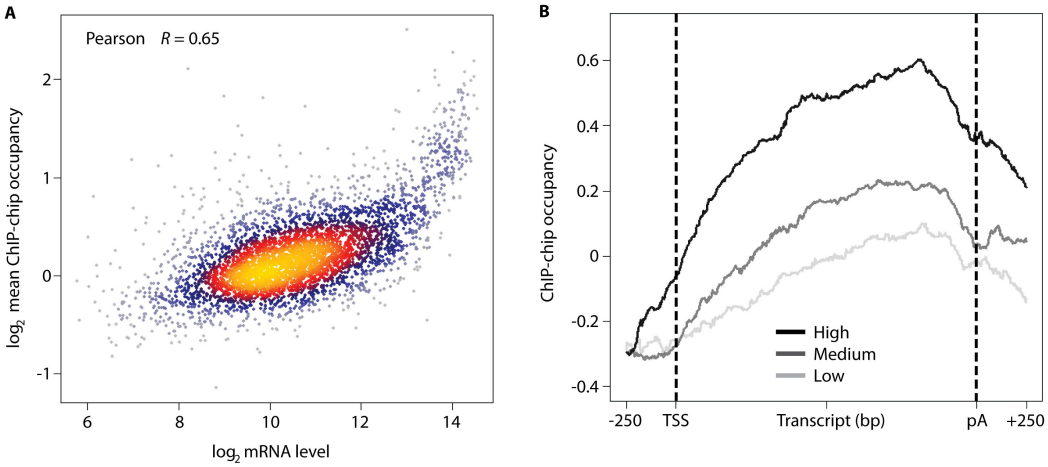

Figure 25: CTD Tyr1 phosphorylation levels correlate with mRNA expression in proliferating yeast. (A) Correlation analysis between the logarithm of the average Tyr1-P levels and the logarithm of mRNA levels [46]. (B) Gene-averaged profiles for genes in three different mRNA expression level classes. The genes were partitioned into three groups: low (25-50% quantile), medium (50-75% quantile) and high (>75% quantile) expression. From this set of genes, those with ORF lengths between 938 and 1538 were selected.

A correlation between levels of Tyr1 phosphorylation and mRNA expression (Figure 25), further indicated that Tyr1 phosphorylation is functionally relevant.

Gene-averaging of ChIP profiles (Section 5.7.3) revealed Tyr1 phosphorylation in the coding region (Figure 26A and B). Whereas Tyr1 phosphorylation signals were low at promoters, they increased downstream of the TSS. The gene-averaged profile resembled that for Ser2 phosphorylation, except that Ser2 phosphorylation signals persist downstream of the pA site for ~200 bp, whereas Tyr1 phosphorylation signals decrease already around 180 bp upstream of the pA site (Figure 26B and D). The point of Tyr1 phosphorylation signal increase was dependent on the TSS, whereas the point of decrease was dependent on the pA site, but not on gene length or expression level (Figures 25B and 26D). These results indicate that Tyr1 phosphorylation marks are set and removed within the transcription cycle.

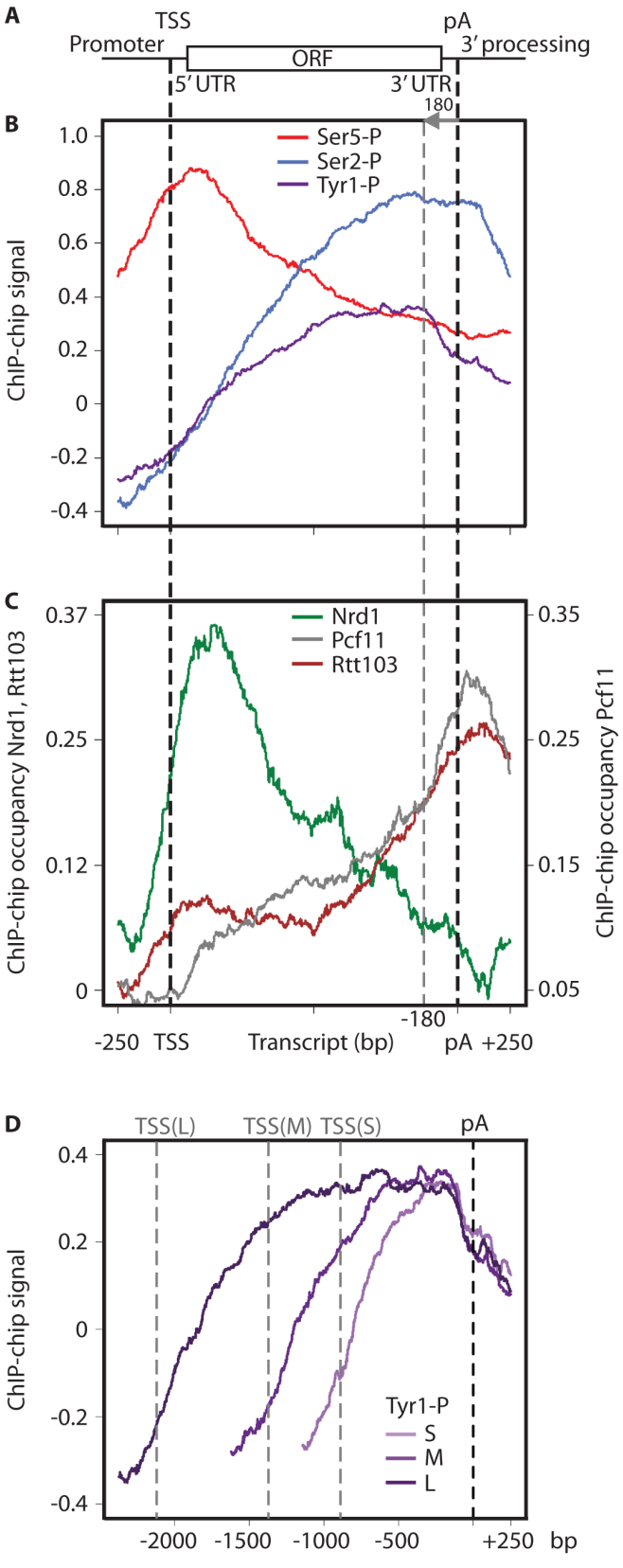

Figure 26: Gene-averaged ChIP-chip profiles for CTD phosphorylations and termination factors. (A) DNA frame with promoter, 5′-untranslated region (UTR), open reading frame (ORF), and 3′-UTR. Dashed black lines indicate the TSS and pA site. The dashed gray line marks the position 180 bp upstream of the pA site. (B) Gene-averaged profiles for Ser5, Ser2, and Tyr1 phosphorylation for 339 genes of 'medium/M' length (1,238 ± 300 bp). (C) Gene-averaged profiles for Nrd1, Pcf11, and Rtt103. ChIP-chip occupancy of Nrd1 and Rtt103 is on the left y axis, Pcf11 occupancy on the right y axis. (D) Gene-averaged Tyr1 phosphorylation ChIP profiles for 'small/S' (725 ± 213 bp, 266 genes), 'medium/M' (B), and 'long/L' (2,217 ± 679 bp, 299 genes) gene length classes, aligned at the pA site.

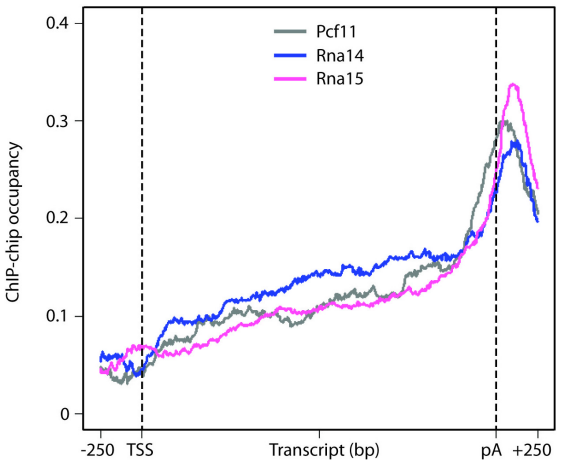

Figure 27: Genome-wide ChIP occupancy profiling of CFI subunits in yeast.
Gene-averaged profiles for the long gene-length class L (2,217 ± 679 bp, 299 genes) for Pcf11 (Section 6.2), Rna14, and Rna15. Profiles of other length classes are generally similar [147]. Dashed black lines indicate the TSS and pA site.

8.2 Genome-wide occupancy profiling of termination factors

8.2.1 CID-containing termination factors

To investigate whether Tyr1 phosphorylation influences factor recruitment to Pol II, we determined genome-wide occupancy profiles for termination factors Nrd1, Rtt103, and Pcf11, which contain a CTD-interacting domain (CID). The gene-averaged Nrd1 occupancy peaked at the beginning of the transcribed region, 193 ± 44 bp downstream of the TSS (Figure 26C). This region also showed maximum signals in Ser5 phosphorylation, and genomic Nrd1 and Ser5 phosphorylation profiles correlate ($R = 0.6$), consistent with Nrd1 binding to the Ser5-phosphorylated CTD [224]. The general presence of Nrd1 at protein-coding genes extends previous results [38, 106] and befits a role of Nrd1 in early transcription termination [19, 107, 191, 210]. Rtt103 showed peak occupancy at the end of genes, 112 ± 27 bp downstream of the pA site, where peak levels of Pcf11 were also observed (Figure 26C). Since this region shows the maximum difference between Ser2 and Tyr1 phosphorylation signals, Tyr1 phosphorylation may impair recruitment of Rtt103 and Pcf11 upstream of the pA site. Consistent with this, genome-wide occupancies of Rtt103 and Pcf11 do not correlate well with Ser2 phosphorylation signals ($R = 0.4$, for both), although both proteins bind the Ser2-phosphorylated CTD [136, 150].

8.2.2 Other termination factors

To investigate whether non CID-containing termination factors show a similar binding profile, we performed ChIP-chip for Rna14 and Rna15. These two proteins are components of the yeast cleavage factor I (CFI) that also contains Pcf11. This analysis revealed CFI recruitment at all protein-coding genes that are occupied by Pol II and its elongation factors (Section 6). The gene-averaged ChIP-chip profiles showed sharp occupancy peaks for Rna14 and Rna15, ∼100 bp downstream of the pA site (Figure 27). These profiles were independent of gene length and expression level [147]. Comparison with previous profiles (Section 6.2) revealed that the peak occupancies of non CID-containing Rna14 and Rna15 coincided with CID-containing Pcf11 occupancy, which peaked ∼60 bp downstream of the pA site (Figure 27 and Table 16).

8.3 Tyr1 phosphorylation impairs CTD binding of termination but not elongation factors

Gene-averaged occupancy profiles of CID-containing termination factors indicated that Tyr1 phosphorylation might impair CTD binding (Figure 26). To test whether Tyr1 phosphorylation impairs CTD binding of termination factors *in vitro*, Andreas Mayer determined the affinity of purified recombinant CIDs of yeast Nrd1, Pcf11, and Rtt103 for various CTD diheptad phosphopeptides using fluorescence anisotropy [145]. Consistent with previous results [136, 224], Pcf11-CID and Rtt103-CID bound to the Ser2-phosphorylated CTD, whereas the Nrd1-CID preferentially bound to a Ser5-phosphorylated CTD peptide. In contrast, none of the CIDs bound Tyr1-phosphorylated CTD peptides, regardless of whether additional phosphorylations were present or not. Thus Tyr1 phosphorylation blocks CID binding to the CTD *in vitro*, consistent with the hypothesis that it impairs termination factor recruitment *in vivo*.
Elongation factor Spt6 contains a tandem SH2 domain that also binds the CTD. This domain binds the Ser2-phosphorylated CTD [31, 49, 131, 213] and is required for high Spt6 occupancy on transcribed genes (Section 6.7), suggesting that Tyr1 phosphorylation does not interfere with its CTD binding. Indeed the recombinant domain (residues 1250-1444) bound very well to CTD peptides phosphorylated at Tyr1, Tyr1 and Ser2, or Tyr1 and Ser5, but not to unphosphorylated CTD [145]. These results were consistent with recent data [31, 49] and showed that interactions with Tyr1-phosphorylated CTD peptides were even stronger than for peptides with phosphorylations at Ser2 or Ser5 alone. This shows that Tyr1 phosphorylation stimulates CTD binding of a *bona fide* elongation factor, whereas it blocks binding of CID-containing termination factors.

8.4 Glc7 dephosphorylates Tyr1

Our results show that the transition from transcription elongation to termination involves CTD Tyr1 dephosphorylation by an unknown Tyr1 phosphatase (Section 8.3 and Figure 26). We reasoned that the Tyr1 phosphatase could be part of the large mRNA 3´ processing and termination machinery. In particular, two of the fifteen subunits of cleavage and polyadenylation factor (CPF), Ssu72 [116, 152] and Glc7 [161], are candidate phosphatases. Ssu72 is known to dephosphorylate CTD residue Ser5 [9, 116, 241], whereas Glc7 has no known CTD-related activity.
Amelie Schreieck and Stefanie Etzold in collaboration with the group of Lori Passmore could show that Glc7 dephosphorylates Tyr1 *in vitro* (manuscript submitted). To investigate whether Glc7 acts as a CTD Tyr1 phosphatase also *in vivo*, we depleted Glc7 from the nucleus using the anchor-away method [76] and monitored changes in genome-wide occupancy with Tyr1-phosphorylated Pol II by chromatin immunoprecipitation profiling (Sections 5.6 and 5.7). For metagene analysis of ChIP data, we aligned genes at their pA sites [159] and excluded genes flanked by a neighboring gene within 400 bp downstream of their pA site. Glc7 nuclear depletion led to a strong defect in Tyr1 dephosphorylation at pA sites (Figure 28A). Defective Tyr1 dephosphorylation was not due to rapamycin treatment that is used in the anchor-away procedure (Figure 28B). Additional ChIP data indicated that Glc7 is required for transcription termination. The ChIP signal for total Pol II (Rpb3 subunit) normally decreased about 200 bp downstream of the pA site, indicating transcription termination and Pol II release from DNA (Figure 28C, −Rapa). However, upon nuclear depletion of Glc7, strong Pol II ChIP signals remained further downstream (Figure 28C, +Rapa). These data show that Glc7 acts globally and is required for effective CTD Tyr1 dephosphorylation and normal transcription termination *in vivo*.

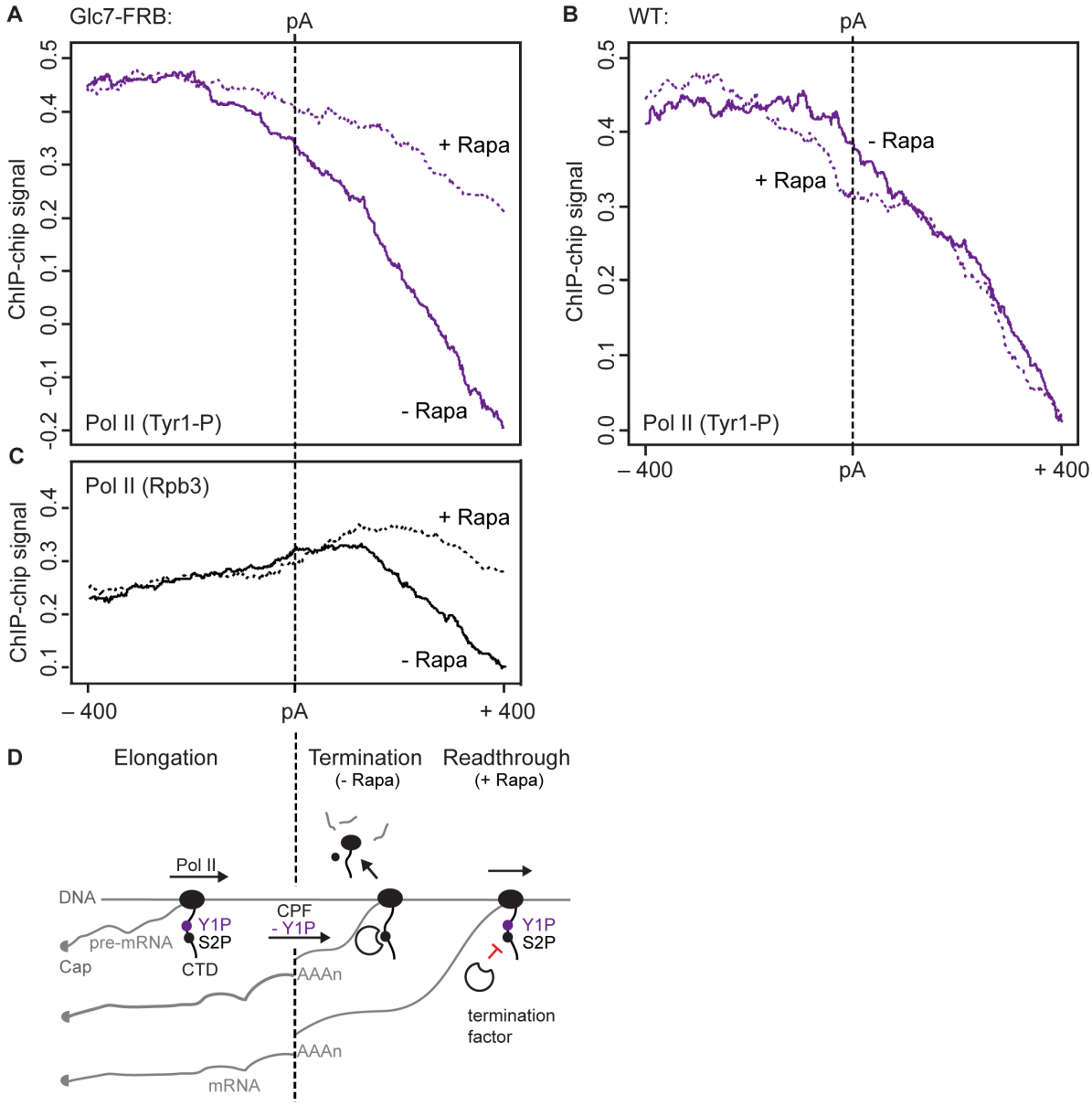

Figure 28: Glc7 is required for Tyr1 dephosphorylation and transcription termination *in vivo*. (A, B) Metagene analysis for genome-wide ChIP occupancy of Tyr1-phosphorylated Pol II around pA sites in the Glc7-FRB anchor-away strain (A) and in the anchor-away wild-type strain (B). 619 genes were aligned at the pA site (Section 5.7.3). (A) Occupancy for Tyr1-phosphorylated Pol II decreases sharply at the pA site (−Rapa, violet line), but Glc7 depletion largely abolishes this decrease (+Rapa, violet dotted line). (B) Rapamycin treatment has no effect on Tyr1-phosphorylated Pol II occupancy in the anchor-away wild-type strain (compare −Rapa, violet line and +Rapa, violet dotted line). (C) Metagene analysis as in A but for total Pol II (subunit Rpb3). Occupancy of Pol II normally decreases downstream of the pA site, indicating transcription termination (−Rapa, black line). Nuclear depletion of Glc7 impairs termination (+Rapa, black dotted line). (D) Model for the Pol II elongation-termination transition. DNA and RNA are depicted as gray lines, Pol II and its CTD are in black. For details see text.

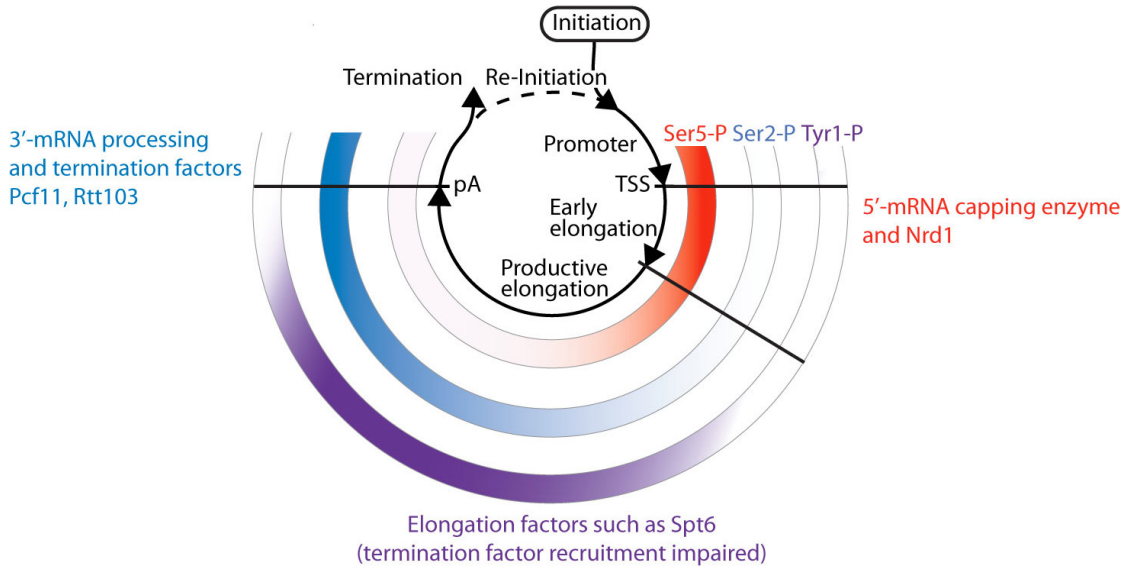

Figure 29: Extended CTD code for transcription cycle coordination. During the cycle, levels of CTD phosphorylation at Tyr1, Ser2, and Ser5 residues change differently, as illustrated by color gradients.

8.5 Discussion & Summary

Our results extend the previously proposed CTD code [18, 33, 51] (for details see Section 1.2.1), leading to an extended CTD code for the coordination of the transcription cycle with factor recruitment (Figure 29), and a new model for the elongation-termination transition of Pol II at the 3´-ends of protein-coding genes (Figure 28D).

During initiation and early elongation, the CTD is phosphorylated on Ser5, which facilitates recruitment of the capping enzyme and Nrd1. 150-200 bp downstream of the TSS, peak occupancy levels are reached for Nrd1 and Pol II, likely marking a decision point where Pol II transiently pauses and either terminates or continues elongation [19]. When Tyr1 and Ser2 phosphorylation levels rise, Pol II binds elongation factors stably and continues elongation. Tyr1 phosphorylation releases Nrd1 and impairs recruitment of Rtt103 and Pcf11, thereby suppressing termination during elongation. Our results indicate that Tyr1 CTD phosphorylation is a target to activate transcription by suppressing Pol II termination, and explain why mutation of Tyr1 to phenylalanine, which lacks the oxygen atom required for phosphorylation, is lethal [232].

When Pol II reaches the pA site, the Glc7 subunit of CPF dephosphorylates Tyr1, whereas Ser2 phosphorylation remains. This allows for binding of Pcf11 and Rtt103, termination factors that are not part of CPF and show peak occupancy ∼50-100 bp downstream of the pA site. Transcription terminates further downstream and releases Pol II from genes (Figure 28D, −Rapa). When Glc7 is depleted from the nucleus, Tyr1 phosphorylation levels remain high downstream of the pA site, impairing termination factor recruitment and Pol II release (Figure 28D, +Rapa). Our results not only establish Glc7 as the Pol II CTD Tyr1 phosphatase, but they also unravel CPF-triggered Pol II dephosphorylation as a key event that couples pA-dependent 3´ RNA processing to transcription termination.

A link between the pA site and transcription termination was established over 25 years ago [133]. In the 'anti-terminator' model, transcription of the pA site triggers a change in the Pol II machinery that allows for termination [17, 19, 155]. In the 'torpedo' model, pA-dependent RNA cleavage results in a new RNA 5´-end that is recognized by the Rat1/Rai1/Rtt103 exonuclease complex that degrades nascent RNA and triggers termination [17, 19, 155]. Our data are consistent with a combination of both models. Tyr1 phosphorylation serves as an anti-terminator that is released at the pA site by Glc7 dephosphorylation, which then allows recruitment of the torpedo nuclease complex.

9 A global view of the RNA polymerase II transcription cycle

In this work we used high-resolution ChIP-chip occupancy profiling of the RNA polymerase (Pol) II transcription machinery to obtain new insights into the mRNA trancription cycle. To integrate all the data we carried out a correlation analysis using all genome-wide profiles reported here (Sections 6 to 8). Pearson correlation coefficients between profiles were provided as a similarity metric to calculate a correlation-based two-dimensional network using the GraphViz's Neato algorithm [56] (Figure 30). In the resulting network, elongation factors cluster with Pol II, indicating that the elongating form of Pol II dominates its occupancy profile. The three groups of elongation factors observed in Section 6 form sub-groups of the elongation cluster. Initiation and termination factors form distinct clusters, reflecting different phases of the transcription cycle. Profiles for the capping enzymes lie between clusters for initiation and elongation factors. This unbiased analysis highlights the central role of pre-mRNA capping during the initiation-elongation transition (Section 7). Profiles of the Pol II isoforms phosphorylated at Ser5 and Ser7 fall between the initiation and elongation clusters, consistent with the role of these modifications in the transition [41, 114, 149, 199, 209]. Profiles of the Pol II isoforms phosphorylated at Tyr1 and Ser2 fall between the elongation and termination clusters, highlighting their role during the 3´ transition (Section 8) [3, 127, 136, 150, 251].

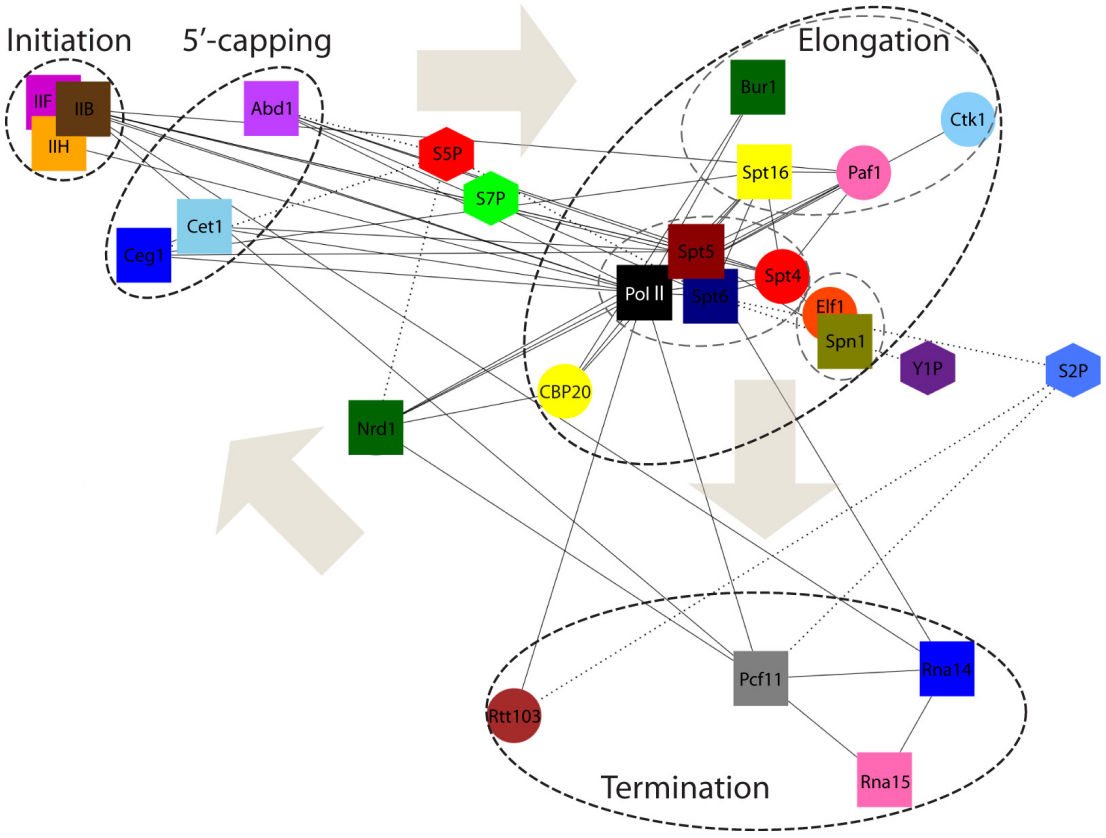

Figure 30: Correlation-based network of occupancy profiles reflects different phases of the transcription cycle. Correlation analysis of genome-wide ChIP-chip profiles reported in Sections 6 to 8. Pairwise Pearson correlation coefficients were calculated between concatenated gene profiles ranging from TSS−250 bp to pA+250 bp and provided as a similarity metric to calculate a two-dimensional network using the GraphViz's Neato algorithm (see Methods 5.7.5). Solid lines represent known direct interactions between factors. Dashed lines represent known direct interactions between factors and the phosphorylated CTD of Pol II. Essential and non-essential factors are represented as boxes and circles, respectively. Pol II phospho-isoforms are represented as hexagons.

Part IV
Future perspectives

Precise coordination of gene expression is essential for proper cell functioning and depends on the interaction of proteins with DNA. The capability of studying genome-wide occupancy of transcription factors ([72, 122, 226]), the general RNA polymerase (Pol) II transcription machinery (this work and [9, 106, 158, 188, 218, 225, 249]), as well as chromatin states and modifications ([67, 78, 143, 144, 200]), have greatly aided our understanding of global gene regulation. Some of the future challenges arising from the results presented here are discussed in the following paragraphs:

Mechanisms that modulate the composition of the transcribing Pol II complex

This work started by generating high-resolution genome-wide ChIP occupancy profiles of Pol II, its phosphorylated CTD isoforms, as well as components of the initiation, elongation, termination, and RNA processing machinery in proliferating *S. cerevisiae* cells. Statistical data analysis provided evidence for a general Pol II transcription complex that mediates transcription and mRNA processing at all actively transcribed Pol II genes. At both ends of genes this complex undergoes uniform transitions in factor composition (Section 6). We could identify which factors are exchanged during these transitions, and furthermore determine the precise genomic locations of these events.
The challenge ahead is to completely unravel the molecular mechanisms involved in orchestrating the composition of the transcribing Pol II complex throughout the transcription cycle.
We started to address this issue by providing evidence that mRNA 5´ cap completion underlies the initiation-elongation transition and triggers productive Pol II elongation (Section 7), and that Pol II CTD tyrosine 1 dephosphorylation underlies the elongation-termination transition and triggers mRNA 3´ processing and termination (Section 8). These observations together with additional data by other groups [2, 57, 141, 147, 182, 181, 213, 239] provide mechanistic insights into the triggers involved in driving transcription factor exchange during the 5' and 3' transitions. However, more research is needed to completely understand the molecular mechanisms involved.
For instance, it remains to be fully understood how the main CTD Ser2 kinase Ctk1 associates with Pol II during the initiation-elongation transition. Although we could show that the cap methyltransferase Abd1 and the cap-binding complex (CBC) contribute to Ctk1 recruitment, these factors alone cannot explain the entire Ctk1 occupancy along genes (Section 7). Moreover, elongation factor Spt6, which is known to bind the phosphorylated CTD *in vitro* [145], is still partly recruited to the transcribing polymerase when its CTD interaction domain is deleted (Section 6). Thus, additional recruitment mechanisms must exist to ensure proper association of these factors with Pol II.

Emerging roles of the Spt5 CTR and its phosphorylation

It has become apparent that not only the Pol II CTD but also the Spt5 CTR plays an important role as a platform to recruit accessory factors to the transcribing Pol II complex. As it is the case for the CTD, factor recruitment by the CTR also occurs in coordination with the transcription and RNA processing cycle. The Spt5 CTR contributes to capping enzyme recruitment (Section 7), Paf1 recruitment [132], and cleavage factor I recruitment [147], during early elongation, productive elongation, and termination, respectively. Considering that the CTR can be phosphorylated by Bur1 [132, 254], and since our data suggest that this phosphorylation of the CTR might release capping enzymes from Spt5, it will be fascinating to study the role of the phosphorylation state of the CTR in factor recruitment in coordination with the transcription cycle. Similar to the proposed "CTD code" (Section 1.2.1) the changing phosphorylation status of the CTR during the transcription cycle could allow binding of different transcription and RNA processing factors

in an ordered manner. Experimentally this could be addressed by generating an antibody that specifically recognizes the phospho-CTR. Such an antibody could then be used to perform ChIP experiments revealing possible changes in CTR phosphorylation during the transcription cycle. Moreover, correlation analysis between phosphorylation level and factor occupancies could provide insights about the phosphorylation state of the CTR in known protein-CTR interactions and also contribute to the identification of novel CTR binders. In addition, an unbiased approach to identify novel CTR binders would be to combine an affinity-purification assay with mass spectrometry (MS).

Novel and additional roles of CTD modifications

For various CTD phospho-modifications there are now highly specific antibodies available, which allowed us to investigate the genome-wide distribution of these marks (Sections 6 and 8).
Correlation analysis between genome-wide phosphorylation levels and occupancy of proteins known to bind to the phosphorylated CTD *in vitro*, revealed that the *in vivo* situation of CTD binding might be more complex than initially thought. For instance, based on the ChIP profiles observed for CTD phosphorylation marks and CTD-interacting termination factors, we hypothesized and verified in the lab that Tyr1-P impairs binding of these termination factors to the CTD during the elongation phase of transcription (Section 8, [145]). Thus, Tyr1-P is used to fine-tune the binding of termination factors to the Ser5-P and Ser2-P CTD. It will be interesting to see whether additional mechanisms exist that allow fine-tuning of the "CTD code".
Tyr1-P could also modulate the binding of the capping enzymes Cet1-Ceg1 to the Ser5-P CTD, since Cet1-Ceg1 is released from the CTD prior to Ser5 dephosphorylation (Section 7) and the ChIP-chip profiles of Cet1-Ceg1 occupancy and Tyr1-P levels are mutually exclusive (Sections 7 and 8). One way to investigate this further would be to determine the *in vitro* affinity of purified recombinant Cet1-Ceg1 for CTD diheptad phosphopeptides (Ser5-P vs. Tyr1-P+Ser5-P) by using fluorescence anisotropy [145]. Additionally, once the Tyr1 kinase has been identified, ChIP experiments could be used to investigate what effect conditional depletion of the kinase has on Cet1-Ceg1 release *in vivo*.

Although the genome-wide profile of Ser7-P has been mapped (Section 6) and its kinase discovered [4], its function at protein coding genes remains unclear and has to be investigated. Since the Ser7-P profile correlates well with occupancy of the capping enzymes (Sections 6 and 7), one could imagine a role in recruiting capping enzymes in addition to Ser5-P or stimulation of capping. Furthermore, the functions and binders of Thr4-P in yeast as well as of other CTD modifications like glycosylations need to be investigated. Future advances in the development of *in vivo* cross-linking technologies coupled with MS approaches might be of use to determine the complete set of CTD binding proteins and to eventually understand how they coordinately interact with the CTD to regulate transcription and/or RNA processing.

Identification of missing CTD kinases and phosphatases

We obtained genome-wide levels of all possible CTD phosphorylations (Sections 6 and 8), however, the yeast kinases and the phosphatase required for Tyr1/Thr4 phosphorylation and Thr4 dephosphorylation, respectively, remain to be identified. One way to find respective kinase(s) and phosphatase(s) *in vivo* would be to conditionally deplete potential candidates and observe whether the Tyr1/Thr4 phosphorylation level is altered by ChIP experiments. Conditional nuclear depletion of the factors of interest can be achieved by using the anchor-away technique (Section 7, [76]). Moreover, kinases can also be identified by the use of analog-sensitive yeast strains [14]. Furthermore, *in vitro* phosphatase assays with purified phosphatases and Thr4 phosphorylated CTD as substrates could contribute to the identification of the phosphatases.

Analysis of the CTD by mass spectrometry

One current limitation of CTD research is that most knowledge relies on the use of monoclonal antibodies that specifically recognize certain modified CTD residues. Binding of these antibodies requires that the epitopes are accessible, which is often prevented by neighboring modifications. Moreover, these antibodies cannot distinguish on which individual repeat a certain mark is present along the linear CTD. To circumvent some of the limitations imposed by using antibodies, the challenge ahead will be to use novel MS approaches to decipher the CTD state. Since the lack of positively charged amino acids makes the phospho-CTD patterns difficult to read out via MS, genetically engineered CTDs could be designed to facilitate proper MS analysis [77]. This would allow a more direct observation of the CTD modification status (i.e. which modifications occur, in which combinations, in which parts of the CTD).

The role of nascent RNA in recruiting Pol II associated factors

The nascent RNA adds an additional layer of complexity to the assembly/disassembly events of the transcribing Pol II complex. It has been shown for some transcription and RNA processing factors that they can bind the nascent RNA [38, 47, 147, 170, 175], and for some it was demonstrated that they are (at least partially) recruited to the transcription machinery in an RNA dependent manner [47, 147]. Thus, the nascent RNA can also contribute to modulate the factor composition of the transcribing Pol II complex. In the work presented here we used ChIP to investigate the interactions of the Pol II machinery with DNA. To further complement the resulting occupancy data, it would be beneficial to investigate whether and if so how, the Pol II machinery interacts with nascent RNA during the transcription cycle. PAR-CLIP, a novel method to identify transcriptome-wide protein-RNA interactions *in vivo* [70], can be used to address this issue.

Controlling of pervasive transcription

The use of tiling microarrays and, more recently, RNA sequencing for studying trancriptomes has led to the dicovery that most of the DNA of eukaryotic genomes is transcribed by Pol II [89]. This pervasive transcription gives rise to a wealth of novel non-coding and potentially non-functional RNA species, and implies a requirement for massive RNA quality surveillance in the nucleus. Despite the question to what extent these novel transcripts (or their synthesis per se) are biologically functional [228], or whether the majority simply accounts for transcriptional noise [211], another future challenge is to unravel the underlying molecular mechanisms that are employed to distinguish functional from non-functional RNAs and degrade the latter.
In yeast, cryptic unstable transcripts (CUTs) are likely to be enriched for RNA binding motifs that are recognized by the Nrd1–Nab3–Sen1 complex, which usually facilitates termination of sno/snRNAs [38, 217]. This complex can also induce termination of CUT transcription and couples it to degradation by the nuclear exosome. A comparable mechanism in higher eukaryotes remains to be identified, and it is likely that additional characteristics exist to distinguish the products of pervasive transcription from stable RNAs, including chromatin modifications or the status/composition of the transcribing Pol II complex. A combination of genome-wide ChIP, PAR-CLIP, as well as nascent RNA sequencing experiments under various conditions will be needed to further address these questions and elucidate the underlying mechanisms.

References

[1] K. Adelman and J. T. Lis. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.*, 13(10):720–731, 2012. 1.3, 7

[2] S. H. Ahn, M.-C. Keogh, and S. Buratowski. Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II. *EMBO J.*, 28(3):205–12, Feb. 2009. 6.13, 8, IV

[3] S. H. Ahn, M. Kim, and S. Buratowski. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell*, 13(1):67–76, Jan. 2004. 1.2.1, 6.13, 7, 8, 9

[4] M. S. Akhtar, M. Heidemann, J. R. Tietjen, D. W. Zhang, R. D. Chapman, D. Eick, and A. Z. Ansari. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol. Cell*, 34(3):387–93, May 2009. 6.8, IV

[5] O. Aparicio, J. V. Geisberg, E. Sekinger, A. Yang, Z. Moqtaderi, and K. Struhl. Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. *Curr. Protoc. Mol. Biol.*, Chapter 21:Unit 21.3, Feb. 2005. 2

[6] D. Barillà, B. A. Lee, and N. J. Proudfoot. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci.*, 98(2):445–50, Jan. 2001. 6.10

[7] B. Bartkowiak, P. Liu, H. P. Phatnani, N. J. Fuda, J. J. Cooper, D. H. Price, K. Adelman, J. T. Lis, and A. L. Greenleaf. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.*, 24(20):2303–16, Oct. 2010. 1.2.1, 7, 7.10

[8] R. Baskaran, M. E. Dahmus, and J. Y. Wang. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc. Natl. Acad. Sci.*, 90(23):11167–71, Dec. 1993. 1.2.1, 8

[9] A. R. Bataille, C. Jeronimo, P.-E. Jacques, L. Laramée, M.-E. Fortin, A. Forest, M. Bergeron, S. D. Hanes, and F. Robert. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol. Cell*, 45(2):158–70, Jan. 2012. 1.2.1, 8.4, IV

[10] R. Belotserkovskaya, S. Oh, V. A. Bondarenko, G. Orphanides, V. M. Studitsky, and D. Reinberg. FACT facilitates transcription-dependent nucleosome alteration. *Science*, 301(5636):1090–3, Aug. 2003. 1.1.2, 6.13

[11] T. Benoukraf, P. Cauchy, R. Fenouil, A. Jeanniard, F. Koch, S. Jaeger, D. Thieffry, J. Imbert, J.-C. Andrau, S. Spicuglia, et al. CoCAS: a ChIP-on-chip analysis suite. *Bioinformatics*, 25(7):954–955, 2009. 2.1.2

[12] E. Birney, J. A. Stamatoyannopoulos, A. Dutta, R. Guigó, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, R. E. Thurman, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146):799–816, 2007. I

[13] C. Birse, L. Minvielle-Sebastia, B. Lee, W. Keller, and N. Proudfoot. Coupling Termination of Transcription to Messenger RNA Maturation in Yeast. *Science*, 280(5361):298–301, Apr. 1998. 1.2.4

[14] A. C. Bishop, J. A. Ubersax, D. T. Petsch, D. P. Matheos, N. S. Gray, J. Blethrow, E. Shimizu, J. Z. Tsien, P. G. Schultz, M. D. Rose, et al. A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature*, 407(6802):395–401, 2000. IV

[15] S. Boeing, C. Rigault, M. Heidemann, D. Eick, and M. Meisterernst. RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. *J. Biol. Chem.*, 285(1):188–96, Jan. 2010. 6.8

[16] A. Bortvin and F. Winston. Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science*, 272(5267):1473–1476, 1996. 1.1.2

[17] K. Brannan and D. L. Bentley. Control of transcriptional elongation by RNA polymerase II: a retrospective. *Genet. Res. Int.*, 2012:170173, Jan. 2012. 8.5

[18] S. Buratowski. The CTD code. *Nat. Struct. Mol. Biol.*, 10(9):679–680, 2003. 8.5

[19] S. Buratowski. Progression through the RNA polymerase II CTD cycle. *Mol. Cell*, 36(4):541–6, Nov. 2009. 6, 7, 8, 8.2.1, 8.5

[20] C. B. Burge, T. Tuschl, and P. A. Sharp. 20 Splicing of Precursors to mRNAs by the Spliceosomes. *Cold Spring Harbor Monograph Archive*, 37:525–560, 1999. 1.2.3

[21] G. Calero, K. F. Wilson, T. Ly, J. L. Rios-Steiner, J. C. Clardy, and R. A. Cerione. Structural basis of m7GpppG binding to the nuclear cap-binding protein complex. *Nat. Struct. Mol. Biol.*, 9(12):912–7, Dec. 2002. 1.2.2, 7.6

[22] R. D. Chapman, M. Heidemann, T. K. Albert, R. Mailhammer, A. Flatley, M. Meisterernst, E. Kremmer, and D. Eick. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science*, 318(5857):1780–2, Dec. 2007. 1.2.1, 6.7

[23] R. D. Chapman, M. Heidemann, C. Hintermair, and D. Eick. Molecular evolution of the RNA polymerase II CTD. *Trends Genet.*, 24(6):289–96, June 2008. 8

[24] M. Chen and J. L. Manley. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.*, 10(11):741–754, 2009. 1.2.3

[25] B. Cheng and D. H. Price. Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *J. Biol. Chem.*, 282(30):21901–12, July 2007. 1.2.1, 7, 7.10

[26] A. C. Cheung, S. Sainsbury, and P. Cramer. Structural basis of initial RNA polymerase II transcription. *EMBO J.*, 30(23):4755–4763, 2011. 1.1.1

[27] Y.-L. Chiu, C. K. Ho, N. Saha, B. Schwer, S. Shuman, and T. M. Rana. Tat stimulates cotranscriptional capping of HIV mRNA. *Mol. Cell*, 10(3):585–597, 2002. 7.10

[28] E. J. Cho, M. S. Kobor, M. Kim, J. Greenblatt, and S. Buratowski. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev.*, 15(24):3319–29, Dec. 2001. 1.1.2, 1.2.1, 6.8, 7, 7.8, 7.10

[29] H. Cho, T.-K. Kim, H. Mancebo, W. S. Lane, O. Flores, and D. Reinberg. A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev.*, 13(12):1540–1552, 1999. 1.2.1

[30] L. S. Churchman and J. S. Weissman. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, 469(7330):368–373, 2011. 2.2

[31] D. Close, S. J. Johnson, M. a. Sdano, S. M. McDonald, H. Robinson, T. Formosa, and C. P. Hill. Crystal structures of the S. cerevisiae Spt6 core and C-terminal tandem SH2 domain. *J. Mol. Biol.*, 408(4):697–713, May 2011. 8.3

[32] T. A. Cooper, L. Wan, and G. Dreyfuss. RNA and Disease. *Cell*, 136(4):777–793, Feb. 2009. 1.3

[33] J. L. Corden. Transcription. Seven ups the code. *Science*, 318(5857):1735–6, Dec. 2007. 8, 8.5

[34] L. Core, J. Waterfall, and J. Lis. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 322(5909):1845–1848, 2008. 1.3, 2.2, 6.11

[35] V. Cowling. Regulation of mRNA cap methylation. *Biochem. J.*, 302:295–302, 2010. 1.2.2, 7

[36] P. Cramer, K.-J. Armache, S. Baumli, S. Benkert, F. Brueckner, C. Buchen, G. Damsma, S. Dengl, S. Geiger, A. Jasiak, et al. Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.*, 37:337–352, 2008. I, 1.2.1

[37] P. Cramer, D. A. Bushnell, and R. D. Kornberg. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science (New York, N.Y.)*, 292(5523):1863–1876, June 2001. 1.2

[38] T. J. Creamer, M. M. Darby, N. Jamonnak, P. Schaughency, H. Hao, S. J. Wheelan, and J. L. Corden. Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet.*, 7(10):e1002329, Oct. 2011. 8.2.1, IV, IV

[39] F. Crick. Central dogma of molecular biology. *Nature*, 227, 1970. 1

[40] N. Custodio, M. Carmo-Fonseca, F. Geraghty, H. S. Periera, F. Grosveld, and M. Antoniou. Inefficient processing impairs release of RNA from the site of transcription. *EMBO J.*, 18(10):2855–2866, 1999. 1.3

[41] N. Czudnochowski, C. A. Bösken, and M. Geyer. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun.*, 3(may):842, Jan. 2012. 9

[42] S. Danckwardt, M. W. Hentze, and A. E. Kulozik. 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.*, 27(3):482–498, 2008. 1.3

[43] C. G. Danko, N. Hah, X. Luo, A. L. Martins, L. Core, J. T. Lis, A. Siepel, and W. L. Kraus. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol. Cell*, 2013. 2.2

[44] L. David, W. Huber, M. Granovskaia, J. Toedling, C. J. Palm, L. Bofkin, T. Jones, R. W. Davis, and L. M. Steinmetz. A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci.*, 103(14):5320–5, Apr. 2006. I, 2.1.1, 2.1.2, 2.2, 3, 5.6.3, 6.1, 6

[45] S. de Almeida and M. Carmo-Fonseca. Cotranscriptional RNA checkpoints. *Epigenomics*, 2:449–455, 2010. 1.3

[46] S. Dengl, A. Mayer, M. Sun, and P. Cramer. Structure and in vivo requirement of the yeast Spt6 SH2 domain. *J. Mol. Biol.*, 389(1):211–25, May 2009. 5.7.3, 6.1, 13, 6.10, 6.12, 18, 25

[47] J. L. Dermody and S. Buratowski. Leo1 subunit of the yeast paf1 complex binds RNA and contributes to complex recruitment. *J. Biol. Chem.*, 285(44):33671–9, Oct. 2010. IV

[48] P. D'haeseleer et al. How does gene expression clustering work? *Nat. Biotechnol.*, 23(12):1499–1502, 2005. 2.1.2

[49] M.-L. Diebold, E. Loeliger, M. Koch, F. Winston, J. Cavarelli, and C. Romier. Noncanonical tandem SH2 enables interaction of elongation factor Spt6 with RNA polymerase II. *J. Biol. Chem.*, 285(49):38389–98, Dec. 2010. 8.3

[50] A. Droit, C. Cheung, and R. Gottardo. rMAT-an R/Bioconductor package for analyzing ChIP-chip experiments. *Bioinformatics*, 26(5):678–679, 2010. 2.1.2

[51] S. Egloff and S. Murphy. Cracking the RNA polymerase II CTD code. *Trends Genet.*, 24(6):280–8, June 2008. 8.5

[52] S. Egloff, D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick, and S. Murphy. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, 318(5857):1777–9, Dec. 2007. 1.2.1, 8

[53] X. Fan, N. Lamarre-Vincent, Q. Wang, and K. Struhl. Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments. *Nucleic Acids Res.*, 36(19):e125, Nov. 2008. 2, 5.5.4

[54] Y. W. Fong and Q. Zhou. Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, 414(6866):929–933, 2001. 1.3

[55] N. J. Fuda, M. B. Ardehali, and J. T. Lis. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*, 461(7261):186–192, Sept. 2009. 1.3

[56] S. Gansner, E; North. Improved force-directed layouts. In *Graph Drawing 98*, pages 364–373. Springer, 1998. 5.7.5, 9

[57] L. Gao and D. S. Gross. Sir2 silences gene transcription by targeting the transition between RNA polymerase II initiation and elongation. *Mol. Cell. Biol.*, 28(12):3979–94, June 2008. 8, IV

[58] R. C. Gentleman, V. J. Carey, D. M. Bates, and others. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, 5:R80, 2004. 2.1.2, 5.7.1

[59] A. Ghosh and C. D. Lima. Enzymology of RNA cap synthesis. *Wiley Interdiscip. Rev. RNA*, 1(1):152–72, July 2010. 1.2.2, 7

[60] D. A. Gilchrist, D. C. Fargo, and K. Adelman. Using ChIP-chip and ChIP-seq to study the regulation of gene expression: genome-wide localization studies reveal widespread regulation of transcription elongation. *Methods*, 48(4):398–408, 2009. 2.2

[61] D. A. Gilchrist, S. Nechaev, C. Lee, S. K. B. Ghosh, J. B. Collins, L. Li, D. S. Gilmour, and K. Adelman. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev.*, 22(14):1921–1933, 2008. 1.3

[62] D. S. Gilmour and J. T. LiS. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. *Proc. Natl. Acad. Sci.*, 81(14):4275–4279, 1984. 2

[63] D. S. Gilmour and J. T. Lis. In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. *Mol. Cell. Biol.*, 5(8):2009–2018, 1985. 2

[64] K. Glover-Cutter, S. Kim, J. Espinosa, and D. L. Bentley. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat. Struct. Mol. Biol.*, 15(1):71–8, Jan. 2008. 6.13

[65] D. Grohmann, J. Nagy, A. Chakraborty, D. Klose, D. Fielden, R. H. Ebright, J. Michaelis, and F. Werner. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol. Cell*, 43(2):263–74, July 2011. 1.1.1, 1.1.2, 6.13, 7, 7.10

[66] M. Gu, K. Rajashankar, and C. Lima. Structure of the Saccharomyces cerevisiae Cet1-Ceg1 mRNA capping apparatus. *Structure*, 18(2):216–227, 2010. 7.3, 7.10

[67] M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, and R. A. Young. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, 130(1):77–88, 2007. 2.1, IV

[68] A. Guiguen, J. Soutourina, M. Dewez, L. Tafforeau, M. Dieu, M. Raes, J. Vandenhaute, M. Werner, and D. Hermand. Recruitment of P-TEFb (Cdk9-Pch1) to chromatin by the cap-methyl transferase Pcm1 in fission yeast. *EMBO J.*, 26(6):1552–9, Mar. 2007. 1.3, 7, 7.10

[69] M. Guo, F. Xu, J. Yamada, T. Egelhofer, Y. Gao, G. A. Hartzog, M. Teng, and L. Niu. Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. *Structure*, 16(11):1649–58, Nov. 2008. 6.13

[70] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano Jr, A.-C. Jungkamp, M. Munschauer, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 141(1):129–141, 2010. IV

[71] S. Hahn and E. T. Young. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics*, 189(3):705–736, 2011. 1.1.1, 1.3

[72] C. T. Harbison, D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J.-B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, and R. A. Young. Transcriptional regulatory code of a eukaryotic genome. *Nature*, 431(7004):99–104, Sept. 2004. 2.1, IV

[73] H. Hartmann, E. W. Guthöhrlein, M. Siebert, S. Luehr, and J. Söding. P-value-based regulatory motif discovery using positional weight matrices. *Genome Res.*, 23(1):181–194, 2013. 2.1.2

[74] G. A. Hartzog and J. Fu. The Spt4-Spt5 complex: A multi-faceted regulator of transcription elongation. *Biochim. Biophys. Acta*, 2012. 1.1.2

[75] G. A. Hartzog, T. Wada, H. Handa, and F. Winston. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes Dev.*, 12(3):357–69, Feb. 1998. 1.1.2

[76] H. Haruki, J. Nishikawa, and U. K. Laemmli. The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol. Cell*, 31(6):925–32, Sept. 2008. 5.2.1, 5.4, 7.9, 8.4, IV

[77] M. Heidemann, C. Hintermair, K. Voß, and D. Eick. Dynamic phosphorylation patterns of RNA Polymerase II CTD during transcription. *Biochim. Biophys. Acta*, Sept. 2012. I, 1.1.2, 1.2, 1.2.1, IV

[78] N. D. Heintzman, R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.*, 39(3):311–318, 2007. 2.1, IV

[79] J. Hesselberth, X. Chen, and Z. Zhang. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat. Methods*, 6(4):283–289, 2009. 12

[80] P. Hilleren, T. McCarthy, M. Rosbash, R. Parker, and T. H. Jensen. Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature*, 413(6855):538–542, 2001. 1.3

[81] Y. Hirose and J. Manley. RNA polymerase II and the integration of nuclear events. *Genes Dev.*, 14:1415–1429, 2000. 6

[82] A. Hirtreiter, G. E. Damsma, A. C. M. Cheung, D. Klose, D. Grohmann, E. Vojnic, A. C. R. Martin, P. Cramer, and F. Werner. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.*, 38(12):4040–51, July 2010. 6.13

[83] C. K. Ho, V. Sriskanda, S. McCracken, D. Bentley, B. Schwer, and S. Shuman. The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.*, 273(16):9577–85, Apr. 1998. 7.10

[84] J. W. Ho, E. Bishop, P. V. Karchenko, N. Nègre, K. P. White, and P. J. Park. ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. *BMC Genomics*, 12(1):134, 2011. 2.2

[85] M. A. Hossain, C. Chung, S. K. Pradhan, and T. L. Johnson. The Yeast Cap Binding Complex Modulates Transcription Factor Recruitment and Establishes Proper Histone H3K36 Trimethylation during Active Transcription. *Mol. Cell. Biol.*, 33(4):785–99, Feb. 2013. 7, 7.10

[86] J.-P. Hsin and J. L. Manley. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.*, 26(19):2119–37, Oct. 2012. 1.1.2, 1.2, 1.2.1, 7

[87] J.-P. Hsin, A. Sheth, and J. L. Manley. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science*, 334(6056):683–6, Nov. 2011. 1.2.1, 8

[88] V. R. Iyer, C. E. Horak, C. S. Scafe, D. Botstein, M. Snyder, and P. O. Brown. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, 409(6819):533–538, 2001. 2

[89] A. Jacquier. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.*, 10(12):833–844, 2009. I, 1.3, IV

[90] J. A. Jaehning. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim. Biophys. Acta*, 1799(5-6):379–88, 2010. 1.1.2, 7.10

[91] A. J. Jasiak, H. Hartmann, E. Karakasili, M. Kalocsay, A. Flatley, E. Kremmer, K. Strässer, D. E. Martin, J. Söding, and P. Cramer. Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. *J. Biol. Chem.*, 283(39):26423–7, Sept. 2008. 6.1

[92] C. Jeronimo, A. R. Bataille, and F. Robert. The Writers, Readers, and Functions of the RNA Polymerase II C-Terminal Domain Code. *Chem. Rev.*, 2013. 1.2.1

[93] H. Ji, H. Jiang, W. Ma, D. S. Johnson, R. M. Myers, and W. H. Wong. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat. Biotechnol.*, 26(11):1293–1300, 2008. 2.1.2

[94] C. Jiang and B. F. Pugh. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.*, 10(3):161–72, Mar. 2009. 6.11

[95] X. Jiao, S. Xiang, C. Oh, C. E. Martin, L. Tong, and M. Kiledjian. Identification of a quality-control mechanism for mRNA 5'-end capping. *Nature*, 467(7315):608–11, Sept. 2010. 1.3, 7.10

[96] D. S. Johnson, W. Li, D. B. Gordon, A. Bhattacharjee, B. Curry, J. Ghosh, L. Brizuela, J. S. Carroll, M. Brown, P. Flicek, et al. Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets. *Genome Res.*, 18(3):393–403, 2008. 2.1.1

[97] W. E. Johnson, W. Li, C. A. Meyer, R. Gottardo, J. S. Carroll, M. Brown, and X. S. Liu. Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl. Acad. Sci.*, 103(33):12457–12462, 2006. 2.1.2, 5.7.1

[98] M. S. Jurica and M. J. Moore. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell*, 12(1):5–14, 2003. 1.2.3

[99] T. Juven-Gershon and J. T. Kadonaga. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.*, 339(2):225–229, 2010. 1.3

[100] C. D. Kaplan, M. J. Holland, and F. Winston. Interaction between transcription elongation factors and mRNA 3'-end formation at the Saccharomyces cerevisiae GAL10-GAL7 locus. *J. Biol. Chem.*, 280(2):913–22, Jan. 2005. 6.13

[101] C. D. Kaplan, L. Laprade, and F. Winston. Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, 301(5636):1096–9, Aug. 2003. 1.1.2, 6.13

[102] P. Kapranov, J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermüller, I. L. Hofacker, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 316(5830):1484–1488, 2007. I

[103] M. Keogh, V. Podolny, and S. Buratowski. Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Mol. Cell. Biol.*, 23(19):7005–18, 2003. 1.2.1, 6.13, 7

[104] H. Kettenberger, K.-J. Armache, and P. Cramer. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*, 114(3):347–57, Aug. 2003. 1.1.2

[105] H. Kettenberger, K.-J. Armache, and P. Cramer. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell*, 16(6):955–965, 2004. 1.1.1

[106] H. Kim, B. Erickson, W. Luo, D. Seward, J. H. Graber, D. D. Pollock, P. C. Megee, and D. L. Bentley. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat. Struct. Mol. Biol.*, 17(10):1279–86, Oct. 2010. 8.2.1, IV

[107] K.-Y. Kim and D. E. Levin. Mpk1 MAPK association with the Paf1 complex blocks Sen1-mediated premature transcription termination. *Cell*, 144(5):745–56, Mar. 2011. 8.2.1

[108] M. Kim, S.-H. Ahn, N. J. Krogan, J. F. Greenblatt, and S. Buratowski. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J.*, 23(2):354–64, Jan. 2004. 3, 6.13

[109] M. Kim, N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedea, J. F. Greenblatt, and S. Buratowski. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, 432(7016):517–522, Nov. 2004. 1.1.3

[110] M. Kim, H. Suh, E.-J. Cho, and S. Buratowski. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J. Biol. Chem.*, 284(39):26421–6, Sept. 2009. 1.2.1, 6.8

[111] M. Kim, L. Vasiljeva, O. J. Rando, A. Zhelkovsky, C. Moore, and S. Buratowski. Distinct pathways for snoRNA and mRNA termination. *Mol. Cell*, 24(5):723–734, 2006. 1.2.4

[112] K. Kizer, H. Phatnani, and Y. Shibata. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol. Cell. Biol.*, 25(8):3305–16, 2005. 7.10

[113] M. S. Kobor, J. Archambault, W. Lester, F. C. Holstege, O. Gileadi, D. B. Jansma, E. G. Jennings, F. Kouyoumdjian, A. R. Davidson, R. A. Young, et al. An Unusual Eukaryotic Protein Phosphatase Required for Transcription by RNA Polymerase II and CTD Dephosphorylation in S. cerevisiae. *Mol. Cell*, 4(1):55–62, 1999. 1.2.1

[114] P. Komarnitsky, E. Cho, and S. Buratowski. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.*, 14:2452–2460, 2000. 1.2.1, 1.2.2, 1.2.4, 6, 6.7, 7, 7.1, 7.10, 8, 9

[115] D. Kostrewa, M. E. Zeller, K.-J. Armache, M. Seizl, K. Leike, M. Thomm, and P. Cramer. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, 462(7271):323–30, Nov. 2009. 1.1.1, 6.2

[116] S. Krishnamurthy, X. He, M. Reyes-Reyes, C. Moore, and M. Hampsey. Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol. Cell*, 14(3):387–94, May 2004. 8.4

[117] N. Krogan, M. Kim, and S. Ahn. RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Mol. Cell. Biol.*, 22(20):6979–92, 2002. 1.1.2, 6.13

[118] J. N. Kuehner and D. A. Brow. Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. *J. Biol. Chem.*, 281(20):14119–28, May 2006. 6.2

[119] U. K. Laemmli et al. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259):680–685, 1970. 5.3.2

[120] R. N. Laribee, N. J. Krogan, T. Xiao, Y. Shibata, T. R. Hughes, J. F. Greenblatt, and B. D. Strahl. BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex. *Curr. Biol.*, 15(16):1487–93, Aug. 2005. 6.13

[121] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science*, 332(6028):475–8, Apr. 2011. 8

[122] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, et al. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science*, 298(5594):799–804, 2002. 2, 2.1, IV

[123] W. Lee, D. Tillo, N. Bray, R. H. Morse, R. W. Davis, T. R. Hughes, and C. Nislow. A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.*, 39(10):1235–1244, 2007. 2.1

[124] T. Lenasi, B. M. Peterlin, and M. Barboric. Cap-binding protein complex links pre-mRNA capping to transcription elongation and alternative splicing through positive transcription elongation factor b (P-TEFb). *J. Biol. Chem.*, 286(26):22758–68, July 2011. 7, 7.10

[125] J. Lewis and E. Izaurralde. The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.*, 247:461–469, 1997. 1.2.2, 7

[126] B. Li, M. Carey, and J. L. Workman. The role of chromatin during transcription. *Cell*, 128(4):707–719, 2007. 1.1.2

[127] D. D. Licatalosi, G. Geiger, M. Minet, S. Schroeder, K. Cilli, J. B. McNeil, and D. L. Bentley. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol. Cell*, 9(5):1101–11, May 2002. 1.2.4, 8, 9

[128] M. Lidschreiber, K. Leike, and P. Cramer. Cap completion and CTD kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. *Mol. Cell. Biol.*, 2013. 7, 7.3, 7.4, 17, 7.6, 7.8, 7.9

[129] S. Lin, G. Coutinho-Mansfield, D. Wang, S. Pandit, and X.-D. Fu. The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.*, 15(8):819–826, 2008. 1.3

[130] D. Lindstrom and S. Squazzo. Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell. Biol.*, 23(4):1368–1378, 2003. 1.1.2, 1.1.2, 1.3, 6.13, 7, 7.1, 7.10

[131] J. Liu, J. Zhang, Q. Gong, P. Xiong, H. Huang, B. Wu, G. Lu, J. Wu, and Y. Shi. Solution structure of tandem SH2 domains from Spt6 protein and their binding to the phosphorylated RNA polymerase II C-terminal domain. *J. Biol. Chem.*, 286(33):29218–26, Aug. 2011. 8.3

[132] Y. Liu, L. Warfield, C. Zhang, J. Luo, J. Allen, W. H. Lang, J. Ranish, K. M. Shokat, and S. Hahn. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol. Cell. Biol.*, 29(17):4852–63, Sept. 2009. 1.1.2, 6.13, 7.1, 7.2, 7.10, IV

[133] J. Logan, E. Falck-Pedersen, J. E. Darnell, and T. Shenk. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc. Natl. Acad. Sci.*, 84(23):8306–10, Dec. 1987. 1.2.4, 8.5

[134] H. Lu, O. Flores, R. Weinmann, and D. Reinberg. The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc. Natl. Acad. Sci.*, 88(22):10004–10008, 1991. 1.2.1

[135] S. Lu and B. R. Cullen. Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA*, 9(5):618–630, 2003. 1.2

[136] B. M. Lunde, S. L. Reichow, M. Kim, H. Suh, T. C. Leeper, F. Yang, H. Mutschler, S. Buratowski, A. Meinhart, and G. Varani. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.*, 17(10):1195–201, Oct. 2010. 8, 8.2.1, 8.3, 9

[137] D. S. Luse. Promoter clearance by RNA polymerase II. *Biochim. Biophys. Acta*, 2012. 1.1.1

[138] S. Lykke-Andersen and T. H. Jensen. Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie*, 89(10):1177–1182, 2007. 1.1.3

[139] S. S. Mandal, C. Chu, T. Wada, H. Handa, A. J. Shatkin, and D. Reinberg. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc. Natl. Acad. Sci.*, 101(20):7572–7, May 2004. 1.2, 1.3, 7, 7.10

[140] N. Marshall and D. Price. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.*, 270(21):12335–8, 1995. 1.2.1, 7, 7.10

[141] F. W. Martinez-Rucobo, S. Sainsbury, A. C. M. Cheung, and P. Cramer. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.*, 30(7):1302–10, Apr. 2011. 1, 1.1.1, 1.1.2, 6.13, 7, 7.10, IV

[142] P. B. Mason and K. Struhl. The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Mol. Cell. Biol.*, 23(22):8323–8333, 2003. 1.1.2

[143] T. N. Mavrich, I. P. Ioshikhes, B. J. Venters, C. Jiang, L. P. Tomsho, J. Qi, S. C. Schuster, I. Albert, and B. F. Pugh. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.*, 18(7):1073–83, July 2008. IV

[144] T. N. Mavrich, C. Jiang, I. P. Ioshikhes, X. Li, B. J. Venters, S. J. Zanton, L. P. Tomsho, J. Qi, R. L. Glaser, S. C. Schuster, et al. Nucleosome organization in the Drosophila genome. *Nature*, 453(7193):358–362, 2008. IV

[145] A. Mayer, M. Heidemann, M. Lidschreiber, A. Schreieck, M. Sun, C. Hintermair, E. Kremmer, D. Eick, and P. Cramer. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science*, 336(6089):1723–5, June 2012. 8, 8.3, IV, IV

[146] A. Mayer, M. Lidschreiber, M. Siebert, K. Leike, J. Söding, and P. Cramer. Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.*, 17(10):1272–8, Oct. 2010. 5.5.2, 5.7.6, 6, 6.1, 6, 6.6, 12, 6.11, 17

[147] A. Mayer, A. Schreieck, M. Lidschreiber, K. Leike, D. E. Martin, and P. Cramer. The Spt5 C-terminal region recruits yeast 3' RNA cleavage factor I. *Mol. Cell. Biol.*, 32(7):1321–31, Apr. 2012. 7.1, 8, 27, 8.2.2, IV, IV, IV

[148] C. Mazza, A. Segref, I. W. Mattaj, and S. Cusack. Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.*, 21(20):5548–57, Oct. 2002. 1.2.2, 7.6

[149] S. McCracken, N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, A. E. Program, S. Shuman, and D. L. Bentley. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.*, 11(24):3306–3318, Dec. 1997. 1.2.2, 7, 7.1, 7.4, 7.10, 9

[150] A. Meinhart and P. Cramer. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, 430(6996):223–6, July 2004. 2, 1.2.1, 1.2.4, 7.10, 8, 8.2.1, 9

[151] A. Meinhart, T. Kamenski, S. Hoeppner, S. Baumli, and P. Cramer. A structural perspective of CTD function. *Genes Dev.*, 19(12):1401–15, June 2005. 2, 1.2.1, 6

[152] A. Meinhart, T. Silberzahn, and P. Cramer. The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase. *J. Biol. Chem.*, 278(18):15917–21, May 2003. 8.4

[153] C. Miller, B. Schwalb, K. Maier, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken, D. E. Martin, A. Tresch, and P. Cramer. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Molecular systems biology*, 7, Jan. 2011. 2.2

[154] L. Minvielle-Sebastia, P. J. Preker, and W. Keller. RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. *Science*, 266(5191):1702–5, Dec. 1994. 1.2.4

[155] H. E. Mischo and N. J. Proudfoot. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim. Biophys. Acta*, 1829(1):174–85, Jan. 2013. 1.1.3, 1.2.4, 8.5

[156] A. L. Mosley, S. G. Pattenden, M. Carey, S. Venkatesh, J. M. Gilmore, L. Florens, J. L. Workman, and M. P. Washburn. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol. Cell*, 34(2):168–178, 2009. 1.2.1

[157] S. Moteki and D. Price. Functional coupling of capping and transcription of mRNA. *Mol. Cell*, 10(3):599–609, 2002. 7.10

[158] G. W. Muse, D. A. Gilchrist, S. Nechaev, R. Shah, J. S. Parker, S. F. Grissom, J. Zeitlinger, and K. Adelman. RNA polymerase is poised for activation across the genome. *Nat. Genet.*, 39(12):1507–11, Dec. 2007. 1.3, 2.1, IV

[159] U. Nagalakshmi, Z. Wang, K. Waern, and C. Shou. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881):1344–1349, 2008. I, 2.1.2, 2.2, 5.7.3, 5.7.5, 5.7.6, 6.1, 7.3, 17, 8.4

[160] S. Nechaev, D. C. Fargo, G. dos Santos, L. Liu, Y. Gao, and K. Adelman. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science*, 327(5963):335–8, Jan. 2010. 6.11

[161] E. Nedea, X. He, M. Kim, J. Pootoolal, G. Zhong, V. Canadien, T. Hughes, S. Buratowski, C. L. Moore, and J. Greenblatt. Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *J. Biol. Chem.*, 278(35):33000–10, Aug. 2003. 8.4

[162] C. Neuvéglise, C. Marck, and C. Gaillardin. The intronome of budding yeasts. *C. R. Biol.*, 334(8):662–670, 2011. 1.2.3

[163] Z. Ni, A. Saunders, N. J. Fuda, J. Yao, J.-R. Suarez, W. W. Webb, and J. T. Lis. P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo. *Mol. Cell. Biol.*, 28(3):1161–70, Feb. 2008. 1.2.1, 7, 7.10

[164] A. Nott, S. H. Meislin, and M. J. Moore. A quantitative analysis of intron effects on mammalian gene expression. *RNA*, 9(5):607–617, 2003. 1.2

[165] H. O'Geen, C. M. Nicolet, K. Blahnik, R. Green, and P. J. Farnham. Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques*, 41(5):577, 2006. 2.1.1, 5.6.2

[166] G. Orphanides, G. LeRoy, C.-H. Chang, D. S. Luse, and D. Reinberg. FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell*, 92(1):105–116, 1998. 1.1.2

[167] G. Orphanides and D. Reinberg. RNA polymerase II elongation through chromatin. *Nature*, 407(September):471–5, 2000. 1.1.1, 6, 7

[168] G. Orphanides and D. Reinberg. A unified theory of gene expression. *Cell*, 108:439–451, 2002. 1.1.1, 1.3, 6, 7, 7.10

[169] M. Pal, A. S. Ponticelli, and D. S. Luse. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell*, 19(1):101–10, July 2005. 1.1.1, 7

[170] C. Pancevac, D. C. Goldstone, A. Ramos, and I. a. Taylor. Structure of the Rna15 RRM-RNA complex reveals the molecular basis of GU specificity in transcriptional 3'-end processing factors. *Nucleic Acids Res.*, 38(9):3119–32, May 2010. IV

[171] P. J. Park. ChIP–seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, 10(10):669–680, 2009. 2.2

[172] Y. Pei, B. Schwer, and S. Shuman. Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. *J. Biol. Chem.*, 278(9):7180–8, Feb. 2003. 1.3, 7, 7.10

[173] Y. Pei and S. Shuman. Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.*, 277(22):19639–48, May 2002. 1.1.2, 1.3, 7, 7.1, 7.10

[174] R. Perales and D. Bentley. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell*, 36(2):178–91, Oct. 2009. 6, 7

[175] J. M. Pérez-Cañadillas. Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. *EMBO J.*, 25(13):3167–78, July 2006. IV

[176] D. Pokholok, N. Hannett, and R. Young. Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol. Cell*, 9:799–809, 2002. 1.1.1, 3, 6, 7, 8

[177] D. K. Pokholok, C. T. Harbison, S. Levine, M. Cole, N. M. Hannett, T. I. Lee, G. W. Bell, K. Walker, P. A. Rolfe, E. Herbolsheimer, et al. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, 122(4):517–527, 2005. 1.1.2, 2.1, 3

[178] D. Prather and N. Krogan. Identification and characterization of Elf1, a conserved transcription elongation factor in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, 25:10122–35, 2005. 6.13

[179] N. Proudfoot and G. Brownlee. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature*, 1976. 1.2.4

[180] N. J. Proudfoot. Ending the message: poly (A) signals then and now. *Genes Dev.*, 25(17):1770–1782, 2011. 1.2.4

[181] H. Qiu, C. Hu, N. A. Gaur, and A. G. Hinnebusch. Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *EMBO J.*, 31(16):3494–505, Aug. 2012. 7.10, IV

[182] H. Qiu, C. Hu, and A. G. Hinnebusch. Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol. Cell*, 33(6):752–62, Mar. 2009. 1.1.2, 1.2.1, 6.8, 6.13, 7, 7.8, 7.10, IV

[183] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0. 2.1.2, 5.7.1

[184] M. Radonjic, J.-C. Andrau, P. Lijnzaad, P. Kemmeren, T. T. J. P. Kockelkorn, D. van Leenen, N. L. van Berkum, and F. C. P. Holstege. Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon S. cerevisiae stationary phase exit. *Mol. Cell*, 18(2):171–83, Apr. 2005. 6.11

[185] P. B. Rahl, C. Y. Lin, A. C. Seila, R. A. Flynn, S. McCuine, C. B. Burge, P. A. Sharp, and R. A. Young. c-Myc regulates transcriptional pause release. *Cell*, 141(3):432–45, Apr. 2010. 6.11

[186] B. Ren, F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, et al. Genome-wide location and function of DNA binding proteins. *Science*, 290(5500):2306–2309, 2000. 2

[187] D. B. Renner, Y. Yamaguchi, T. Wada, H. Handa, and D. H. Price. A highly purified RNA polymerase II elongation control system. *J. Biol. Chem.*, 276(45):42601–9, Nov. 2001. 6.2

[188] H. S. Rhee and B. F. Pugh. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, 483(7389):295–301, Mar. 2012. 2.2, 7, 7.4, IV

[189] P. Richard and J. L. Manley. Transcription termination by nuclear RNA polymerases. *Genes & Development*, 23(11):1247–1269, June 2009. 1.1.3

[190] C. Rodriguez and E. Cho. Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol. Cell. Biol.*, 20:104–12, 2000. 6.9

[191] A. G. Rondón, H. E. Mischo, J. Kawauchi, and N. J. Proudfoot. Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. *Mol. Cell*, 36(1):88–98, Oct. 2009. 8.2.1

[192] A. E. Rougvie and J. T. Lis. The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell*, 54(6):795–804, 1988. 1.3

[193] M. Sadowski, B. Dichtl, W. Hübner, and W. Keller. Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *EMBO J.*, 22(9):2167–77, May 2003. 1.1.3

[194] S. Sainsbury, J. Niesser, and P. Cramer. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, 493(7432):437–40, Jan. 2013. 1.1.1, 7

[195] A. Saunders, L. J. Core, and J. T. Lis. Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol.*, 7(8):557–567, 2006. 1.1.2, 1

[196] A. Saunders, L. J. Core, C. Sutcliffe, J. T. Lis, and H. L. Ashe. Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. *Genes Dev.*, 27(10):1146–1158, 2013. 2.2

[197] A. Saunders, J. Werner, E. D. Andrulis, T. Nakayama, S. Hirose, D. Reinberg, and J. T. Lis. Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. *Science*, 301(5636):1094–1096, 2003. 1.1.2

[198] S. Schneider, Y. Pei, S. Shuman, and B. Schwer. Separable functions of the fission yeast Spt5 carboxyl-terminal domain (CTD) in capping enzyme binding and transcription elongation overlap with those of the RNA polymerase II CTD. *Mol. Cell. Biol.*, 30(10):2353–64, May 2010. 1.1.2, 1.3, 7, 7.1, 7.10

[199] S. C. Schroeder, B. Schwer, S. Shuman, and D. Bentley. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev.*, 14(19):2435–2440, 2000. 1.2.1, 1.2.2, 6, 6.7, 7, 7.1, 7.10, 8, 9

[200] J. M. Schulze, T. Hentrich, S. Nakanishi, A. Gupta, E. Emberly, A. Shilatifard, and M. S. Kobor. Splitting the task: Ubp8 and Ubp10 deubiquitinate different cellular pools of H2BK123. *Genes Dev.*, 25(21):2242–2247, 2011. 2.1.2, IV

[201] B. Schwer, X. Mao, and S. Shuman. Accelerated mRNA decay in conditional mutants of yeast mRNA capping enzyme. *Nucleic Acids Res.*, 26(9):2050–2057, 1998. 1.2.2, 7

[202] B. Schwer and S. Shuman. Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing in vivo. *RNA*, 2:574–583, 1996. 1.2, 1.2.2, 7

[203] A. Shatkin. Capping of eucaryotic mRNAs. *Cell*, 9(December):645–53, 1976. 1.2.2, 7

[204] S. Shuman. What messenger RNA capping tells us about eukaryotic evolution. *Nat. Rev. Mol. Cell Biol.*, 3(8):619–625, 2002. 4

[205] T. W. Sikorski and S. Buratowski. The basal initiation machinery: beyond the general transcription factors. *Current Opinion in Cell Biology*, 21(3):344–351, June 2009. 1.1.1, 1.3

[206] R. J. Sims, R. Belotserkovskaya, and D. Reinberg. Elongation by RNA polymerase II: the short and long of it. *Genes Dev.*, 18(20):2437–2468, 2004. 1.1.2, 1

[207] M. Smolle and J. L. Workman. Transcription-associated histone modifications & cryptic transcription. *Biochim. Biophys. Acta*, 2012. 1.1.2

[208] N. Sonenberg and A. Hinnebusch. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, 136(4):731–745, 2009. 1.2.2, 7

[209] C. V. St Amour, M. Sansó, C. A. Bösken, K. M. Lee, S. Larochelle, C. Zhang, K. M. Shokat, M. Geyer, and R. P. Fisher. Separate domains of fission yeast Cdk9 (P-TEFb) are required for capping enzyme recruitment and primed (Ser7-phosphorylated) Rpb1 carboxyl-terminal domain substrate recognition. *Mol. Cell. Biol.*, 32(13):2372–83, July 2012. 7, 9

[210] E. J. Steinmetz, N. K. Conrad, D. A. Brow, and J. L. Corden. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature*, 413(6853):327–31, Sept. 2001. 8.2.1

[211] K. Struhl. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.*, 14(2):103–105, 2007. I, IV

[212] T. Stuwe, M. Hothorn, E. Lejeune, V. Rybin, M. Bortfeld, K. Scheffzek, and A. G. Ladurner. The FACT Spt16 "peptidase" domain is a histone H3-H4 binding module. *Proc. Natl. Acad. Sci.*, 105(26):8884–9, July 2008. 6.13

[213] M. Sun, L. Larivière, S. Dengl, A. Mayer, and P. Cramer. A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). *J. Biol. Chem.*, 285(53):41597–603, Dec. 2010. 1.2.1, 7.8, 7.10, 8.3, IV

[214] M. Sun, B. Schwalb, D. Schulz, N. Pirkl, S. Etzold, L. Larivière, K. C. Maier, M. Seizl, A. Tresch, and P. Cramer. Comparative Dynamic Transcriptome Analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Research*, Mar. 2012. 2.2

[215] M. Swanson. SPT5, an essential gene important for normal transcription in Saccharomyces cerevisiae, encodes an acidic nuclear protein with a carboxy-terminal repeat. *Mol. Cell. Biol.*, 11:3009–19, 1991. 1.1.2

[216] M. Swanson and F. Winston. SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in Saccharomyces cerevisiae. *Genetics*, 132(2):325–36, 1992. 1.1.2, 1.1.2

[217] M. Thiebaut, E. Kisseleva-Romanova, M. Rougemaille, J. Boulay, and D. Libri. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol. Cell*, 23(6):853–864, 2006. IV

[218] J. R. Tietjen, D. W. Zhang, J. B. Rodríguez-Molina, B. E. White, M. S. Akhtar, M. Heidemann, X. Li, R. D. Chapman, K. Shokat, S. Keles, D. Eick, and A. Z. Ansari. Chemical-genomic dissection of the CTD code. *Nat. Struct. Mol. Biol.*, 17(9):1154–61, Sept. 2010. 1.2.1, IV

[219] J. Toedling and W. Huber. Analyzing ChIP-chip data using bioconductor. *PLoS Comput. Biol.*, 4(11):e1000227, 2008. 2.1.2

[220] J. Toedling, O. Sklyar, and W. Huber. Ringo–an R/Bioconductor package for analyzing ChIP-chip readouts. *BMC Bioinformatics*, 8(1):221, 2007. 2.1.2, 5.7.2

[221] B. N. Tomson and K. M. Arndt. The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. *Biochim. Biophys. Acta*, 2012. 1.1.2

[222] S. Vagner, C. Vagner, and I. W. Mattaj. The carboxyl terminus of vertebrate poly (A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev.*, 14(4):403–413, 2000. 1.2

[223] E. Van Dijk, C. Chen, Y. d'Aubenton Carafa, S. Gourvennec, M. Kwapisz, V. Roche, C. Bertrand, M. Silvain, P. Legoix-Ne, S. Loeillet, et al. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature*, 475(7354):114–117, 2011. I

[224] L. Vasiljeva, M. Kim, H. Mutschler, S. Buratowski, and A. Meinhart. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.*, 15(8):795–804, Aug. 2008. 1.2.1, 8.2.1, 8.3

[225] B. J. Venters and B. F. Pugh. A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome Res.*, 19(3):360–71, Mar. 2009. 2.1.2, 3, 6, IV

[226] B. J. Venters, S. Wachi, T. N. Mavrich, B. E. Andersen, P. Jena, A. J. Sinnamon, P. Jain, N. S. Rolleri, C. Jiang, C. Hemeryck-Walsh, and B. F. Pugh. A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. *Mol. Cell*, 41(4):480–92, Feb. 2011. IV

[227] T. Wada, T. Takagi, Y. Yamaguchi, A. Ferdous, T. Imai, S. Hirose, S. Sugimoto, K. Yano, G. A. Hartzog, F. Winston, et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.*, 12(3):343–356, 1998. 1.1.2

[228] W. Wei, V. Pelechano, A. I. Järvelin, and L. M. Steinmetz. Functional consequences of bidirectional promoters. *Trends Genet.*, 27(7):267–276, 2011. I, IV

[229] R. O. J. Weinzierl. *Mechanisms of Gene Expression: Structure, Function and Evolution of the Basal Transcriptional Machinery.* World Scientific Pub. Co. Inc., 1 edition, 8 1999. I

[230] Y. Wen and A. J. Shatkin. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.*, 13(14):1774–9, July 1999. 1.1.2, 1.3, 7, 7.1, 7.10

[231] F. Werner and D. Grohmann. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.*, 9(2):85–98, Feb. 2011. 1.1.2

[232] M. West and J. Corden. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics*, 140:1223–1233, 1995. 8.5

[233] S. West, N. Gromak, and N. J. Proudfoot. Human 5' -> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016):522–525, 2004. 1.1.3

[234] S. West, N. J. Proudfoot, and M. J. Dye. Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Mol. Cell*, 29(5):600–610, 2008. 1.1.3

[235] C. L. Will and R. Lührmann. Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology*, 3(7), 2011. 1.2.3

[236] F. Winston, D. T. Chaleff, B. Valent, and G. R. Fink. Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. *Genetics*, 107(2):179–97, June 1984. 1.1.2, 1.1.2

[237] F. Winston and P. Sudarsanam. The SAGA of Spt proteins and transcriptional analysis in yeast: past, present, and future. In *Cold Spring Harbor symposia on quantitative biology*, volume 63, pages 553–562. Cold Spring Harbor Laboratory Press, 1998. 1.1.2

[238] C. Wittenberg and S. I. Reed. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. *Oncogene*, 24(17):2746–2755, 2005. 1.3

[239] C.-M. Wong, H. Qiu, C. Hu, J. Dong, and A. G. Hinnebusch. Yeast cap binding complex impedes recruitment of cleavage factor IA to weak termination sites. *Mol. Cell. Biol.*, 27(18):6520–31, Sept. 2007. 1.2.2, 7.7, 7.10, IV

[240] K. Xiang, J. L. Manley, and L. Tong. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. *Nat. Commun.*, 3:946, 2012. 1.2.1

[241] K. Xiang, T. Nagaike, S. Xiang, T. Kilic, M. M. Beh, J. L. Manley, and L. Tong. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature*, 467(7316):729–33, Oct. 2010. 8.4

[242] Z. Xu, W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Münster, J. Camblong, E. Guffanti, F. Stutz, W. Huber, and L. M. Steinmetz. Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 457(7232):1033–7, Feb. 2009. I, 5.7.6, 6.1

[243] Y. Yamaguchi, H. Shibata, and H. Handa. Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochim. Biophys. Acta*, 2012. 1.3

[244] Y. Yamaguchi, T. Takagi, T. Wada, K. Yano, A. Furuya, S. Sugimoto, J. Hasegawa, and H. Handa. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, 97(1):41–51, 1999. 1.3, 7

[245] S. M. Yoh, H. Cho, L. Pickle, R. M. Evans, and K. A. Jones. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev.*, 21(2):160–74, Jan. 2007. 6.10, 6.13

[246] G.-C. Yuan, Y.-J. Liu, M. F. Dion, M. D. Slack, L. F. Wu, S. J. Altschuler, and O. J. Rando. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science*, 309(5734):626–30, July 2005. 2.1, 12

[247] N. Yudkovsky, J. A. Ranish, and S. Hahn. A transcription reinitiation intermediate that is stabilized by activator. *Nature*, 408(6809):225–229, Nov. 2000. 1.1.3

[248] B. Zacher, P. Kuan, and A. Tresch. Starr: Simple Tiling ARRay analysis of Affymetrix ChIP-chip data. *BMC Bioinformatics*, 11(1):194+, 2010. 2.1.2, 5.7.1

[249] J. Zeitlinger, A. Stark, M. Kellis, J.-W. Hong, S. Nechaev, K. Adelman, M. Levine, and R. A. Young. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat. Genet.*, 39(12):1512–6, Dec. 2007. 1.3, 2.1, 6.11, IV

[250] D. W. Zhang, J. B. Rodríguez-Molina, J. R. Tietjen, C. M. Nemec, and A. Z. Ansari. Emerging views on the CTD code. *Genet. Res. Int.*, 2012:347214, Jan. 2012. 1.1.2, 1.2, 1.2.1, 3

[251] Z. Zhang, J. Fu, and D. S. Gilmour. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev.*, 19(13):1572–80, July 2005. 1.2.4, 8, 9

[252] Z. Zhang and D. S. Gilmour. Pcf11 Is a Termination Factor in Drosophila that Dismantles the Elongation Complex by Bridging the CTD of RNA Polymerase II to the Nascent Transcript. *Mol. Cell*, 21(1):65–74, 2006. 1.2.4

[253] M. Zheng, L. O. Barrera, B. Ren, and Y. N. Wu. ChIP-chip: Data, Model, and Analysis. *Biometrics*, 63(3):787–796, 2007. 2.1.2

[254] K. Zhou, W. H. W. Kuo, J. Fillingham, and J. F. Greenblatt. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc. Natl. Acad. Sci.*, 106(17):6956–61, Apr. 2009. 1.1.2, 6.13, 7.2, IV

List of Figures

List of Tables

Publications

2013 **Cap completion and CTD kinase recruitment underlie the initiation-elongation transition of RNA polymerase II**
M. Lidschreiber, K. Leike, and P. Cramer
Molecular and Cellular Biology 13 (Epub ahead of print, July 22).

2013 **RNA polymerase II termination involves CTD tyrosine dephosphorylation by CPF subunit Glc7**
A. Schreieck*, A. Easter*, S. Etzold, K. Wiederhold, **M. Lidschreiber**, P. Cramer, and L. Passmore (* contributed equally)
Manuscript submitted.

2013 **Reciprocal regulation of protein synthesis and carbon metabolism for thylakoid membrane biogenesis**
A. Bohne, C. Schwarz, M. Schottkowski, **M. Lidschreiber**, M. Piotrowski, W. Zerges, and J. Nickelsen
PLoS Biology 11(2):e1001482

2012 **Real-time transcriptional profiling of cellular and viral gene expression during lytic cytomegalovirus infection**
L. Marcinowski, **M. Lidschreiber**, L. Windhager, M. Rieder, B. Bosse, B. Rädle, T. Bonfert, I. György, M. de Graaf, O. Prazeres da Costa, P. Rosenstiel, C. Friedel, R. Zimmer, Z. Ruzsics, and L. Dölken
PLoS Pathogens 8(9):e1002908

2012 **CTD tyrosine phosphorylation impairs termination factor recruitment to transcribing RNA polymerase II**
A. Mayer*, M. Heidemann*, **M. Lidschreiber**, A. Schreieck, M. Sun, C. Hintermair, E. Kremmer, D. Eick, and P. Cramer (* contributed equally)
Science 336, 1723.

2012 **MC EMiNEM maps the interaction landscape of the Mediator**
T. Niederberger, S. Etzold, **M. Lidschreiber**, K. Maier, D. Martin, H. Fröhlich, P. Cramer, and A. Tresch
PLoS Computational Biology 8(6):e1002568

2012 **The Spt5 C-terminal region recruits yeast 3´-RNA cleavage factor I**
A. Mayer*, A. Schreieck*, **M. Lidschreiber**, K. Leike, D. Martin, and P. Cramer (* contributed equally)
Molecular and Cellular Biology 32, 1321.

2010 **Uniform transitions of the general RNA polymerase II transcription complex**
A. Mayer*, **M. Lidschreiber***, M. Siebert*, K. Leike, S. Söding, and P. Cramer (* contributed equally)
Nature Structural and Molecular Biology 17, 1272.