

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

RNA-dependent chromatin
association of transcription elongation
factors and Pol II CTD kinases

Sofia Luciana Battaglia

aus

San Pedro, Argentinien

2017

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 07.04.2017

Sofia Battaglia

Dissertation eingereicht am 12.04.2017

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

Mündliche Prüfung am 29.05.2017

Acknowledgments

First of all, I would like to thank my supervisor Professor Patrick Cramer for giving me the opportunity to work in his group. He gave me freedom to explore the field, but was also present for discussion and troubleshooting any time when needed. I thank him for all the support, advice and positivity during the last four years. His trust was essential for my development as a scientist.

I am very thankful to have worked with Michael Lidschreiber, who was an amazing collaborator. He is a smart co-worker and a good friend now. He was very calm even when the science was not doing what we expected. Finally, we managed to have a super nice story!

I thank Margaux Michel for troubleshooting together during the optimization of the ChIP protocol and many other general discussions. I also thank her for her wonderful atmosphere. Seychelle Vos helped me with one very important and interesting experiment for this thesis. She gave me a lot of smart advices and was always very motivated with our work. I also thank Amelie Schreieck for sharing her expertise in several techniques at the beginning of my PhD. I also want to give Kerstin Maier a big thank you for constant discussions and advise as a desk colleague and an expert in the yeast field. Petra Rus was always very nice and gave our little yeast lab the perfect atmosphere to work in a good mood. She was also very flexible to help me with experiments when needed. Mark Böhning was a perfect join in the lab, we often stayed late discussing about our projects and trying to help one another. I thank Carlo Baejen for teaching me the PAR-CLIP protocol. Christian Roth was very patient showing me how to deal with sequencing data and perform basic analysis myself. In this regard, I also want to thank Johannes Söding for discussions about PAR-CLIP data.

I also would like to specially thank Luis Fonseca, Marc Böhning, Anna Sawicka, Kerstin Maier, Michael Lidschreiber and Quinn Klinge for critical reading of parts of this thesis. Special thanks to my committee members PD Dr. Dietmar Martin, Professor Peter Becker, Professor Andreas Ladurner, Dr. Fabiana Perocchi and Professor Klaus Förstemann for their time. And of course a big thanks to all the members of the Cramer group. It was an enormous pleasure to work with such a good team.

I deeply thank Luis Fonseca, my best friend. Without his company and support this thesis would have been a lot harder for me. His good mood and positivity were a big help for me in difficult times. It was also very productive and fun having science discussions with him.

Many thanks to my dear girl friends for all the great moments together. Spending time with them was a way to recharge my energies. I have been very blessed in my life to have their love and friendship. I thank my siblings, Alba y Francisco, for being part of my family and sharing so many wonderful things in this life. I love laughing with them.

And my very special thanks go to my parents, Irene and Horacio, that even being far away they were always close and supportive any day and any time I needed. I thank them for the person who I am today. Gracias. I also thank my lovely Oma. I am sure she would be celebrating this moment with me.

Summary

For transcription through chromatin, RNA polymerase (Pol) II associates with transcription factors. Recent work revealed that transcription factors also interact with the nascent RNA to regulate gene expression. The focus of this thesis relies on the characterization of this form of interactions during the process of transcription elongation. I used our recently optimized PAR-CLIP protocol to show that many elongation factors (EFs) crosslink to RNA emerging from transcribing Pol II in the yeast *Saccharomyces cerevisiae*. These *in vivo* direct interactions were most notable for the kinases Ctk1 and Bur1 that phosphorylate the C-terminal repeat domain (CTD) of the largest Pol II subunit and for the histone H3 methyltransferases Set1 and Set2. Bioinformatic analysis indicated that most EFs crosslink preferentially to mRNAs, rather than to unstable non-coding RNAs, consistent with their recruitment to transcribed protein-coding genes. Furthermore, I developed an RNA degradation assay to test whether the observed RNA-protein interactions affect protein-chromatin binding under native conditions. Comparing protein-chromatin binding in the presence and absence of RNA revealed that RNA contributes to chromatin association in particular of the CTD serine 2 kinases Ctk1 and Bur1 and the histone H3 methyltransferases Set1, Set2 and Dot1. Additionally, I confirmed the *in vivo* observed EF-RNA interactions for an active CTDK-I kinase complex *in vitro* using fluorescence anisotropy. Finally, I optimized our ChIP protocol for high-throughput sequencing and performed ChIP-Seq experiments of most Pol II EFs and histone marks involved in this study. Comparison of factor occupancies on DNA (ChIP-Seq) and on RNA (PAR-CLIP) revealed that interactions of EFs with nascent RNA are established before EFs are recruited to chromatin. Taken together, these studies argue for a role of nascent RNA in EF recruitment. In this model, EF-RNA interactions facilitate assembly of the elongation complex on transcribed genes when RNA emerges from Pol II, and loss of EF-RNA interactions upon RNA cleavage at the polyadenylation site trigger disassembly of the elongation complex.

Publications

Part of this work has been published or is in the process of being published.

2017 **RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases**

Sofia Battaglia^{*}, Michael Lidschreiber^{*}, Carlo Baejen, Phillipp Torkler, Seychelle M. Vos, Patrick Cramer (^{*} joint first authorship).

Under revision.

Author contributions: **SB** and **CB** performed **PAR-CLIP** experiments. **SB** and **SMV** purified recombinant **CTDK-I** and performed *in vitro* assays. **SB** performed **ChIP-Seq** and **chromatin association** experiments. **ML** and **PT** carried out PAR-CLIP data analysis. **ML** carried out PAR-CLIP normalization and ChIP-Seq data analysis. **PC** supervised research. **SB**, **ML** and **PC** wrote the manuscript.

2017 **Genome-wide analysis of RNA polymerase II termination at protein-coding genes**

Carlo Baejen, Jessica Andreani, Phillipp Torkler, **Sofia Battaglia**, Bjoern Schwalb, Michael Lidschreiber, Kerstin C. Maier, Andrea Boltendahl, Petra Rus, Stephanie Esslinger, Johannes Söding, and Patrick Cramer

Molecular Cell

Author contributions: **CB** and **PC** designed the study. **CB** planned and coordinated experiments. **KCM**, **SE** and **PR** cloned strains and performed anchor-away and 4tU-Seq experiments. **KCM**, **PR** and **CB** carried out growth rate experiments. **AB** and **CB** performed PAR-CLIP experiments. **SB** performed **ChIP-Seq** and **ChIP-qPCR** experiments. **KCM** performed ChIP-qPCR experiments. **JA**, **PT**, **BS**, **JS**, **CB** and **PC** planned data analyses. **JA** and **BS** carried out 4tU-Seq data analyses. **JA** carried out ChIP-Seq data analyses. **PT**, **CB** and **ML** carried out PAR-CLIP data analyses. **JS** supervised data analysis and modelling. **PC** supervised research. **CB** and **PC** wrote the manuscript with input from all authors.

2017 **Intergenic non-coding transcription utilises an RNA-based mechanism to drive transcriptional interference**

Beth R. Watts, Sina Wittmann, **Sofia Battaglia**, Cornelia Kilchert, Dong-hyuk Heo, Patrick Cramer, and Lidia Vasiljeva

Under revision.

Author contributions: **BRW** and **LV** conceived and designed experiments. **BRW** performed all experiments except: **PAR-CLIP (SB)**, Mmi1 purification (**D-HH**), analysis of ChIP-seq, RNA-seq and **PAR-CLIP analyses (SB** and **SW**). **CK** and **SW** helped with generating strains. **BRW** and **LV** wrote the paper and all authors edited the manuscript.

Additional ongoing collaborations:

2017 The stress responsive kinase Hog1 binds mRNA in response to osmotic stress

Research group of Francesc Posas (UPF Departament de Ciències Experimentals i de la Salut, Barcelona, Spain.

Manuscript in preparation.

My contributions: **SB performed conditional PAR-CLIP after salt stress**

Table of Contents

Erklärung	3
Eidesstattliche Versicherung	3
Acknowledgments	5
Summary	7
Publications	9
1 Introduction	13
1.1 RNA synthesis by RNA polymerases	13
1.2 The RNA polymerase II transcription cycle	14
1.2.1 Transcription initiation and promoter clearance	16
1.2.2 Transcription elongation.....	17
1.2.3 Transcription termination and re-initiation	18
1.3 Regulation of transcription during elongation	19
1.3.1 Elongation factors in <i>Saccharomyces cerevisiae</i>	20
1.3.2 Elongation factor recruitment.....	22
1.4 Nascent RNA as a transcription regulator	23
1.5 In vivo methods for detection of RNA-protein interactions	24
1.5.1 RIP methods	25
1.5.2 CLIP methods.....	25
1.6 Aims and scope of this thesis	27
2 Materials and Methods	29
2.1 Materials	29
2.1.1 Bacterial and yeast strains.....	29
2.1.2 Cell lines	30
2.1.3 Growth media and media additives	30
2.1.4 Oligonucleotide sequences	31
2.1.5 Antibodies	33
2.1.6 Plasmids	34
2.1.7 Buffers and solutions.....	34
2.2 Methods	35
2.2.1 <i>S. cerevisiae</i> strain validation	35
2.2.2 General DNA methods.....	36
2.2.3 General protein methods.....	36

2.2.4	PAR-CLIP.....	37
2.2.5	ChIP-Seq.....	43
2.2.6	Chromatin association assay.....	46
2.2.7	Purification of recombinant CTDK-I protein complex.....	46
2.2.8	CTDK-I kinase activity assays.....	49
2.2.9	Fluorescence anisotropy assays with CTDK-I.....	50
3	Results.....	53
3.1	Elongation factors directly crosslink to RNA <i>in vivo</i>.....	53
3.2	Comparisons of PAR-CLIP data require normalization.....	56
3.3	Differences in EF occupancy along RNAs.....	58
3.3.1	EF localization along mRNA transcripts.....	58
3.3.2	EFs bind nascent pre-mRNA.....	60
3.3.3	Most EFs preferentially interact with coding transcripts.....	62
3.4	Chromatin association of EFs depends on RNA.....	66
3.5	Ctk1 kinase complex binds RNA <i>in vitro</i>.....	67
3.6	Evidence that RNA contributes to EF recruitment.....	69
4	Discussion.....	73
5	Future perspectives.....	77
6	Additional unpublished data.....	81
6.1	Investigation of novel factors in chromatin transcription.....	81
6.2	Genome-wide occupancy profiles of Pol II CTD phosphorylation marks..	87
	References.....	89
	Abbreviations.....	107

1 Introduction

Transcription is a fundamental process in living cells and part of the central dogma of molecular biology, which states that genetic information encoded in DNA is transcribed to RNA, and RNA is translated to protein (Crick 1970, Crick 1958). Transcription is the first step in gene expression, and malfunctions within this process have devastating pleiotropic effects in all living systems. In humans, for example, most tumorigenic processes that lead to different types of cancer are associated with defects in transcription machinery (Villard 2004). Therefore, understanding this system is of paramount concern for the scientific community.

1.1 RNA synthesis by RNA polymerases

Multisubunit DNA-dependent RNA Polymerase enzymes catalyze the transcription of DNA into RNA in all kingdoms of life (reviewed in (Thomas and Chiang 2006)). RNA polymerases developed early in evolution, and their 5 subunits core structure is conserved from bacteria to humans (Ebright 2000), deeply emphasizing the importance of this process during evolution and development. Since the year 2000, many high-resolution structures of bacterial and eukaryotic RNA polymerases and complexes of RNA polymerases with nucleic acid scaffolds or basal transcription factors have been obtained, providing insights into the molecular mechanisms that govern transcription (Armache, Kettenberger, and Cramer 2003, Armache et al. 2005, Bushnell and Kornberg 2003, Bushnell et al. 2004, Cramer, Bushnell, and Kornberg 2001, Wang et al. 2006, Westover, Bushnell, and Kornberg 2004). More recently, also structures of the archaeal RNA polymerase were solved (Hirata, Klein, and Murakami 2008, Kusser et al. 2008).

A single RNA polymerase performs prokaryotic transcription. In contrast, RNA synthesis in eukaryotic organisms is mainly performed by three distinct RNA polymerases (Pol), namely Pol I, Pol II and Pol III in the nucleus of cells (Kusser et al. 2008). Pol I transcribes the ribosomal RNA (rRNA) precursor (35S rRNA in yeast; 45S rRNA in human) (Moss et al. 2007, Venema and Tollervey 1999), and Pol III produces short untranslated RNAs such as transfer RNAs (tRNAs) and 5S ribosomal RNA (rRNA) (White 2011). Pol II carries out transcription of all protein-coding genes

to produce messenger RNAs (mRNAs), as well as several non-coding RNAs, including small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and cryptic unstable transcripts (CUTs) (Barrandon, Spiluttini, and Bensaude 2008, Kusser et al. 2008, Xu et al. 2009). Pol I, Pol II and Pol III consist of a structurally conserved 10-subunit core and specific additional subunits located on the periphery (Kusser et al. 2008).

Eukaryotes have an additional, single-subunit RNA polymerase that transcribes the mitochondrial genome (mitochondrial RNA polymerase or mitoPol). The mitoPol, which is distantly related to the RNA polymerase of the bacteriophage T7, produces polycistronic transcripts that are subsequently cleaved to generate individual mRNAs, tRNAs and rRNAs (Asin-Cayuela and Gustafsson 2007, Mercer et al. 2011, Schwinghammer et al. 2013). Furthermore, two non-essential RNA polymerases have been recently identified in plants: Pol IV and Pol V. These polymerases generate non-coding RNAs with specific functions in RNA-directed DNA methylation and transcriptional silencing (Pikaard et al. 2008).

1.2 The RNA polymerase II transcription cycle

The transcription cycle consists of three well-defined stages: initiation, elongation and termination (Figure 1). During these stages, the 12-subunit RNA polymerase (Pol) II enzyme is highly regulated (Thomas and Chiang 2006). Although all eukaryotic RNA Polymerases and the prokaryotic RNA polymerase are homolog to each other, Pol II possesses a peculiar C-terminal extension on its largest subunit (Rpb1), referred to as the C-terminal domain (CTD) (Hsin and Manley 2012). The CTD is composed of tandem heptad repeats that constitutes a unique feature of Pol II and distinguishes it from all other polymerases. It is an unstructured domain consisting of 26 (yeast), 42 (*Drosophila*) and 52 (mammals) heptapeptide repeats of the consensus sequence: tyrosine-serine-proline-threonine-serine-proline-serine (YSPTSPS) (Chapman et al. 2008). The amino acids in the CTD heptapeptide can be phosphorylated at serines (positions 2, 5 and 7: Ser2-P, Ser5-P and Ser7-P), at threonine (position 4: Thr4-P) and at tyrosine (position 1: Tyr1-P) (Jeronimo, Bataille, and Robert 2013). Furthermore, both prolines (Pro3 and Pro6) can be found in either *cis* or *trans* conformation (Eick and Geyer 2013). In higher eukaryotes, many of the heptad

repeats diverge from the consensus sequence and are acetylated at position 7 on lysine (Lys7) (Schroder et al. 2013). These posttranslational modifications in the CTD and the CTD in its entirety play a key role in transcription regulation throughout the transcription process as described below.

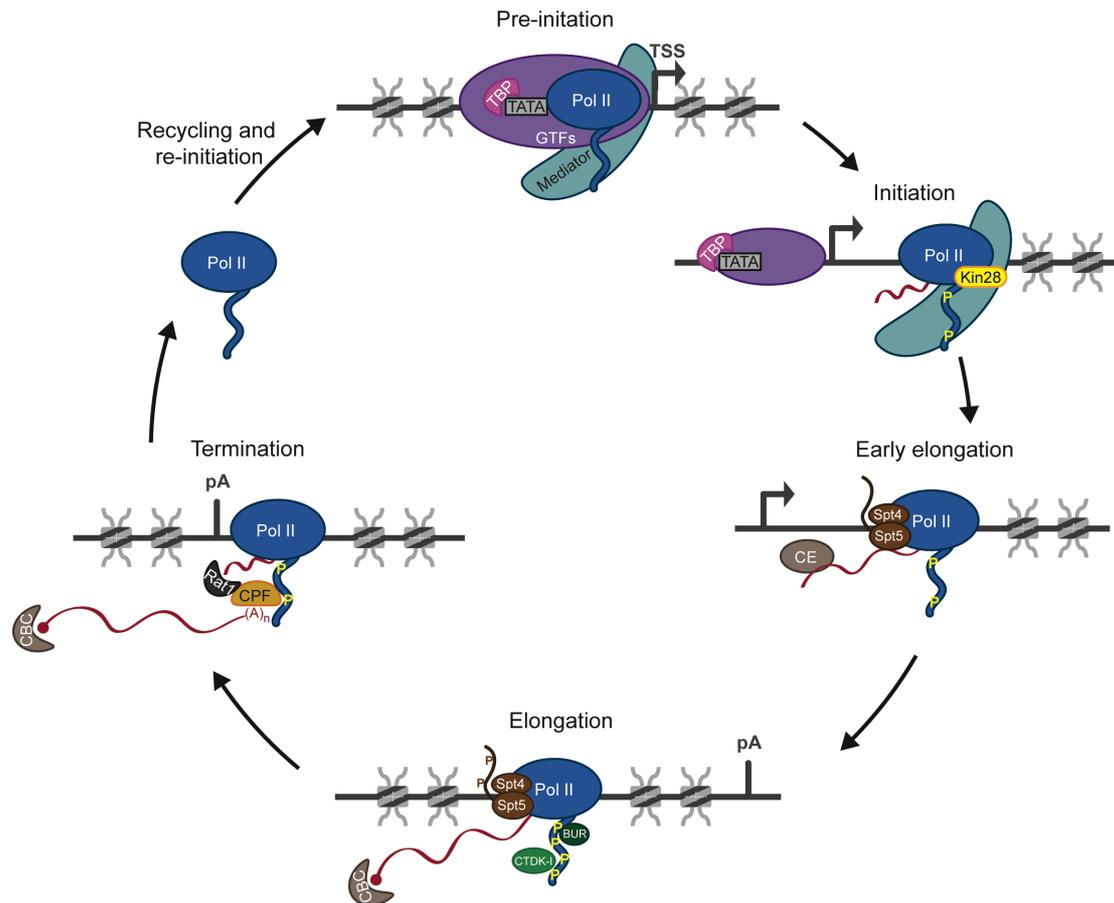


Figure 1 Simplified overview of the Pol II transcription cycle in the yeast *Saccharomyces cerevisiae*. Pre-initiation: Pol II binds to the promoter region (TATA) via general transcription factors (GTFs), TATA box binding protein (TBP), and Mediator to form the preinitiation complex (PIC). Initiation: Kin28 (a subunit of the general transcription factor TFIIF) phosphorylates the C-terminal domain (CTD) of Pol II on Ser5 and Ser7 leading to promoter clearance. Early elongation: capping enzymes (CE) and the complex containing the transcription elongation factors Spt4 and Spt5 are recruited to Pol II. Elongation: the cap binding complex (CBC) binds the capped nascent RNA (red); the kinase complexes CTDK-I and BUR further phosphorylate the CTD of Pol II on Ser2 and the C-terminal repeat (CTR) of Spt5; the CTD gets partially dephosphorylated at Ser5. Termination: the cleavage and polyadenylation factor (CPF) is recruited to the CTD, the RNA is cleaved and polyadenylated at its 3' end, followed by degradation of the downstream RNA by Rat1 and release of Pol II from the template. Recycling and re-initiation: the CTD is completely dephosphorylated and the free polymerase can initiate another round of transcription. For simplicity, all CTD phosphorylation sites are depicted in yellow. TSS: transcription start site; pA: polyadenylation site.

The transcribing DNA is packed in form of chromatin, which enables activators and repressors of transcription to access specific regions in a correct temporal and spatial manner (reviewed in (Smolle and Workman 2013)). The histone units of chromatin are important targets for modifications that act as a recruitment platform for proteins involved in the synthesis of mRNAs. Some of the writers of active histone modifications are part of this work and will be discussed below.

1.2.1 Transcription initiation and promoter clearance

For initiation of eukaryotic gene transcription, general transcription factors (TFII) and Pol II assemble in a stepwise manner at promoter DNA to form the preinitiation complex (PIC). General transcription factors in eukaryotes include TFIIA, TFIIB, TFIID, TFIIE, TFIIIF and TFIIH (Hahn 2004). Promoters are sequence specific regions on the DNA that are located upstream of a transcribed region. Many Pol II promoters contain a TATA element (or TATA box) that is recognized by the TATA box-binding protein (TBP), a subunit of TFIID (Davison et al. 1983, Nakajima, Horikoshi, and Roeder 1988, Parker and Topol 1984, Pelham 1982). In multicellular organisms, cis-regulatory DNA regions called enhancers act in gene activation. Enhancers increase transcription independently of their orientation, position and distance to a promoter (Banerji, Rusconi, and Schaffner 1981). Enhancers are prevalently found in higher eukaryotes and are beyond the scope of this thesis (for a detailed review see: (Andersson et al. 2014, Li, Notani, and Rosenfeld 2016)).

The PIC comprises closed, double-stranded promoter DNA that is first unwound by the ATPase/helicase subunit of TFIIH (Ssl2 in yeast; XBP in human) to form an open complex enabling transcription to take place (Fishburn, Galburt, and Hahn 2016). In the catalytically active open complex, a single-stranded DNA “transcription bubble” is formed from -9 to -2 position relative to the transcription start site (TSS) (Holstege, Fiedler, and Timmers 1997, Wang, Carey, and Gralla 1992). Transcription begins at the $+1$ position on the TSS with the formation of the first phosphodiester bond. At this stage, transcription often last only up to the $+8/+10$ position resulting in the production and release of numerous abortive short RNAs (Holstege, Fiedler, and Timmers 1997). Stability in the synthesis of the transcript is obtained when Pol II reaches the $+9$ position (Margaritis and Holstege 2008). General

transcription factors dissociate from Pol II, and elongation factors (EFs) are recruited (Orphanides and Reinberg 2000, 2002, Pokholok, Hannett, and Young 2002). Pol II escapes the abortive cycle of RNA synthesis and is capable of productive RNA synthesis. This transition of Pol II from an initiation complex to an elongation complex is known as “promoter clearance” or “promoter escape” (Luse 2013, Pal, Ponticelli, and Luse 2005).

1.2.2 Transcription elongation

For promoter clearance and productive mRNA elongation, the kinase subunit of the multisubunit factor TFIIF (Kin28 in yeast; CDK7 in human) phosphorylates the CTD at Ser5 and Ser7 positions leading to the recruitment of RNA 5' capping enzymes (Hsin and Manley 2012, McCracken et al. 1997). It is worth mentioning that the Mediator complex, a transcriptional coactivator, stimulates the kinase activity of TFIIF (Plaschka et al. 2015, Sogaard and Svejstrup 2007). Capping enzymes (Cet1, Ceg1, and Abd1 in yeast) catalyze the formation of a 7-methyl-guanosine (m7G) cap on the 5' end of the growing mRNA (Wei and Moss 1977, Wei and Moss 1975). The 5' cap prevents degradation of the RNA by 5' exonucleases and promotes translation during protein synthesis (Schwer, Mao, and Shuman 1998, Sonenberg and Hinnebusch 2009). It also associates with the cap-binding complex (CBC), which functions in pre-mRNA splicing and mRNA export (Lewis and Izaurralde 1997, Schwer and Shuman 1996).

During the past decade, genome-wide sequencing studies have detected high occupancy levels of Pol II during early steps of elongation (Adelman and Lis 2012, Kwak et al. 2013, Rahl et al. 2010). This phenomenon, primarily observed in *Drosophila* and humans, is currently subject of intensive research and refers to the promoter-proximal pausing of Pol II. Pol II pauses 30–60 nucleotides downstream of the TSS and is released via phosphorylation of negative elongation factor (NELF) and DRB-sensitivity-inducing factor (DISF) by the CDK9 kinase subunit of positive transcription elongation factor-b (P-TEFb) (Nechaev and Adelman 2011, Peterlin and Price 2006, Zhou, Li, and Price 2012). Several studies showed that pausing and release of Pol II at promoter-proximal regions is a key step in transcription regulation (reviewed in (Jonkers and Lis 2015)). CDK9 (yeast Bur1) also phosphorylates the

CTD of Pol II at Ser2 position (Marshall et al. 1996, Ramanathan et al. 2001) and triggers productive elongation (Cheng and Price 2007, Marshall and Price 1995, Ni et al. 2008). Phosphorylation of Pol II CTD on the Ser2 position by Bur1 (human CDK9) and Ctk1 (human CDK12) helps to recruit factors important for transcription elongation, termination and pre-mRNA splicing as well as histone modifiers and remodelers (Bartkowiak et al. 2010, Henikoff 2008, Smolle and Workman 2013).

During transcription elongation, the spliceosome is recruited to the splice sites of transcripts containing introns. Intron sequences of pre-mRNAs are removed before transcripts are exported to the cytoplasm and translated into proteins. The splicing machinery can either catalyze the splicing reactions during transcription (co-transcriptional splicing) or immediately after transcription termination (Carrillo Oesterreich et al. 2016). Despite the fact that the *S. cerevisiae* genome only contains about 4% of genes with introns, they represent an important aspect of gene expression (Qin et al. 2016).

1.2.3 Transcription termination and re-initiation

At the end of genes, Pol II discontinues elongation of the RNA product and releases the DNA template, entering the transcription termination process. Termination involves endonucleolytic cleavage of the nascent RNA followed by synthesis of a poly-A tail on the new 3' end of the cleaved product by the poly-A polymerase Pap1 (Baejen et al. 2017, Richard and Manley 2009).

The release of DNA and mRNA at the end of transcripts requires an exchange of Pol II associating factors. When Pol II transcribes over the polyadenylation (pA) site –which marks the end of protein-coding genes– the nascent RNA displays the pA sequence. This signal then leads to the recruitment of the protein machinery for pre-mRNA cleavage and polyadenylation (CPF and CF complexes), formally promoting the termination of transcription (Keller and Minvielle-Sebastia 1997, Manley and Takagaki 1996, Mischo and Proudfoot 2013, Porrua and Libri 2015, Proudfoot 2011). Proteins involved in these processes are recruited through binding to the phosphorylated CTD of Pol II at Ser2 position (Komarnitsky, Cho, and Buratowski 2000, Mayer, Heidemann, et al. 2012, Meinhart and Cramer 2004), or to the C-

terminal region (CTR) of the Pol II-associated general elongation factor Spt5 (Mayer, Schrieck, et al. 2012).

After assembly of the 3'-processing machinery, the RNA is cleaved and the resulting RNA 3'-end is poly-adenylated to complete the mRNA transcript. The new 5'-end of the nascent RNA is unprotected and prone to degradation by the exonuclease Rat1/XRN2 (yeast/human). The exonuclease degrades nascent RNA to chase the transcribing polymerase and promote its release from DNA and RNA (Connelly and Manley 1988, Proudfoot 1989). This termination mechanism is now commonly referred to as 'torpedo' termination because the exonuclease acts like a torpedo to dismantle the Pol II elongation complex (Baejen et al. 2017, Luo and Bentley 2004).

Despite extensive research, it is not clear where in the yeast genome termination occurs and whether Pol II termination at protein-coding genes is generally achieved by the torpedo mechanism. In one of our recent studies, we used various functional genomics techniques that allowed to distinguish Pol II termination from RNA degradation downstream of the pA site, and to provide evidence that Rat1 is generally required for Pol II termination at the end of protein-coding genes (Baejen et al. 2017).

Finally, remaining of a subset of general transcription factors and the Mediator complex at promoters facilitates transcription re-initiation of genes and following rounds of Pol II transcription (Yudkovsky, Ranish, and Hahn 2000). Additionally, the formation of gene loops brings the promoter and the 3' end of genes in physical proximity and helps re-initiation (Hampsey et al. 2011).

1.3 Regulation of transcription during elongation

For productive transcription through chromatin, RNA polymerase (Pol) II associates with general elongation factors (EFs) (Perales and Bentley 2009, Shilatifard 2004, Shilatifard, Conaway, and Conaway 2003, Sims, Belotserkovskaya, and Reinberg 2004) that are recruited to the body of transcribed genes in yeast (Mayer et al. 2010).

1.3.1 Elongation factors in *Saccharomyces cerevisiae*

EFs in the yeast *Saccharomyces cerevisiae* include Spt5 (a subunit of human DSIF), the histone chaperone Spt6, and the Paf1 complex (Paf1C). The Pol II CTD kinases Bur1 (human CDK9) and Ctk1 (human CDK12), and their cyclin partners Bur2 and Ctk2, respectively, can also be classified as EFs. In addition, the histone methyltransferases Set1 (a subunit of the COMPASS complex), Set2, and Dot1, are recruited to elongating Pol II to set the ‘active’ histone marks H3K4me3, H3K36me3, and H3K79me3, respectively. A simplified schematic representation of the active transcription elongation phase is illustrated in Figure 2.

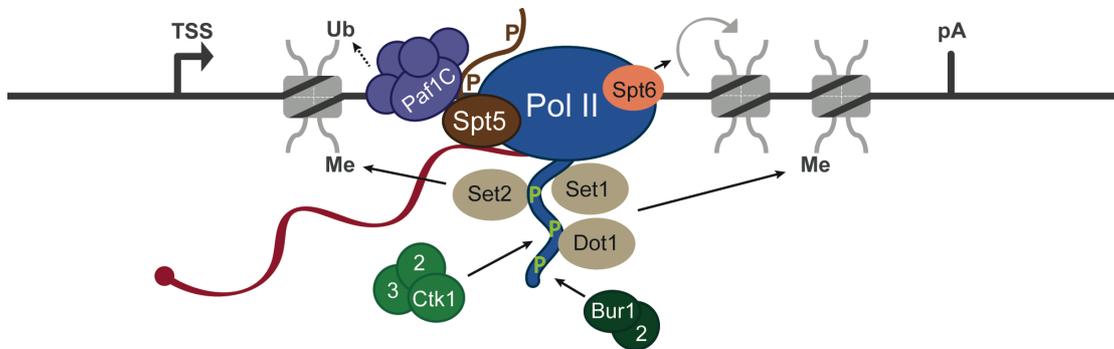


Figure 2 Schematic representation of transcribing Pol II and its interaction with transcription elongation factors (EFs). Spt5 (brown) binds the body of Pol II and its phosphorylated C terminal repeat (CTR) recruits the five subunit Paf1 complex (Paf1C; purple). Paf1C facilitates ubiquitination of histone H2B on Lys123 by the monoubiquitinase complex Rad6/Bre1 (not shown). The three and two subunit kinase complexes CTDK-I (light green) and BUR (dark green), respectively, phosphorylate the Ser5 pre-phosphorylated CTD of Pol II on Ser2. The three histone methyltransferases Set1, Set2 and Dot1 that set the ‘active’ histone marks H3K4me3, H3K36me3, and H3K79me3, respectively, are depicted in beige. Nascent pre-mRNA: red; nucleosome units: light gray; chromatin: dark gray. TSS: transcription start site; pA: polyadenylation site.

The transcription elongation factor Spt5 is a subunit of the heterodimeric complex Spt4-Spt5 and functions as part of the stable Pol II elongation machinery mediating interactions between Pol II and other proteins (Hartzog and Fu 2013). Spt5 is conserved in all domains of life; however, eukaryotic Spt5 contains an unstructured region at its C terminus (the C-terminal repeated region [CTR]), which comprises a set of short repeats whose consensus motifs vary across species. The CTR is subject to phosphorylation and participates in Spt5’s mediation of nuclear activities, such as

recruitment of additional elongation factors, as shown for the Paf1 complex ((Qiu et al. 2006); Figure 2).

The Paf1 (Pol II-associated factor 1) complex (Paf1C) is composed of five subunits in yeast, namely Cdc73, Ctr9, Leo1, Paf1 and Rtf1. Paf1C has a direct role in transcription elongation consistent with accumulation of Paf1 over the transcribed region of genes revealed by chromatin immunoprecipitation (ChIP; (Mayer et al. 2010)). Recently, Paf1C has been shown to have a regulatory function determining differential transcript fate at selected genes (Fischl et al. 2017). It is also required for deposition of ubiquitin on Lys123 of histone H2B (H2Bub) by the Rad/Bre1 E2/E3 ubiquitin ligase complex (Robzyk, Recht, and Osley 2000, Wood et al. 2003) that in turn facilitates recruitment of nuclear-cytoplasmic export factors (Tuck and Tollervey 2013). Furthermore, H2Bub enables recruitment of the Set1/COMPASS complex that methylates histone 3 at Lys4 (H3K4), a hallmark of active promoters and the 5' end of ORFs that correlates well with increased levels of gene expression (Pokholok et al. 2005). Methylation of H3K4 is a highly conserved pathway in eukaryotes, though, in yeast, the Set1 complex is the only H3K4 methyltransferase, whereas methylation of H3K4 in *Drosophila* is carried out by three methyltransferase complexes and in mammals by at least six such complexes (Smolle and Workman 2013). Dot1 mediates the methylation of H3K79, which also requires prior ubiquitination of histone H2B and it has been linked to transcription elongation (Nguyen and Zhang 2011).

The essential kinase Bur1 and its associated cyclin Bur2 form the BUR kinase complex that phosphorylates the CTR of Spt5 and the CTD of Pol II at Ser2 position (Liu et al. 2009, Murray et al. 2001). The Pol II CTD is further phosphorylated on Ser2 by the CTDK-I kinase complex (Bartkowiak et al. 2010). CTDK-I contains the cyclin-dependent kinase Ctk1, the cyclin Ctk2, and the yeast-specific subunit Ctk3, which is required for CTDK-I stability and activity (Muhlbacher et al. 2015). Pol II CTD phosphorylation on Ser2 by Ctk1 and Bur1 promotes recruitment of the Set2 methyltransferase catalyzing mono-, di- and trimethylation of H3K36 (Krogan, Kim, et al. 2003, Li et al. 2003, Li, Moazed, and Gygi 2002, Youdell et al. 2008). This methylation mark leads to recruitment of the Rpd3s histone deacetylase complex, a key step in preventing cryptic transcription initiation within open reading frames (Carrozza, Li, et al. 2005). Phosphorylated Ser2 CTD also leads to the recruitment of

Spt6, an essential histone chaperone and remodeler of chromatin in eukaryotes (Bortvin and Winston 1996, Youdell et al. 2008). Besides recruiting RNA processing factors to Pol II, Spt6 is required to establish a repressive chromatin environment that prevents initiation of long non-coding RNAs (lncRNAs) within coding regions (Ard and Allshire 2016).

1.3.2 Elongation factor recruitment

Despite extensive research, it remains unclear for several EFs how they are recruited to active genes. EFs may be recruited by interactions with the body of transcribing Pol II, or by contacts with the tail-like CTD of Pol II, or they may bind via other Pol II-associated EFs. Spt5 binds the body of the Pol II elongation complex (Grohmann et al. 2011, Klein et al. 2011, Martinez-Rucobo et al. 2011), whereas Bur1, Spt6 and Set2 bind the CTD (Dengl et al. 2009, Kizer et al. 2005, Li et al. 2003, Phatnani, Jones, and Greenleaf 2004, Sun et al. 2010, Yoh et al. 2007, Qiu, Hu, and Hinnebusch 2009, Li, Moazed, and Gygi 2002). Interaction of Paf1C with Pol II involves Spt5 (Liu et al. 2009, Mayekar, Gardner, and Arndt 2013, Wier et al. 2013, Zhou et al. 2009, Qiu et al. 2012, Qiu, Hu, and Hinnebusch 2009) and the CTD (Qiu et al. 2012), whereas interaction of Set1 with Pol II involves Paf1C (Krogan, Dover, et al. 2003, Ng et al. 2003).

However, it is likely that other recruitment mechanisms exist because mutations in EFs that prevent their interactions with Pol II do not abolish gene occupancy of such factors, including Bur1, Paf1C subunits, Spt6, and Set2 (Ng et al. 2003, Qiu et al. 2012, Qiu, Hu, and Hinnebusch 2009, Mayer et al. 2010, Zhou et al. 2009, Krogan, Kim, et al. 2003). Further, it remains unknown how the yeast CTD serine 2 (Ser2) kinase Ctk1 is recruited, which is apparently a prerequisite for recruitment of Spt6 and Set2, since these factors bind the Ser2-phosphorylated CTD (Dengl et al. 2009, Kizer et al. 2005, Li et al. 2003, Phatnani, Jones, and Greenleaf 2004, Sun et al. 2010, Yoh et al. 2007). More generally, it is unknown whether and how EFs can distinguish transcribing Pol II from free or initiating polymerase based on polymerase interactions alone, in particular at an early stage of elongation when Ser2 phosphorylation is absent.

An alternative mechanism of EF recruitment would involve interactions with the nascent pre-mRNA. Such RNA interactions are well established for RNA processing factors that are recruited during Pol II elongation (Perales and Bentley 2009, Bentley 2005, Baejen et al. 2014, Tuck and Tollervey 2013) for co-transcriptional capping (Martinez-Rucobo et al. 2015), splicing (Bentley 2005, Saldi et al. 2016), and 3'-processing (Proudfoot 2011, Shi and Manley 2015) of the pre-mRNA. Some observations indeed suggest that nascent RNA contributes to the recruitment of EFs to Pol II. Spt5 and Set1 bind RNA *in vitro* (Meyer et al. 2015, Missra and Gilmour 2010, Tresaugues et al. 2006, Halbach et al. 2009), Ctk1 and Bur1 *in vivo* occupancy at active genes depends on the cap-binding complex, which binds 5'-capped RNA (Hossain et al. 2013, Lidschreiber, Leike, and Cramer 2013), and Paf1C binds RNA, which is required for full gene occupancy (Dermody and Buratowski 2010).

1.4 Nascent RNA as a transcription regulator

The movement of Pol II along genes is a key mechanism for the control of gene expression regulated by proteins that bind to DNA, nucleosomes and Pol II itself. That nascent non-coding RNA can also regulate transcription and chromatin function was first demonstrated for the transactivation response element (TAR; reviewed in (Peterlin and Price 2006)). TAR is an RNA stem-loop formed at the 5' end of nascent HIV transcripts that recruits the viral protein transactivator of transcription (Tat) and the positive transcription elongation factor b (P-TEFb; (Wei et al. 1998, Zhu et al. 1997)).

A rapidly accumulating wealth of studies are identifying and functionally characterizing diverse types of non-coding RNAs that directly bind to regulatory proteins and modulate their recruitment to genes. These ncRNAs, including lncRNAs and enhancer RNAs (eRNAs), contribute to gene regulation in either their nascent or mature forms (reviewed in (Skalska et al. 2017)). However, recent studies revealed that also nascent pre-mRNAs directly interact with activators and repressors of transcription during initiation, elongation, termination and RNA processing (Beltran et al. 2016, Di Ruscio et al. 2013). For instance, for co-transcriptional splicing, splice-site sequences in the nascent pre-mRNA recruit the spliceosome, and Pol II pauses at

these sequences, thereby enhancing splicing fidelity (Alexander et al. 2010). In higher eukaryotes, the polycomb repressive complex 2 (PRC2) methylates H3K27 to maintain gene repression during development. *In vivo* sequencing methods demonstrated that PRC2 directly interact with several nascent pre-mRNAs, and RNA degradation experiments revealed that loss of RNA leads to increased interactions of PRC2 with chromatin at active genes (Beltran et al. 2016, Davidovich et al. 2013, Kaneko et al. 2014, Kaneko et al. 2013, Zhao et al. 2010). Thus, indicating a role for nascent RNA in preventing PRC2 association with chromatin at transcriptionally active genes (reviewed in (Skalska et al. 2017)).

Nascent pre-mRNAs can also interact with activators of transcription. The WD repeat-containing 5 (WDR5) protein, a subunit of the mammal Set1/COMPASS complex, was originally shown to bind lncRNAs and recently revealed to also associate with pre-mRNAs and mRNAs (Hendrickson et al. 2016, Wang et al. 2011, Yang et al. 2014). The interaction of WDR5 with these RNAs positively correlates with di- and trimethylation levels of H3K4 (H3K4me2 and H3K4me3) in *cis* (Hendrickson et al. 2016). Thus, suggesting that nascent pre-mRNA can form a positive feedback loop that promotes gene expression.

Similarly, the human transcription factor yin and yang 1 (YY1) was first found to bind to the mature lncRNA *XIST* (Jeon and Lee 2011) and recent crosslinking and immunoprecipitation (CLIP) data detected that YY1 also interacts with nascent pre-mRNA and nascent eRNA (Sigova et al. 2015). Interestingly, YY1 binding to chromatin was largely decreased when cells were treated with RNase, and binding to chromatin increased when RNA was tethered to DNA (Sigova et al. 2015). This data supports a model where nascent RNA functions to recruit and maintain transcription factors near the DNA.

1.5 *In vivo* methods for detection of RNA-protein interactions

It is likely that nascent RNAs conduct their regulatory functions in the form of RNA-protein complexes. Therefore, techniques to study RNA-protein interactions are essential to further uncover the mechanisms governing biological processes, such as transcription, chromatin remodeling and splicing. Proteins interact with RNA in a

similar manner as with DNA. In both cases, electrostatic and hydrophobic interactions, hydrogen bonding, and base stacking are the driving forces that govern this complex association. During the last decade, the role of RNA in biological processes has been recognized as an area of deep interest, and a variety of methods to study protein-RNA interactions have been –and continue to be– developed. The most relevant and latest reported methods for identifying RNA-protein interactions *in vivo* consist of a combination of genetic, biochemical and computational approaches.

1.5.1 RIP methods

RNA immunoprecipitation (RIP) is used to detect the interaction between individual proteins and specific RNA molecules *in vivo* (Gilbert and Svejstrup 2006, Keene, Komisarow, and Friedersdorf 2006). Following immunoprecipitation of the protein of interest, co-precipitated RNAs are isolated and quantitated by qRT-PCR. The basic mechanisms of RIP are very similar to ChIP. RIP-Chip and RIP-Seq combine RIP with microarray profiling of RNAs and high-throughput sequencing, respectively (Jain et al. 2011, Zhao et al. 2010). These techniques have been used to find interactions between one specific protein and RNAs at transcriptome level. RIP methods, however, are limited to the characterization of kinetically stable interactions and are prone to detecting nonspecific interactions (Hendrickson et al. 2016, Mili and Steitz 2004). Recently, an improved formaldehyde crosslinking RNA immunoprecipitation technique followed by deep sequencing (fRIP-Seq) was used to detect RNA interactions of several proteins containing RNA binding domains, but also proteins that lack classically defined RNA binding domains (Hendrickson et al. 2016). fRIP-Seq uses lower formaldehyde concentrations than ChIP and RIP methods and requires smaller amounts of input RNA than CLIP methods. However, similar to ChIP and RIP, it can lead to detection of indirect interactions between proteins and nucleic acids.

1.5.2 CLIP methods

Crosslinking and immunoprecipitation (CLIP) combines ultraviolet light (UV)-crosslinking with immunoprecipitation of a specific protein followed by the isolation of crosslinked RNA segments and cDNA sequencing (Ule et al. 2003). Here, UV

irradiation of cells or tissues at the wavelength of 254 nm generates covalent bonds between RNA and protein when they are in close contact. In the original CLIP protocol, amplified cDNAs are subjected to Sanger sequencing and the resulting sequences mapped to the reference genome to reveal protein-binding sites within the corresponding transcripts. The combination of CLIP with high-throughput sequencing (HITS-CLIP or CLIP-Seq) is used now for transcriptome-wide studies (Licatalosi et al. 2008).

The binding site resolution of traditional CLIP corresponds to the length of the fragmented RNAs. Recent methods allow now single-nucleotide resolution of RNA-binding proteins. Individual nucleotide resolution CLIP (iCLIP) makes use of the covalently bound polypeptide fragment at the cross-link site on the RNA (Konig et al. 2010). During reverse transcription, the reverse transcriptase enzyme produces truncated cDNAs that are used for the identification of binding sites genome-wide at nucleotide resolution.

Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) is an improved method that also allows direct identification of RNA-binding proteins transcriptome-wide at single-nucleotide resolution (Hafner et al. 2010). However, PAR-CLIP uses photoreactive ribonucleoside analogs, such as 4-thiouridine (4sU), that are readily taken up by cells and become incorporated into U-containing regions of newly synthesized transcripts (Schwalb et al. 2016). Additionally, PAR-CLIP involves crosslinking with 365 nm UV light (instead of 254 nm) to establish covalent crosslinks between aromatic amino acids of the protein of interest and the thio group of 4sU. During reverse transcription, reverse transcriptase reads through the crosslink site and mistakenly incorporates a guanosine (G) instead of an adenosine (A), which allows to precisely identify binding sites by scoring for thymidine (T) to cytidine (C) transitions in the sequenced cDNA (Hafner et al. 2010). Due to the use of photoreactive nucleoside analogs and a longer wavelength, PAR-CLIP induces specific crosslinks between labeled RNAs and interacting proteins, thus allowing detection of direct RNA-protein interactions and reducing the amount of false positive results.

1.6 Aims and scope of this thesis

Several recent studies revealed that proteins known to interact with transcribing chromatin and regulate gene expression also interact with the nascent RNA (see section 1.4). Furthermore, in most of the reports this interaction was shown to have either a positive or a negative effect on gene transcription (reviewed in (Skalska et al. 2017)), suggesting a model in which RNA, together with the chromatin structure and proteins, is a main player in regulating the highly conserved process of transcription.

The aims of this study focus on the characterization of RNA-protein interactions during transcription elongation and their possible roles in gene expression. We took advantage of the powerful technique PAR-CLIP, to investigate interactions between EFs in the yeast *S. cerevisiae* and the nascent transcribed RNA *in vivo* (Figure 3). Based on our results, we further aimed to analyze if binding preferences can differ for coding RNAs and for non-coding (nc) RNAs, such as CUTs. Moreover, we aimed to investigate whether these RNA-protein interactions have an implication on the protein-DNA interactions. To answer this question, we established RNA degradation assays to compare protein-chromatin binding in the presence and absence of RNA.

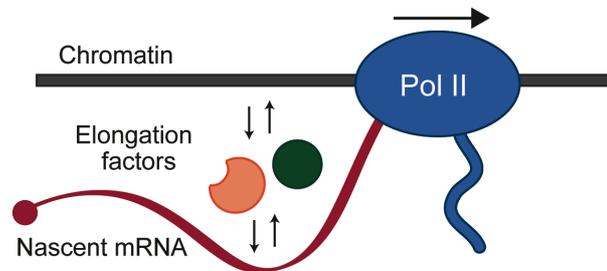


Figure 3 Representation of EFs interaction with transcribed chromatin and nascent RNA during Pol II transcription. The nascent RNA could have a role in establishing and also maintaining EF-chromatin interaction.

Another aim during this work was to investigate whether the *in vivo* observed interactions are also detectable *in vitro*. In order to do so, we tested one prominent EF, the CTDK-I kinase complex, for RNA binding *in vitro*. We established a protocol to purify recombinant CTDK-I containing all three subunits and performed activity assays to test the kinase activity of the purified complex. Finally, we used

fluorescence anisotropy to test whether the CTD Ser2 kinase complex binds RNA *in vitro*.

To further study recruitment mechanisms of EFs, we compared factor occupancies on DNA and on RNA. Our laboratory and others have previously obtained genome-wide occupancy profiles for EFs by chromatin immunoprecipitation followed by tiling microarray analysis (ChIP-chip; (Liu et al. 2005, Mayer et al. 2010, Ng et al. 2003, Pokholok et al. 2005, Weiner et al. 2015)). However, this data is not sufficiently suitable for comparison with our high-resolution genome-wide PAR-CLIP data. In this regard, we established and performed ChIP-Seq experiments of most *S. cerevisiae* Pol II EFs and histone marks involved in this study. With both PAR-CLIP and ChIP-Seq data sets, we can now directly compare when specific transcription factors are recruited to RNA and when to DNA.

Based on our results, we were able to suggest a model where nascent RNA actively participates in the assembly and stability of the Pol II elongation complex. RNA-EF interactions provide a missing link for understanding the coordination and fine-tuning of the transcription cycle.

2 Materials and Methods

2.1 Materials

2.1.1 Bacterial and yeast strains

Table 1 Bacterial strains.

Strain	Genotype	Source
XL1-Blue	<i>recA1; endA1; gyrA96; thi-1; hsdR17; supE44; relA1; lac[F' proAB lacIqZΔM15 Tn10(Tet^r)]</i>	Stratagene
BL21(DE3)pLysS	B; F ⁻ ; <i>ompT; hsdS_B(r_B⁻, m_B⁻); dcm; gal; λ(DE3); pLysS; Cm^R</i>	Stratagene

Table 2 Yeast strains.

Strain	Genotype	Source
BY4741 (wild-type)	MA ⁺ Ta; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	Euroscarf
Bur1-TAP	BY4741; <i>BUR1-TAP::HIS3MX6</i>	Euroscarf
Bur2-TAP	BY4741; <i>BUR2-TAP::HIS3MX6</i>	Euroscarf
Cdc73-TAP	BY4741; <i>CDC73-TAP::HIS3MX6</i>	Euroscarf
Ctk1-TAP	BY4741; <i>CTK1-TAP::HIS3MX6</i>	Euroscarf
Ctk2-TAP	BY4741; <i>CTK2-TAP::HIS3MX6</i>	Euroscarf
Ctr9-TAP	BY4741; <i>CTR9-TAP::HIS3MX6</i>	Euroscarf
Paf1-TAP	BY4741; <i>PAF1-TAP::HIS3MX6</i>	Euroscarf
Rpb1-TAP	BY4741; <i>RPB1-TAP::HIS3MX6</i>	Euroscarf
Rpb3-TAP	BY4741; <i>RPB3-TAP::HIS3MX6</i>	Euroscarf
Rtf1-TAP	BY4741; <i>RTF1-TAP::HIS3MX6</i>	Euroscarf
Spt5-TAP	BY4741; <i>SPT5-TAP::HIS3MX6</i>	Euroscarf
Spt6-TAP	BY4741; <i>SPT6-TAP::HIS3MX6</i>	Euroscarf
TFIIB-TAP	BY4741; <i>TFIIB-TAP::HIS3MX6</i>	Euroscarf

2.1.2 Cell lines

Table 3 Cell lines used for expression of recombinant CTDK-I (2.2.7.2).

Cell line	Description
DH10EMBacY	Modified <i>Autographa californica multicapsid nucleopolyhedrovirus</i> (AcMNPV) genome as Bacmid; helper plasmid for Tn7 transposon enzyme; YFP reporter gene in virus backbone; LacZ for blue-white screening.
High five (Hi5)	Insect cell line used for expression of recombinant proteins; <i>Trichoplusia ni</i> embryonic tissue.
Sf9	Insect cell line used for isolation and propagation of recombinant baculoviral stocks; <i>Spodoptera frugiperda</i> ovarian tissue.
Sf21	Insect cell line used for isolation and propagation of recombinant baculoviral stocks; <i>Spodoptera frugiperda</i> ovarian tissue.

2.1.3 Growth media and media additives

Table 4 Growth media.

Name	Description/source	Species/cell line
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)	<i>E. coli</i>
YPD	2% (w/v) peptone; 2% (w/v) glucose; 1.5% (w/v) yeast extract (+1.8% (w/v) agar for solid media plates)	<i>S. cerevisiae</i>
Minimal medium	6.9 g/L yeast nitrogen base; 2% (w/v) glucose; 0.4 g/L CSM drop-out: complete; 0.4 g/L CSM drop-out: -URA	<i>S. cerevisiae</i>
Sf-900 III SFM	Growth and maintenance medium/ThermoFisher	Sf9 cells
ESF921 medium	Expression medium/Expression Technologies	Hi5 and Sf21 cells

Table 5 Growth media additives.

Name	Description	Working concentration
Ampicillin (Amp)	Antibiotic	100 µg/mL
4-thiouracil (4tU)	Metabolic labeling of nascent RNAs	1 mM
X-tremeGENE9	Transfection reagent	-

2.1.4 Oligonucleotide sequences

Table 6 Primers used for CTDK-I cloning (2.2.7.1). Lowercase sequences: overhangs for homologous recombination with the desired vector.

Name	Sequence (5'-3')
Ctk1_438-A fwd	tactccaatccaatcgATGTCCTACAATAATGGCAATACTTATTCAAAG AGTTATAGC
Ctk1_438-A rev	ttatccactccaatgttattaTTATCATCATCGTCATTATTATTATTATTAT TATTATTACTATTACCATTACCCTTATTG
Ctk2_438-C fwd	tactccaatccaatgcaCCTAGCACGTTTGAATCCCAGC
Ctk2_438-C rev	ttatccactccaatgttattaTGCATGTCTTGTAGAACTATTTATGCTGGAC
Ctk3_438-A fwd	tactccaatccaatcgATGGACTCTCTTGAAGCTAGATTACAATTCATT C
Ctk3_438-A rev	ttatccactccaatgttattaATATATGTAAGATGCCTTCGCAATGTCATTT AAAGC

Table 7 Adapters, primers and barcodes used for PAR-CLIP (2.2.4.1).

Name	Sequence (5'-3')
3'-adapter	/5rApp/TGGAATTCTCGGGTGCCAAGG/3ddC/
5'-adapter	/5InvddT/rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrAr UrCrNrNrNrNrN
RT primer	CCTTGGCACCCGAGAATTCCA
NEXTflex barcode primer (fwd):	CAAGCAGAAGACGGCATAACGAGA-Barcode-GTGACTGGAGTTC CTTGGCACCCGAGAATTCCA
Barcode 1	TCGTGAT
Barcode 2	TACATCG
Barcode 3	TGCCTAA
Barcode 4	TTGGTCA
Universal primer (rev)	AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTAC AGTCCGA
Nextera primer 1	AATGATACGGCGACCACCGA
Nextera primer 2	CAAGCAGAAGACGGCATAACGA

Table 8 Primers used for ChIP-qPCR (2.2.5.2).

Gene	Sequence (5'-3') fwd primer	Sequence (5'-3') rev primer	Binding site
<i>ADHI</i>	CTACGAATCCCACGGTAAGT	ACAATTCGTTGGCCTTTG	5'-end
<i>ADHI</i>	AGCCGCTCACATTCCTCAAG	ACGGTGATACCAGCACACAAGA	ORF
<i>ADHI</i>	ACCTTGCCAGAAATTTACGA	CAACAACGTATCTACCAACGA	3'-end
<i>ALD5</i>	AAAAGCCAAAGAAGAAGAAATGC	GAATATTCTGGAATTCGGAGCTG	5'-end
<i>ALD5</i>	CTTTTCGCTTCCCAGTTGTG	GGAACCCGGAAGGATATTGA	ORF
<i>ALD5</i>	CCCGAGTTGACTGCTCATTG	TGTCACGAGGCATTTTTTCATT	pA site
<i>ILV5</i>	ACCCAGTATTTTCCCTTCC	TTGTCTATATGTTTTTGCTTGC	Prom.
<i>ILV5</i>	CTATCAAGCCATTGTTGACC	CTTGAAGACTGGGGAGAAAC	ORF
<i>ILV5</i>	ACACCATCAGAAACATGGAA	TGGTTTTCTGGTCTCAACTTT	3'-end
<i>MUP1</i>	AGCAATGTCGGAAGGAAGAA	TGATAGTTCTCCTTGTGAAGACA	5'-end
<i>MUP1</i>	AGGACGTGAAGGTGTGTTGC	TTGGGGAGTTGAATGGCTTA	ORF
<i>MUP1</i>	TTTAAGATGCTATTCGGTTTCG	TTGATTCGCGCAATAGTAGG	pA site
<i>PDC1</i>	TGTTCGAAAGATTAAGCAAGTCA	TTGAAGTCACCTGGCAAACC	5'-end
<i>PDC1</i>	CACTTTCCAGGTGTCCAAA	CTTAGCGGCGTCAGCAATAG	ORF
<i>PDC1</i>	GTTGCCAGTCTTCGATGCTC	AGCGTTGGTAGCAGCAGTCA	3'-end
<i>PMA1</i>	TGACTGATACATCATCCTCTT	TTGGCTGATGAGCTGAAACAGAA	5'-end
<i>PMA1</i>	GGGTTCCGTGCTTTAGGTGT	CACCCAAGATTTCCCAGTGA	ORF
<i>PMA1</i>	TTGATGAACGGTAAGCCAATG	GCAGCCATGAAGTCTTCGAC	3'-end
<i>YER</i>	TGCGTACAAAAAGTGCAAGAGATT	ATGCGCAAGAAGGTGCCTAT	Chr5

YER: Heterochromatic region on chromosome V.

Table 9 5'-FAM labeled ssRNA, ssDNA and dsDNA sequences used for fluorescence anisotropy experiments (2.2.9).

Name	Sequence (5'-3')
ssRNA, 24% GC, A-rich	/56-FAM/rArArUrArUrUrCrArArGrArCrGrArUrUrUrArGrArCrGrArUrArArUrArUrUrCrArUrA
ssRNA, 24% GC, U-rich	/56-FAM/rArUrGrUrUrGrUrArUrGrArUrArUrCrUrUrGrCrUrArArCrUrUrArArUrUrUrGrArU
ssRNA, 45% GC, A-rich	/56-FAM/rArArGrCrArGrCrCrArArArCrArArGrCrArGrUrCrArArCrArUrCrArArGrUrCrGrU
ssRNA, 45% GC, U-rich	/56-FAM/rUrUrCrGrUrCrGrGrUrUrUrGrUrGrCrGrUrCrArGrUrUrGrUrArGrUrUrCrArUrCrA
dsDNA, 45% GC, A-rich	/56-FAM/AAGCAGCCAAACAAGCAGTCAACATCAAGTCGT

2.1.5 Antibodies

Table 10 Antibodies.

Name	Dilution	Application(s)	Source
Anti-Histone H3 (HRP), ab21054	1:5000	Western blot	Abcam
Peroxidase Anti-Peroxidase (PAP), P1291	1:2000	Western blot	Sigma
Anti-rat IgG (HRP), A9037	1:3000	Western blot	Sigma
Anti-Ser2-P, 3E10	1:14; 25 µL/IP	Western blot; ChIP	Dirk Eick lab
Anti-Ser5-P, 3E8	1:14; 20 µL/IP	Western blot; ChIP	Dirk Eick lab
Anti-Tyr1-P, 3D12	1:7; 100 µL/IP	Western blot; ChIP	Dirk Eick lab
Anti-Histone H3 (tri methyl K4), ab8580	30 µg/IP	ChIP	Abcam
Anti-Histone H3 (tri methyl K36), ab9050	15 µg/IP	ChIP	Abcam
Anti-Histone H3 (tri methyl K79), ab2621	10 µg/IP	ChIP	Abcam
IgG from rabbit serum	100 µg/IP	ChIP, PAR-CLIP	Sigma

2.1.6 Plasmids

Table 11 Series-438 *MacroBac* Expression Vectors used for CTDK-I cloning (2.2.7.1).

Vector name	Description	Addgene
438-A	Modified pFastBac cloning vector with BioBrick PolyPromoter LIC Subcloning; yORF	55218
438-C	Modified pFastBac His6-MBP-Asn10-TEV cloning vector with BioBrick PolyPromoter LIC Subcloning; His6-MBP-N10-TEV-yORF	55220

2.1.7 Buffers and solutions

Table 12 Buffers and solutions.

Name	Description	Application(s)
Amylose elution buffer	400 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0, 10 g maltose	2.2.7.3
ChIP lysis buffer	150 mM NaCl, 50 mM HEPES-KOH pH 7.5, 1 mM EDTA, 1% Triton X-100, 0.1% Na deoxycholate, 0.1% SDS	2.2.5, 2.2.6
ChIP wash buffer	100 mM Tris-HCl at pH 7.5, 500 mM LiCl, 1% NP-40, 1% Na deoxycholate	2.2.5
CLIP lysis buffer	50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.5% sodium deoxycholate, 0.1% SDS, 0.5% NP-40	2.2.4
CLIP wash buffer	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.5% sodium deoxycholate, 0.1% SDS, 0.5% NP-40	2.2.4
CTDK-I dilution buffer	200 mM NaCl, 20 mM Na•HEPES pH 7.4, 1 mM DTT and 10% glycerol	2.2.9
CTDK-I lysis buffer	400 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0	2.2.7.2, 2.2.7.3
High salt wash buffer	800 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0	2.2.7.3
HU buffer	5% (w/v) SDS, 0.2 M Tris-HCl pH 6.8, 10 mM EDTA, 215 mM β -mercaptoethanol, 8 M urea, 0.01% (w/v) bromophenolblue	2.2.3.1
Nickel elution buffer	400 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% glycerol (v/v), 1 mM DTT, 500 mM imidazole pH 8.0	2.2.7.3
1 \times PBS	140 mM NaCl, 3 mM KCl, 4 mM Na ₂ HPO ₄ , 2 mM KH ₂ PO ₄ ; pH 7.4	2.2.4
PBS-T	1% (for Western blot) or 0.02% (for IPs) Tween 20 in 1 \times PBS	2.2.3.3, 2.2.5, 2.2.6

Phosphatase reaction buffer	50 mM Tris-HCl pH 7.0, 1 mM MgCl ₂ , 0.1 mM ZnCl ₂	2.2.4
Phosphatase wash buffer	50 mM Tris-HCl pH 7.5, 20 mM EGTA, 0.5% NP-40	2.2.4
PNK buffer	50 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM MgCl ₂	2.2.4
1× protease-inhibitor mix	1 mM leupeptin, 2 mM pepstatin A, 100 mM PMSF, 280 mM benzamidine in 100% ethanol	2.2.4, 2.2.5, 2.2.6, 2.2.7
Proteinase K buffer	50 mM Tris-HCl pH 7.5, 6.25 mM EDTA, 75 mM NaCl	2.2.4.1
RNase storage buffer	10 mM HEPES pH 7.5, 20 mM NaCl, 0.1% Triton X-100, 1 mM EDTA, 50% glycerol	2.2.6
SE buffer	400 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, 1 mM DTT	2.2.7.3
T1 buffer	50 mM Tris-HCl pH 7.5, 2 mM EDTA	2.2.4
1× TAE	40 mM Tris, 20 mM acetic acid, 1 mM EDTA pH 8.0	2.2.2.2
1× TBS	150 mM NaCl, 50 mM Tris-HCl; pH 7.5	2.2.5, 2.2.6
TE buffer	10 mM Tris-HCl at pH 7.5, 1 mM EDTA	2.2.5

2.2 Methods

2.2.1 *S. cerevisiae* strain validation

Saccharomyces (S.) cerevisiae strains containing C-terminally TAP-tagged ORFs were obtained from Euroscarf and were isogenic to BY4741 wild-type strain. All TAP strains used in this work are listed in Table 2. These strains were validated before using them for experiments. Firstly, colony PCR (2.2.2) using a gene-specific upstream primer and a reverse primer within the TAP tag was performed to confirm that the DNA coding for the TAP-tag was at the correct genomic position. Secondly, the expression level of TAP-tagged proteins was controlled by Western blotting (2.2.3), using an anti-TAP antibody (PAP; Table 10).

2.2.2 General DNA methods

2.2.2.1 Amplification of genomic *S. cerevisiae* DNA (colony PCR)

A pinhead-sized amount of cells from an YPD plate (Table 4) was dissolved in 100 μ L of 20 mM NaOH. Approximately 50 μ L of 0.5 mm glass beads (Roth) were added to the solution and samples were boiled for 5 min at 95°C and 1,400 rpm. Samples were then centrifuged at 13,000 rpm for 1 min and 3 μ L of the supernatant was used as template for PCR amplification. Colony PCRs were performed using Taq polymerase (NEB), 0.25 μ M primers, 2 mM MgCl₂, 0.1 mM dNTPs, and 35 cycles of a 3-step PCR reaction: 94°C for 1 min, 50-60°C for 30 sec, 72°C for 30 sec. Annealing temperatures and elongation times were optimized for specific primer pairs. For validation of TAP-tagged yeast strains, suggested primer pairs from Euroscarf were used. DNA sequences for other primers are given in Table 6.

2.2.2.2 Electrophoretic separation of DNA

PCR samples were analyzed by agarose gel electrophoreses. Samples were mixed with Orange DNA Loading Dye (ThermoFisher) and loaded onto 1 \times TAE (Table 12) agarose gels containing 1% agarose and 1 \times SYBR Safe (Invitrogen). Along PCR samples, 5 μ L of 100 bp or 1 kb O'GeneRuler Plus DNA Ladder, ready-to-use (ThermoFisher) were loaded onto the gel. DNA was separated at 120 V for approximately 45 min and visualized under UV light. DNA bands were excised with a scalpel and the DNA was purified using the QIAquick gel extraction kit (Qiagen) following the manufacturer's instructions.

2.2.3 General protein methods

2.2.3.1 Protein extraction from *S. cerevisiae* cells

A pinhead-sized amount of cells from an YPD plate was resuspended in 1 mL of cold water. 150 μ L of 7.5% β -mercaptoethanol in 1.85 M NaOH were added and cells were incubated for 15 min on ice. Afterwards, 150 μ L of 55% TCA were added and the sample incubated for 10 min on ice. Next, cells were centrifuged for 10 min at 14,000 rpm and 4°C, the supernatant discarded and the pellet resuspended in 50 μ L HU buffer (Table 12). The sample was neutralized using 5-15 μ L of 1 M Tris base

and incubated for 10 min at 1,400 rpm and 65°C. Cells were centrifuged for 5 min at 14,000 rpm at room temperature and the supernatant, containing denatured proteins, was collected.

2.2.3.2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein were separated by SDS-PAGE according to their size using 4-15% Mini-PROTEAN[®] TGX[™] precast protein gels and the Mini-PROTEAN[®] Tetra Cell Systems chambers in 1× Tris/glycine/SDS running buffer (Bio-Rad). Along protein samples, 3 μL PageRuler Prestained Protein Ladder (ThermoFisher) was loaded onto the gel. Proteins were separated at 200 V for approximately 45 min. Protein gels were either stained using InstantBlue[™] (expedeon) for 20 min and destained with H₂O overnight or used for Western blot analysis (2.2.3.3).

2.2.3.3 Western blot analysis

Proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane using the Trans-Blot Turbo Transfer System (Bio-Rad). The membrane was blocked with 2% milk powder (Roth) in PBS-T (Table 12) for 1 h at RT. Primary antibodies or HRP-coupled antibodies were diluted accordingly (Table 10) in a 2% milk/PBS-T solution and incubated with the membrane for 1 hr at RT or at 4°C overnight. Afterwards, the membrane was washed three times for 10 min with PBS-T and, if necessary, incubated for 1 hr with a secondary antibody diluted accordingly in a 2% milk/PBS-T solution (Table 10). If incubated with a secondary antibody, membranes were washed again three times for 10 min with PBS-T. Antibody detection was performed using the SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher). The membrane was imaged with an Advanced Fluorescent Imager (Intas).

2.2.4 PAR-CLIP

PAR-CLIP experiments were performed as previously described (Baejen et al. 2014, Schulz et al. 2013), with some modifications. *S. cerevisiae* cells expressing the TAP-tagged protein were grown overnight in YPD medium (Table 4) at 30°C and 160 rpm. Overnight cultures were diluted to an OD₆₀₀ of ~ 0.1 in 1 L minimal medium (Table 4) supplemented with 100 mM 4-thiouracil (4tU). Here, we used 4tU instead of 4-thiouridine (4sU) for RNA labeling, because *S. cerevisia* do not express a nucleoside

transporter and cannot take up 4sU. 4tU is taken up by *S. cerevisiae* without expression of a nucleoside transporter (Sun et al. 2012). Yeast cultures were then grown to $OD_{600} \sim 0.5$ before 4tU was added to a final concentration of 1 mM and cells grown further for 4 hr ($OD_{600} \sim 1.2$). Following RNA labeling, cells were harvested, washed with $1 \times$ PBS, resuspended in 20 mL $1 \times$ PBS and UV-irradiated on ice with an energy dose of 12 J/cm^2 at 365 nm under continuous shaking (Bio-Link BLX-365, Vilber Lourmat). Cells were harvested, flash frozen in liquid nitrogen and stored at -80°C . All subsequent steps were performed at 4°C with precooled buffers and in the presence of $1 \times$ protease-inhibitor mix (Table 12). Cells pellets were resuspended in 3 mL CLIP lysis buffer (Table 12) and divided into three 2 mL FastPrep tubes. Cell disruption was performed by bead beating (FastPrep®-24 Instrument, MP Biomedicals, LLC.) in the presence of 1 mL of 0.5 mm zirconia/silica beads (Roth) for 40 sec at 4 m/s, followed by an incubation of the sample for 1 min on ice. This was repeated 8 times. The success of the cell lysis was monitored by photometric measurements and the cell lysis efficiency was usually $>80\%$. Samples were solubilized for 1 min via sonication with a Covaris S220 instrument (COVARIS, INC.) using following parameters: Peak Incident Power (W): 140; Duty Factor: 5%; Cycles per Burst: 200. Samples were pooled and the lysate was cleared by centrifugation. Immunoprecipitation (IP) was performed on a rotating wheel overnight at 4°C with rabbit IgG-conjugated Protein G magnetic beads (Invitrogen). After IP, beads were washed twice in CLIP wash buffer (Table 12) and once in T1 buffer (Table 12). Immunoprecipitated and crosslinked RNA was partially digested with 50 U of RNase T1 per mL for 20 min at 25°C and 400 rpm. Beads were washed twice in T1 buffer and once in phosphatase reaction buffer (Table 12). For dephosphorylation, $1 \times$ antarctic phosphatase reaction buffer (NEB) with 1 U/ μL of antarctic phosphatase and 1 U/ μL of RNase OUT (Invitrogen) were added and the suspension was incubated at 37°C for 30 min and 800 rpm. Beads were washed once in phosphatase wash buffer (Table 12) and twice in polynucleotide kinase (PNK) buffer (Table 12). Beads were resuspended in $1 \times$ T4 PNK reaction buffer A (Fermentas) with a final concentration of 1 U/ μL T4 PNK and 1 U/ μL RNase OUT. Phosphorylation of PAR-CLIP samples was performed using either 1 mM ATP per mL (cold-labeling) or 0.5 μCi of gamma- ^{32}P -ATP per mL (radioactive labeling). The bead suspension was incubated for 1 hr at 37°C and 800 rpm and washed subsequently five times with PNK buffer. For visualization of protein-RNA

interactions, the radioactively labeled samples were subjected to SDS-PAGE analysis. Radioactive RNA-protein bands were detected with the Typhoon FLA 9500 instrument.

2.2.4.1 PAR-CLIP library preparation and high-throughput sequencing

For 3' adapter ligation, beads were resuspended in 1× T4 RNA ligase buffer (NEB) containing 10 U/μL T4 RNA ligase 2 (KQ) (NEB, M0373), 10 μM 3' adapter (Table 7), 1 U/μL RNase OUT, and 15% (w/v) PEG 8000. The bead suspension was incubated for 18 hr at 16°C and 600 rpm. Beads were washed in PNK buffer to remove unligated adapters. For 5' adapter ligation, beads were resuspended in 1× T4 RNA ligase buffer (NEB) containing 6 U/μL T4 RNA ligase 1 (NEB), 10 μM 5' adapter (Table 7), 1 mM ATP, 1 U/μL RNase OUT, 5% (v/v) DMSO, and 10% (w/v) PEG 8000. The suspension was incubated for 4 hr at 24°C and 600 rpm. Beads were washed twice in PNK buffer, and twice in proteinase K buffer (Table 12). Beads were boiled twice at 95°C for 5 min in proteinase K buffer containing 1% SDS and eluted RNA-protein complexes were treated with 1.5 mg/mL proteinase K (NEB) for 2 hr at 55°C. RNA was recovered by acidic phenol/chloroform extraction followed by ethanol precipitation supported by addition of 0.5 μL GlycoBlue (Invitrogen) and 100 μM RT primer (Table 7). Reverse transcription was performed for 1 hr at 44°C using SuperScript III RT (Invitrogen). For PCR amplification, NEXTflex barcode primer and universal primer (Table 7) and Phusion HF master mix (NEB) were added. After PCR amplification, cDNA was size-selected on a precast 4% E-Gel® EX Agarose Gel (Invitrogen) and DNA fragments between ~170 bp and 350 bp were extracted from the gel using the MinElute gel extraction kit (Qiagen) following the manufacturer's instructions. Concatemers and other PCR artifacts in the generated cDNA were eliminated through an additional PCR cycle (One-Step-PCR) using the KAPAHiFi™ PCR Kit (Peqlab Biotechnologie GmbH). PCR was performed using Nextera primers 1 and 2 (Table 7) with a 3 min denaturing step at 94°C, followed by 30 sec at 55°C and 4 min at 72°C. PCR products were subsequently purified using AMPure XP beads with a 1.8x ratio (Beckman Coulter, Inc.), quantified on an Agilent 2200 TapeStation instrument, and sequenced on an Illumina HiSeq 1500 sequencer.

Three independent biological replicates for Bur1, Cdc73, Leo1 and Rtf1 and two TFIIB, Spt6, Set2, Dot1, Set1, Paf1, Ctr9, Ctk2, Ctk1 and Bur2 were performed for this study.

2.2.4.2 PAR-CLIP data processing and analysis

PAR-CLIP data was analyzed as described (Baejen et al. 2014, Schulz et al. 2013), with some modifications by Dr. Michael Lidschreiber. mRNA and CUT transcript annotations were taken from (Pelechano, Wei, and Steinmetz 2013) and (Xu et al. 2009), respectively. Unless stated otherwise, for PAR-CLIP analyses mRNA transcripts were selected to be at least 150 nt away from neighboring transcripts on the same strand. Unless stated otherwise, mRNAs and CUTs were selected to be 800-5000 nt and 350-1500 nt long, respectively. Bidirectional promoters were selected as follows: distance between TSS of mRNAs and divergent CUTs was smaller than 350 bp. Moreover, only mRNAs and CUTs that did not overlap with any other transcripts in the region from their TSS to 400 nt downstream on the same strand were considered. PAR-CLIP processing indices and colocalization measures were calculated essentially as described (Baejen et al. 2014, Schulz et al. 2013).

Data quality control and mapping was essentially performed as described (Baejen et al. 2014). Briefly, quality-trimmed reads are aligned to the *S. cerevisiae* genome (sacCer3, version 64.2.1) using the short read aligner STAR (version 2.5.2b; options: --outFilterMultimapNmax 1, --outFilterMismatchNmax 1, --scoreDelOpen -10000, --scoreInsOpen -10000, --alignSJoverhangMin 10000, --alignSJstitchMismatchNmax 0 0 0 0 (Dobin et al. 2013)). The resulting SAM files are then converted into BAM and PileUp files using SAMTools (Li et al. 2009).

We calculated the P-values for true crosslinking sites as described (Baejen et al. 2014). Briefly, we had to quantitatively model the null hypothesis, i.e., the probability that the T-to-C mismatches observed in reads covering a certain T nucleotide in the genome were not caused by crosslinks between the immunoprecipitated factor and RNA but are due to the other sources of mismatches. Owing to the exquisite sensitivity of our experimental PAR-CLIP procedure, we could set a very stringent P-value cut-off of 0.005 and a minimum coverage threshold of 2. For true crosslinking sites passing our stringent thresholds, the PAR-CLIP-

induced T-to-C transitions strongly dominate over the contributions by sequencing errors and SNPs. For any given T site in the transcriptome, the number of reads showing the T-to-C transition is proportional to the occupancy of the factor on the RNA times the concentration of RNAs covering the T site. Therefore, the occupancy of the factor on the RNA is proportional to the number of reads showing the T-to-C transition divided by the concentration of RNAs covering the T site. This concentration was estimated either from the RNA-Seq read coverage measured under comparable conditions as described (Baejen et al. 2014) or by the read coverage obtained from a Rpb1 PAR-CLIP experiment (this study) and was used to obtain normalized occupancies. We compared RNA and Pol II (Rpb1) normalized occupancy profiles and found that the latter were less prone to biases introduced due to difficulties in measuring unstable RNA species, including CUTs, introns and nascent transcripts downstream of the pA site.

For transcript annotation, we used the recent TIF-Seq data from (Pelechano, Wei, and Steinmetz 2013) to derive TSS and pA site annotations for 5,578 coding genes. TSS and TTS positions of non-coding RNAs were taken from (Xu et al. 2009) for CUTs and from the *Saccharomyces Genome Database* (SGD, version=R64-2-1) for snoRNAs. Annotated transcripts were distance-filtered for downstream analysis to reduce ambiguous signals from overlapping transcripts.

To generate transcript class-averaged heat maps and profiles, transcripts were aligned at their 5'-end ('TSS') and pA sites and either scaled to the same length (median) or cut around the TSS and pA sites before taking the average RNA-binding occupancy at each genomic position. Average occupancies were smoothed (sliding window averaging, window half size of 30 nt) and for each factor individually re-scaled between 0 (0% occupancy) and 1 (100% occupancy) for all figures but Figure 6A, for which all factors were globally scaled to show the relative strength of factor binding. To compare averaged RNA-binding occupancies between transcript classes, they were scaled together by setting min (transcript class 1, transcript class 2) to 0 and max (transcript class 1, transcript class 2) to 1 (Figure 10 and Figure 11).

For generation of non-averaged heat maps of filtered mRNAs (Figure 7 and Figure 9A) transcripts were sorted by length and aligned at their 5'-end ('TSS'). Smoothed occupancies were binned in cells of 20 nucleotide positions times 10

transcripts to avoid aliasing effects due to limited resolution of the plots. The color code displays the occupancy of the PAR-CLIPped factor (with the 97% quantile of these bins scaled to 1). In Figure 9B, all introns (SGD annotation) with lengths between 150 and 650 nt were aligned at the 5'-splice site (5'SS) and the occupancy of each intron is displayed without binning in either x or y direction.

To calculate processing indices (PIs) (Figure 8B) we assume that read counts (not crosslinking sites) N^{down} downstream of a pA site can only occur from pre-mRNAs, $N^{down} = N^{prem}$, whereas read counts N^{up} upstream of a pA site are a mixture of mature mRNA counts N^{mat} and pre-mRNA counts N^{prem} . Therefore, $N^{up} = N^{mat} + N^{prem}$. For increased robustness with regard to different transcript isoforms and uncertainties in the exact location of pA sites, we computed N_i^{up} and N_i^{down} as average of the read counts for each transcript i of a given annotation A :

$$N_i^{up} = 1/50 \sum_{j=pA-75}^{pA-25} readcounts_j$$

$$N_i^{down} = 1/50 \sum_{j=pA+25}^{pA+75} readcounts_j$$

Transcriptome wide averages of N^{up} and N^{down} are defined as

$$N_A^{up} = 1/|A| \sum_i^{|A|} N_i^{up}$$

$$N_A^{down} = 1/|A| \sum_i^{|A|} N_i^{down}$$

Finally the processing index is given by

$$PI = \log_2 \left(\frac{N_A^{down}}{\max(1, (N_A^{up} - N_A^{down}))} \right)$$

Colocalization analysis was done as described (Baejen et al. 2014), with modifications. Briefly, to calculate the tendency of pairs of factors A and B to bind

locations in the transcriptome near each other, we computed the average occupancy of factor B within ± 20 nt of occupancy peaks of factor A (unsmoothed occupancy data). First, crosslink sites of factor A are sorted according to their occupancy and the strongest $n=3000$ sites are selected. For each crosslink site a_i of this selection the maximum occupancy value of factor B m_i^B is identified based on the occupancies of factor B 20 nt \pm around a_i . The average colocalization c is then given by $1/n \sum_i^n m_i^B$. Next, the background binding of factor B is defined as the median of all occupancies of factor B. The colocalization is defined as $\log_2(c/b)$. Finally, we constructed a data matrix containing the calculated colocalization values between all EF pairs. After data normalization the derived colocalization dissimilarity matrix (Euclidean distance) was subjected to average-linkage hierarchical clustering (Figure 8C).

2.2.5 ChIP-Seq

Yeast strains were grown in 600 mL YPD medium (Table 4) to mid-log phase (OD_{600} , ~ 0.8) and treated with formaldehyde (1%, Sigma F1635) for 20 min at 20°C. Crosslinking was quenched with 75 mL of 3 M glycine for 5 min at 20°C. All subsequent steps were performed at 4°C with precooled buffers and in the presence of 1 \times protease-inhibitor mix (Table 12). Cells were collected by centrifugation, washed with 1 \times TBS (Table 12) and twice with ChIP lysis buffer (Table 12). Cell pellets were flash frozen in liquid nitrogen and stored at -80°C . Cells were thawed, resuspended in 2 mL ChIP lysis buffer and lysed as described in 2.2.4. Chromatin was washed with ChIP lysis buffer (15 min at 14,000 rpm) and solubilized via sonication with a Covaris S220 instrument (COVARIS, INC.). Parameters were optimized to yield the average DNA fragment size of 200 bp. This was achieved by sonicating the sample for 18 min using the following parameters: Peak Incident Power (W): 140; Duty Factor: 5%; Cycles per Burst: 200. After chromatin shearing, samples were centrifuged for 40 min at 14,000 rpm. 30 μL of the supernatant were saved as input material and also used as a control for the average chromatin fragment size (2.2.5.1). Magnetic Dynabeads® Protein G (life technologies) were prewashed two times with PBS-T with 1 \times protease-inhibitor mix (Table 12), coated with the respective antibody for 30 min at 4°C and washed again three times. The remaining chromatin sample was immunoprecipitated with antibody-coated beads at 4°C for 3 hr (ChIP of TAP-tagged proteins) or overnight (ChIP with protein-specific antibodies) on a turning wheel.

Immunoprecipitated chromatin was washed 5 times with ChIP wash buffer and one time with TE buffer (Table 12). Chromatin was eluted from the beads for 10 min at 70°C with 120 µL ChIP elution buffer (Table 12). 80 µL TE buffer were added to the IP samples; input samples were mixed with 110 µL ChIP elution buffer and 60 µL TE buffer. IP and input material were incubated with 10 µL RNase A (10 mg/mL) at 37°C for 30 min and subsequently subjected to Proteinase K (20 µL of 20 mg/mL Proteinase K) digestion (37°C for 2 hr) and reversal of crosslinks (65°C overnight). IP DNA and input samples were purified with the QIAquick MinElute PCR Purification Kit (Qiagen) according to the manufacturer's instructions. Elution was performed adding three times 15 µL H₂O to the columns with a 5 min incubation time in between. The DNA concentration of the IP and input samples was determined with Qubit 2.0, dsDNA HS (Invitrogen).

2.2.5.1 Fragment size control

Chromatin fragment size was determined for each experiment. 1 µL of the purified input samples was analyzed on an Agilent 2200 TapeStation instrument using a D1000 ScreenTape (Agilent Technologies, Figure 16A).

2.2.5.2 ChIP-qPCR (quantitative real-time PCR)

Input and immunoprecipitated (IP) samples from ChIP experiments were analyzed via qPCR on 4 housekeeping genes to assess the extent of protein occupancy at different genomic regions. Primer pairs directed against promoter, coding and terminator regions of the housekeeping genes *ADHI*, *ALD5*, *ILV5*, *MUPI*, *PDC1* and *PMA1* as well as against a heterochromatic control region of chromosome V (YER) (Table 8) were designed and the corresponding PCR efficiencies determined. All primer pairs used in this study had PCR efficiencies in the range of 95-100%. PCR reactions contained 1 µL DNA template, 0.8 µM of each primer and 10 µL 2× SensiFAST SYBR No-ROX Mix (BIOLINE). Quantitative real-time PCR was performed on a qTOWER 2.2 Real-Time System (Analytik Jena AG) using a 2 min denaturing step at 95°C, followed by 40 cycles of 5 sec at 95°C, 10 sec at 61°C and 15 sec at 72°C. Threshold cycle (Ct) values were determined using the corresponding qPCRsoft 3.1 software. The enrichment of specific DNA sequences in the IP sample over the input sample (percent input) was determined as following:

$$\% \text{ Input} = 100 \times 2^{((Ct(\text{Input}) - \log_2(100)) - Ct(\text{IP}))}$$

2.2.5.3 Library preparation and high-throughput sequencing of ChIP samples

For Illumina sequencing of ChIP samples, 1–10 ng of IP or input DNA were used for library preparation according to the manufacturer's recommendations using the ThruPLEX® DNA-Seq Kit. Libraries were size selected to exclude adapters and adapter dimers using AMPure XP beads with a 1.0x ratio. Samples were qualified on an Agilent 2200 TapeStation instrument (Figure 16B). ChIP libraries were quantified with Qubit 2.0, pooled and sequenced on an Illumina HiSeq 1500 sequencer.

2.2.5.4 ChIP-Seq data processing and analysis

ChIP-Seq data was processed and analyzed by Dr. Michael Lidschreiber. Paired-end 50 bp reads were aligned to the *S. cerevisiae* genome (sacCer3, version 64.2.1) using the short read aligner Bowtie (version 2.2.3) (Langmead and Salzberg 2012). SAMTools was used to quality filter SAM files (Li et al. 2009). Alignments with MAPQ smaller than 7 (-q 7) were skipped and only proper pairs (-f99, -f147, -f83, -f163) were selected. The BEDTools toolset (Quinlan and Hall 2010) was used to obtain coverage tracks that were subsequently imported into R/Bioconductor where further processing of the data was carried out. Normalization between IP and input was done using the signal extraction scaling (SES) factor obtained with the estimateScaleFactor function from deepTools (Ramirez et al. 2014) with options: -l 100 -n 100000 and the median fragment size (-f) estimated from the data (around 200 bp). ChIP enrichments were obtained by dividing SES-normalized IP intensities by the corresponding input intensities: $\log_2(\text{IP}/\text{Input})$.

The same transcript annotations as for PAR-CLIP data analysis (see above) were used for ChIP-Seq data analysis, except that filtering criteria had to be more stringent due to the lack of strand-specificity and lower resolution of ChIP-Seq data. Thus, for Figure 15, the distance filtering between transcripts was increased to 200 bp and transcripts on both strands were considered.

2.2.6 Chromatin association assay

Yeast cultures were grown in 200 mL of YPD medium (Table 4) at 30°C to mid-log phase (OD_{600} , ~ 0.8). Subsequent steps were performed at 4°C with precooled buffers and in the presence of 1× protease-inhibitor mix (Table 12). Cells were collected by centrifugation, washed with 1× TBS buffer and with ChIP lysis buffer (Table 12). Cell pellets were flash frozen in liquid nitrogen and stored at -80°C. Pellets were thawed, resuspended in 1 mL lysis buffer, and disrupted via beat beating (2.2.4). The lysate was divided into two samples. One half was treated with 7.5 U of RNase A and 300 U of RNase T1 (Ambion); the other half was treated with the same volume of RNase storage buffer (Table 12). After 30 min incubation at room temperature, chromatin was isolated by centrifugation at 15,000 rpm for 15 min. Chromatin was solubilized in 1 mL ChIP lysis buffer via sonication with a Covaris S220 instrument (COVARIS, INC.). Chromatin solutions were then analyzed by SDS-PAGE and Western blotting (2.2.3.2, 2.2.3.3) against the C-terminal TAP tag of the analyzed factor and against H3, which served as loading control. Three independent biological replicates for TFIIB, Rpb1, Rtf1, Paf1, Ctk2, Bur1, Set2 and Spt6 and two independent biological replicates for Leo1, Ctk1, Bur1, Set1, Dot1 and Spt5 were performed. Band intensities were quantified using ImageJ software (1.49v). For statistical analysis, multiple group comparisons were done by one-way ANOVA with Dunnett post-hoc test. Data are presented as mean ± SD. Differences were considered significant when $p < 0.5$ (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; n.s. = not significant).

2.2.7 Purification of recombinant CTDK-I protein complex

2.2.7.1 Cloning of CTDK-I

The full-length subunits of the CTDK-I complex, Ctk1, Ctk2 and Ctk3, were amplified from genomic *S. cerevisiae* DNA (BY4741 strain) as described in 2.2.2.1 using primers listed in Table 6. Gel purified DNA inserts (2.2.2.2) were cloned into modified pFastBac vectors (Series-438 *MacroBac* Expression Vectors; Table 11; a gift of Scott Gradia, UC Berkeley) via ligation independent cloning (LIC). Prior ligation, vectors were linearized with SspI and gel purified. Purified vectors and DNA inserts were treated with T4 DNA polymerase in the presence of dGTP or dCTP, respectively, and ligated for 30 min at RT. Ctk1 and Ctk3 were cloned into the 438-A

vector; Ctk2 was cloned into the 438-C vector bearing an N-terminal 6xHis-MBP-tag followed by a tobacco etch virus protease cleavage site (TEV site). After ligation, plasmids were transformed into *E. coli* XL1Blue cells (Table 1). Ligation reaction was added to 50 μ L competent cells and incubated on ice for 10 min. Suspension was heat-shocked at 42°C for 45 sec and then cooled down on ice for 2 min. 450 μ L LB medium was added and cells were incubated for 4 hr at 37°C and 400 rpm. Cells were harvested by centrifugation for 5 min at 4,000 rpm, resuspended in 150 μ L LB medium and plated on ampicillin selective LB plates. Plates were incubated overnight at 37°C and single colonies were picked and grown overnight at 37°C in LB medium containing ampicillin. Plasmid DNA was isolated using QIAprep spin miniprep Kit (Qiagen) and sequence-verified (Seqlab, Göttingen). Ctk1 and Ctk3 subunits, encoded in the 438-A plasmids, were combined into the Ctk2_438-C plasmid by successive rounds of LIC reactions (Figure 4) using *Swa*I and *Pme*I for restriction digests. The plasmid containing all three subunits of the CTDK-I complex was amplified and verified as described above. Each subunit is preceded by a PolH promoter and followed by an SV40 termination site.

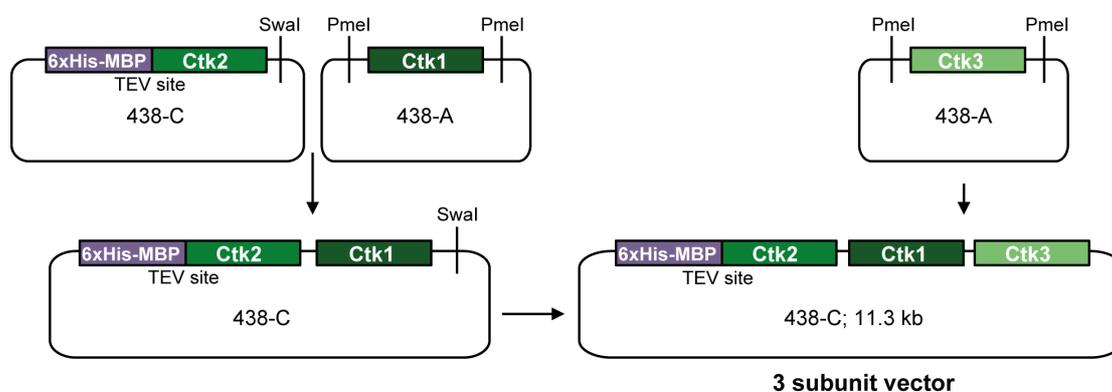


Figure 4 Schematic procedure of LIC reactions to generate a plasmid encoding all three subunits of the CTDK-I kinase complex.

2.2.7.2 Expression of CTDK-I

Insect cell culture work was performed by Dr. Seychelle M. Vos. Purified plasmid DNA (0.5 μ g) was electroporated into DH10EMBacY cells to generate bacmids (Berger, Fitzgerald, and Richmond 2004). Bacmids were prepared from positive clones by isopropanol precipitation and transfected into Sf9 cells grown in Sf-900 III

SFM (ThermoFisher) with X-tremeGENE9 transfection reagent (Sigma) to generate V0 virus. V0 virus was harvested 72 hr after transfection. V1 virus was produced by infecting 25 mL of Sf21 cells grown at 27°C, 300 rpm with V0 virus (1E6 cell/mL, 1:50 (v/v) cells:virus). V1 viruses were harvested 48 hr after proliferation arrest and stored at 4°C. For protein expression, 600 mL of Hi5 cells (1E6/mL) grown in ESF921 medium (Expression Technologies) were infected with 200 µL of V1 virus and grown for 72 hr at 27°C. Cells were harvested by centrifugation (238xg, 4°C, 30 min), resuspended in CTDK-I lysis buffer at 4°C containing 1× protease-inhibitor mix (Table 12), snap frozen, and stored at -80°C.

2.2.7.3 Purification of complete CTDK-I

Protein purification steps were performed at 4°C. Frozen cell pellets were thawed and lysed by sonication with CTDK-I lysis buffer (Table 12) containing 1× protease-inhibitor mix. Lysates were clarified by centrifugation in an A27 rotor (ThermoFisher) (26,195 g, 4°C, 30 min), followed by ultracentrifugation in a Type 45 Ti rotor (Beckman Coulter) (235,000 g, 4°C, 60 min). Clarified lysates were filtered through 0.8 mm syringe filters (Millipore) and applied to a 5 mL HisTrap column (GE Healthcare) equilibrated in CTDK-I lysis buffer. HisTrap columns were washed with 10CV of CTDK-I lysis buffer followed by 5CV of high salt wash buffer (Table 12) containing 1× protease-inhibitor mix and 5CV of CTDK-I lysis buffer. An amylose column (New England Biolabs) equilibrated in CTDK-I lysis buffer was directly coupled to the HisTrap column. Protein was eluted from the HisTrap column by a gradient from 0–100% nickel elution buffer (Table 12) containing 1× protease-inhibitor mix, after which the HisTrap and amylose column were decoupled. The amylose column was washed with 5CV of CTDK-I lysis buffer and protein was eluted with amylose elution buffer (Table 12) containing 1× protease-inhibitor mix. Peak fractions were analyzed by SDS-PAGE (2.2.3.2). Amylose column fractions containing CTDK-I were combined with 1.5 mg of His6-TEV protease and dialyzed overnight at 4°C in a Slide-A-Lyzer (3–12 mL 10 kDa MWCO) (ThermoFisher) against 1 L of lysis buffer. Protein was removed from the Slide-A-Lyzer cassette and applied to a 5 mL HisTrap column to remove uncleaved protein and TEV protease. Protein was concentrated in an Amicon 15 mL centrifugal 30 kDa MWCO concentrator (Millipore) to 500 µL. The protein was applied to a Superdex 200

Increase 10/300 column (GE Healthcare) equilibrated in SE buffer (Table 12). Peak fractions were analyzed by SDS-PAGE. Pure fractions were concentrated as described above to 500 μ L, aliquoted, flash frozen, and stored at -80°C . Protein preparation yielded ~ 4 mg of full-length CTDK-I from 3.8 L of insect cell culture.

2.2.8 CTDK-I kinase activity assays

Purified human GST-CTD (see 2.2.8.1) and *S. cerevisiae* Pol II (Plaschka et al. 2016) were used as kinase substrates. *In vitro* kinase assays were performed in 20 μ L reactions. Recombinant purified CTDK-I (0.4 μ M) was mixed together with either GST-CTD (10 μ M) or Pol II (2 μ M) and the following final conditions: 1 mM DTT, 100 mM NaCl, 30 mM Na•HEPES pH 7.5, 4% (v/v) glycerol and 3 mM MgCl_2 . Reactions were incubated at 30°C for 5 min; 1 μ L “0 min” time point was taken before adding 3 mM ATP. Reactions were then incubated at 30°C with constant shaking at 300 rpm. 2 μ L were taken for each time point (reaction with CTD substrate: 5, 10, 20, 30, 60 and 120 min; reaction with Pol II substrate: 2, 4, 6, 10, 20, 30, 60 and 90 min). The reactions were stopped with LDS-loading buffer and heated at 95°C for 5 min before SDS-PAGE (2.2.3.2). Western blotting was performed as described above (2.2.3.3) using anti-Ser2-P as primary antibody and anti-rat IgG (HRP) as secondary antibody (Table 10).

2.2.8.1 His6-TEV-GST 52x human CTD expression and purification

Dr. Seychelle M. Vos purified His6-TEV-GST 52x human CTD. For expression, B121(DE3)pLysS *E. coli* cells were used. Cells were grown in 2xYT to OD_{600} of 0.5 (37°C , 160 rpm), changed to 18°C , and induced overnight with 0.5 mM IPTG. Protein purification steps were performed at 4°C . 4L B121(DE3)pLysS) frozen cells were thawed and lysed by sonication. Lysates were clarified by centrifugation in an A27 rotor (ThermoFisher) (26,195 g, 4°C , 30 min). Clarified lysate was applied to a 5 mL HisTrap nickel column (GE Healthcare) equilibrated in lysis buffer (150 mM NaCl, 20 mM Na•HEPES pH 7.4, 1 mM DTT, 10% glycerol, 30 mM imidazole) with $1\times$ protease-inhibitor mix (Table 12). Column was then washed in high salt wash buffer (800 mM NaCl, 20 mM Na•HEPES pH 7.4, 1 mM DTT, 10% glycerol, 30 mM imidazole) containing $1\times$ protease-inhibitor mix followed by lysis buffer. A 5 mL HiTrap Q column was equilibrated in lysis buffer and added in line after the nickel

column. Protein was eluted with a gradient of 0-100% nickel elution buffer (150 mM NaCl, 20 mM Na•HEPEs pH 7.4, 1 mM DTT, 10% glycerol, 500 mM imidazole) containing 1× protease-inhibitor mix. Following imidazole elution, the nickel column was removed and the Q column washed in 150 mM NaCl, 20 mM Na•HEPEs pH 7.4, 1 mM DTT, 10% glycerol and 1× protease-inhibitor mix. Gradient elution was performed with 1 M NaCl, 20 mM Na•HEPEs pH 7.4, 1 mM DTT and 1× protease-inhibitor mix. Peak fractions from Q column elution were concentrated in a 5 mL 30 kDa MWCO concentrator to 500 μL. Protein was centrifuged for 10 min at 15k prior loading to a Superdex 200 Increase 10/300 column (GE Healthcare), equilibrated in SE buffer (300 mM NaCl, 20 mM Na•HEPEs pH 7.4, 1 mM DTT, 10% glycerol). Peak fractions were analyzed by SDS-PAGE (2.2.3.2), pure fractions concentrated in 5 mL 30 kDa MWCO Amicon Millipore concentrator to 40 μM, aliquoted, flash frozen, and stored at −80°C.

2.2.9 Fluorescence anisotropy assays with CTDK-I

5'-FAM labeled ssRNA and dsDNA were obtained from Integrated DNA Technologies and dissolved in water to 100 μM. ssRNA sequences had different GC-contents and were either A- or U-rich (Table 9). The dsDNA sequence corresponds to the 45% GC, A-rich ssRNA sequence. 24% GC, A-rich, 24% GC, U-rich and 45% GC, A-rich sequences correspond to natural coding sequences in *S. cerevisiae*. RNA oligos were unfolded by incubating the RNA at 95°C for 1 min and transferring to ice for 10 min. Oligonucleotides were diluted in water for all experiments.

Purified CTDK-I was serially diluted in two fold steps in dilution buffer (Table 12). Nucleic acids (8 nM final concentration) were added on ice and the reaction was incubated for 10 min. The assay was brought to a final volume of 30 μL and incubated for 20 min at RT in the dark (final conditions: 100 mM NaCl, 2 mM MgCl₂, 20 mM Na•HEPES pH 7.4, 1 mM TCEP, 4% glycerol, 0.01 mg/mL BSA and 5 μg/mL yeast tRNA (Sigma) as a competitor for non-specific binding). 18 μL of each reaction were transferred to a Greiner 384 Flat Bottom Black Small volume plate.

Fluorescence anisotropy (FA) was measured at 30°C with an Infinite M1000Pro reader (Tecan) with an excitation wavelength of 470 nm (± 5 nm), an emission wavelength of 518 nm (± 20 nm) and a gain of 63. All experiments were done in triplicate and analyzed with GraphPad Prism Version 7. Binding curves were fitted with a single site quadratic binding equation:

$$FA = \left(\frac{B_{max} * ([P] + [L] + K_{d,app}) - \sqrt{([P] + [L] + K_{d,app})^2 - 4 ([P] * [L])}}{2 * [L]} \right)$$

where B_{max} is the maximum specific binding, L is the concentration of nucleic acid, P is the concentration of CTDK-I, $K_{d,app}$ is the apparent disassociation constant for CTDK-I and nucleic acid. Error bars represent the standard deviation from the mean of three experimental replicates.

3 Results

Some of the results presented here were obtained in collaboration with colleagues. For detailed author contributions see publications (page 7) and section 2.2.

3.1 Elongation factors directly crosslink to RNA *in vivo*

To investigate whether elongation factors (EFs) interact with RNA *in vivo*, we used photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) (Hafner et al. 2010), a method that detects and maps direct protein-RNA interactions without chemical crosslinkers. We applied our recently optimized PAR-CLIP protocol (Baejen et al. 2014) to 14 EFs of the yeast *Saccharomyces (S.) cerevisiae* (Figure 5, Table 13, 2.2.4). These EFs included Spt5, Spt6, the five Paf1C subunits Cdc73, Ctr9, Leo1, Rtf1, and Paf1, the kinases Bur1 and Ctk1, the cyclins Bur2 and Ctk2, and the histone methyltransferases Set1, Set2, and Dot1.

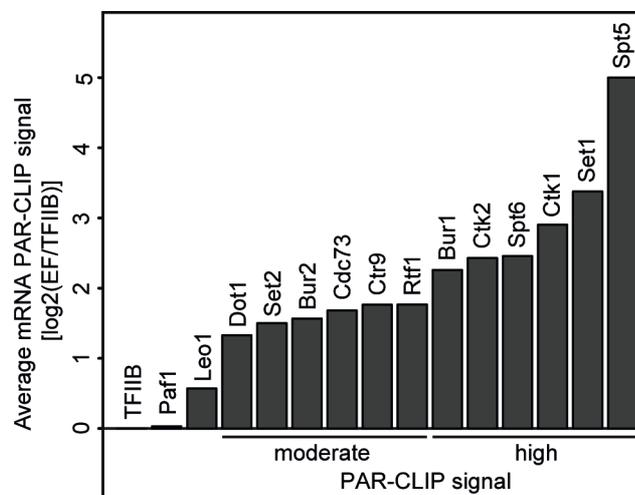


Figure 5 Many elongation factors (EFs) bind RNA *in vivo*. PAR-CLIP signal strength for EFs varies. The bar plots show log₂ fold-enrichments of transcript-averaged PAR-CLIP signals over the averaged PAR-CLIP signal for initiation factor TFIIB, which shows background RNA binding. Averaged PAR-CLIP signals were calculated by taking mean transcript PAR-CLIP signals averaged over all mRNAs, which were filtered to be 800-5,000 nt long and at least 150 nt away from neighboring transcripts (2,532 mRNAs).

For 12 of the above mentioned 14 EFs we obtained PAR-CLIP signals that were more than two-fold above background, showing that these EFs interact with RNA *in vivo* (Figure 5, Figure 6A). We obtained between 42,000 and 520,000 high-confidence protein-RNA crosslinking sites per factor with p-values below 0.005 (Table 13). The collected data sets were highly reproducible (Figure 6B). To estimate background RNA binding, we collected PAR-CLIP data for the transcription initiation factor TFIIB that is recruited to promoter DNA before nascent RNA is made (Sainsbury, Bernecky, and Cramer 2015). Only very low levels of background binding were observed, further emphasizing the significance of EF-RNA interactions detected by UV crosslinking.

Table 13 PAR-CLIP analysis of elongation factors (EFs).

EF	Complex ^a	Number of crosslink sites ^b
Bur1	BUR kinase complex	77,931
Bur2		46,293
Ctk1	CTDK-I	129,352
Ctk2		98,993
Cdc73		57,603
Ctr9		55,807
Leo1	Paf1C	27,665
Paf1		20,742
Rtf1		60,068
Set1	COMPASS	189,723
Set2		68,875
Dot1		42,848
Spt5 ^c	DSIF	517,568
Spt6		93,902
TFIIB ^d		16,686

^aDSIF, DRB sensitivity inducing factor; CTDK, C-terminal domain kinase; Paf1C, Paf1 complex; COMPASS, Complex Proteins Associated with Set1

^bAverage number of crosslink sites with p-values < 0.005

^c(Baejen et al. 2017)

^dInitiation factor, used to determine the level of RNA background crosslinking

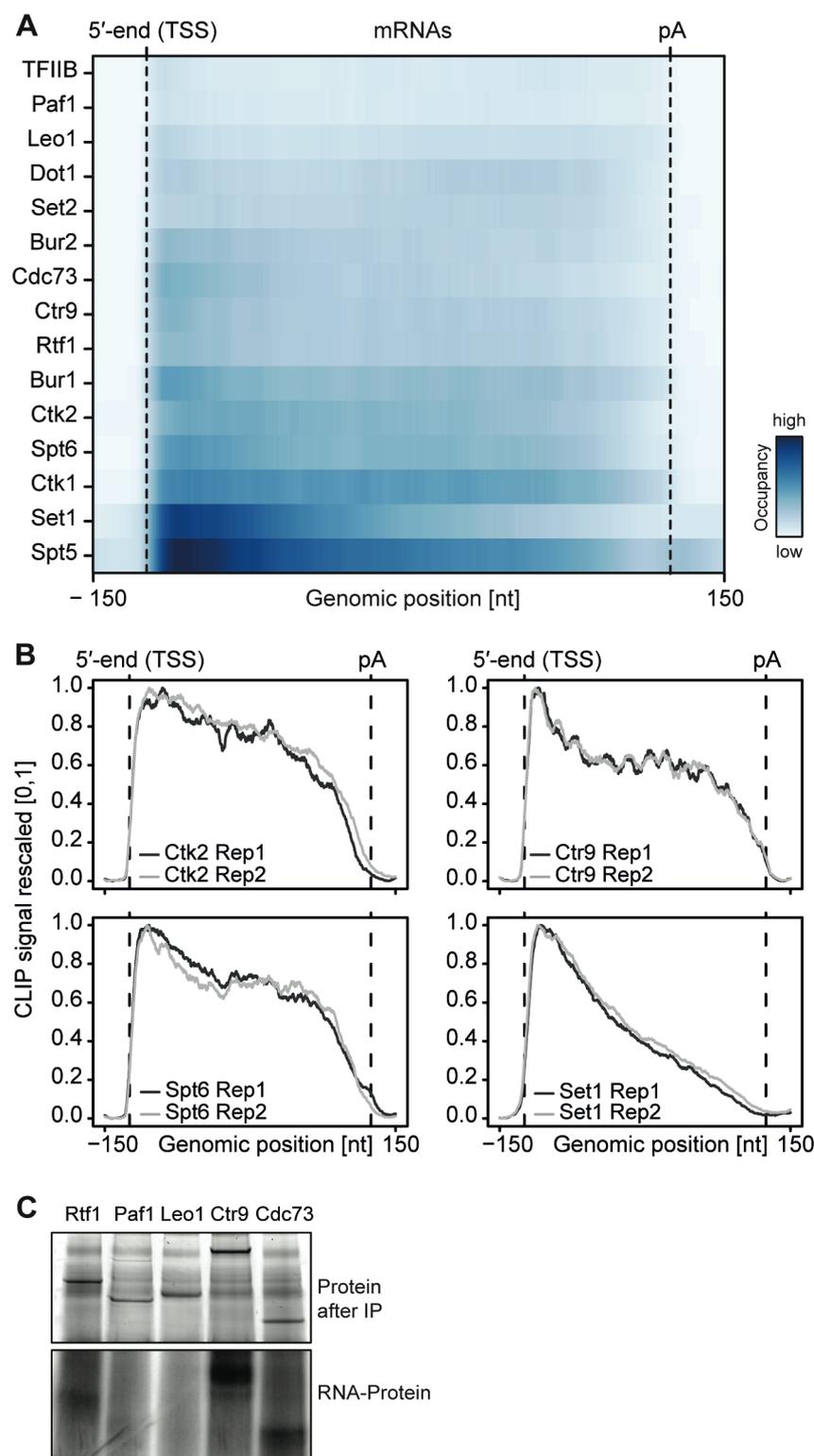


Figure 6 (A) Smoothed, raw PAR-CLIP signals (as measured by the number of PAR-CLIP U-to-C transitions per U site) over a set of 2,532 selected mRNAs were aligned at their 5'-end (TSS), scaled to a common length, then averaged (2.2.4.2). The color code shows the PAR-CLIP signal relative to the maximum PAR-CLIP signal of all profiles (dark blue). Since PAR-CLIP signals of Set1 and Spt5 were much higher than those of other EFs, Set1 and Spt5 occupancies were divided by a factor of 1.5 and 3, respectively, for visualization purposes. (B) Replicate (Rep) measurements show very high correlations. Smoothed, averaged PAR-CLIP profiles of replicate measurements for Ctk2, Cdc73, Ctr9 and Set1. (C) Paf1C interacts

with RNA through its subunits Rtf1, Ctr9 and Cdc73. (Top) SDS-PAGE analysis of Paf1C subunits after IP (InstantBlue stain). Paf1C subunits can be individually pulled down with IgG coated beads against the C-terminal TAP tag of each of the five Paf1C subunits. (Bottom) Phosphorimage of SDS-gel with Paf1C subunits after IP and radioactive labeling of co-precipitated RNA. RNA co-precipitated only with Rtf1, Ctr9 and Cdc73.

We then classified EFs into factors with moderate and high PAR-CLIP signals, based on their fold enrichments (>2 and >4 -fold, respectively) over background TFIIB signals (Figure 5). Spt5, Set1, Ctk1, Spt6, Ctk2 and Bur1 showed high PAR-CLIP signals (Figure 5, Figure 6A, Table 13). EFs with moderate signals included Rtf1, Ctr9, Cdc73, Bur2, Set2 and Dot1. PAR-CLIP signals were clearly specific for individual subunits of known complexes. For instance, only the Paf1C subunits Rtf1, Cdc73 and Ctr9 bound RNA according to the PAR-CLIP results, and the same subunits were detected after IP and radioactive labeling of co-precipitated RNA (Figure 6C). A low background signal was observed for other subunits, whereas the enriched bands were due to the protein of interest. These data revealed that many EFs directly bind RNA *in vivo*, including the Pol II Ser2 kinases Ctk1 and Bur2 and the histone H3 methyltransferases Set1, Set2 and Dot1.

3.2 Comparisons of PAR-CLIP data require normalization

We have previously noted the importance of normalizing the raw PAR-CLIP signal, as measured by the number of U-to-C transitions per U site, to account for differences in RNA abundance (Baejen et al. 2014). Briefly, the raw PAR-CLIP signal is proportional to the occupancy of the factor on RNA and to the concentration of RNAs covering the U site. Therefore, normalization is crucial to enable comparison of PAR-CLIP signals between individual transcripts and transcript classes. Relative occupancies can be estimated by dividing the observed PAR-CLIP signal by RNA-Seq reads that have been obtained under the same experimental conditions (Baejen et al. 2014). An alternative approach is to divide the observed PAR-CLIP signal by a PAR-CLIP signal obtained for Pol II (Baejen et al. 2017), although this is only suitable for proteins that associate with nascent RNA during transcription, which is the case for the EFs studied here.

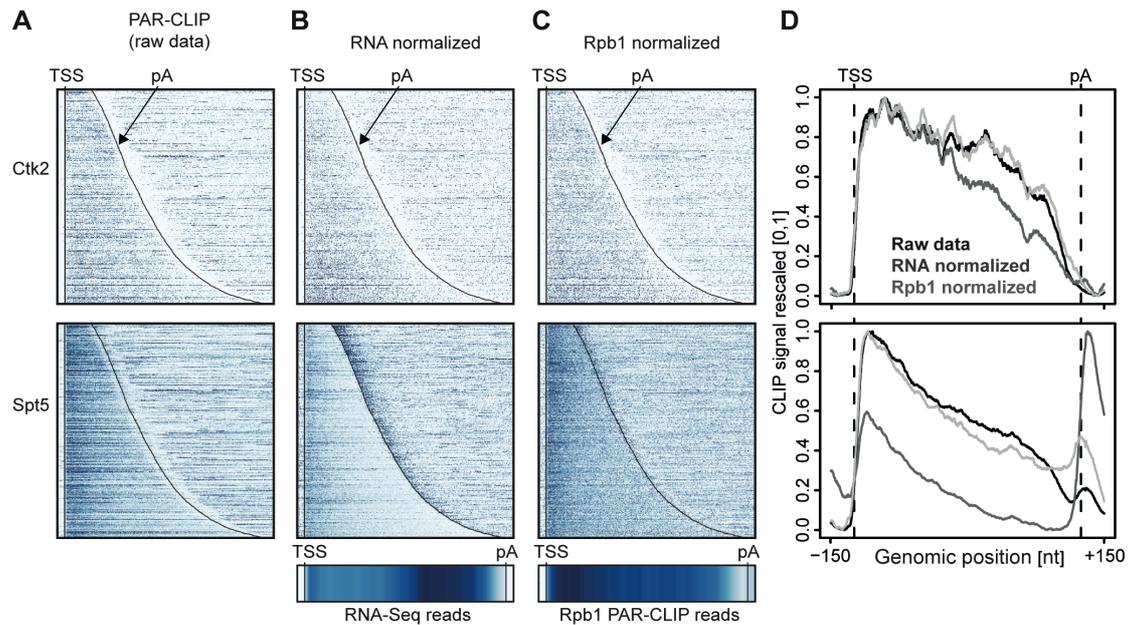


Figure 7 Normalization of PAR-CLIP data shown for two representative EFs, Ctk2 (top) and Spt5 (bottom). (A) Smoothed, raw RNA-binding strength as measured by the number of PAR-CLIP U-to-C transitions per U site for all mRNAs sorted by length and aligned at their RNA 5'-end (transcription start site, TSS). (B) Relative occupancy estimated by dividing the number of U-to-C transitions for each U site by the RNA-Seq signal at the corresponding genomic position for all mRNAs. A heat map showing the transcript-averaged RNA-Seq reads for all mRNAs scaled to the same length is shown below. (C) Relative occupancy estimated by dividing the number of U-to-C transitions for each U site by the Rpb1 PAR-CLIP signal at the corresponding genomic position for all mRNAs. A heat map showing the transcript-averaged Rpb1 PAR-CLIP reads for all mRNAs scaled to the same length is shown below. (D) Smoothed, raw and normalized PAR-CLIP signals as shown in A-C, but averaged over all mRNAs. Before averaging RNA occupancy profiles were aligned at the RNA 5'-end and length-scaled such that the 5'-ends and pA sites coincided.

Figure 7 shows how the two different normalization methods affect EF occupancy profiles on mRNA transcripts. For two representative EFs, Ctk2 and Spt5, the raw data (Figure 7A) was either normalized with RNA-Seq reads (Figure 7B) or with reads from Pol II (Rpb1 subunit) PAR-CLIP data (Figure 7C). Meta-transcript profiles are shown in Figure 7D. In the case of Ctk2, the raw data profile and the Pol II normalized profile look very similar, whereas the RNA-normalized profile shows slightly less occupancy of Ctk2 in the 3' part of the transcripts, due to the slightly higher RNA-Seq signal in this region (Figure 7B, bottom). The PAR-CLIP signal for Spt5 is enriched around the 5'-end of mRNAs, decreases towards the 3'-end, and this was independent of the normalization approach (Figure 7D, bottom). However, Spt5 signals peak just downstream of the pA site, and the size of this peak varies in function of the normalization approach. This is due to the intrinsic instability

of transcripts downstream of the pA site, which reduces the number of RNA-Seq reads and artificially increases the PAR-CLIP peak after RNA-Seq-based normalization.

Taken together, the PAR-CLIP metagene profiles over stable transcripts were largely independent of the type of normalization used. However, proper normalization becomes critical when crosslinking to unstable RNAs is investigated. Since we were interested in comparing EF occupancies between transcript classes, including unstable RNAs, we used Pol II PAR-CLIP normalization to calculate normalized EF PAR-CLIP occupancies and used these for further analysis.

3.3 Differences in EF occupancy along RNAs

3.3.1 EF localization along mRNA transcripts

To localize EFs on transcripts, we mapped the Pol II normalized PAR-CLIP occupancies onto transcripts in different classes (2.2.4.2). We then calculated factor occupancies for 2,532 mRNA transcripts that were filtered to reduce ambiguous signals from overlapping transcripts. We calculated heat maps with occupancies averaged around the transcript 5'-end, which corresponds to the transcription start site (TSS), and around the polyadenylation (pA) site (Figure 8A). The obtained profiles were also visible on individual transcripts (Figure 9A).

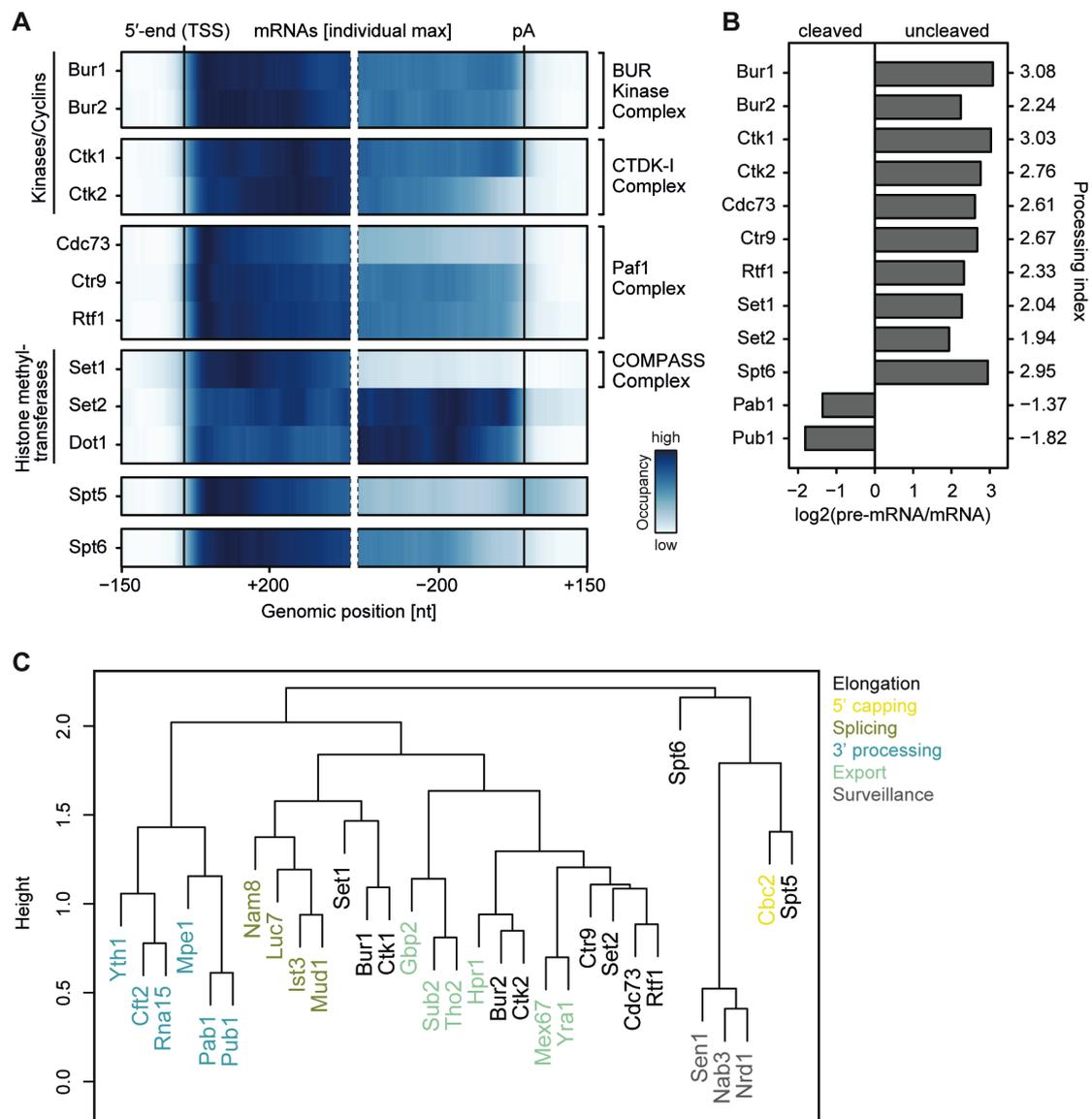


Figure 8 (A) mRNA-binding profiles of EFs. Smoothed, transcript-averaged Pol II normalized PAR-CLIP occupancy profiles of EFs centered around the transcript 5'-end (transcription start site, TSS) [-150 nt to +400 nt] and pA site [-400 nt to +150 nt] of a set of 2,532 filtered mRNAs (compare **Figure 5**). Only factors with average RNA-binding occupancies >2-fold above background are shown. The Spt5 PAR-CLIP profile reveals a peak downstream of the pA site that is discussed in detail elsewhere (Baejen et al. 2017). The color code shows the occupancy relative to the maximum occupancy per profile (dark blue). **(B) EFs bind to pre-mRNA.** Processing indices (PIs) measure preferential binding of factors to uncleaved pre-mRNA with respect to cleaved RNA, computed as \log_2 odds ratios uncleaved versus cleaved RNA bound by the factor (2.2.4.2). The PIs for Pab1 and Pub1, as typical factors binding mature mRNA (Baejen et al. 2014), are shown for comparison. **(C) Colocalization of factor crosslinking sites on transcripts.** Euclidean distances between pairwise colocalization measures were subjected to average-linkage hierarchical clustering (2.2.4.2). The cluster dendrogram shows similarities in crosslinking locations on transcripts between EFs and published RNA processing factors (Baejen et al. 2014, Schulz et al. 2013).

Generally, PAR-CLIP occupancies were high at the 5'-end of mRNAs and decreased shortly before the pA site, with few exceptions (Figure 3A). First, the histone methyltransferases Set2 and Dot1, for which the corresponding methylation marks accumulate in gene bodies (Bannister et al. 2005, Pokholok et al. 2005), showed more RNA-binding sites over transcript bodies. Second, Set1 crosslinked to mRNAs mainly near the beginning of transcripts, which was expected since Set1 and its methylation mark, H3K4me3, are observed in promoter-proximal regions of genes (Ng et al. 2003). Third, the kinases Ctk1 and Bur1 and their cyclin partners Ctk2 and Bur2 were enriched near the 5'-end but also in the transcript body. The 5'-peak for Bur1-Bur2 slightly preceded that of Ctk1-Ctk2. The three Paf1C subunits Cdc73, Ctr9 and Rtf1 showed similar occupancy profiles as the kinases but with a focused peak at the 5'-end. Fourth, Spt5 and Spt6 showed high PAR-CLIP occupancy at the 5'-end of mRNAs and decreased occupancy towards the pA site. This analysis revealed specific differences in EF localization on mRNAs, and additionally suggested that EFs bind nascent RNA during transcription.

3.3.2 EFs bind nascent pre-mRNA

To test whether EFs interact with nascent pre-mRNA or with spliced, mature mRNA, we measured factor occupancies at introns, which are co-transcriptionally spliced out and subsequently degraded (Carrillo Oesterreich et al. 2016). All EFs crosslinked to introns (Figure 9B), indicating that they bind pre-mRNA. Most EFs bound to introns with a frequency that was comparable to that at exons, although Spt5 and Set1 showed slightly higher occupancy within introns, whereas Bur2, Set2 and Dot1 showed lower occupancy (Figure 9B). Taking into account that splicing generally occurs co-transcriptionally (Kornblihtt et al. 2004, Tennyson, Klamut, and Worton 1995, Listerman, Sapra, and Neugebauer 2006), our data show that EFs interact with nascent pre-mRNA. However, only ~4% of yeast genes contain introns (Qin et al. 2016), preventing general statements related to all pre-mRNAs. We therefore calculated a processing index (PI) that measures preferential binding of factors to uncleaved pre-mRNA with respect to cleaved RNA (2.2.4.2) (Baejen et al. 2014). All EFs showed positive PIs, indicating binding to pre-mRNA, in contrast to the negative PIs that we previously obtained for typical RNA binders of processed, mature mRNA,

such as Pab1 and Pub1 (Figure 8B) (Baejen et al. 2014). We conclude that EFs preferentially interact with nascent pre-mRNA.

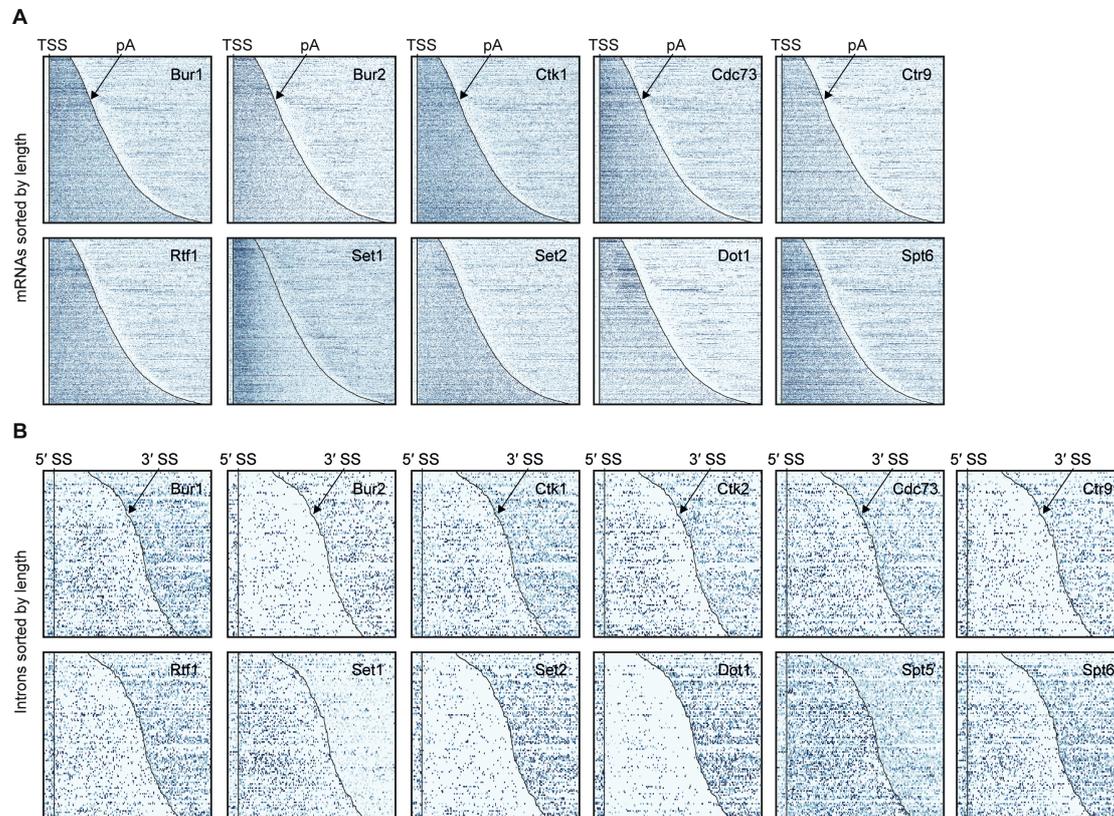


Figure 9 Non-averaged elongation factor RNA occupancies over mRNAs and introns. **(A)** PAR-CLIP occupancies of Pol II normalized EFs transcript-wise. Smoothed occupancy profiles derived from PAR-CLIP data for a set of 2,532 selected mRNAs. Transcripts were sorted by length and aligned at their 5'-end (transcription start site, TSS). Plots for Ctk2 and Spt5 are shown in Figure 2C. **(B)** Smoothed Pol II normalized PAR-CLIP occupancy profiles over all introns. Each line represents an intron, and introns were sorted by length and aligned at their 5' splice site (5' SS). Only introns of lengths between 150 and 650 nt are shown.

We next investigated where EFs localize on RNAs in relation to previously mapped mRNA biogenesis factors (Baejen et al. 2014). We determined the extent of factor colocalization by computing the average occupancy of factor A within ± 20 nucleotides (nt) around RNA-binding sites of factor B and subjected the pairwise colocalization measures to hierarchical clustering (Figure 8C, 2.2.4.2). We found that Spt5 colocalizes with the Cbc2 subunit of the cap-binding complex, consistent with its recruitment during early elongation. Both Ctk1 and Bur1 colocalized with binding sites of Set1 and splicing factors. Paf1C subunits colocalized with Set2, whereas RNA 3'-processing and surveillance factors formed separate groups (Figure 8C). Together

these data show a distinct distribution of EFs over RNAs, and suggested that EFs cooperate with other mRNA biogenesis factors during pre-mRNA binding.

3.3.3 Most EFs preferentially interact with coding transcripts

We next analyzed our PAR-CLIP data for EF binding to non-coding Pol II transcripts including short-lived cryptic unstable transcripts (CUTs), which often arise from upstream antisense transcription of bidirectional promoters (Wyers et al. 2005, Xu et al. 2009). We selected CUTs with a minimum length of 350 nt and compared transcript-averaged RNA-binding occupancies between CUTs and mRNAs (Figure 10A). This revealed that EFs bind to these transcript classes with distinct preferences. Spt5 was equally distributed between CUTs and mRNAs whereas Set1 preferentially bound mRNAs. All other EFs were depleted at CUTs relative to their mRNA occupancies (Figure 10A). This was essentially independent of RNA length (Figure 11A). Thus, most EFs preferentially crosslink to coding RNAs.

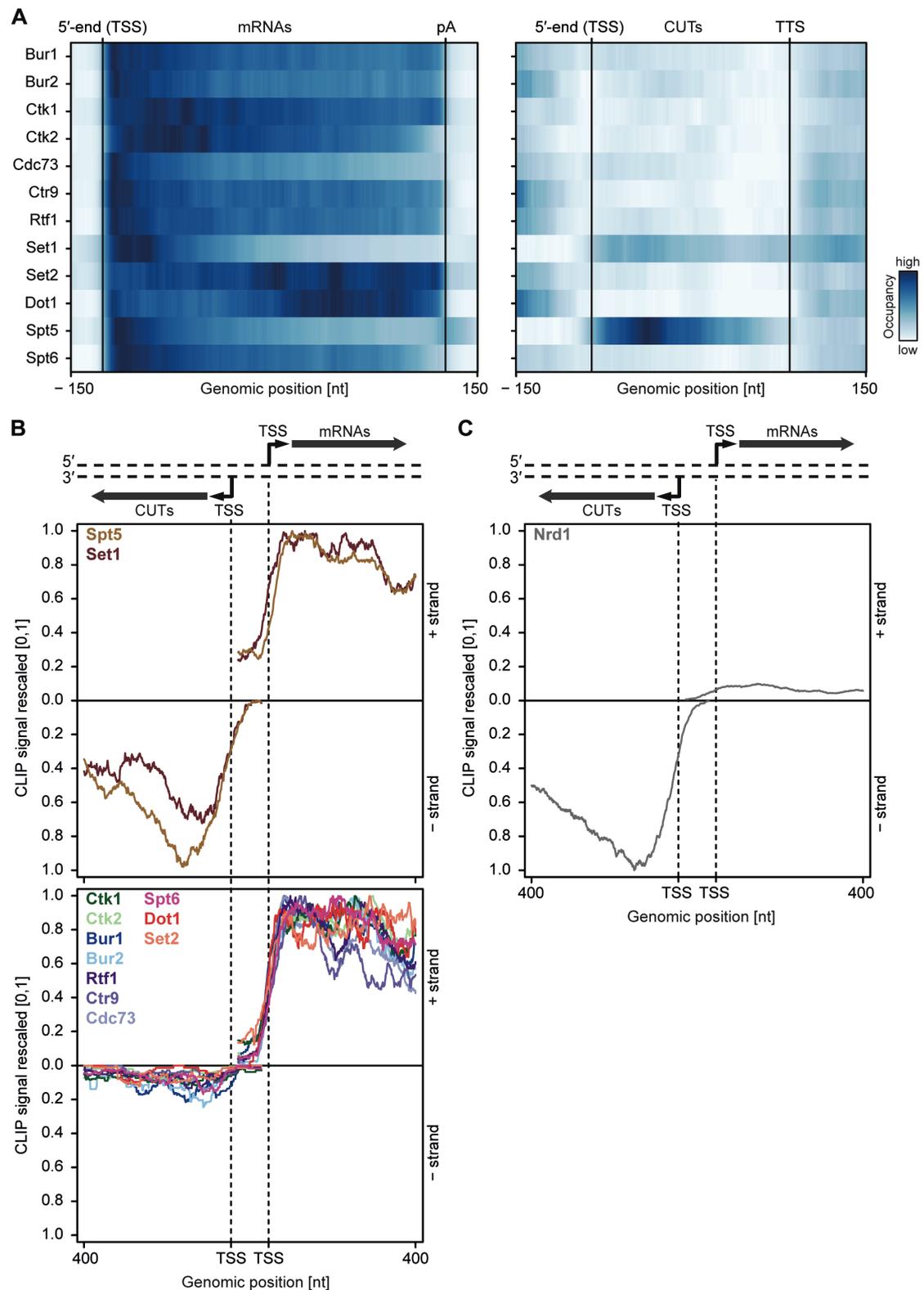


Figure 10 Asymmetric distribution of EFs at coding and non-coding transcripts. (A) PAR-CLIP occupancies over mRNAs (left) and non-coding CUTs (right). Smoothed, averaged Pol II normalized RNA occupancy profiles were aligned at the RNA 5'-end (transcription start site, TSS) and scaled to a common length. The color code shows the occupancy relative to the maximum occupancy per factor over both transcript classes (dark blue). (B) and (C) PAR-CLIP occupancies at selected bidirectional promoters. Smoothed, averaged Pol II normalized RNA occupancy profiles for sense mRNA (right) and divergent

CUT (left) were centered around their 5'-end (TSS) [-75 nt to +400 nt]. We considered only bidirectional promoters producing mRNAs and CUTs that did not overlap with any other transcripts in the depicted region. After normalization, average mRNA and CUT profiles were rescaled, setting the maximum occupancy to 1 and the minimum occupancy to 0.

We then analyzed PAR-CLIP signals at bidirectional promoters, which produce mRNA in one direction and a CUT in the divergent direction (Figure 10B). We observed clear differences in PAR-CLIP signals for divergent directions. As in Figure 10A, Set1 and Spt5 showed high signals on CUTs and mRNAs (Figure 10B, top) whereas all other EFs bound exclusively to mRNAs (Figure 10B, bottom). These differences were also observed when the analysis was restricted to bidirectional promoters producing CUTs and mRNAs of similar lengths (Figure 11B).

How can some EFs distinguish between CUTs and mRNAs? We carried out motif analysis around the strongest PAR-CLIP sites for each EF using XXmotif (Luehr, Hartmann, and Söding 2012) and could not find any significantly enriched motifs, indicating that EFs bind RNA in a non-specific manner. We hypothesize that another RNA-binding factor blocks binding of EFs to CUTs. CUTs are rapidly degraded by a surveillance system, which includes Nrd1 (Schulz et al. 2013, Vasiljeva et al. 2008, Steinmetz and Brow 1996). Nrd1 selectively binds to CUTs (Figure 10C) via motifs that are enriched in CUTs compared to mRNAs (Schulz et al. 2013). Binding of Nrd1 to CUTs might hinder RNA binding of some EFs, especially those which possess lower RNA binding affinity. This may explain how stable elongation complexes preferentially assemble on mRNAs.

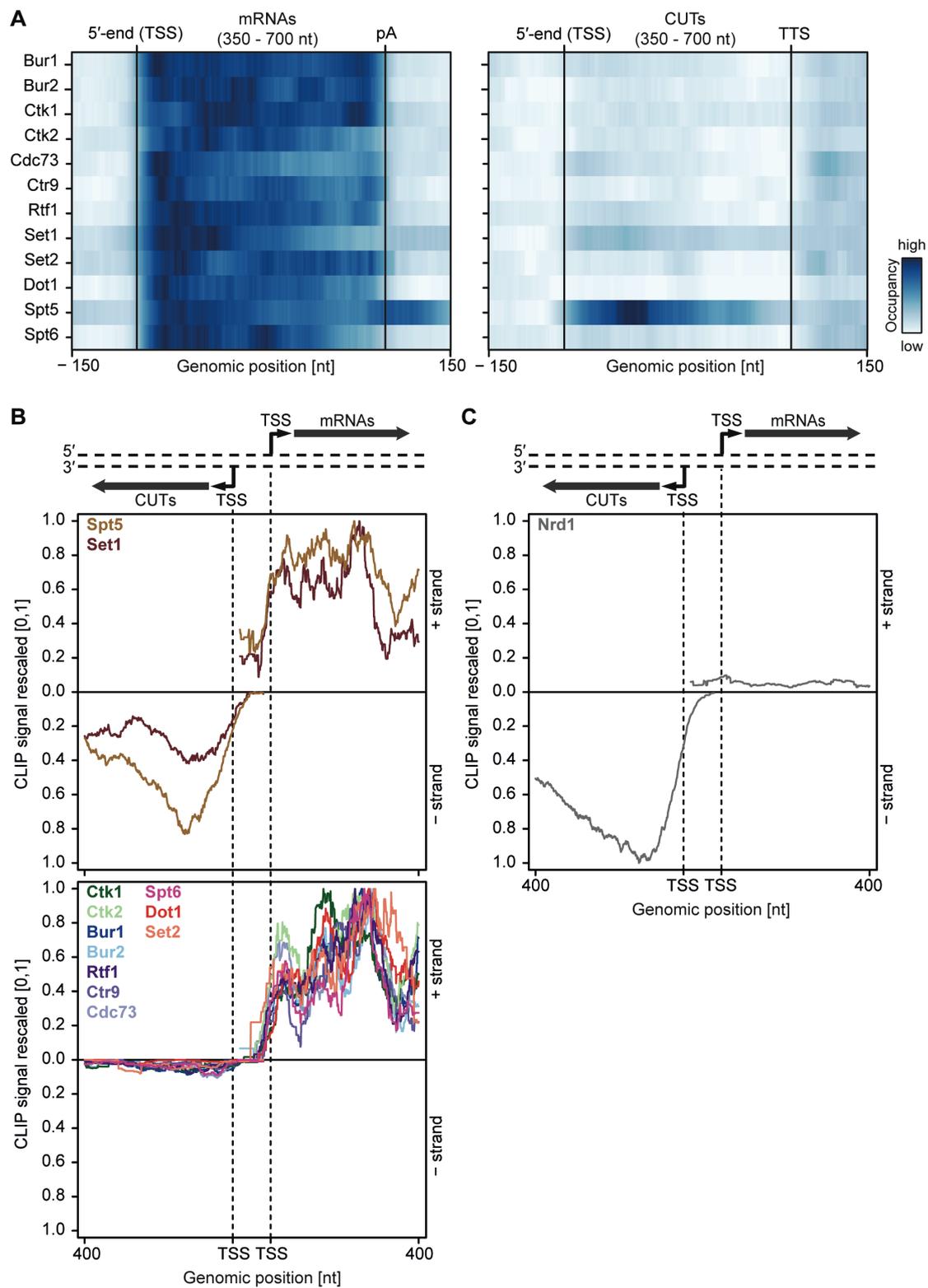


Figure 11 Asymmetric distribution of EFs at coding and non-coding transcripts of similar length. (A) PAR-CLIP occupancy heat map similar to that shown in **Figure 10A**, but with mRNAs and CUTs selected to be of similar lengths, 350-700 nt. **(B)** and **(C)** PAR-CLIP occupancy profiles for elongation factors as in **Figure 10B**, with sense mRNAs and divergent antisense CUTs of similar lengths, 350-700 nt, selected from bidirectional promoters.

3.4 Chromatin association of EFs depends on RNA

We next investigated whether RNA binding of EFs contributes to their association with chromatin. Yeast cells were lysed and incubated with buffer containing RNases or with buffer only. Chromatin was isolated and associated protein factors were detected by Western blotting (see 2.2.6). We found that RNase treatment strongly decreased the levels of chromatin-associated enzymes Set1, Set2, Dot1, Bur1, Ctk1, and the cyclins Bur2 and Ctk2 (Figure 12). Thus, RNA stabilizes chromatin association of these factors. Chromatin association of two non-enzymatic EFs also depended on RNA, although less strongly. With respect to Paf1C, Rtf1 was partially lost upon RNase treatment, whereas Leo1 and Paf1 were not significantly affected. Spt5 binding to chromatin also depended on RNA, whereas Spt6 was not significantly affected by RNase treatment (Figure 12). These few discrepancies between chromatin association and PAR-CLIP results are explained by additional EF interactions, and by the dependence of the PAR-CLIP signal on the concentration of the RNA-interacting protein in the cell (Chong et al. 2015, Kulak et al. 2014).

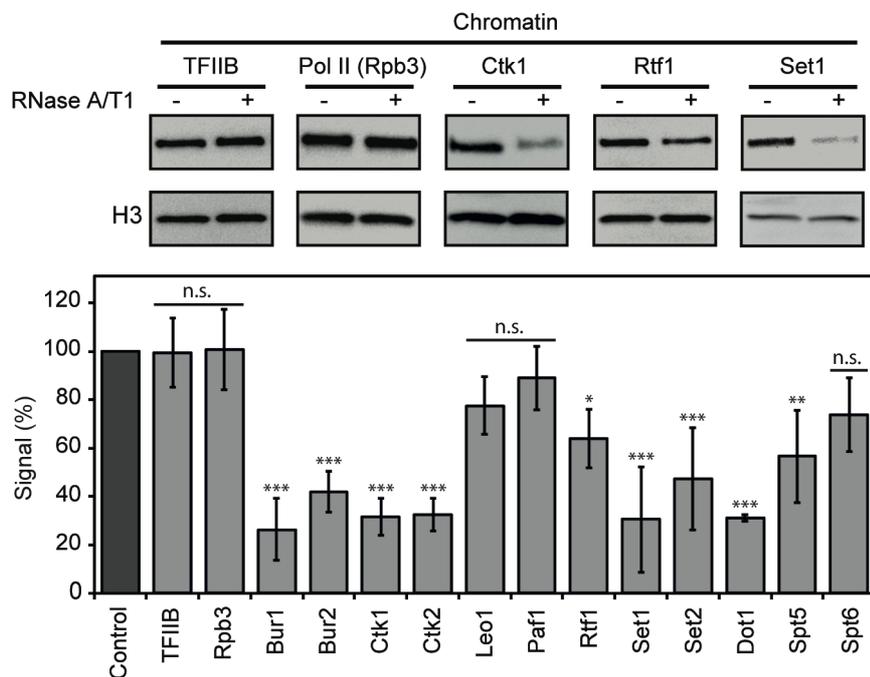


Figure 12 Chromatin association of EFs depends on RNA. Western blot analysis (top) and quantitative densitometry (bottom) of exemplary EFs bound to chromatin before and after treatment with RNase A/T1 mix. H3 was used as loading control. Densitometry data are expressed as mean \pm SD from two to three independent experiments. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; n.s. = not significant (one-way ANOVA Dunnett post-hoc test).

As a negative control, we subjected TFIIB to the RNase assay. We observed no differences in chromatin binding after RNase treatment (Figure 12), consistent with recruitment of TFIIB to DNA during transcription initiation (Sainsbury, Bernecky, and Cramer 2015). Also as expected, RNase treatment did not affect association of Pol II with chromatin, showing that the observed losses of EFs from chromatin upon RNase treatment were not due to a loss of Pol II (Figure 12). The above results show that the association of many EFs with chromatin depends on RNA.

3.5 Ctk1 kinase complex binds RNA *in vitro*

The observed RNA-EF crosslinking *in vivo* and the RNA-dependent chromatin association data strongly suggested that EFs can directly bind RNA. To investigate this *in vitro*, we prepared one EF complex in recombinant form. We chose the prominent Ser2 kinase complex CTDK-I that comprises the subunits Ctk1, Ctk2, and Ctk3 (Muhlbacher et al. 2015, Sterner et al. 1995). CTDK-I is the main yeast kinase responsible for phosphorylating the Pol II CTD at Ser2 (Patturajan et al. 1999, Cho et al. 2001), and this is a decisive event in establishing the mature Pol II elongation complex. Further, RNA-dependent chromatin association of Ctk1 and Ctk2 were most unexpected, as for several other EFs RNA interactions were already reported (compare introduction).

We co-expressed recombinant Ctk1, Ctk2, and Ctk3 in insect cells and purified a complete, intact CTDK-I complex (2.2.7, Figure 13A). We then tested the purified CTDK-I complex for its kinase activity using a purified GST-CTD construct and dephosphorylated full-length *S. cerevisiae* Pol II (2.2.8). Both the GST-CTD and the Rpb1 subunit of Pol II were readily phosphorylated by CTDK-I at the Ser2 position *in vitro* (Figure 13B,C), showing that our purified CTDK-I complex is active.

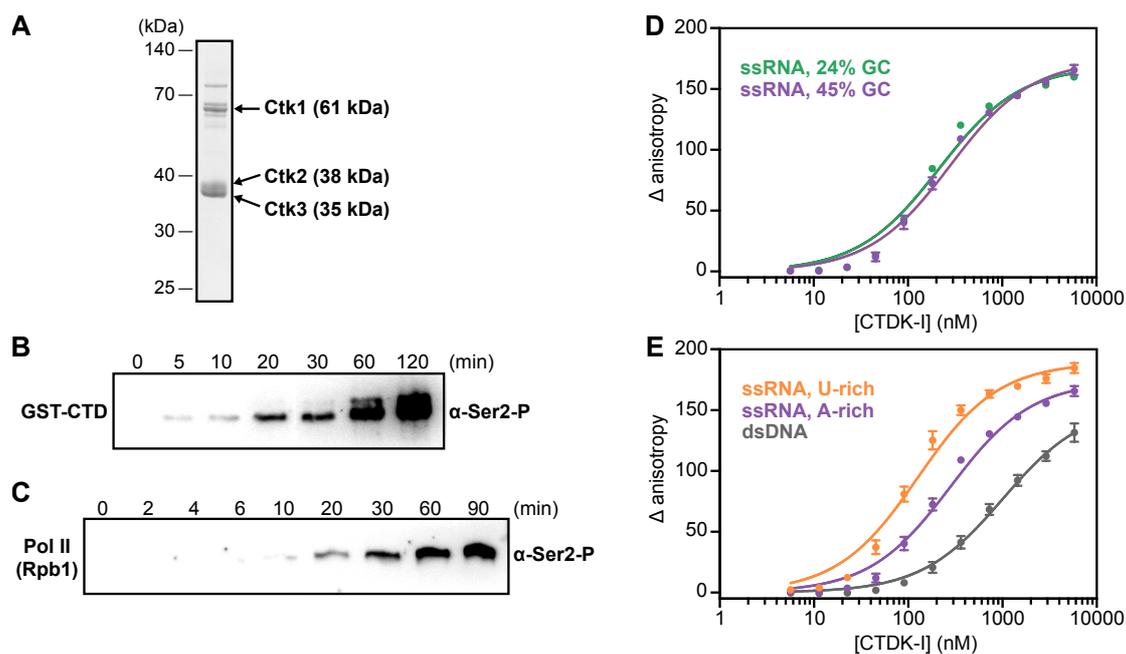


Figure 13 Recombinant CTDK-I complex is active and binds RNA *in vitro*. (A) The three-subunit CTDK-I complex from *S. cerevisiae* was recombinantly expressed in insect cells and purified to homogeneity. The purified complex was run on a 4–15% gradient SDS-PAGE and stained with InstantBlue™. (B) Purified human GST-CTD (10 μ M) was incubated with 0.4 μ M CTDK-I and 3 mM ATP. Time points were taken at 0 (no ATP), 5, 10, 20, 30, 60 and 120 min and CTDK-I activity was determined by Western blot analysis using an antibody that recognizes the Ser2 phosphorylated form of the CTD of Pol II. Molecular mass of GST-CTD is \sim 70 kDa. (C) Purified and dephosphorylated Pol II (2 μ M) from *S. cerevisiae* was incubated with 0.4 μ M CTDK-I and 3 mM ATP. Time points were taken at 0 (no ATP), 2, 4, 6, 10, 20, 30, 60 and 90 min and CTDK-I activity was determined by Western blotting as in (B). Molecular mass of the CTD containing subunit of Pol II, Rpb1, is \sim 200 kDa. (D) Increasing concentrations (0–5.8 μ M) of the complete CTDK-I kinase complex were incubated with 8 nM of a 24% GC (green line; $K_{d,app}$ (nM) = 210 ± 18) and with a 45% GC (purple line; $K_{d,app}$ (nM) = 277 ± 21) ssRNA sequences. Binding was determined by relative change in fluorescence anisotropy. Data was fit with a single site binding equation. Error bars reflect the standard deviation from three experimental replicates. (E) Increasing concentrations (0–5.8 μ M) of the complete CTDK-I kinase complex were incubated with 8 nM of a U-rich ssRNA (orange line; $K_{d,app}$ (nM) = 123 ± 10), an A-rich ssRNA (purple line; $K_{d,app}$ (nM) = 277 ± 21) and a dsDNA (grey line; $K_{d,app}$ (nM) = 1007 ± 67) sequences. Binding strength, data fitting and standard deviation was determined as in (D).

We then tested the purified CTDK-I complex for RNA binding *in vitro*. We performed fluorescence anisotropy titration experiments using single-stranded (ss) RNA oligonucleotides with 45% or 24% GC-content and bearing a 5' FAM label (Figure 13D,E). CTDK-I bound both ssRNAs with similar affinities (Figure 13D). We also tested U- or A-rich sequences for association with CTDK-I and found some preference for U-rich RNA (Figure 13E, Figure 14). Fitting the data with binding curves by linear regression resulted in apparent K_d 's in the nanomolar range (Figure

13D,E, Figure 14). All experiments were done in the presence of tRNA as competitor, indicating that flexible, single-stranded nucleic acids are preferentially bound. Consistent with this, CTDK-I bound to duplex DNA much more weakly (dsDNA, Figure 13E). These experiments show that the EF complex CTDK-I binds to single-stranded RNA *in vitro*, consistent with direct EF-RNA interactions *in vivo*.

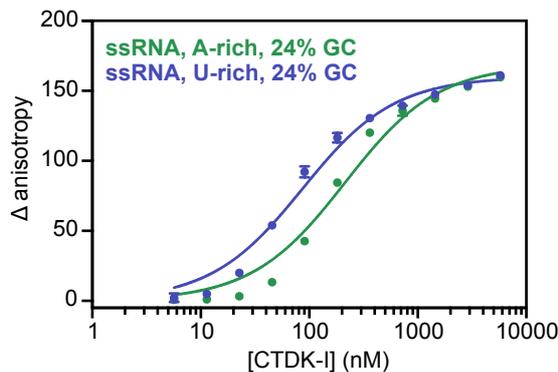


Figure 14 Recombinant and active CTDK-I complex binds preferentially U-rich ssRNA *in vitro*. Binding of CTDK-I to U- and A-rich ssRNA sequences with 24% GC. Increasing concentrations (0–5.8 μ M) of the full-length CTDK-I kinase complex were incubated with 8 nM of a U-rich ssRNA (blue line; $K_{d,app}$ (nM) = 83 ± 6) and an A-rich ssRNA (green line; $K_{d,app}$ (nM) = 210 ± 18) sequences. Binding strength, data fitting and standard deviation was determined as in **Figure 13D**.

3.6 Evidence that RNA contributes to EF recruitment

We also measured gene occupancies of EFs using ChIP-Seq and compared them with our PAR-CLIP occupancies (Figure 15). To do so, we first optimized our ChIP protocol for high-throughput sequencing (Figure 16). The obtained ChIP-Seq data sets were highly reproducible (Figure 17). For comparability with PAR-CLIP data, we collected ChIP-Seq data, although ChIP data are available for single genes or genome-wide using various other techniques or set-ups (Keogh, Podolny, and Buratowski 2003, Kim et al. 2004, Liu et al. 2005, Pokholok et al. 2005, Weiner et al. 2015, Kim et al. 2010, Mayer et al. 2010, Ng et al. 2003, Kizer et al. 2005, Krogan, Kim, et al. 2003). Metagene analysis of our ChIP-Seq data revealed that EF occupancy increased within 100-600 bp downstream of the TSS, and was generally high in gene bodies (Figure 15, red lines). In contrast, PAR-CLIP results showed that EFs interacted with RNA already from around 20 nt downstream of the capped 5'-end of mRNAs (Figure 15, blue lines). This difference was most pronounced for Set2,

which occupies transcripts at the 5'-end but showed peak levels of genome association only in the downstream region, with peak levels 450-300 bp upstream of the pA site. These results are consistent with the idea that RNA contributes to EF recruitment to transcribed genes, and that the contribution of RNA-based recruitment differs for different EFs.

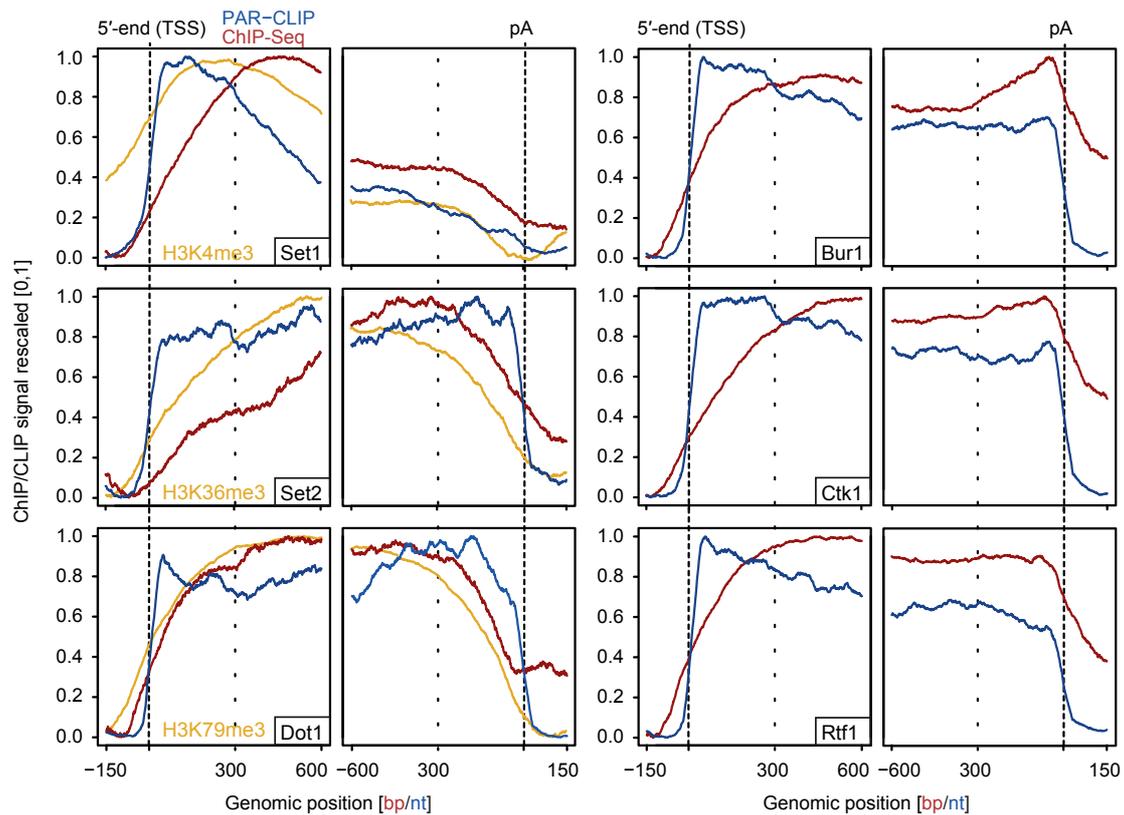


Figure 15 Comparison of PAR-CLIP and ChIP-Seq occupancy profiles. Averaged ChIP-Seq (red) and PAR-CLIP (blue) occupancy profiles of EFs and ChIP-Seq of the histone marks H3K4me3, H3K79me3 and H3K36me3 (yellow) centered around TSSs [-150 bp to +600 bp] and pA sites [-600 bp to +150 bp] individually normalized to range between 0 and 1.

Comparison of our histone methyltransferase PAR-CLIP data sets with ChIP-Seq data of the corresponding methylation marks (Figure 15, left, orange lines) provides further support of the model that RNA binding can contribute to EF recruitment to transcribed regions. In the direction of transcription, the PAR-CLIP signals for methyltransferases increased first, followed by an onset of ChIP-Seq signals for the respective histone methylation marks, which in turn preceded the increase in ChIP-Seq signals for the enzymes (Figure 15, left). This sequence of signal onsets is consistent with the model that these EFs are recruited by nascent RNA

and then modify histones as Pol II moves downstream. Taken together, these results suggest that interactions of EFs with nascent RNA contribute to EF recruitment to actively transcribed genes *in vivo*.

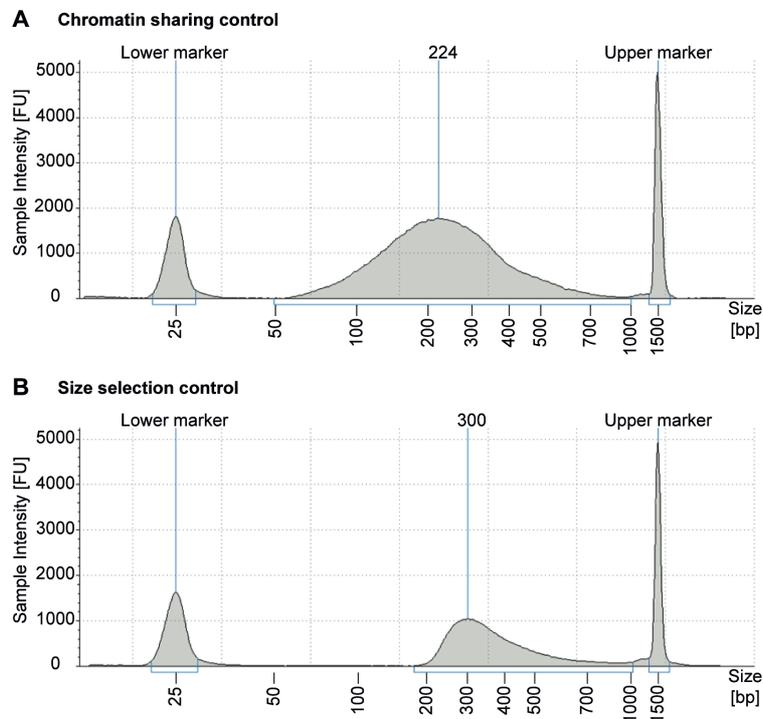


Figure 16 Optimization of two major steps for ChIP followed by high-throughput sequencing (ChIP-Seq). Example of electropherogram acquired on a TapeStation showing DNA fragment size distribution after chromatin sharing (**A**) and fragment size distribution of a DNA library after size selection (**B**).

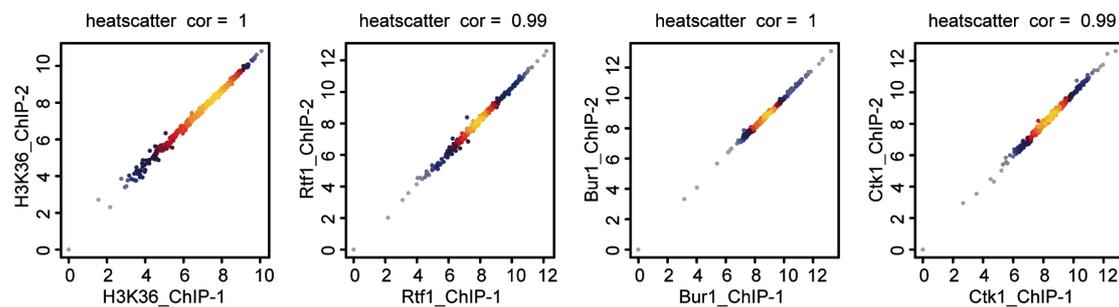


Figure 17 Comparison of replicate measurements for ChIP-Seq. Comparison of ChIP-Seq IP replicate measurements for elongation factors Bur1, Ctk1 and Rtf1 and histone methylation mark H3K36me3. The scatterplots compare average log₂ read counts of all transcripts shown in Figure 7 using Spearman correlation.

4 Discussion

Here we present a large set of system-wide occupancy data for yeast transcription elongation factors on RNA (PAR-CLIP) and DNA (ChIP-Seq), and complementary biochemical data. The remarkable finding from our work is that many elongation factors (EFs) interact with nascent RNA *in vivo*. Additional *in vitro* results support these findings and indicate that RNA can contribute to EF recruitment and the stability of the transcription elongation complex. These results extend our understanding of how the transcription elongation complex is assembled and maintained on active genes. The emerging view from our data is that nascent RNA contributes to EF recruitment and elongation complex stability to different extents for different EFs. We note that our results do not reveal whether and which EFs are initially recruited by RNA, and which EFs establish RNA interactions only after they have been recruited by alternative interactions, although EF binding in the very 5'-region of transcripts argues for an RNA-based recruitment model.

Our PAR-CLIP data revealed that EFs studied here bind nascent transcripts without the need of a specific sequence motif –as shown for traditionally studied RBPs, such as Nrd1 (Schulz et al. 2013). Additionally, the RNA binding affinities of EFs are relatively low when compared to RNA binding proteins involved in RNA export or RNA degradation pathways (Baejen et al. 2014). By contrast, elongation factors bind DNA with high affinity, as shown by ChIP-Seq. The relative affinities of proteins for RNA versus DNA may reflect the main function of the protein. For instance, the transfer of an EF from RNA to DNA will be energetically favorable only if the factor has a higher affinity for DNA than for RNA. All the above leads us to believe that nascent RNA can act as a general binding platform for transcription elongation factors that are then transferred to the transcribing chromatin.

Our results also elucidate the long-standing question how the yeast CTD Ser2 kinases Ctk1 and Bur1, which are essential for transcription elongation, are recruited to transcribing Pol II. The Pol II Ser2 kinases give rise to strong PAR-CLIP signals and their chromatin association is strongly dependent on RNA. In addition, we show that purified CTDK-I complex strongly binds to RNA *in vitro*. This all indicates that nascent RNA plays an important role in recruiting Ser2 kinases to transcribing Pol II.

Binding of the Ser2 kinases near the RNA 5'-end is consistent with stabilization of these kinases on the elongation complex by the cap-binding complex (Hossain et al. 2013, Lidschreiber, Leike, and Cramer 2013). A model of kinase recruitment by capped RNA predicts that these enzymes are lost from the transcribing enzyme upon RNA cleavage at the pA site, and this is indeed observed by ChIP-Seq. In conclusion, RNA-based recruitment of Ser2 kinases explains why Ser2 phosphorylation of the CTD is restricted to transcribing polymerases, whereas free or initiating polymerases are not phosphorylated at Ser2 residues.

How can some EFs bind both RNA and Pol II? EFs are generally modular and contain multiple domains that can be involved in RNA or protein interactions. However, the same domain can mediate both RNA and protein interactions, as documented for the RNA export factor Yra1, which contains an RNA recognition motif (RRM) domain that binds both RNA and the phosphorylated CTD (MacKellar and Greenleaf 2011). Set1 contains two adjacent RRM domains (Tresaugues et al. 2006), and Set2 contains a SRI domain that binds the phosphorylated CTD (Dengl et al. 2009, Sun et al. 2010, Yoh et al. 2007, MacKellar and Greenleaf 2011), but may also bind RNA. The three Paf1C subunits that bind RNA *in vivo*, namely Cdc73, Ctr9 and Rtf1, also interact with the phosphorylated CTD and the phosphorylated C-terminal region (CTR) of Spt5 *in vitro* (Qiu et al. 2012). Rtf1 contains a positively charged Plus-3 domain (Finn et al. 2014) that binds the phosphorylated CTR (Wier et al. 2013) and single-stranded nucleic acids (de Jong et al. 2008). We predict that many EFs contain domains that can interact with RNA or with the phosphorylated CTD or CTR, which resemble RNA in its flexible nature and negative charge. Whereas for some EFs binding to RNA or the CTD may be mutually exclusive, others can bind both Pol II and RNA at the same time, for example Spt5. Due to a lack of solubility of individually expressed EF subunits, and the difficulty of preparing EF complexes in recombinant and pure form in large quantities, we had to limit our *in vitro* RNA-binding analysis to CTDK-I.

Finally, we predict that RNA-based recruitment of EFs provides a missing link in our understanding of how the transcription cycle is coordinated. When the initiation complex assembles at the promoter, TFIIF phosphorylates Ser5 residues in the CTD and this enables recruitment of the capping enzyme (Cho et al. 1997, Fabrega et al.

2003, Rodriguez et al. 2000, Schroeder et al. 2000, Schwer and Shuman 2011). The nascent RNA then receives a 5'-cap (Martinez-Rucobo et al. 2015), and capped RNA could then help to recruit EFs. The requirement for a cap on RNA befits the observation that Ser5 phosphorylation is needed for high gene occupancy with some EFs (Qiu et al. 2012, Qiu, Hu, and Hinnebusch 2009, Qiu et al. 2006, Ng et al. 2003). RNA-based recruitment of the major Ser2 kinase, Ctk1 would then lead to CTD phosphorylation on Ser2 residues and stable binding of other EFs. Eventually, transcription of a pA site triggers RNA cleavage, and this would facilitate loss of many RNA-bound EFs and render the polymerase prone to transcription termination. Thus, the transcribing Pol II complex may be viewed as a self-organizing system that is encoded in the DNA, but only realized on the level of RNA, which plays crucial roles in complex assembly and disassembly.

5 Future perspectives

Standard models suggest that regulation of gene expression first occurs during transcription and is mainly mediated by DNA binding proteins (e.g., transcription factors) (Cosma, Tanaka, and Nasmyth 1999, Felsenfeld and Groudine 2003, Mitchell and Tjian 1989). Subsequently, RNA binding proteins (RBPs) bind the nascent transcripts to dictate post-transcriptional mechanisms, such as splicing, stability, localization and translation (Tuck and Tollervy 2013, Zhao et al. 2010). However, accumulating evidence suggests that nascent pre-mRNAs possess both protein-coding and transcription regulatory functions (reviewed in (Skalska et al. 2017)). Recent RNA-protein interaction studies revealed that many DNA binding proteins also associate with RNA (Di Ruscio et al. 2013, Hendrickson et al. 2016, Sigova et al. 2015). Furthermore, as presented here, the recruitment of transcription elongation factors (EFs) to the nascent RNA seems to be a broadly acting cellular mechanism for regulating gene transcription. Some of the future challenges and open questions arising from our results are discussed below:

Discovery of new RNA binding proteins in yeast and human cells

Here we report *in vivo* interactions of the nascent pre-mRNA with transcription elongation factors in the yeast *S. cerevisiae*. Interestingly, many of the EFs included in our study lack classically defined RNA binding domains. Thus, our findings indicate the existence of unrevealed RNA-protein interactions that are yet to be discovered. More PAR-CLIP experiments and other RNA-protein interaction methods remain to be applied to proteins whose function and/or recruitment are not fully understood. In addition, these findings will help to further elucidate the role of the nascent RNA during its own transcription.

Moreover, establishing our PAR-CLIP protocol for the human system would be the next logical step in our laboratory. For instance, PAR-CLIP or similar experiments are required to confirm direct interactions of regulatory factors, like p-TEFb and splicing factors, with RNAs in living cells and to identify where on the RNAs these interactions take place. Additionally, the findings reported in this thesis, particularly the binding of CTD kinases and histone methyltransferases to RNA, can

be now explored in the human system, providing results for which no extrapolation will be necessary.

Nascent RNA and regulation of divergent transcription

Divergent transcription refers to the production of coding mRNAs from the sense direction and non-coding RNAs upstream and antisense to the annotated gene (Marquardt et al. 2014). This phenomenon is seen at many promoters in both yeast and metazoans (Duttke et al. 2015, Weiner et al. 2015). In *S. cerevisiae* several cryptic unstable transcripts (CUTs) emerge from such promoters and are rapidly degraded by the exosome in a pathway involving Nrd1 (Schulz et al. 2013). However, regulation and function of CUTs are still poorly understood.

Recent studies compared chromatin states between coding and non-coding regions during divergent transcription. They show that histone marks characteristic for 5' ends of coding regions (e.g. H3K4me3) appear in both directions. In contrast, gene body marks (e.g. H3K36me3) correlate mostly with sense transcription (Weiner et al. 2015). Complementarily, our ChIP-Seq data revealed that elongation factor occupancies differ for sense and antisense transcription (unpublished). Here, we showed that also RNA-EF interactions differ for sense and antisense transcripts, especially for late EFs like Set2 and Spt6 (Figure 10 and Figure 11). Together, these data likely reflect rapid termination of antisense CUTs by the Nrd1–Nab3–Sen1 (NNS) termination pathway (Porrua and Libri 2015). Early termination of antisense transcripts would prevent the transition of Pol II from initiation to elongation, thus stopping elongation factors to be recruited and histone marks to be deposited. To confirm this hypothesis, however, more experimental evidence is necessary. For instance PAR-CLIP and ChIP-Seq data of EFs in an Nrd1 depletion background will show if DNA and RNA interactions with these factors can then occur in the absence of Nrd1. These results will indicate whether Nrd1 RNA-binding or early-termination by Nrd1 prevent DNA/RNA binding of some EFs and thus the transition to productive elongation.

Role of RNA in maintaining protein-chromatin interactions

With our chromatin-binding assay after RNA degradation we provide strong evidence for a direct role of RNA in maintaining EF-chromatin interactions. We showed that

levels of EFs bound to chromatin were significantly decreased in the absence of RNA (Figure 12). It will be very interesting to extend these findings and test human homologs of the factors analyzed here in a similar assay to investigate how transcription factors in higher eukaryotes are recruited to transcribed genes.

***In vitro* validation of *in vivo* RNA-protein interactions**

We purified recombinant CTDK-I kinase complex to demonstrate its binding to RNA *in vitro*, and thus validate our *in vivo* findings (Figure 13). Similarly, binding of additional factors, like the BUR kinase complex, the Spt6 histone chaperone and the Set2 histone methyltransferase, to RNA remain to be tested *in vitro*. Once having these complexes in a purified form, we will be able to perform different mutations to uncover the RNA binding domains of the studied proteins, given that many of them lack classically defined RNA binding domains.

Nascent RNA as a recruitment platform for transcription factors

We compared elongation factor occupancies on DNA (ChIP-Seq) and on RNA (PAR-CLIP) (Figure 15) and showed that interactions of EFs with nascent RNA are established before EFs are recruited to chromatin. These results provide additional evidence for an RNA-based EF recruitment. Nonetheless, it was previously shown that protein-protein and chromatin-protein interactions are important for EF recruitment. For instance, Paf1C is required for H2B ubiquitination (H2Bub), which in turns enables recruitment of the Set1-containing COMPASS complex (Krogan, Dover, et al. 2003, Ng et al. 2003).

To extend the finding that nascent RNA is also involved in the recruitment of Set1, ChIP experiments directly comparing Set1 occupancy on genes in strains lacking Paf1C will be important. Also, ChIP experiments of Set1 lacking its RNA recognition motif (RRM) can be performed. In a third experiment, both deletion of Paf1C and the RRM domain would be compared with the wild-type gene occupancy of Set1. If both Paf1C and RNA are important for Set1 recruitment, then impairing these mechanisms will abolish gene occupancy. Whereas Set1 ChIP in a Paf1C knockout background or in a strain lacking the RRM domain of Set1, should result only in a partial decrease of gene occupancy.

6 Additional unpublished data

6.1 Investigation of novel factors in chromatin transcription

Transcription regulation is intimately linked to changes in chromatin structure and modification (Smolle and Workman 2013). Factors that associate with the Pol II transcription machinery can remodel (Chd1) or modify (Set1, Set2) chromatin (Carrozza, Li, et al. 2005, Ng et al. 2003, Smolle et al. 2012). Specific chromatin structures and modifications can be associated with either active or inactive transcription. Histone H3K4 trimethylation and histone H3K9 methylation are typical examples of this (reviewed in (Smolle and Workman 2013)). These histone marks are recognized by specific factors that “read” this modification in order to perform their function (histone readers). Despite a large body of work, there still exist putative histone marks and readers that have not been sufficiently characterized to reveal their functions. Furthermore, specific enzymes that remove the histone mark remain to be identified, such as the demethylase of the active histone mark H3K79me3 (Nguyen and Zhang 2011).

This project focuses on the discovery of new readers of the active histone marks H3K4me3, H3K36me3 and H3K79me3 that are written by the methyltransferases Set1, Set2 and Dot1, respectively. Readers of histone marks contain specific domains that recognize the target mark (Table 14; reviewed in (Smolle and Workman 2013)).

Table 14 Transcription activating histone methylations (blue). Known methyltransferases and demethylases are shown in red and purple, respectively. Recognition modules (reader domains) that bind specific modifications are indicated in green (see **Table 19**, **Table 20**, **Table 21** and **Table 22** for references).

Methyltransferase	Histone Methylation	Reader Domain	Demethylase
Set1	H3K4 (me1, me2, me3)	PHD, Chromo, MBT, Zf-CW	Jhd2 (PHD)
Dot1	H3K79 (me1, me2, me3)	Tudor	Unknown
Set2	H3K36 (me1, me2, me3)	PHD, Chromo, PWWP	Jhd1 (PHD)

We started our investigations by using the open-source software HHpred (Söding, Biegert, and Lupas 2005) to screen the *S. cerevisiae* genome for proteins containing reader domains known to bind these active histone marks (Table 14). This analysis provided us with a list of known and potential new readers of the specific histone marks that we were interested in. With this information, we first performed ChIP-qPCR experiments for the new potential readers to select for chromatin binders (Table 15 and Table 16). We then collected a large set of functional genome-wide data of the most promising candidates. First, we obtained genomic occupancy profiles (ChIP-Seq) and second, we monitored changes in newly synthesized transcription upon factor deletion (knockout; KO) or depletion (anchor-away; AA) (Haruki, Nishikawa, and Laemmli 2008) by 4tU-Seq. 4tU-Seq is a method that applies a short metabolic RNA labeling pulse to sequencing of the newly synthesized RNA and enables extraction of both RNA synthesis and degradation rates genome-wide (Schulz et al. 2013).

Table 15 Factors containing reader domains with positive ChIP-qPCR signal.

Complex	Factor	Domain	Available data		
			ChIP-qPCR	ChIP-Seq	4tU-Seq
-	Asr1	PHD1; PHD2	✓	✓	✓ AA
-	Bye1	PHD		✓	✓ KO
NuA3	Nto1	PHD	✓	✓	
Set3C	Set3	PHD	✓		
-	Set4	PHD	✓	✓	✓ AA
COMPASS	Spp1	PHD	✓		
NuA3	Yng1	PHD	✓	✓	
NuA4	Yng2	PHD	✓		
-	Rad9	Tudor	✓	✓	
SAGA/SILK	Sgf29	Tudor	✓		
Isw1b	Ioc4	PWWP	✓	✓	
NuA3b	Pdp3	PWWP	✓	✓	
	Chd1	Chromo	✓		
NuA4	Esa1	Chromo	✓		

AA: anchor-away

KO: knockout

Table 16 Factors containing reader domains without positive ChIP-qPCR signal.

Complex	Factor	Domain	Available data
			ChIP-qPCR
Rpd3L	Pho23	PHD	✓
Rpd3L	Cti6	PHD	✓
-	Jhd1	PHD	✓
-	Jhd2	PHD	✓
Snt2C	Ecm5	PHD	✓
Snt2C	Snt2	PHD	✓
Rpd3S	Rco1	PHD	✓
Rpd3S/NuA4	Eaf3	Tudor	✓

As a reference and positive control, we also collected such data sets for the histone methyltransferases Set1, Set2 and Dot1 (Table 17). Additionally, genome-wide distributions were also obtained for the chromatin marks H3K4me3, H3K36me3 and H3K79me3 (Table 18). We partially extended our data collection for some of the factors using PAR-CLIP, which can map RNA-associated factors over the transcriptome (Table 17; (Hafner et al. 2010)). See Table 15, Table 16, Table 17 and Table 18 for an overview of the data produced for this project during this thesis; see Table 19, Table 20, Table 21 and Table 22 for an overview of the available literature regarding writers, histone marks and histone readers discussed in this section.

Table 17 Histone methyltransferases (writers) of active histone marks.

Complex	Factor	Domain	Available data			
			ChIP-qPCR	ChIP-Seq	4tU-Seq	PAR-CLIP
-	Dot1		✓	✓		✓
COMPASS	Set1	SET	✓	✓	✓ AA&KO	✓
-	Set2	SET	✓	✓	✓ AA&KO	✓

AA: anchor-away

KO: knockout

Table 18 Histone marks related to active transcription.

Histone mark	Available data	
	ChIP-qPCR	ChIP-Seq
H3K4me3	✓	✓
H3K36me3	✓	✓
H3K79me3	✓	✓

Our genome-wide data is currently under bioinformatics analysis. With the combination of these data sets, we aim to address the questions of where the novel chromatin factors are associated with the yeast genome, which genes require them for normal transcription, and how their occupancy and functional specificity is correlated with previously known, well-studied components of the transcription machinery. We thereby aim to further characterize the relationship between chromatin and transcription in yeast and thus extend the so called “histone code” (Jenuwein and Allis 2001). Bioinformatics analysis will indicate which further wet lab work will be necessary to elucidate our questions. *In vitro* binding assays of the reader’s domain using histone peptide arrays, as well as ChIP-Seq experiments of new potential readers in a writer knockout background are experiments to be considered.

Table 19 Available information for factors containing reader domains with positive ChIP-qPCR signal.

Factor	Protein function	References
Asr1	Ubiquitin ligase; binds Pol II CTD on Ser5-P	(Daulny et al. 2008)
Bye1	Binds H3K4me3 and Pol II	(Kinkelin et al. 2013)
Nto1	Subunit of the NuA3 histone acetyltransferase complex that acetylates H3	(Shi et al. 2007)
Set3	Part of the Set3C histone deacetylase complex	(Krogan, Kim, et al. 2003)
Set4	Function unknown; contains both a SET and a PHD domain	(Pijnappel et al. 2001)
Spp1	Subunit of the COMPASS (Set1C) complex that methylates H3K4	(Miller et al. 2001)
Yng1	Subunit of the NuA3 histone acetyltransferase complex that acetylates H3	(Martin et al. 2006)
Yng2	Subunit of the NuA4 histone acetyltransferase complex that acetylates H4 and H2B	(Chittuluru et al. 2011)
Rad9	DNA damage-dependent checkpoint protein	(Wang et al. 2012)
Sgf29	Component of the transcription regulatory histone acetylation (HAT) complexes SAGA, SLIK, and ADA	(Bian et al. 2011)
Ioc4	Binds H3K36me3; recruits the remodeling complex Isw1b to ORFs	(Smolle et al. 2012)
Pdp3	Binds H3K36me3; part of the new histone acetyltransferase complex NuA3b	(Gilbert et al. 2014)
Chd1	Chromatin-remodeling factor	(Smolle et al. 2012)
Esa1	Catalytic subunit of the NuA4 histone acetyltransferase complex	(Ginsburg et al. 2014)

Table 20 Available information for factors containing reader domains without ChIP-qPCR signal.

Factor	Protein function	References
Pho23	Subunit of the Rpd3L histone deacetylase complex	(Terzi et al. 2011)
Cti6	Component of the Rpd3L histone deacetylase complex	(Carrozza, Florens, et al. 2005)
Jhd1	Histone demethylase that specifically demethylates H3K36	(Tsukada et al. 2006)
Jhd2	Histone demethylase specific for H3K4	(Liang et al. 2007)
Ecm5	Possible histone demethylase; recruits Rpd3p to a small number of promoters	(Baker et al. 2013)
Snt2	Involved in ubiquitylation; recruits Rpd3p to a small number of promoters	(Baker et al. 2013, Singh et al. 2012)
Rco1	Catalytic component of the Rpd3S histone deacetylase complex	(Carrozza, Li, et al. 2005)
Eaf3	Component of the Rpd3S histone deacetylase complex and also part of NuA4 acetyltransferase	(Carrozza, Li, et al. 2005, Eisen et al. 2001)

Table 21 Function of histone methyltransferases (writers).

Factor	Protein function	References
Set1	Mono-, di- and trimethylation of H3K4; involved in initiation	(Ng et al. 2003)
Set2	Mono-, di- and trimethylation of H3K36; involved in elongation	(Krogan, Kim, et al. 2003)
Dot1	Mono-, di- and trimethylation of H3K79; involved in elongation	(Nguyen and Zhang 2011)

Table 22 Function of histone marks.

Histone mark	Function of histone mark	References
H3K4me3	Positive regulation of gene expression; present at promoters and 5' region of genes	(Mulder et al. 2007)
H3K36me3	Positive regulation of gene expression; present at gene-body	(Kizer et al. 2005)
H3K79me3	Positive regulation of gene expression; present at gene-body	(Nguyen and Zhang 2011)

6.2 Genome-wide occupancy profiles of Pol II CTD phosphorylation marks

Specific modifications on the C-terminal domain (CTD) of Pol II are linked to certain stages of gene transcription and RNA processing in various organisms (Heidemann et al. 2013). This refers to the hypothesis of the “CTD code”, where combinations of modifications are written and read by specific factors, thus coordinating the entire transcription cycle (Buratowski 2003). Correlation of gene occupancy profiles of transcription factors, as well as of possible new factors involved in transcription, with the different states of the Pol II CTD, should be a good starting point to address the genome-wide location or even the function of a given factor.

Available genome-wide data of the distribution of CTD phosphorylation marks in *S. cerevisiae* was obtained by ChIP followed by tiling microarray analysis (Mayer et al. 2010, Tietjen et al. 2010). Here, we performed higher resolution ChIP-Seq experiments for three phosphorylation states of the Pol II CTD that will support our investigations in the discovery and characterization of new transcription factors by correlation analysis (Figure 18). Additionally, this high-resolution data is of great interest for the group for current and future research regarding gene transcription.

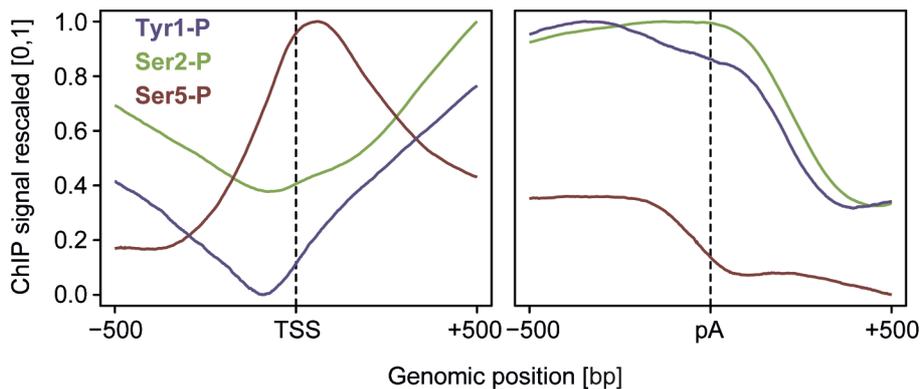


Figure 18 Genome-wide average occupancy profiles of the CTD phosphorylation marks Tyr1-P, Ser2-P and Ser5-P revealed by ChIP-Seq experiments using specific monoclonal antibodies ((Chapman et al. 2007); **Table 10**). While Ser5-P peaks at the transcription start site (TSS) of genes, Tyr1-P and Ser2-P signals increase toward the 3' end and polyadenylation site.

References

- Adelman, K., and J. T. Lis. 2012. "Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans." *Nat Rev Genet* 13 (10):720-31. doi: 10.1038/nrg3293.
- Alexander, R. D., S. A. Innocente, J. D. Barrass, and J. D. Beggs. 2010. "Splicing-dependent RNA polymerase pausing in yeast." *Mol Cell* 40 (4):582-93. doi: 10.1016/j.molcel.2010.11.005.
- Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jorgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhashi, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Muller, Fantom Consortium, A. R. Forrest, P. Carninci, M. Rehli, and A. Sandelin. 2014. "An atlas of active enhancers across human cell types and tissues." *Nature* 507 (7493):455-61. doi: 10.1038/nature12787.
- Ard, R., and R. C. Allshire. 2016. "Transcription-coupled changes to chromatin underpin gene silencing by transcriptional interference." *Nucleic Acids Res* 44 (22):10619-10630. doi: 10.1093/nar/gkw801.
- Armache, K. J., H. Kettenberger, and P. Cramer. 2003. "Architecture of initiation-competent 12-subunit RNA polymerase II." *Proc Natl Acad Sci U S A* 100 (12):6964-8. doi: 10.1073/pnas.1030608100.
- Armache, K. J., S. Mitterweger, A. Meinhart, and P. Cramer. 2005. "Structures of complete RNA polymerase II and its subcomplex, Rpb4/7." *J Biol Chem* 280 (8):7131-4. doi: 10.1074/jbc.M413038200.
- Asin-Cayuela, J., and C. M. Gustafsson. 2007. "Mitochondrial transcription and its regulation in mammalian cells." *Trends Biochem Sci* 32 (3):111-7. doi: 10.1016/j.tibs.2007.01.003.
- Baejen, C., J. Andreani, P. Torkler, S. Battaglia, B. Schwalb, M. Lidschreiber, K. C. Maier, A. Boltendahl, P. Rus, S. Esslinger, J. Söding, and P. Cramer. 2017. "Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes." *Mol Cell*. doi: 10.1016/j.molcel.2017.02.009.
- Baejen, C., P. Torkler, S. Gressel, K. Essig, J. Söding, and P. Cramer. 2014. "Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition." *Mol Cell* 55 (5):745-57. doi: 10.1016/j.molcel.2014.08.005.
- Baker, L. A., B. M. Ueberheide, S. Dewell, B. T. Chait, D. Zheng, and C. D. Allis. 2013. "The yeast Snt2 protein coordinates the transcriptional response to hydrogen peroxide-mediated oxidative stress." *Mol Cell Biol* 33 (19):3735-48. doi: 10.1128/MCB.00025-13.
- Banerji, J., S. Rusconi, and W. Schaffner. 1981. "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences." *Cell* 27 (2 Pt 1):299-308.

- Bannister, A. J., R. Schneider, F. A. Myers, A. W. Thorne, C. Crane-Robinson, and T. Kouzarides. 2005. "Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes." *J Biol Chem* 280 (18):17732-6. doi: 10.1074/jbc.M500796200.
- Barrandon, C., B. Spiluttini, and O. Bensaude. 2008. "Non-coding RNAs regulating the transcriptional machinery." *Biol Cell* 100 (2):83-95. doi: 10.1042/BC20070090.
- Bartkowiak, B., P. Liu, H. P. Phatnani, N. J. Fuda, J. J. Cooper, D. H. Price, K. Adelman, J. T. Lis, and A. L. Greenleaf. 2010. "CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1." *Genes Dev* 24 (20):2303-16. doi: 10.1101/gad.1968210.
- Beltran, M., C. M. Yates, L. Skalska, M. Dawson, F. P. Reis, K. Viiri, C. L. Fisher, C. R. Sibley, B. M. Foster, T. Bartke, J. Ule, and R. G. Jenner. 2016. "The interaction of PRC2 with RNA or chromatin is mutually antagonistic." *Genome Res* 26 (7):896-907. doi: 10.1101/gr.197632.115.
- Bentley, D. L. 2005. "Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors." *Curr Opin Cell Biol* 17 (3):251-6. doi: 10.1016/j.ceb.2005.04.006.
- Berger, I., D. J. Fitzgerald, and T. J. Richmond. 2004. "Baculovirus expression system for heterologous multiprotein complexes." *Nat Biotechnol* 22 (12):1583-7. doi: 10.1038/nbt1036.
- Bian, C., C. Xu, J. Ruan, K. K. Lee, T. L. Burke, W. Tempel, D. Barsyte, J. Li, M. Wu, B. O. Zhou, B. E. Fleharty, A. Paulson, A. Allali-Hassani, J. Q. Zhou, G. Mer, P. A. Grant, J. L. Workman, J. Zang, and J. Min. 2011. "Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation." *EMBO J* 30 (14):2829-42. doi: 10.1038/emboj.2011.193.
- Bortvin, A., and F. Winston. 1996. "Evidence that Spt6p controls chromatin structure by a direct interaction with histones." *Science* 272 (5267):1473-6.
- Buratowski, S. 2003. "The CTD code." *Nat Struct Biol* 10 (9):679-80. doi: 10.1038/nsb0903-679.
- Bushnell, D. A., and R. D. Kornberg. 2003. "Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription." *Proc Natl Acad Sci U S A* 100 (12):6969-73. doi: 10.1073/pnas.1130601100.
- Bushnell, D. A., K. D. Westover, R. E. Davis, and R. D. Kornberg. 2004. "Structural basis of transcription: an RNA polymerase II-TFIIB cocystal at 4.5 Ångstroms." *Science* 303 (5660):983-8. doi: 10.1126/science.1090838.
- Carrillo Oesterreich, F., L. Herzel, K. Straube, K. Hujer, J. Howard, and K. M. Neugebauer. 2016. "Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II." *Cell* 165 (2):372-81. doi: 10.1016/j.cell.2016.02.045.
- Carrozza, M. J., L. Florens, S. K. Swanson, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman. 2005. "Stable incorporation of sequence specific repressors Ash1 and Ume6 into the Rpd3L complex." *Biochim Biophys Acta* 1731 (2):77-87; discussion 75-6. doi: 10.1016/j.bbaexp.2005.09.005.

- Carrozza, M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman. 2005. "Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription." *Cell* 123 (4):581-92. doi: 10.1016/j.cell.2005.10.023.
- Chapman, R. D., M. Heidemann, T. K. Albert, R. Mailhammer, A. Flatley, M. Meisterernst, E. Kremmer, and D. Eick. 2007. "Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7." *Science* 318 (5857):1780-2. doi: 10.1126/science.1145977.
- Chapman, R. D., M. Heidemann, C. Hintermair, and D. Eick. 2008. "Molecular evolution of the RNA polymerase II CTD." *Trends Genet* 24 (6):289-96. doi: 10.1016/j.tig.2008.03.010.
- Cheng, B., and D. H. Price. 2007. "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. doi: 10.1074/jbc.M702936200.
- Chittuluru, J. R., Y. Chaban, J. Monnet-Saksouk, M. J. Carrozza, V. Sapountzi, W. Selleck, J. Huang, R. T. Utley, M. Cramet, S. Allard, G. Cai, J. L. Workman, M. G. Fried, S. Tan, J. Cote, and F. J. Asturias. 2011. "Structure and nucleosome interaction of the yeast NuA4 and Piccolo-NuA4 histone acetyltransferase complexes." *Nat Struct Mol Biol* 18 (11):1196-203. doi: 10.1038/nsmb.2128.
- Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt, and S. Buratowski. 2001. "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. doi: 10.1101/gad.935901.
- Cho, E. J., T. Takagi, C. R. Moore, and S. Buratowski. 1997. "mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain." *Genes Dev* 11 (24):3319-26.
- Chong, Y. T., J. L. Koh, H. Friesen, S. K. Duffy, M. J. Cox, A. Moses, J. Moffat, C. Boone, and B. J. Andrews. 2015. "Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis." *Cell* 161 (6):1413-24. doi: 10.1016/j.cell.2015.04.051.
- Connelly, S., and J. L. Manley. 1988. "A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II." *Genes Dev* 2 (4):440-52.
- Cosma, M. P., T. Tanaka, and K. Nasmyth. 1999. "Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter." *Cell* 97 (3):299-311.
- Cramer, P., D. A. Bushnell, and R. D. Kornberg. 2001. "Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution." *Science* 292 (5523):1863-76. doi: 10.1126/science.1059493.
- Crick, F. 1970. "Central dogma of molecular biology." *Nature* 227 (5258):561-3.
- Crick, F. H. 1958. "On protein synthesis." *Symp Soc Exp Biol* 12:138-63.
- Daulny, A., F. Geng, M. Muratani, J. M. Geisinger, S. E. Salghetti, and W. P. Tansey. 2008. "Modulation of RNA polymerase II subunit composition by ubiquitylation." *Proc Natl Acad Sci U S A* 105 (50):19649-54. doi: 10.1073/pnas.0809372105.

- Davidovich, C., L. Zheng, K. J. Goodrich, and T. R. Cech. 2013. "Promiscuous RNA binding by Polycomb repressive complex 2." *Nat Struct Mol Biol* 20 (11):1250-7. doi: 10.1038/nsmb.2679.
- Davison, B. L., J. M. Egly, E. R. Mulvihill, and P. Chambon. 1983. "Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences." *Nature* 301 (5902):680-6.
- de Jong, R. N., V. Truffault, T. Diercks, E. Ab, M. A. Daniels, R. Kaptein, and G. E. Folkers. 2008. "Structure and DNA binding of the human Rtf1 Plus3 domain." *Structure* 16 (1):149-59. doi: 10.1016/j.str.2007.10.018.
- Dengl, S., A. Mayer, M. Sun, and P. Cramer. 2009. "Structure and in vivo requirement of the yeast Spt6 SH2 domain." *J Mol Biol* 389 (1):211-25. doi: 10.1016/j.jmb.2009.04.016.
- Dermody, J. L., and S. Buratowski. 2010. "Leo1 subunit of the yeast paf1 complex binds RNA and contributes to complex recruitment." *J Biol Chem* 285 (44):33671-9. doi: 10.1074/jbc.M110.140764.
- Di Ruscio, A., A. K. Ebralidze, T. Benoukraf, G. Amabile, L. A. Goff, J. Terragni, M. E. Figueroa, L. L. De Figueiredo Pontes, M. Alberich-Jorda, P. Zhang, M. Wu, F. D'Alo, A. Melnick, G. Leone, K. K. Ebralidze, S. Pradhan, J. L. Rinn, and D. G. Tenen. 2013. "DNMT1-interacting RNAs block gene-specific DNA methylation." *Nature* 503 (7476):371-6. doi: 10.1038/nature12598.
- Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. "STAR: ultrafast universal RNA-seq aligner." *Bioinformatics* 29 (1):15-21. doi: 10.1093/bioinformatics/bts635.
- Duttke, S. H., S. A. Lacadie, M. M. Ibrahim, C. K. Glass, D. L. Corcoran, C. Benner, S. Heinz, J. T. Kadonaga, and U. Ohler. 2015. "Human promoters are intrinsically directional." *Mol Cell* 57 (4):674-84. doi: 10.1016/j.molcel.2014.12.029.
- Ebright, R. H. 2000. "RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II." *J Mol Biol* 304 (5):687-98. doi: 10.1006/jmbi.2000.4309.
- Eick, D., and M. Geyer. 2013. "The RNA polymerase II carboxy-terminal domain (CTD) code." *Chem Rev* 113 (11):8456-90. doi: 10.1021/cr400071f.
- Eisen, A., R. T. Utley, A. Nourani, S. Allard, P. Schmidt, W. S. Lane, J. C. Lucchesi, and J. Cote. 2001. "The yeast NuA4 and Drosophila MSL complexes contain homologous subunits important for transcription regulation." *J Biol Chem* 276 (5):3484-91. doi: 10.1074/jbc.M008159200.
- Fabrega, C., V. Shen, S. Shuman, and C. D. Lima. 2003. "Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II." *Mol Cell* 11 (6):1549-61.
- Felsenfeld, G., and M. Groudine. 2003. "Controlling the double helix." *Nature* 421 (6921):448-53. doi: 10.1038/nature01411.
- Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. Sonnhammer, J. Tate, and M. Punta. 2014.

- "Pfam: the protein families database." *Nucleic Acids Res* 42 (Database issue):D222-30. doi: 10.1093/nar/gkt1223.
- Fischl, H., F. S. Howe, A. Furger, and J. Mellor. 2017. "Paf1 Has Distinct Roles in Transcription Elongation and Differential Transcript Fate." *Mol Cell* 65 (4):685-698 e8. doi: 10.1016/j.molcel.2017.01.006.
- Fishburn, J., E. Galburt, and S. Hahn. 2016. "Transcription Start Site Scanning and the Requirement for ATP during Transcription Initiation by RNA Polymerase II." *J Biol Chem* 291 (25):13040-7. doi: 10.1074/jbc.M116.724583.
- Gilbert, C., and J. Q. Svejstrup. 2006. "RNA immunoprecipitation for determining RNA-protein associations in vivo." *Curr Protoc Mol Biol* Chapter 27:Unit 27 4. doi: 10.1002/0471142727.mb2704s75.
- Gilbert, T. M., S. L. McDaniel, S. D. Byrum, J. A. Cades, B. C. Dancy, H. Wade, A. J. Tackett, B. D. Strahl, and S. D. Taverna. 2014. "A PWWP domain-containing protein targets the NuA3 acetyltransferase complex via histone H3 lysine 36 trimethylation to coordinate transcriptional elongation at coding regions." *Mol Cell Proteomics* 13 (11):2883-95. doi: 10.1074/mcp.M114.038224.
- Ginsburg, D. S., T. E. Anlembom, J. Wang, S. R. Patel, B. Li, and A. G. Hinnebusch. 2014. "NuA4 links methylation of histone H3 lysines 4 and 36 to acetylation of histones H4 and H3." *J Biol Chem* 289 (47):32656-70. doi: 10.1074/jbc.M114.585588.
- Grohmann, D., J. Nagy, A. Chakraborty, D. Klose, D. Fielden, R. H. Ebright, J. Michaelis, and F. Werner. 2011. "The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation." *Mol Cell* 43 (2):263-74. doi: 10.1016/j.molcel.2011.05.030.
- Hafner, M., M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, A. C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan, and T. Tuschl. 2010. "PAR-Clip--a method to identify transcriptome-wide the binding sites of RNA binding proteins." *J Vis Exp* (41). doi: 10.3791/2034.
- Hahn, S. 2004. "Structure and mechanism of the RNA polymerase II transcription machinery." *Nat Struct Mol Biol* 11 (5):394-403. doi: 10.1038/nsmb763.
- Halbach, A., H. Zhang, A. Wengi, Z. Jablonska, I. M. Gruber, R. E. Halbeisen, P. M. Dehe, P. Kemmeren, F. Holstege, V. Geli, A. P. Gerber, and B. Dichtl. 2009. "Cotranslational assembly of the yeast SET1C histone methyltransferase complex." *EMBO J* 28 (19):2959-70. doi: 10.1038/emboj.2009.240.
- Hampsey, M., B. N. Singh, A. Ansari, J. P. Laine, and S. Krishnamurthy. 2011. "Control of eukaryotic gene expression: gene loops and transcriptional memory." *Adv Enzyme Regul* 51 (1):118-25. doi: 10.1016/j.advenzreg.2010.10.001.
- Hartzog, G. A., and J. Fu. 2013. "The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation." *Biochim Biophys Acta* 1829 (1):105-15. doi: 10.1016/j.bbagr.2012.08.007.
- Haruki, H., J. Nishikawa, and U. K. Laemmli. 2008. "The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes." *Mol Cell* 31 (6):925-32. doi: 10.1016/j.molcel.2008.07.020.

- Heidemann, M., C. Hintermair, K. Voss, and D. Eick. 2013. "Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription." *Biochim Biophys Acta* 1829 (1):55-62. doi: 10.1016/j.bbagr.2012.08.013.
- Hendrickson, D. G., D. R. Kelley, D. Tenen, B. Bernstein, and J. L. Rinn. 2016. "Widespread RNA binding by chromatin-associated proteins." *Genome Biol* 17:28. doi: 10.1186/s13059-016-0878-3.
- Henikoff, S. 2008. "Nucleosome destabilization in the epigenetic regulation of gene expression." *Nat Rev Genet* 9 (1):15-26. doi: 10.1038/nrg2206.
- Hirata, A., B. J. Klein, and K. S. Murakami. 2008. "The X-ray crystal structure of RNA polymerase from Archaea." *Nature* 451 (7180):851-4. doi: 10.1038/nature06530.
- Holstege, F. C., U. Fiedler, and H. T. Timmers. 1997. "Three transitions in the RNA polymerase II transcription complex during initiation." *EMBO J* 16 (24):7468-80. doi: 10.1093/emboj/16.24.7468.
- Hossain, M. A., C. Chung, S. K. Pradhan, and T. L. Johnson. 2013. "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. doi: 10.1128/mcb.00947-12.
- Hsin, J. P., and J. L. Manley. 2012. "The RNA polymerase II CTD coordinates transcription and RNA processing." *Genes Dev* 26 (19):2119-37. doi: 10.1101/gad.200303.112.
- Jain, R., T. Devine, A. D. George, S. V. Chittur, T. E. Baroni, L. O. Penalva, and S. A. Tenenbaum. 2011. "RIP-Chip analysis: RNA-Binding Protein Immunoprecipitation-Microarray (Chip) Profiling." *Methods Mol Biol* 703:247-63. doi: 10.1007/978-1-59745-248-9_17.
- Jenuwein, T., and C. D. Allis. 2001. "Translating the histone code." *Science* 293 (5532):1074-80. doi: 10.1126/science.1063127.
- Jeon, Y., and J. T. Lee. 2011. "YY1 tethers Xist RNA to the inactive X nucleation center." *Cell* 146 (1):119-33. doi: 10.1016/j.cell.2011.06.026.
- Jeronimo, C., A. R. Bataille, and F. Robert. 2013. "The writers, readers, and functions of the RNA polymerase II C-terminal domain code." *Chem Rev* 113 (11):8491-522. doi: 10.1021/cr4001397.
- Jonkers, I., and J. T. Lis. 2015. "Getting up to speed with transcription elongation by RNA polymerase II." *Nat Rev Mol Cell Biol* 16 (3):167-77. doi: 10.1038/nrm3953.
- Kaneko, S., J. Son, R. Bonasio, S. S. Shen, and D. Reinberg. 2014. "Nascent RNA interaction keeps PRC2 activity poised and in check." *Genes Dev* 28 (18):1983-8. doi: 10.1101/gad.247940.114.
- Kaneko, S., J. Son, S. S. Shen, D. Reinberg, and R. Bonasio. 2013. "PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells." *Nat Struct Mol Biol* 20 (11):1258-64. doi: 10.1038/nsmb.2700.
- Keene, J. D., J. M. Komisarow, and M. B. Friedersdorf. 2006. "RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts." *Nat Protoc* 1 (1):302-7. doi: 10.1038/nprot.2006.47.

- Keller, W., and L. Minvielle-Sebastia. 1997. "A comparison of mammalian and yeast pre-mRNA 3'-end processing." *Curr Opin Cell Biol* 9 (3):329-36.
- Keogh, M. C., V. Podolny, and S. Buratowski. 2003. "Bur1 kinase is required for efficient transcription elongation by RNA polymerase II." *Mol Cell Biol* 23 (19):7005-18.
- Kim, H., B. Erickson, W. Luo, D. Seward, J. H. Graber, D. D. Pollock, P. C. Megee, and D. L. Bentley. 2010. "Gene-specific RNA polymerase II phosphorylation and the CTD code." *Nat Struct Mol Biol* 17 (10):1279-86. doi: 10.1038/nsmb.1913.
- Kim, M., S. H. Ahn, N. J. Krogan, J. F. Greenblatt, and S. Buratowski. 2004. "Transitions in RNA polymerase II elongation complexes at the 3' ends of genes." *Embo j* 23 (2):354-64. doi: 10.1038/sj.emboj.7600053.
- Kinkelin, K., G. G. Wozniak, S. B. Rothbart, M. Lidschreiber, B. D. Strahl, and P. Cramer. 2013. "Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue." *Proc Natl Acad Sci U S A* 110 (38):15277-82. doi: 10.1073/pnas.1311010110.
- Kizer, K. O., H. P. Phatnani, Y. Shibata, H. Hall, A. L. Greenleaf, and B. D. Strahl. 2005. "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. doi: 10.1128/mcb.25.8.3305-3316.2005.
- Klein, B. J., D. Bose, K. J. Baker, Z. M. Yusoff, X. Zhang, and K. S. Murakami. 2011. "RNA polymerase and transcription elongation factor Spt4/5 complex structure." *Proc Natl Acad Sci U S A* 108 (2):546-50. doi: 10.1073/pnas.1013828108.
- Komarnitsky, P., E. J. Cho, and S. Buratowski. 2000. "Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription." *Genes Dev* 14 (19):2452-60.
- Konig, J., K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombe, and J. Ule. 2010. "iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution." *Nat Struct Mol Biol* 17 (7):909-15. doi: 10.1038/nsmb.1838.
- Kornblihtt, A. R., M. de la Mata, J. P. Fededa, M. J. Munoz, and G. Nogues. 2004. "Multiple links between transcription and splicing." *Rna* 10 (10):1489-98. doi: 10.1261/rna.7100104.
- Krogan, N. J., J. Dover, A. Wood, J. Schneider, J. Heidt, M. A. Boateng, K. Dean, O. W. Ryan, A. Golshani, M. Johnston, J. F. Greenblatt, and A. Shilatifard. 2003. "The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation." *Mol Cell* 11 (3):721-9.
- Krogan, N. J., M. Kim, A. Tong, A. Golshani, G. Cagney, V. Canadien, D. P. Richards, B. K. Beattie, A. Emili, C. Boone, A. Shilatifard, S. Buratowski, and J. Greenblatt. 2003. "Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II." *Mol Cell Biol* 23 (12):4207-18.
- Kulak, N. A., G. Pichler, I. Paron, N. Nagaraj, and M. Mann. 2014. "Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells." *Nat Methods* 11 (3):319-24. doi: 10.1038/nmeth.2834.

- Kusser, A. G., M. G. Bertero, S. Naji, T. Becker, M. Thomm, R. Beckmann, and P. Cramer. 2008. "Structure of an archaeal RNA polymerase." *J Mol Biol* 376 (2):303-7. doi: 10.1016/j.jmb.2007.08.066.
- Kwak, H., N. J. Fuda, L. J. Core, and J. T. Lis. 2013. "Precise maps of RNA polymerase reveal how promoters direct initiation and pausing." *Science* 339 (6122):950-3. doi: 10.1126/science.1229386.
- Langmead, B., and S. L. Salzberg. 2012. "Fast gapped-read alignment with Bowtie 2." *Nat Methods* 9 (4):357-9. doi: 10.1038/nmeth.1923.
- Lewis, J. D., and E. Izaurralde. 1997. "The role of the cap structure in RNA processing and nuclear export." *Eur J Biochem* 247 (2):461-9.
- Li, B., L. Howe, S. Anderson, J. R. Yates, 3rd, and J. L. Workman. 2003. "The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II." *J Biol Chem* 278 (11):8897-903. doi: 10.1074/jbc.M212134200.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. "The Sequence Alignment/Map format and SAMtools." *Bioinformatics* 25 (16):2078-9. doi: 10.1093/bioinformatics/btp352.
- Li, J., D. Moazed, and S. P. Gygi. 2002. "Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation." *J Biol Chem* 277 (51):49383-8. doi: 10.1074/jbc.M209294200.
- Li, W., D. Notani, and M. G. Rosenfeld. 2016. "Enhancers as non-coding RNA transcription units: recent insights and future perspectives." *Nat Rev Genet* 17 (4):207-23. doi: 10.1038/nrg.2016.4.
- Liang, G., R. J. Klose, K. E. Gardner, and Y. Zhang. 2007. "Yeast Jhd2p is a histone H3 Lys4 trimethyl demethylase." *Nat Struct Mol Biol* 14 (3):243-5. doi: 10.1038/nsmb1204.
- Licatalosi, D. D., A. Mele, J. J. Fak, J. Ule, M. Kayikci, S. W. Chi, T. A. Clark, A. C. Schweitzer, J. E. Blume, X. Wang, J. C. Darnell, and R. B. Darnell. 2008. "HITS-CLIP yields genome-wide insights into brain alternative RNA processing." *Nature* 456 (7221):464-9. doi: 10.1038/nature07488.
- Lidschreiber, M., K. Leike, and P. Cramer. 2013. "Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II." *Mol Cell Biol* 33 (19):3805-16. doi: 10.1128/mcb.00361-13.
- Listerman, I., A. K. Sapra, and K. M. Neugebauer. 2006. "Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells." *Nat Struct Mol Biol* 13 (9):815-22. doi: 10.1038/nsmb1135.
- Liu, C. L., T. Kaplan, M. Kim, S. Buratowski, S. L. Schreiber, N. Friedman, and O. J. Rando. 2005. "Single-nucleosome mapping of histone modifications in *S. cerevisiae*." *PLoS Biol* 3 (10):e328. doi: 10.1371/journal.pbio.0030328.
- Liu, Y., L. Warfield, C. Zhang, J. Luo, J. Allen, W. H. Lang, J. Ranish, K. M. Shokat, and S. Hahn. 2009. "Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex." *Mol Cell Biol* 29 (17):4852-63. doi: 10.1128/mcb.00609-09.

- Luehr, S., H. Hartmann, and J. Söding. 2012. "The XXmotif web server for eXhaustive, weight matriX-based motif discovery in nucleotide sequences." *Nucleic Acids Res* 40 (Web Server issue):W104-9. doi: 10.1093/nar/gks602.
- Luo, W., and D. Bentley. 2004. "A ribonucleolytic rat torpedo RNA polymerase II." *Cell* 119 (7):911-4. doi: 10.1016/j.cell.2004.11.041.
- Luse, D. S. 2013. "Promoter clearance by RNA polymerase II." *Biochim Biophys Acta* 1829 (1):63-8. doi: 10.1016/j.bbagr.2012.08.010.
- MacKellar, A. L., and A. L. Greenleaf. 2011. "Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II." *J Biol Chem* 286 (42):36385-95. doi: 10.1074/jbc.M111.268144.
- Manley, J. L., and Y. Takagaki. 1996. "The end of the message--another link between yeast and mammals." *Science* 274 (5292):1481-2.
- Margaritis, T., and F. C. Holstege. 2008. "Poised RNA polymerase II gives pause for thought." *Cell* 133 (4):581-4. doi: 10.1016/j.cell.2008.04.027.
- Marquardt, S., R. Escalante-Chong, N. Pho, J. Wang, L. S. Churchman, M. Springer, and S. Buratowski. 2014. "A chromatin-based mechanism for limiting divergent noncoding transcription." *Cell* 157 (7):1712-23. doi: 10.1016/j.cell.2014.04.036.
- Marshall, N. F., J. Peng, Z. Xie, and D. H. Price. 1996. "Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase." *J Biol Chem* 271 (43):27176-83.
- Marshall, N. F., and D. H. Price. 1995. "Purification of P-TEFb, a transcription factor required for the transition into productive elongation." *J Biol Chem* 270 (21):12335-8.
- Martin, D. G., K. Baetz, X. Shi, K. L. Walter, V. E. MacDonald, M. J. Wlodarski, O. Gozani, P. Hieter, and L. Howe. 2006. "The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3." *Mol Cell Biol* 26 (21):7871-9. doi: 10.1128/MCB.00573-06.
- Martinez-Rucobo, F. W., R. Kohler, M. van de Waterbeemd, A. J. Heck, M. Hemann, F. Herzog, H. Stark, and P. Cramer. 2015. "Molecular Basis of Transcription-Coupled Pre-mRNA Capping." *Mol Cell* 58 (6):1079-89. doi: 10.1016/j.molcel.2015.04.004.
- Martinez-Rucobo, F. W., S. Sainsbury, A. C. Cheung, and P. Cramer. 2011. "Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity." *Embo j* 30 (7):1302-10. doi: 10.1038/emboj.2011.64.
- Mayekar, M. K., R. G. Gardner, and K. M. Arndt. 2013. "The recruitment of the *Saccharomyces cerevisiae* Paf1 complex to active genes requires a domain of Rtf1 that directly interacts with the Spt4-Spt5 complex." *Mol Cell Biol* 33 (16):3259-73. doi: 10.1128/mcb.00270-13.
- Mayer, A., M. Heidemann, M. Lidschreiber, A. Schrieck, M. Sun, C. Hintermair, E. Kremmer, D. Eick, and P. Cramer. 2012. "CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II." *Science* 336 (6089):1723-5. doi: 10.1126/science.1219651.

- Mayer, A., M. Lidschreiber, M. Siebert, K. Leike, J. Söding, and P. Cramer. 2010. "Uniform transitions of the general RNA polymerase II transcription complex." *Nat Struct Mol Biol* 17 (10):1272-8. doi: 10.1038/nsmb.1903.
- Mayer, A., A. Schreieck, M. Lidschreiber, K. Leike, D. E. Martin, and P. Cramer. 2012. "The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I." *Mol Cell Biol* 32 (7):1321-31. doi: 10.1128/MCB.06310-11.
- McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, S. Shuman, and D. L. Bentley. 1997. "5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II." *Genes Dev* 11 (24):3306-18.
- Meinhart, A., and P. Cramer. 2004. "Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors." *Nature* 430 (6996):223-6. doi: 10.1038/nature02679.
- Mercer, T. R., S. Neph, M. E. Dinger, J. Crawford, M. A. Smith, A. M. Shearwood, E. Haugen, C. P. Bracken, O. Rackham, J. A. Stamatoyannopoulos, A. Filipovska, and J. S. Mattick. 2011. "The human mitochondrial transcriptome." *Cell* 146 (4):645-58. doi: 10.1016/j.cell.2011.06.051.
- Meyer, P. A., S. Li, M. Zhang, K. Yamada, Y. Takagi, G. A. Hartzog, and J. Fu. 2015. "Structures and Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5." *Mol Cell Biol* 35 (19):3354-69. doi: 10.1128/mcb.00520-15.
- Mili, S., and J. A. Steitz. 2004. "Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses." *RNA* 10 (11):1692-4. doi: 10.1261/rna.7151404.
- Miller, T., N. J. Krogan, J. Dover, H. Erdjument-Bromage, P. Tempst, M. Johnston, J. F. Greenblatt, and A. Shilatifard. 2001. "COMPASS: a complex of proteins associated with a trithorax-related SET domain protein." *Proc Natl Acad Sci U S A* 98 (23):12902-7. doi: 10.1073/pnas.231473398.
- Mischo, H. E., and N. J. Proudfoot. 2013. "Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast." *Biochim Biophys Acta* 1829 (1):174-85. doi: 10.1016/j.bbagr.2012.10.003.
- Missra, A., and D. S. Gilmour. 2010. "Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex." *Proc Natl Acad Sci U S A* 107 (25):11301-6. doi: 10.1073/pnas.1000681107.
- Mitchell, P. J., and R. Tjian. 1989. "Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins." *Science* 245 (4916):371-8.
- Moss, T., F. Langlois, T. Gagnon-Kugler, and V. Stefanovsky. 2007. "A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis." *Cell Mol Life Sci* 64 (1):29-49. doi: 10.1007/s00018-006-6278-1.
- Muhlbacher, W., A. Mayer, M. Sun, M. Remmert, A. C. Cheung, J. Niesser, J. Söding, and P. Cramer. 2015. "Structure of Ctk3, a subunit of the RNA polymerase II CTD kinase

- complex, reveals a noncanonical CTD-interacting domain fold." *Proteins* 83 (10):1849-58. doi: 10.1002/prot.24869.
- Mulder, K. W., A. B. Brenkman, A. Inagaki, N. J. van den Broek, and H. T. Timmers. 2007. "Regulation of histone H3K4 tri-methylation and PAF complex recruitment by the Ccr4-Not complex." *Nucleic Acids Res* 35 (7):2428-39. doi: 10.1093/nar/gkm175.
- Murray, S., R. Udupa, S. Yao, G. Hartzog, and G. Prelich. 2001. "Phosphorylation of the RNA polymerase II carboxy-terminal domain by the Bur1 cyclin-dependent kinase." *Mol Cell Biol* 21 (13):4089-96. doi: 10.1128/MCB.21.13.4089-4096.2001.
- Nakajima, N., M. Horikoshi, and R. G. Roeder. 1988. "Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID." *Mol Cell Biol* 8 (10):4028-40.
- Nechaev, S., and K. Adelman. 2011. "Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation." *Biochim Biophys Acta* 1809 (1):34-45. doi: 10.1016/j.bbagr.2010.11.001.
- Ng, H. H., F. Robert, R. A. Young, and K. Struhl. 2003. "Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity." *Mol Cell* 11 (3):709-19.
- Nguyen, A. T., and Y. Zhang. 2011. "The diverse functions of Dot1 and H3K79 methylation." *Genes Dev* 25 (13):1345-58. doi: 10.1101/gad.2057811.
- Ni, Z., A. Saunders, N. J. Fuda, J. Yao, J. R. Suarez, W. W. Webb, and J. T. Lis. 2008. "P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo." *Mol Cell Biol* 28 (3):1161-70. doi: 10.1128/MCB.01859-07.
- Orphanides, G., and D. Reinberg. 2000. "RNA polymerase II elongation through chromatin." *Nature* 407 (6803):471-5. doi: 10.1038/35035000.
- Orphanides, G., and D. Reinberg. 2002. "A unified theory of gene expression." *Cell* 108 (4):439-51.
- Pal, M., A. S. Ponticelli, and D. S. Luse. 2005. "The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II." *Mol Cell* 19 (1):101-10. doi: 10.1016/j.molcel.2005.05.024.
- Parker, C. S., and J. Topol. 1984. "A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene." *Cell* 37 (1):273-83.
- Patturajan, M., N. K. Conrad, D. B. Bregman, and J. L. Corden. 1999. "Yeast carboxyl-terminal domain kinase I positively and negatively regulates RNA polymerase II carboxyl-terminal domain phosphorylation." *J Biol Chem* 274 (39):27823-8.
- Pelechano, V., W. Wei, and L. M. Steinmetz. 2013. "Extensive transcriptional heterogeneity revealed by isoform profiling." *Nature* 497 (7447):127-31. doi: 10.1038/nature12121.
- Pelham, H. R. 1982. "A regulatory upstream promoter element in the Drosophila hsp 70 heat-shock gene." *Cell* 30 (2):517-28.

- Perales, R., and D. Bentley. 2009. "'Cotranscriptionality': the transcription elongation complex as a nexus for nuclear transactions." *Mol Cell* 36 (2):178-91. doi: 10.1016/j.molcel.2009.09.018.
- Peterlin, B. M., and D. H. Price. 2006. "Controlling the elongation phase of transcription with P-TEFb." *Mol Cell* 23 (3):297-305. doi: 10.1016/j.molcel.2006.06.014.
- Phatnani, H. P., J. C. Jones, and A. L. Greenleaf. 2004. "Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome." *Biochemistry* 43 (50):15702-19. doi: 10.1021/bi048364h.
- Pijnappel, W. W., D. Schaft, A. Roguev, A. Shevchenko, H. Tekotte, M. Wilm, G. Rigaut, B. Seraphin, R. Aasland, and A. F. Stewart. 2001. "The *S. cerevisiae* SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program." *Genes Dev* 15 (22):2991-3004. doi: 10.1101/gad.207401.
- Pikaard, C. S., J. R. Haag, T. Ream, and A. T. Wierzbicki. 2008. "Roles of RNA polymerase IV in gene silencing." *Trends Plant Sci* 13 (7):390-7. doi: 10.1016/j.tplants.2008.04.008.
- Plaschka, C., M. Hantsche, C. Dienemann, C. Burzinski, J. Plitzko, and P. Cramer. 2016. "Transcription initiation complex structures elucidate DNA opening." *Nature* 533 (7603):353-8. doi: 10.1038/nature17990.
- Plaschka, C., L. Lariviere, L. Wenzek, M. Seizl, M. Hemann, D. Tegunov, E. V. Petrotchenko, C. H. Borchers, W. Baumeister, F. Herzog, E. Villa, and P. Cramer. 2015. "Architecture of the RNA polymerase II-Mediator core initiation complex." *Nature* 518 (7539):376-80. doi: 10.1038/nature14229.
- Pokholok, D. K., N. M. Hannett, and R. A. Young. 2002. "Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo." *Mol Cell* 9 (4):799-809.
- Pokholok, D. K., C. T. Harbison, S. Levine, M. Cole, N. M. Hannett, T. I. Lee, G. W. Bell, K. Walker, P. A. Rolfe, E. Herbolzheimer, J. Zeitlinger, F. Lewitter, D. K. Gifford, and R. A. Young. 2005. "Genome-wide map of nucleosome acetylation and methylation in yeast." *Cell* 122 (4):517-27. doi: 10.1016/j.cell.2005.06.026.
- Porrua, O., and D. Libri. 2015. "Transcription termination and the control of the transcriptome: why, where and how to stop." *Nat Rev Mol Cell Biol* 16 (3):190-202. doi: 10.1038/nrm3943.
- Proudfoot, N. J. 1989. "How RNA polymerase II terminates transcription in higher eukaryotes." *Trends Biochem Sci* 14 (3):105-10.
- Proudfoot, N. J. 2011. "Ending the message: poly(A) signals then and now." *Genes Dev* 25 (17):1770-82. doi: 10.1101/gad.17268411.
- Qin, D., L. Huang, A. Wlodaver, J. Andrade, and J. P. Staley. 2016. "Sequencing of lariat termini in *S. cerevisiae* reveals 5' splice sites, branch points, and novel splicing events." *Rna* 22 (2):237-53. doi: 10.1261/rna.052829.115.

- Qiu, H., C. Hu, N. A. Gaur, and A. G. Hinnebusch. 2012. "Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex." *EMBO J* 31 (16):3494-505. doi: 10.1038/emboj.2012.188.
- Qiu, H., C. Hu, and A. G. Hinnebusch. 2009. "Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters." *Mol Cell* 33 (6):752-62. doi: 10.1016/j.molcel.2009.02.018.
- Qiu, H., C. Hu, C. M. Wong, and A. G. Hinnebusch. 2006. "The Spt4p subunit of yeast DSIF stimulates association of the Paf1 complex with elongating RNA polymerase II." *Mol Cell Biol* 26 (8):3135-48. doi: 10.1128/MCB.26.8.3135-3148.2006.
- Quinlan, A. R., and I. M. Hall. 2010. "BEDTools: a flexible suite of utilities for comparing genomic features." *Bioinformatics* 26 (6):841-2. doi: 10.1093/bioinformatics/btq033.
- Rahl, P. B., C. Y. Lin, A. C. Seila, R. A. Flynn, S. McCuine, C. B. Burge, P. A. Sharp, and R. A. Young. 2010. "c-Myc regulates transcriptional pause release." *Cell* 141 (3):432-45. doi: 10.1016/j.cell.2010.03.030.
- Ramanathan, Y., S. M. Rajpara, S. M. Reza, E. Lees, S. Shuman, M. B. Mathews, and T. Pe'ery. 2001. "Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences." *J Biol Chem* 276 (14):10913-20. doi: 10.1074/jbc.M010975200.
- Ramirez, F., F. Dundar, S. Diehl, B. A. Gruning, and T. Manke. 2014. "deepTools: a flexible platform for exploring deep-sequencing data." *Nucleic Acids Res* 42 (Web Server issue):W187-91. doi: 10.1093/nar/gku365.
- Richard, P., and J. L. Manley. 2009. "Transcription termination by nuclear RNA polymerases." *Genes Dev* 23 (11):1247-69. doi: 10.1101/gad.1792809.
- Robzyk, K., J. Recht, and M. A. Osley. 2000. "Rad6-dependent ubiquitination of histone H2B in yeast." *Science* 287 (5452):501-4.
- Rodriguez, C. R., E. J. Cho, M. C. Keogh, C. L. Moore, A. L. Greenleaf, and S. Buratowski. 2000. "Kin28, the TFIIF-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II." *Mol Cell Biol* 20 (1):104-12.
- Sainsbury, S., C. Bernecky, and P. Cramer. 2015. "Structural basis of transcription initiation by RNA polymerase II." *Nat Rev Mol Cell Biol* 16 (3):129-43. doi: 10.1038/nrm3952.
- Saldi, T., M. A. Cortazar, R. M. Sheridan, and D. L. Bentley. 2016. "Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing." *J Mol Biol*. doi: 10.1016/j.jmb.2016.04.017.
- Schroder, S., E. Herker, F. Itzen, D. He, S. Thomas, D. A. Gilchrist, K. Kaehlcke, S. Cho, K. S. Pollard, J. A. Capra, M. Schnolzer, P. A. Cole, M. Geyer, B. G. Bruneau, K. Adelman, and M. Ott. 2013. "Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells." *Mol Cell* 52 (3):314-24. doi: 10.1016/j.molcel.2013.10.009.

- Schroeder, S. C., B. Schwer, S. Shuman, and D. Bentley. 2000. "Dynamic association of capping enzymes with transcribing RNA polymerase II." *Genes Dev* 14 (19):2435-40.
- Schulz, D., B. Schwalb, A. Kiesel, C. Baejen, P. Torkler, J. Gagneur, J. Söding, and P. Cramer. 2013. "Transcriptome surveillance by selective termination of noncoding RNA synthesis." *Cell* 155 (5):1075-87. doi: 10.1016/j.cell.2013.10.024.
- Schwalb, B., M. Michel, B. Zacher, K. Fruhauf, C. Demel, A. Tresch, J. Gagneur, and P. Cramer. 2016. "TT-seq maps the human transient transcriptome." *Science* 352 (6290):1225-8. doi: 10.1126/science.aad9841.
- Schwer, B., X. Mao, and S. Shuman. 1998. "Accelerated mRNA decay in conditional mutants of yeast mRNA capping enzyme." *Nucleic Acids Res* 26 (9):2050-7.
- Schwer, B., and S. Shuman. 1996. "Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing in vivo." *RNA* 2 (6):574-83.
- Schwer, B., and S. Shuman. 2011. "Deciphering the RNA polymerase II CTD code in fission yeast." *Mol Cell* 43 (2):311-8. doi: 10.1016/j.molcel.2011.05.024.
- Schwinghammer, K., A. C. Cheung, Y. I. Morozov, K. Agaronyan, D. Temiakov, and P. Cramer. 2013. "Structure of human mitochondrial RNA polymerase elongation complex." *Nat Struct Mol Biol* 20 (11):1298-303. doi: 10.1038/nsmb.2683.
- Shi, X., I. Kachirskaia, K. L. Walter, J. H. Kuo, A. Lake, F. Davrazou, S. M. Chan, D. G. Martin, I. M. Fingerman, S. D. Briggs, L. Howe, P. J. Utz, T. G. Kutateladze, A. A. Lugovskoy, M. T. Bedford, and O. Gozani. 2007. "Proteome-wide analysis in *Saccharomyces cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36." *J Biol Chem* 282 (4):2450-5. doi: 10.1074/jbc.C600286200.
- Shi, Y., and J. L. Manley. 2015. "The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site." *Genes Dev* 29 (9):889-97. doi: 10.1101/gad.261974.115.
- Shilatifard, A. 2004. "Transcriptional elongation control by RNA polymerase II: a new frontier." *Biochim Biophys Acta* 1677 (1-3):79-86. doi: 10.1016/j.bbaexp.2003.11.013.
- Shilatifard, A., R. C. Conaway, and J. W. Conaway. 2003. "The RNA polymerase II elongation complex." *Annu Rev Biochem* 72:693-715. doi: 10.1146/annurev.biochem.72.121801.161551.
- Sigova, A. A., B. J. Abraham, X. Ji, B. Molinie, N. M. Hannett, Y. E. Guo, M. Jangi, C. C. Giallourakis, P. A. Sharp, and R. A. Young. 2015. "Transcription factor trapping by RNA in gene regulatory elements." *Science* 350 (6263):978-81. doi: 10.1126/science.aad3346.
- Sims, R. J., 3rd, R. Belotserkovskaya, and D. Reinberg. 2004. "Elongation by RNA polymerase II: the short and long of it." *Genes Dev* 18 (20):2437-68. doi: 10.1101/gad.1235904.

- Singh, R. K., M. Gonzalez, M. H. Kabbaj, and A. Gunjan. 2012. "Novel E3 ubiquitin ligases that regulate histone protein levels in the budding yeast *Saccharomyces cerevisiae*." *PLoS One* 7 (5):e36295. doi: 10.1371/journal.pone.0036295.
- Skalska, L., M. Beltran-Nebot, J. Ule, and R. G. Jenner. 2017. "Regulatory feedback from nascent RNA to chromatin and transcription." *Nat Rev Mol Cell Biol*. doi: 10.1038/nrm.2017.12.
- Smolle, M., S. Venkatesh, M. M. Gogol, H. Li, Y. Zhang, L. Florens, M. P. Washburn, and J. L. Workman. 2012. "Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange." *Nat Struct Mol Biol* 19 (9):884-92. doi: 10.1038/nsmb.2312.
- Smolle, M., and J. L. Workman. 2013. "Transcription-associated histone modifications and cryptic transcription." *Biochim Biophys Acta* 1829 (1):84-97. doi: 10.1016/j.bbagr.2012.08.008.
- Södning, J., A. Biegert, and A. N. Lupas. 2005. "The HHpred interactive server for protein homology detection and structure prediction." *Nucleic Acids Res* 33 (Web Server issue):W244-8. doi: 10.1093/nar/gki408.
- Sogaard, T. M., and J. Q. Svejstrup. 2007. "Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator." *J Biol Chem* 282 (19):14113-20. doi: 10.1074/jbc.M701345200.
- Sonenberg, N., and A. G. Hinnebusch. 2009. "Regulation of translation initiation in eukaryotes: mechanisms and biological targets." *Cell* 136 (4):731-45. doi: 10.1016/j.cell.2009.01.042.
- Steinmetz, E. J., and D. A. Brow. 1996. "Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1." *Mol Cell Biol* 16 (12):6993-7003.
- Sterner, D. E., J. M. Lee, S. E. Hardin, and A. L. Greenleaf. 1995. "The yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex." *Mol Cell Biol* 15 (10):5716-24.
- Sun, M., L. Lariviere, S. Dengl, A. Mayer, and P. Cramer. 2010. "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. doi: 10.1074/jbc.M110.144568.
- Sun, M., B. Schwalb, D. Schulz, N. Pirkl, S. Etzold, L. Lariviere, K. C. Maier, M. Seizl, A. Tresch, and P. Cramer. 2012. "Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation." *Genome Res* 22 (7):1350-9. doi: 10.1101/gr.130161.111.
- Tennyson, C. N., H. J. Klamut, and R. G. Worton. 1995. "The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced." *Nat Genet* 9 (2):184-90. doi: 10.1038/ng0295-184.
- Terzi, N., L. S. Churchman, L. Vasiljeva, J. Weissman, and S. Buratowski. 2011. "H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway." *Mol Cell Biol* 31 (17):3569-83. doi: 10.1128/MCB.05590-11.

- Thomas, M. C., and C. M. Chiang. 2006. "The general transcription machinery and general cofactors." *Crit Rev Biochem Mol Biol* 41 (3):105-78. doi: 10.1080/10409230600648736.
- Tietjen, J. R., D. W. Zhang, J. B. Rodriguez-Molina, B. E. White, M. S. Akhtar, M. Heidemann, X. Li, R. D. Chapman, K. Shokat, S. Keles, D. Eick, and A. Z. Ansari. 2010. "Chemical-genomic dissection of the CTD code." *Nat Struct Mol Biol* 17 (9):1154-61. doi: 10.1038/nsmb.1900.
- Tresaugues, L., P. M. Dehe, R. Guerois, A. Rodriguez-Gil, I. Varlet, P. Salah, M. Pamblanco, P. Luciano, S. Quevillon-Cheruel, J. Sollier, N. Leulliot, J. Couprie, V. Tordera, S. Zinn-Justin, S. Chavez, H. van Tilbeurgh, and V. Geli. 2006. "Structural characterization of Set1 RNA recognition motifs and their role in histone H3 lysine 4 methylation." *J Mol Biol* 359 (5):1170-81. doi: 10.1016/j.jmb.2006.04.050.
- Tsukada, Y., J. Fang, H. Erdjument-Bromage, M. E. Warren, C. H. Borchers, P. Tempst, and Y. Zhang. 2006. "Histone demethylation by a family of JmjC domain-containing proteins." *Nature* 439 (7078):811-6. doi: 10.1038/nature04433.
- Tuck, A. C., and D. Tollervey. 2013. "A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs." *Cell* 154 (5):996-1009. doi: 10.1016/j.cell.2013.07.047.
- Ule, J., K. B. Jensen, M. Ruggiu, A. Mele, A. Ule, and R. B. Darnell. 2003. "CLIP identifies Nova-regulated RNA networks in the brain." *Science* 302 (5648):1212-5. doi: 10.1126/science.1090095.
- Vasiljeva, L., M. Kim, H. Mutschler, S. Buratowski, and A. Meinhart. 2008. "The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain." *Nat Struct Mol Biol* 15 (8):795-804. doi: 10.1038/nsmb.1468.
- Venema, J., and D. Tollervey. 1999. "Ribosome synthesis in *Saccharomyces cerevisiae*." *Annu Rev Genet* 33:261-311. doi: 10.1146/annurev.genet.33.1.261.
- Villard, J. 2004. "Transcription regulation and human diseases." *Swiss Med Wkly* 134 (39-40):571-9. doi: 2004/39/smw-10191.
- Wang, D., D. A. Bushnell, K. D. Westover, C. D. Kaplan, and R. D. Kornberg. 2006. "Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis." *Cell* 127 (5):941-54. doi: 10.1016/j.cell.2006.11.023.
- Wang, G., X. Tong, S. Weng, and H. Zhou. 2012. "Multiple phosphorylation of Rad9 by CDK is required for DNA damage checkpoint activation." *Cell Cycle* 11 (20):3792-800. doi: 10.4161/cc.21987.
- Wang, K. C., Y. W. Yang, B. Liu, A. Sanyal, R. Corces-Zimmerman, Y. Chen, B. R. Lajoie, A. Protacio, R. A. Flynn, R. A. Gupta, J. Wysocka, M. Lei, J. Dekker, J. A. Helms, and H. Y. Chang. 2011. "A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression." *Nature* 472 (7341):120-4. doi: 10.1038/nature09819.
- Wang, W., M. Carey, and J. D. Gralla. 1992. "Polymerase II promoter activation: closed complex formation and ATP-driven start site opening." *Science* 255 (5043):450-3.

- Wei, C. M., and B. Moss. 1975. "Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA." *Proc Natl Acad Sci U S A* 72 (1):318-22.
- Wei, C., and B. Moss. 1977. "5'-Terminal capping of RNA by guanylyltransferase from HeLa cell nuclei." *Proc Natl Acad Sci U S A* 74 (9):3758-61.
- Wei, P., M. E. Garber, S. M. Fang, W. H. Fischer, and K. A. Jones. 1998. "A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA." *Cell* 92 (4):451-62.
- Weiner, A., T. H. Hsieh, A. Appleboim, H. V. Chen, A. Rahat, I. Amit, O. J. Rando, and N. Friedman. 2015. "High-resolution chromatin dynamics during a yeast stress response." *Mol Cell* 58 (2):371-86. doi: 10.1016/j.molcel.2015.02.002.
- Westover, K. D., D. A. Bushnell, and R. D. Kornberg. 2004. "Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center." *Cell* 119 (4):481-9. doi: 10.1016/j.cell.2004.10.016.
- White, R. J. 2011. "Transcription by RNA polymerase III: more complex than we thought." *Nat Rev Genet* 12 (7):459-63. doi: 10.1038/nrg3001.
- Wier, A. D., M. K. Mayekar, A. Heroux, K. M. Arndt, and A. P. VanDemark. 2013. "Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin." *Proc Natl Acad Sci U S A* 110 (43):17290-5. doi: 10.1073/pnas.1314754110.
- Wood, A., N. J. Krogan, J. Dover, J. Schneider, J. Heidt, M. A. Boateng, K. Dean, A. Golshani, Y. Zhang, J. F. Greenblatt, M. Johnston, and A. Shilatifard. 2003. "Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter." *Mol Cell* 11 (1):267-74.
- Wyers, F., M. Rougemaille, G. Badis, J. C. Rousselle, M. E. Dufour, J. Boulay, B. Regnault, F. Devaux, A. Namane, B. Seraphin, D. Libri, and A. Jacquier. 2005. "Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase." *Cell* 121 (5):725-37. doi: 10.1016/j.cell.2005.04.030.
- Xu, Z., W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Munster, J. Camblong, E. Guffanti, F. Stutz, W. Huber, and L. M. Steinmetz. 2009. "Bidirectional promoters generate pervasive transcription in yeast." *Nature* 457 (7232):1033-7. doi: 10.1038/nature07728.
- Yang, Y. W., R. A. Flynn, Y. Chen, K. Qu, B. Wan, K. C. Wang, M. Lei, and H. Y. Chang. 2014. "Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency." *Elife* 3:e02046. doi: 10.7554/eLife.02046.
- Yoh, S. M., H. Cho, L. Pickle, R. M. Evans, and K. A. Jones. 2007. "The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export." *Genes Dev* 21 (2):160-74. doi: 10.1101/gad.1503107.
- Youdell, M. L., K. O. Kizer, E. Kisseleva-Romanova, S. M. Fuchs, E. Duro, B. D. Strahl, and J. Mellor. 2008. "Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36." *Mol Cell Biol* 28 (16):4915-26. doi: 10.1128/MCB.00001-08.
- Yudkovsky, N., J. A. Ranish, and S. Hahn. 2000. "A transcription reinitiation intermediate that is stabilized by activator." *Nature* 408 (6809):225-9. doi: 10.1038/35041603.

-
- Zhao, J., T. K. Ohsumi, J. T. Kung, Y. Ogawa, D. J. Grau, K. Sarma, J. J. Song, R. E. Kingston, M. Borowsky, and J. T. Lee. 2010. "Genome-wide identification of polycomb-associated RNAs by RIP-seq." *Mol Cell* 40 (6):939-53. doi: 10.1016/j.molcel.2010.12.011.
- Zhou, K., W. H. Kuo, J. Fillingham, and J. F. Greenblatt. 2009. "Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5." *Proc Natl Acad Sci U S A* 106 (17):6956-61. doi: 10.1073/pnas.0806302106.
- Zhou, Q., T. Li, and D. H. Price. 2012. "RNA polymerase II elongation control." *Annu Rev Biochem* 81:119-43. doi: 10.1146/annurev-biochem-052610-095910.
- Zhu, Y., T. Pe'ery, J. Peng, Y. Ramanathan, N. Marshall, T. Marshall, B. Amendt, M. B. Mathews, and D. H. Price. 1997. "Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro." *Genes Dev* 11 (20):2622-32.

Abbreviations

AA	Anchor-away
4sU	4-thiouridine
4tU	4-thiouracil
A	Adenine, adenosine
<i>ADHI</i>	Alcohol Dehydrogenase gene
<i>ALD5</i>	Aldehyde Dehydrogenase gene
ATP	Adenosine triphosphate
bp	Base pairs
BSA	Bovine serum albumin
BUR	Bypass UAS Requirement
C	Cytosine, Cytidine
Cdc73	Cell Division Cycle
CDK	Cyclin-dependent kinase
CF	Cleavage factor
ChIP	Chromatin immunoprecipitation
CPF	Cleavage and polyadenylation factor
CSM	Complete Supplement Mixture
Ct	Cycle threshold
CTD	Carboxy terminal domain (of Pol II)
CTDK	CTD Kinase
CTR	Carboxy terminal repeat (of Spt5)
Ctr9	Cln Three (CLN3) Requiring
CUTs	Cryptic unstable transcripts
DMSO	Dimethyl sulfoxide
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dNTP	Deoxynucleoside triphosphate
Dot1	Disruptor Of Telomeric silencing
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
EF	Elongation factor
eRNA	Enhancer RNA
fwd	Forward primer
G	Guanine, guanosine
GST	Glutathion-S-Transferase
GTF	General transcription factor
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	Horseradish peroxidase
IgG	Immunoglobulin G
<i>ILV5</i>	IsoLeucine-plus-Valine requiring gene
K	Lysine
kb	Kilo base pairs
$K_{d,app}$	Apparent disassociation constant
kDa	Kilodalton
KO	Knockout
LB	Luria-Bertani media
Leo1	LEft Open reading frame
LIC	Ligation independent cloning
lncRNA	Long non-coding RNA
m7G	7-methyl-guanosine
MBP	Maltose binding protein
m/s	Meter per second
mRNA	Messenger RNA
<i>MUP1</i>	Methionine Uptake gene
ncRNA	Non-coding RNA
NP-40	Nonyl phenoxypolyethoxylethanol
nt	Nucleotide

OD ₆₀₀	Optical density at 600 nm wavelength
ORF	Open reading frame
pA	Polyadenylation
Paf1C	Polymerase associated factor 1 complex
PAR-CLIP	Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
<i>PDC1</i>	Pyruvate Decarboxylase gene
PEG	Polyethylene glycol
PI	Processing index
PIC	Preinitiation complex
<i>PMAI</i>	Plasma Membrane ATPase gene
PMSF	Phenylmethylsulfonyl fluoride
PNK	Polynucleotide kinase
Pol	Polymerase
Prom.	Promoter
PVDF	polyvinylidene difluoride
qPCR	Quantitative real-time PCR
rev	Reverse primer
rpm	Rotations per minute
RRM	RNA recognition motif
rRNA	Ribosomal RNA
RT	Room temperature or reverse transcription
Rtf1	Restores TBP Function
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE	Size exclusion
-Seq	Deep-sequencing
Ser2-P, Ser5-P	Serine 2, Serine 5 phosphorylation
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
snRNA	Small nuclear RNA
snoRNA	Small nucleolar RNA
Spt	SuPpressor of Ty's
T	Thymine, thymidine
TAE	Tris-acetate-EDTA
TAP	Tandem affinity purification
TBS	Tris-buffered saline
TCA	Trichloroacetic acid
TEV	Tobacco etch virus
tRNA	Transfer RNA
TSS	Transcription start site
Tris	Tris-(hydroxymethyl)-aminomethan
Tyr1-P	Tyrosine 1 phosphorylation
U	Uracil, uridine
UAS	Upstream activating sequence
yORF	Your ORF
YPD	Yeast extract – peptone – dextrose (glucose) media
v/v %	Volume/volume percent
w/v %	Weight/volume percent