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

Insights into the mRNA transcription cycle from genome-wide occupancy profiling

Andreas Mayer
aus
Augsburg

2011

Erklärung

Diese Dissertation wurde im Sinne von §13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 (in der Fassung der sechsten Änderungssatzung vom 16. August 2010) von Herrn Prof. Dr. Patrick Cramer betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfe erarbeitet.

München, ...

.....
Andreas Mayer

Dissertation eingereicht am	08.12.2011
1. Gutachter	Prof. Dr. Patrick Cramer
2. Gutachter	PD Dr. Dietmar Martin
Mündliche Prüfung am	16.01.2012

Acknowledgements

First of all I want to thank my supervisor Prof. Dr. Patrick Cramer not only for giving me the opportunity to work on challenging projects in a fantastic scientific environment, but also for supporting me all the years. You gave me freedom to realize own ideas and excellent advice whenever needed. You are a great mentor and if I could turn back the clock, I would not hesitate to join your outstanding group again.
I want to thank my other mentors Prof. Dr. Thomas Cremer, Prof. Dr. Dirk Eick and Prof. Dr. Dietmar Martin for constant advice, help and countless discussions about science related topics and beyond. To have such superb mentors is a great honor and an extraordinary gift.
I want to especially thank Mai Sun for being the best bench neighbor who I can imagine. You taught me so many things and it was always a great fun working with you. I will never forget our discussions about the important and not so important topics in life. Thank you Mai for your advice, your help, sharing projects with me, the trips into the surroundings of Munich and many things more. I will keep my fingers crossed that your gigantic book "Baum und Bäume" will get published soon.
I want to thank Kristin Leike for excellent technical assistance in many projects and Michael Lidschreiber for sharing several projects with me, for your untiring help and for thousands of discussions. Without your constant support over several years the establishment of our very powerful ChIP-chip pipeline wouldn't have been possible. It was always a great pleasure for me working with you guys!
I want to thank my collaborators Amelie Schreieck and Wolfgang Mühlbacher for allowing me to take care of their master theses. Both of you guys did a great job and it was a fantastic experience for me seeing you getting better and better each day. I wish you all the best for the future.
I am also very grateful to Claudia Buchen, Stefanie Etzold, Dr. Heidi Feldmann and Kerstin Maier for advice, help, critical comments and for introducing me to countless practical tricks that made my daily hard-core bench work much easier. Many thanks also to Stefan Benkert for providing me with tons of Protease inhibitors and many more things during all the years.
I also want to thank Kathrin Schwinghammer for help and collaboration. It was a great pleasure shearing the Friday lab seminars with you.
Thanks to all the other members of the Cramer lab for help, discussions, collaborations and for the very nice working atmosphere.
I am also greatly thankful to all my other advisors and collaborators including Dr. Martin Heidemann, Prof. Dr. Minkyu Kim, Dr. Philipp Korber, Dr. Dierk Niessing, Matthias Siebert, Dr. Johannes Söding, Dr. Katja Strässer, Dr. Tobias Straub and Dr. Achim Tresch.
I want to thank Natalie Maria Röder for advice, constant support and for many beautiful years.
Many great thanks to my best friends for your support, for your understanding and for reminding me that there is also a very worthwhile life outside of the lab.
Ganz besonders danken möchte ich meiner Familie und hier insbesondere meinem Vater und meinem Onkel Friedrich für die jahrelange Unterstützung. Ohne Euch wäre das alles nicht möglich gewesen.

Summary

Gene transcription begins with the assembly of RNA polymerase (Pol) II and its initiation factors on promoter DNA. Pol II then initiates mRNA synthesis and exchanges initiation factors for elongation factors, which are required for chromatin passage and RNA processing. During transcription elongation nascent RNA is synthesized in a processive manner. Usually, Pol II transcription termination occurs within a region downstream of the polyadenylation (pA) site. In the course of the mRNA transcription cycle the phosphorylation pattern of the C-terminal repeat domain (CTD) of the largest subunit of Pol II changes. The CTD in yeast consists of 26 heptapeptide repeats of the consensus sequence Y1-S2-P3-T4-S5-P6-S7. For selected genes it was shown that the CTD gets phosphorylated on Ser5 residues during initiation and early elongation by site-specific CTD kinases. In the course of transcription elongation Ser2 phosphorylation levels increase, leading to the recruitment of elongation and termination factors. Recently, it was shown that the CTD can also be phosphorylated on Ser7 and Thr4 residues. However, it was not known whether transcription is carried out by a general Pol II transcription complex that functions at all active genes and whether the composition of this complex changes uniformly during the mRNA transcription cycle. Furthermore, it remained to be determined whether the CTD phosphorylation cycle occurs genome-wide, whether the CTD gets also phosphorylated on Tyr1 residues in proliferating yeast and, if so, whether this modification has a functional role.

This work provides strong evidence for a general transcription complex that mediates transcription and mRNA processing at all Pol II genes in proliferating yeast. The genome-wide occupancy profiles for Pol II, its phosphorylated forms, its elongation factors and components of the Pol II initiation and termination machinery were determined by an optimized ChIP-chip protocol. The data reveals that Pol II and its associated transcription factors co-localize at all active genes. The data also shows that the CTD phosphorylation cycle occurs genome-wide, and includes Tyr1-phosphorylation. The results obtained in this work converge on the following model. In the course of a uniform initiation-elongation transition Pol II exchanges initiation factors for elongation factors that is completed 150 nucleotides (nt) downstream of the transcription start site. The resulting elongation complex is composed of all elongation factors and shows high levels of Ser7 and Ser5 phosphorylation on the CTD. In the course of transcription elongation Ser7 and Ser5 phosphorylation decrease and Tyr1 as well as Ser2 phosphorylation levels increase. Whereas Tyr1 phosphorylation starts to decrease ~180 nt upstream of the pA site, Ser2 phosphorylation levels remain high. Low occupancy levels of Tyr1 phosphorylation and high levels of Ser2 phosphorylation triggers the recruitment of termination factors such as Pcf11 and Rtt103 to transcribing Pol II, at a narrow region downstream of the pA site. Tyr1 and Ser2 phosphorylation thus play a role in the uniform two-step elongation-termination transition, where several elongation factors are exchanged for termination factors. Additionally, this work shows that apart from the Pol II CTD, nascent RNA and the C-terminal region of the highly conserved Pol II transcription elongation factor Spt5 contribute to the recruitment of factors at defined regions of genes during the mRNA transcription cycle.

Publications

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

Mayer A*, Heidemann M*, Lidschreiber M, Schreieck A, Sun M, Kremmer E, Eick D, Cramer P. CTD tyrosine phosphorylation impairs termination factor recruitment to transcribing RNA polymerase II. Manuscript in preparation.

* equally contributed: A.M. carried out ChIP-chip experiments; A.M. purified proteins and carried out fluorescence anisotropy experiments; M.H. validated the 3D12 antibody and performed in vitro kinase assays; M.L. analyzed ChIP-chip data; M.S. performed structural modeling; E.K. generated the 3D12 antibody; D.E. and P.C. designed research and supervised the project;

Mayer A*, Schreieck A*, Lidschreiber M, Leike K, Martin DE, Cramer P (2011). The Spt5 C-terminal region recruits yeast 3'-RNA cleavage factor I. Mol Cell Biol, in revision.

* equally contributed: A.M. performed ChIP-chip and RNase-ChIP experiments; A.M., A.S., K.L. performed ChIP experiments; A.S. performed 3'-RACE and readthrough assay; M.L. analyzed ChIP-chip data; P.C. designed research; D.M. and P.C. supervised project;

Blattner C, Jennebach S, Herzog F, **Mayer A**, Cheung AC, Witte G, Lorenzen K, Hopfner KP, Heck AJ, Aebersold R, Cramer P (2011). Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. Genes Dev 25: 2093-2105.

*Müller M, *Heym RG, **Mayer A**, Kramer K, Schmid M, Cramer P, Urlaub H, Jansen RP, Niessing D (2011). A cytoplasmic complex mediates specific mRNA recognition and localization in yeast. PLoS Biol 9: e1000611. * equally contributed

*Miller C, *Schwalb B, *Maier K, Schulz D, Dümcke S, Zacher B, **Mayer A**, Sydow J, Marcinowski L, Dölken L, Martin DE, Tresch A, Cramer P. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol 7: 458. * equally contributed

Clausing E, **Mayer A**, Chanarat S, Müller B, Germann SM, Cramer P, Lisby M, Strässer K (2010). The transcription elongation factor Bur1-Bur2 interacts with replication protein A and maintains genome stability during replication stress. J Biol Chem 285: 41665-41674.

Sun M, Larivière L, Dengl S, **Mayer A**, Cramer P (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: 41597-41603.

Mayer A*, Lidschreiber M*, Siebert M*, Leike K, Söding J, Cramer P (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol 17: 1272-1278.

* equally contributed: A.M established experimental protocols; A.M., M.L., K.L. performed experiments; M.L., M.S. evaluated data sets; J.S. designed and supervised data evaluation; P.C. designed and supervised research;

Röther S*, Burkert C*, Brünger KM*, **Mayer A***, Kieser A, Strässer K (2010). Nucleocytoplasmic shuttling of the La motif-containing protein Sro9 might link its nuclear and cytoplasmic functions. RNA 16: 1393-1401. * equally contributed

Dengl S, **Mayer A**, Sun M, Cramer P (2009). Structure and in vivo requirement of the yeast Spt6 SH2 domain. J Mol Biol 389: 211-225.

Table of contents

Acknowledgements ..... **I**

Summary ..... **II**

Publications ..... **III**

I. General Introduction

1. Gene transcription systems in eukaryotes **1**

2. Gene transcription by RNA polymerase (Pol) II **2**

2.1 The Pol II transcription cycle 2
2.2 Pol II transcription in a chromatin environment 3
2.3 The Pol II C-terminal repeat domain (CTD) 3
2.3.1 Basic aspects 3
2.3.2 The CTD phosphorylation cycle 4

3. Towards systems biology **6**

3.1 A new era of biological research has started 6
3.2 ChIP-chip: a powerful systems-wide approach 6
3.2.1 A short history 6
3.2.2 Key steps of the technique 7
3.2.3 Strengths and limitations of the technique 8
3.2.4 New innovative approaches 9

II. Materials and General methods

1. Materials **12**

1.1 Bacterial and yeast strains 12
1.2 Plasmids and primers 13
1.3 Antibodies 16
1.4 Media and supplements 16
1.5 Buffers and solutions 17

2. General methods **19**

2.1 Chromatin immunoprecipitation with quantitative PCR (ChIP-qPCR) 19
2.1.1 ChIP with TAP-tagged proteins 19
2.1.2 ChIP of Pol II phospho-isoforms 20
2.1.3 Control of average chromatin fragment size 20
2.1.4 Quantitative PCR (qPCR) 21
2.2 Chromatin immunoprecipitation with tiling array hybridization (ChIP-chip) 21
2.2.1 ChIP for ChIP-chip 21

2.2.2 Whole genome amplification....22
2.2.3 DNA labeling, fragmentation and tiling array hybridization....22
2.2.4 Bioinformatics analysis....23
2.2.4.1 ChIP-chip data normalization....23
2.2.4.2 Transcript-wise occupancy profiles....23
2.2.4.3 Gene-averaged occupancy profiles....23
2.2.4.4 Pairwise profile correlations and correlation network....24
2.3 Molecular cloning and preparation of DNA....24
2.4 Preparation of competent cells....26
2.5 Protein expression in *E. coli*....26
2.6 Protein purification....27
2.7 Western Blot analysis....27
2.8 Electrophoretic separation of DNA and proteins....28

III. Uniform transitions of the general RNA polymerase II transcription complex

1. Introduction....29
1.1 Pol II transcription elongation....29
1.1.1 Transcript elongation occurs discontinuously and is highly regulated....29
1.1.2 Transcription elongation through chromatin....30
1.2 Modification of the Pol II CTD during gene transcription....31
1.3 Aims and scope of this work....33

2 Specific procedures....34
2.1 Molecular cloning and phenotyping of epitope-tagged yeast strains....34
2.2 ChIP-chip analysis of the Pol II transcription machinery....35
2.2.1 Genome-wide occupancy profiling of Spt6ΔC and TFIIS....36
2.2.2 ChIP-chip analysis of Pol II and its different phosphorylated forms....36
2.3 Tandem affinity purification of Elf1 and Spn1....37

3 Results....37
3.1 Genome-wide profiling reveals Pol II on a majority of genes....37
3.2 Initiation/termination factors flank the transcribed region....40
3.3 Elongation factors enter during a single 5'-transition....40
3.4 Spn1 and Elf1 interact within a Pol II complex....42
3.5 Elongation factors exit during a two-step 3'-transition....42
3.6 A general Pol II elongation complex for chromatin transcription....43
3.7 CTD phosphorylation profiles depend on TSS location....45
3.8 Recruitment of CTD kinases explains CTD phosphorylations....45
3.9 CTD phosphorylation and factor recruitment....46
3.10 Possible Pol II CTD masking and CTD-independent recruitment....46
3.11 No evidence for promoter-proximally stalled Pol II....47
3.12 General elongation complexes are productive....48
3.13 Unpublished result: Genome-wide occupancy profiling of TFIIS....49

4 Discussion....50
4.1 The initiation-elongation transition of the Pol II transcription cycle 50
4.2 A two-step elongation-termination transition at the 3'-end of Pol II genes 51
4.3 The role of Pol II CTD phosphorylation during 5' and 3' transitions 51
4.4 TFIIS co-localizes with Pol II at promoter and coding regions of genes 51

5 Conclusions and Outlook....52

IV. CTD tyrosine phosphorylation impairs termination factor recruitment to transcribing RNA polymerase II

1. Introduction....55
1.1 Read-out of the Pol II CTD code 55
1.1.1 CTD recognition by Nrd1, Pcf11 and Rtt103 55
1.1.2 CTD recognition by Spt6 56
1.2 Tyrosine-phosphorylation of the mammalian Pol II CTD 56
1.3 Pol II transcription termination 57
1.3.1 Antitermination versus torpedo model 57
1.3.2 Early versus late transcription termination pathway 57
1.4 Aims and scope of this work 58

2. Specific procedures....59
2.1 Generation and characterization of the monoclonal antibody 3D12 59
2.2 ChIP-chip of the Tyr1 Pol II phospho-isoform 60
2.3 Ssu72 depletion in yeast cells 60
2.3.1 Molecular cloning and validation of a Ssu72-degron yeast strain 60
2.3.2 ChIP of Pcf11 and the Tyr1 Pol II phospho-isoform under Ssu72 depletion 61
2.4 Protein sample preparations 61
2.5 Pol II CTD phosphopeptide interaction assay 63
2.6 Pol II CTD phosphatase assay 64

3. Results and Discussion....64
3.1 A monoclonal antibody that recognizes tyrosine phosphorylated Pol II CTD in yeast and mammals 64
3.2 The CTD of genome-associated Pol II is phosphorylated on Tyr1 (Y1P) residues 66
3.3 Y1P is a mark for all active genes in proliferating yeast 67
3.4 Termination factors localize to gene regions with low Y1P levels 69
3.5 Y1P impairs CID-CTD interactions *in vitro* 69
3.6 Molecular basis for the impairment of CID-CTD interaction by Y1P 71
3.7 The Pol II elongation factor Spt6 associates with Y1P 72
3.8 Y1P restricts the recruitment of termination factors to narrow regions at the 5' and 3' end of genes 73
3.9 An extended model of the Pol II CTD code 74
3.10 The CTD phosphatase Ssu72 targets S5P but not Y1P 75

4. Conclusions and Outlook.....77

V. The Spt5 C-terminal region recruits yeast 3'-RNA cleavage factor I

1. Introduction.....79

1.1 The transcription factor Spt5 79
1.1.1 Basic aspects 79
1.1.2 The repetitive C-terminal region (CTR) of Spt5 80
1.2 The cleavage factor (CF) I complex 81
1.3 Aims and scope of this work 81

2. Specific procedures.....82

2.1 Molecular cloning and phenotyping of Spt5 ΔCTR yeast strains 83
2.2 RNase-ChIP assay 83
2.3 Rapid amplification of cDNA 3'-ends (3'-RACE) 84
2.4 Pol II readthrough assay 84
2.5 GST pull-down assay 84
2.6 Quantitative Western blot analysis 85

3. Results.....85

3.1 Investigation of elongation factor recruitment by the Spt5 CTR 85
3.2 Spt5 CTR is required for recruitment of CFI *in vivo* 87
3.3 Spt5 CTR interacts with CFI *in vitro* 89
3.4 RNA contributes to CFI recruitment 89
3.5 CFI co-localizes with the Ser2-phosphorylated CTD downstream of the pA site 90
3.6 CTR deletion does not impair termination 91
3.7 CTR deletion does not alter pA site usage 92

4. Discussion.....93

4.1 Paf1 complex-independent recruitment of CFI 93
4.2 Spt5 CTR is not required for normal pA site usage and transcription termination 93
4.3 A role of Spt5 in coupling transcription and translation 95

5. Conclusions and Outlook.....95

References.....97

Curriculum vitae.....113

I. General Introduction

1. Gene transcription systems in eukaryotes

Gene transcription represents one of the most fundamental processes of life. During gene transcription a nascent RNA molecule is synthesized from a DNA template by DNA dependent RNA polymerases (317). Gene transcription in a eukaryotic cell is carried out by four different DNA dependent RNA polymerases Pol I, Pol II, Pol III and the mitochondrial RNA polymerase (mitoPol) (11, 57).

Pol I, Pol II and Pol III are multi-subunit nuclear enzymes that consist of 14 (589 kDa), 12 (514 kDa) and 17 (693 kDa) subunits, respectively (57). Pol I synthesizes ribosomal RNA (rRNA) and its activity accounts for up to 60% of all nuclear transcription and rRNA accounts for up to 80% of all cellular RNA (141). Pol II produces messenger RNA (mRNA), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) (57), cryptic unstable transcripts (CUTs, (330)), stable unannotated transcripts (SUTs, (330)), Xrn1-sensitive unstable transcripts (XUTs, (307)) and other non-coding RNAs (16). Pol II transcribes up to 85% of the yeast genome (64, 219). Consistently, Pol II occupies large parts of the yeast genome (136, 287). Pol III synthesizes transfer RNAs (tRNAs), 5S rRNA, the spliceosomal U6 snRNA and the signal recognition particle 7SL RNA (323). A recent transcriptome analysis extended the repertoire of transcripts that can be synthesized by Pol III, including several undescribed non-coding RNAs (227).
Until recently it was thought that Pol I, Pol II and Pol III independently transcribe distinct sets of genes. However, a genome-wide study recently revealed that Pol II associates with several Pol III transcribed genes in human cells (256). In addition, factors such as TFIIS that are normally associated with Pol II transcription (45), are also detected at Pol III genes (96). Pol II transcription factors including Fcp1 (84), Spt5 (270) and the Paf1 complex (340) also play a role in Pol I transcription. All these observations argue for a substantial crosstalk between all three nuclear transcription systems.
A typical eukaryotic cell contains a forth transcription system that is located in the mitochondrion (11). Transcription of the mitochondrial genome is carried out by a single-subunit DNA dependent RNA polymerase, the mitoPol, that is distantly related to the RNA polymerase of the bacteriophage T7 (262). The mitoPol produces polycistronic transcripts that are subsequently cleaved to generate individual mRNAs, tRNAs and rRNAs (11). A recent analysis of the human mitochondrial transcriptome revealed unexpected complexity in the expression and processing of mitochondrial RNA, including the identification of several new transcripts (204).
Recently, a fourth and fifth multi-subunit RNA polymerase, Pol IV and Pol V, was discovered in plants. Both enzymes are not essential for cell viability and play a role in small RNA-mediated gene silencing pathways in plants (242).

2. Gene transcription by RNA polymerase (Pol) II

2.1 The Pol II transcription cycle

The Pol II transcription cycle can be divided into different main phases: initiation, elongation, termination and re-initiation (Figure 2) (114, 220). However, the transitions between the different phases are fluent. During initiation Pol II is recruited to the gene promoter and associates with the general transcription initiation factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH to form the preinitiation complex (PIC) (114). A recent quantitative proteomics analysis of the yeast PIC revealed new subunits, including Sub1 and RPA, indicating that the PIC is a more complicated assembly than originally thought (278). At the promoter region Pol II also associates with the Mediator complex. The Mediator complex from yeast is 1.4 MDa multiprotein complex that comprises 25 subunits (33). The Mediator represents a coactivator complex that integrates signals from activators bound upstream of the core promoter into the initiation complex (313).

After opening of the DNA duplex, transcription initiates and the nascent RNA molecule is synthesized (114). The early steps of RNA synthesis do not proceed smoothly. Early transcription up to the +8 position is accompanied by several rounds of abortive initiation, leading to the production of many tiny transcripts (196). Past this region stability in the synthesis of the transcript is obtained and Pol II leaves the promoter region, known as promoter clearance or promoter escape (196).

Pol II enters into the early elongation state. During the last couple of years evidence for early transcription elongation has accumulated to be highly regulated (220). Genome-wide localization analyses of Pol II in higher eukaryotes, primarily in *Drosophila* and humans, revealed peak occupancy levels 20 to 50 nucleotides downstream of the TSS (257). Whether this occupancy peak corresponds to promoter proximally pausing as originally thought, or whether it is due to premature termination is currently under debate (288). With the help of dedicated factors, such as of P-TEFb kinase in higher eukaryotes, Pol II enters into the productive elongation phase where nascent RNA is synthesized by a stable elongation complex in a processive manner (220). Additional aspects of transcription elongation are described in section III 1.1.

At the pA site the transcript is cleaved, but Pol II transcribes past this site (169). At a region downstream of the pA site Pol II falls off the DNA template and transcription terminates. Further aspects of transcription termination are detailed in section IV 1.3 and V 1.2.

After transcription termination a new round of transcription can be started either by initiation or by re-initiation (169). Whereas initiation requires the recruitment of the complete transcription machinery to the gene promoter, re-initiation occurs via a different pathway (336). After the first round of transcription a subset of the transcription machinery, including TFIID, TFIIE, TFIIF and the Mediator complex, remain at the promoter, thus facilitating the following rounds of Pol II transcription (336).

It is currently under debate whether the Pol II transcription cycle occurs on a linear DNA template where the promoter and terminator regions are physically separated, or whether it takes place on a DNA template where the promoter and terminator regions are close to each other via gene looping (115).

2.2 Pol II transcription in a chromatin environment

The simple model of the Pol II transcription cycle outlined above is helpful for getting an idea of how Pol II transcription may works, but it has ignored that transcription occurs in the context of chromatin *in vivo*. Therefore the model needs to be extended.
Pol II transcription takes place on chromatin DNA. The primary structure of chromatin is the 10 nm fiber that appears as "beads on a string" in electron micrographs (228, 258). The "beads" correspond to the nucleosome particles that are connected by linker DNA. The nucleosome consists of an octamer of histone proteins (two H2A-H2B dimers and one H3-H4 tetramer) and 146 base pairs of DNA that is wrapped around the histone octamer (191). Genome-wide studies in various organisms have shown that genomes are usually densely occupied by nucleosomes (138, 175, 335). Nucleosomes limit the access to regulatory DNA sequences, such as gene promoters, and thus can inhibit transcription (258). Before the preinitiation complex can be assembled and Pol II transcription be initiated, promoter DNA has to be made accessible. Gene-specific activators and repressors work in conjunction with protein complexes that remodel and modify chromatin structure in a way that regulatory sequences become either accessible or inaccessible (310). Therefore regulation of chromatin structure is intrinsically tied to the regulation of Pol II transcription initiation.
Furthermore, nucleosomes represent a physical barrier to transcribing Pol II. How Pol II overcomes this barrier is described in section III 1.1.

2.3 The Pol II C-terminal repeat domain (CTD)

2.3.1 Basic aspects

Although Pol I, Pol II and Pol III are similar in structure and subunit configuration, Pol II possesses a unique tail-like CTD that is part of its largest subunit Rpb1 (57). The CTD is flexibly linked to the catalytic core, near the RNA exit pore of the enzyme (58). The CTD seems to be largely flexible, but it shows some tendency to form β-turns (201, 202). In an extended β-turn conformation the yeast CTD would be ~650 Å long and thus could in principle reach any location on the surface of Pol II, which is ~150 Å in diameter (Figure 1) (202).
The CTD consists of heptapeptide repeats of the consensus sequence Y1-S2-P3-T4-S5-P6-S7. However, the sequence can vary considerably from this consensus motif (42, 291). The number of repeats depends on the organism and is 26 and 52 in yeast and humans, respectively (42). In yeast at least eight repeats are required for cell viability (321). The functional unit of the CTD lies within heptapeptide pairs as was shown by site-directed mutagenesis in yeast (289). Mutagenesis approaches also revealed that residues Tyr1, Ser2 and Ser5 are essential for cell viability in yeast, whereas residues Thr4 and Ser7 are not (290, 321).
The Pol II CTD primarily functions as a general platform to recruit proteins involved in transcription, mRNA processing and histone modifications (37). Direct protein-CTD interactions ensure that all of these processes occur cotranscriptionally (237). Post-translational CTD modifications which change during the transcription cycle, allow that the various factors associate with the CTD at the appropriate location.

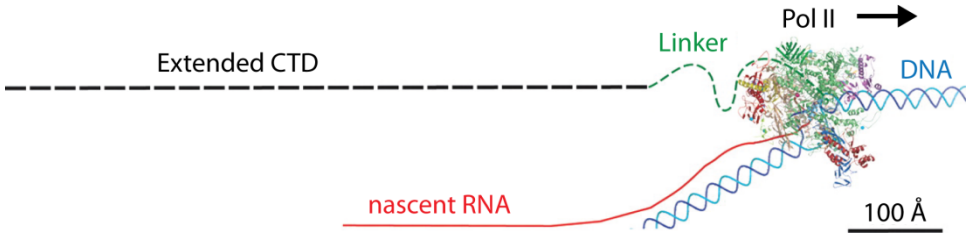

Figure 1: Elongating Pol II and the relative length of the CTD. The complete 12-subunit yeast Pol II is shown as a ribbon diagram. The subunits are indicated in different colors. DNA is in blue and nascent RNA is in red. The extended CTD and the Linker region are illustrated in black and green, respectively. The direction of transcribing Pol II is indicated by an arrow. (modified from (202))

2.3.2 The CTD phosphorylation cycle

For selected genes it was shown that the Pol II CTD undergoes dynamic phosphorylation and dephosphorylation during the transcription cycle (Figure 2) (37). The early studies performed in yeast revealed Ser2 and Ser5 to be the main phosphorylation sites within the CTD heptapeptide repeat (162, 272). These studies also show that (i) Ser5 phosphorylation levels peak at the beginning of the transcribed region, (ii) Ser2 phosphorylation increases during transcription with maximum levels at the 3' end of the transcribed region and (iii) that the CTD is dephosphorylated at the end of gene transcription (162, 272). These observations led to the idea that a CTD phosphorylation cycle exists that is associated with the transcription cycle (Figure 2) (162).

Following the discovery of these CTD modifications, kinases and phosphatases were identified that target Ser5 or Ser2, respectively (161). At that time first proteins were described that preferentially interact with certain CTD phosphorylation patterns. For example, the yeast mRNA capping enzyme was shown to bind directly to Ser5 phosphorylated CTD thus explaining the early recruitment of this enzyme to actively transcribed genes (82). The discovery of several other proteins that directly bind to particular phosphorylated forms of the Pol II CTD (Pol II phospho-isoforms) led to the concept of the "CTD code" (36, 79). According to this concept the different CTD phospho-isoforms that predominate at each stage of the transcription cycle recruit a particular set of transcription, mRNA processing and histone-modifying factors (36, 79).

Recently, Ser7 phosphorylation (157) and more recently Thr4 phosphorylation (128) of the CTD was described in yeast and mammals, and thus extending the CTD code. Whether the Pol II CTD phosphorylation cycle is true on a genome-wide level, whether other types of CTD phosphorylations occur and how Ser7 phosphorylation levels change during the transcription cycle are only some of the questions addressed in this work.

Apart from posphorylation other covalent CTD modifications were described, including glycosylation (147), and more recently methylation (279) and acetylation (Melanie Ott, University of California, San Francisco, personal communication).

Additional details of CTD modifications and of how the CTD code is read are described in section III 1.2 and IV 1.1-1.2.

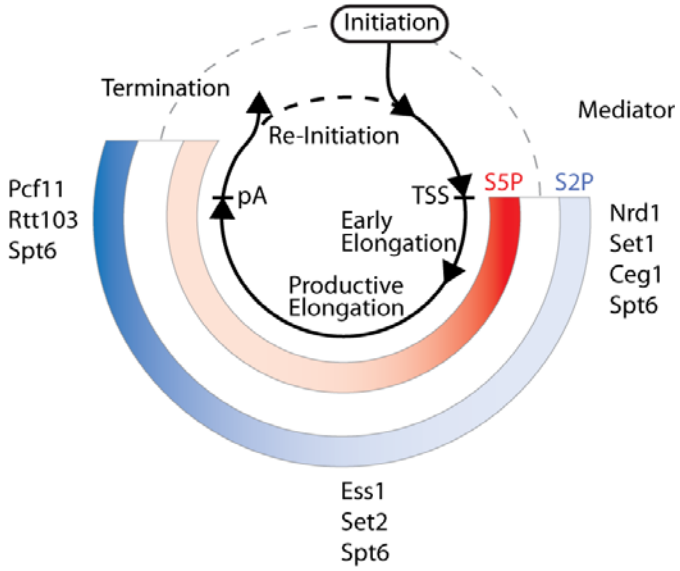

Figure 2: Coordination of the Pol II transcription cycle by CTD phosphorylation. The main stages of the Pol II transcription cycle are indicated by a black circle. The TSS and the pA site are also shown. Regions of the transcription cycle where the Pol II CTD is phosphorylated on Ser5 and Ser2 residues are indicated in red and blue, respectively. The stronger the color saturation, the stronger is the expected CTD phosphorylation level. The region where the Pol II CTD is thought to be unphosphorylated is illustrated as a dashed gray line. A selection of factors that were shown to interact with particular CTD phosphorylation patterns or with unphosphorylated CTD (Mediator complex) is given. Whereas the guanylyltransferase subunit of the mRNA capping enzyme Ceg1, the histone methyltransferase Set1 and the termination factor Nrd1 directly interact with Ser5-phosphorylated CTD, the termination factors Pcf11 and Rtt103 as well as the Pol II elongation factor Spt6 preferentially bind to Ser2-phosphorylated CTD. Please note that although the Ser5-phosphorylation is predominantly observed during initiation and early elongation, low levels also occur throughout elongation and past the pA site. This results in an extended region in which Ser5- and Ser2-phosphorylation overlap. The peptidyprolyl-cis/trans-isomerase Pin1 and the histone methyltransferase Set2 preferentially interact with CTD that is phosphorylated on Ser5 and Ser2 and thus can "read" the doubly phosphorylated CTD.

3 Towards systems biology

3.1 A new era of biological research has started

The availability of DNA sequences of complete genomes as well as technological advances enables researches for the first time to address biological questions at a global and systems level. Until the mid 1990s almost all biological research was focused on the detailed study of individual components, such as single genes and particular proteins (281). This approach changed in the mid 1990s by the development of several large-scale studies that for the first time allowed the analyses of a large number of components systematically and simultaneously (281). The majority of these projects were pioneered in *Saccharomyces cerevisiae* (*S. cerevisiae*) since its genome was one of the first sequenced and due to its facile genetics (172). A selection of early and state-of-the-art systems-wide approaches is given in Table 1 and 2.

Table 1: Systems-wide approaches used in yeast and other organisms. The method, its goal as well as the corresponding reference(s) are given. Please see also Table 2 for complementary genome-wide techniques.

Method and goal	Reference(s)
Gene expression analysis with DNA microarrays (steady state mRNA levels)	(71, 177)
Systematic gene knockouts	(99, 327)
Subcellular protein localization	(129)
Protein profiling by mass spectrometry (protein levels)	(65, 241, 316)
Protein-protein interactions by coupling affinity purification with mass spectrometry	(93, 126, 132, 166, 306, 334)
Genetic interactions	(53, 232, 304, 305)
Mapping of post-translational modifications by mass spectrometry	(47, 85, 95, 106, 173, 250)
Genomic Run-On (GRO) to measure rates of mRNA synthesis	(91)
Transcriptome profiling with DNA microarrays and high-throughput sequencing	(64, 76, 219, 231)
Ribosome profiling	(131)
Dynamic transcriptome analysis (DTA) to measure rates of mRNA synthesis and decay	(205)

3.2 ChIP-chip: a powerful systems-wide approach

3.2.1 A short history

Chromatin immunoprecipitation (ChIP) followed by DNA microarray (chip) hybridization is a very powerful technique to determine DNA binding sites of a protein of interest genome-wide (116). The technique was developed in the laboratories of Patrick O. Brown (Stanford University, Stanford, California), Michael Snyder (Yale University, New Haven; now: Stanford University) and Richard A. Young (Whitehead Institute for Biomedical Research, Cambridge, Massachusetts), and the protocols were published in the years 2000 and 2001 (134, 260). The technique was originally developed for *S. cerevisiae* and was made possible by several breakthroughs (Figure 3).

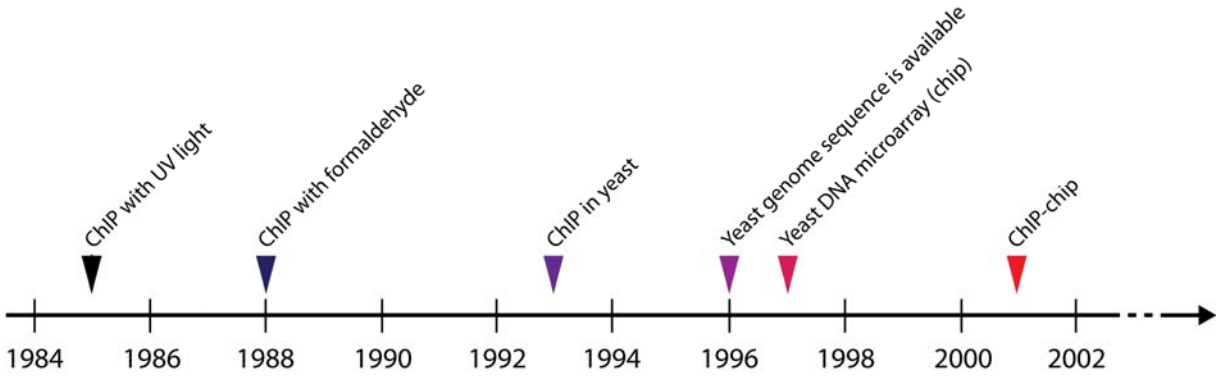

Figure 3: Towards ChIP-chip in yeast. The different milestones on the way to setting up a ChIP-chip protocol as well as a timeline are indicated.

One major breakthrough was the availability of the DNA sequence of the yeast genome in 1996 (51). The DNA sequence information together with new developments in parallel chemical synthesis led to the fabrication of high-density microarrays (64, 87, 269). The first microarray covering a complete eukaryotic genome (*S. cerevisiae*) appeared in 1997 (177). Apart from the developments leading to DNA microarrays, another milestone was the establishment of a reproducible ChIP protocol. The first ChIP protocols appeared in 1984 and 1985 and were used to investigate the DNA binding of bacterial RNA polymerase (*Escherichia coli*) and of RNA polymerase II of *Drosophila melanogaster in vivo*, respectively (101, 102). In the early ChIP protocol, protein-DNA crosslinking was performed with UV light. However, UV light was replaced by formaldehyde as a fixative soon. Formaldehyde is cheap, simple to apply, penetrates biological samples very rapidly, the crosslinks are heat-reversible and in contrast to UV fixation, formaldehyde crosslinking can be used to localize chromatin associated proteins that do not directly bind to DNA (282). In the following years the ChIP protocol was improved and adapted to other model organisms (67). In 1993 the first ChIP study was performed in yeast, investigating the relationship between transcriptional silencing and histone acetylation levels (35). The development of a reproducible ChIP protocol and of microarray technology, together with other experimental and technological advances, such as whole genome amplification methods, finally resulted in ChIP-chip. Although the ChIP-chip protocol was originally applied to investigate protein-DNA interactions in yeast, it is now adapted to many organisms, ranging from bacteria to humans (116).

3.2.2 Key steps of the technique

From an experimental point of view ChIP-chip consists of two main parts, chromatin immunoprecipitation (ChIP) and microarray (chip) hybridization (including the preparatory steps). Whereas the first part of the protocol necessitates classical biochemical approaches that can vary considerably among different laboratories, the second part is rather standardized and includes several ready-to-use kits, depending on the microarray platform.

According to a standard ChIP protocol formaldehyde is used to generate protein-DNA, protein-RNA and protein-protein crosslinks between molecules that are in close proximity *in vivo* (9). The chromatin is prepared and fragmented usually by sonication. The average DNA fragment length is in the range of 200 to 500 base pairs (bp) and is important for the spatial resolution of DNA binding events. Parts of the chromatin sample are kept as "input" sample, the remaining sample is immunoprecipitated by an antibody either directed against the protein of interest or a particular modification (such as acetylation, methylation and phosphorylation) or against a certain epitope-tag ("IP" sample). This leads to a selective enrichment of DNA sequences that directly or indirectly (via other proteins or RNA) crosslink with the protein of interest. After the reversal of crosslinks the amount of a particular DNA sequence in the "IP" and "input" samples are determined by quantitative PCR (qPCR).
A single ChIP experiment does not provide enough DNA for labeling and hybridization to an array and thus has to be amplified. Several protocols are available that allow linear DNA amplification in a standardized manner (226). After amplification the DNA is labeled, for instance by incorporation of a biotinylated nucleotide analog and then loaded onto the array.
Different array platforms are currently available including the single-color array platform of Affymetrix (Santa Clara, USA) and the two-color array systems of Agilent (Palo Alto, USA) and NimbleGen (Madison, USA). Whereas in case of single-color arrays the IP and input sample is hybridized on different arrays, the two-color array systems permit the hybridization of the differently labeled IP and input sample on the same array. However, experiments with two-color arrays necessitate dye-swap replications, in which each hybridization is performed twice, with the dye assignment reversed in the second hybridization in order to reduce systematic color biases (136, 332).
For ChIP-chip analyses high-density tiling arrays are usually used that cover the complete genome. For model organisms with smaller genome sizes such as yeast (12 Mb) the whole genome is covered on a single array. For example, the *S. cerevisiae* Tiling 1.0R Array of Affymetrix comprises over 3.2 million 25-mer probes tiled at an average resolution of 5 bp across the yeast genome. However, for organisms with large genomes such as humans (3,300 Mb) the whole genome is covered by several arrays (seven in case of Human Tiling 2.0R Array, Affymetrix). A new generation of arrays contains isothermal probes, meaning probes that are adjusted in length (45- to 75-mer, NimbleGen) to keep the same melting temperature across the entire probe set. This new development should lead to more uniform results.

3.2.3 Strengths and limitations of the technique

ChIP-chip is a very powerful technique to investigate protein-genome interactions *in vivo*. Standard ChIP-chip protocols are available for various organisms, ranging from bacteria to humans. Protocols are convenient, fast and can be automatized. In addition, ChIP-chip yields highly reproducible and high-resolution data. Until now, the method led to many groundbreaking studies that changed our view of how the expression of genes is regulated. For example, ChIP-chip led to the dissection of transcription regulatory networks (117, 179), revealed the importance of post-initiation gene regulatory mechanisms (196, 217, 339) and also set the ground for a global understanding of the histone code hypothesis (110, 120, 244).

Despite its enormous power ChIP-chip is also associated with some limitations. Firstly, standard ChIP-chip experiments provide no information of how the protein associates with chromatin. Questions like "what other proteins and modifications are required for binding of a protein of interest?" cannot be addressed directly by the standard procedure. Secondly, ChIP-chip cannot distinguish whether a protein of interest binds DNA directly or associates with DNA indirectly via protein-protein or protein-RNA interactions. Thirdly, it does not provide any information about the function of the protein of interest bound to particular genomic regions. Fourthly, the crosslinking efficiency of formaldehyde can vary between different proteins (283). Fifthly, a ChIP-grade antibody is not available against all proteins of interest. The specificity and affinity of the antibody as well as the epitope accessibility are other important issues in this context (78, 116). Sixthly, the final spatial resolution of DNA binding events of a certain protein of interest is limited. The resolution mainly depends on the average chromatin fragment size that is reached by sonication before the IP is performed. Despite extensive optimization trials, DNA average fragment sizes lower than 200 bp could not be realized (83). Seventhly, ChIP-chip consists of several experimental modules that usually include various kits and DNA microarrays, and thus is a quite expensive technique. Finally, ChIP-chip leads to huge data sets whose analyses require advanced bioinformatics expertise and tools.

3.2.4 New innovative approaches

Since the availability of the protocol more than a decade ago, ChIP-chip was the method of choice to study protein-genome interactions for several years. In November 2011 the literature database Pubmed listed 597 publications that contain the term "ChIP-chip" either in their title or abstract, or both. However, the literature search also shows that the numbers of publications that contain results of ChIP-chip studies are currently on the decline (Figure 4). The reason for that trend is that new innovative approaches are gaining ground. An overview of these new powerful techniques is given in Table 2. Among these alternative approaches, ChIP-seq turned out to be very powerful. The technique was developed in the laboratories of Michael Snyder (Yale University, New Haven, now: Stanford University) and of Steven Jones (British Columbia Cancer Agency Genome Science Center, Vancouver), and the protocol was published in 2007 (263). Although the first part of the protocol (ChIP) is very similar to ChIP-chip, the second part is quite different. Instead of hybridizing the amplified immunoprecipitated DNA to microarrays, the DNA is sequenced. This development was made possible by next-generation high-throughput sequencing technologies that allow massively parallel sequencing (324). Comparisons of the data obtained by either ChIP-chip or ChIP-seq have revealed high correlations (125). However, comparisons also have shown that ChIP-seq generally produces occupancy profiles with higher spatial resolution, dynamic range and genomic coverage (125). Another advantage of ChIP-seq is that its applicability is not limited by the availability of an organism-specific microarray and thus can be used to analyze virtually any species. These advantages together with the fact that the costs for sequencing are declining (223) have led to the rapid adaptation of this technique (Figure 4). Since 2010, ChIP-seq led to more publications than ChIP-chip (Figure 4).

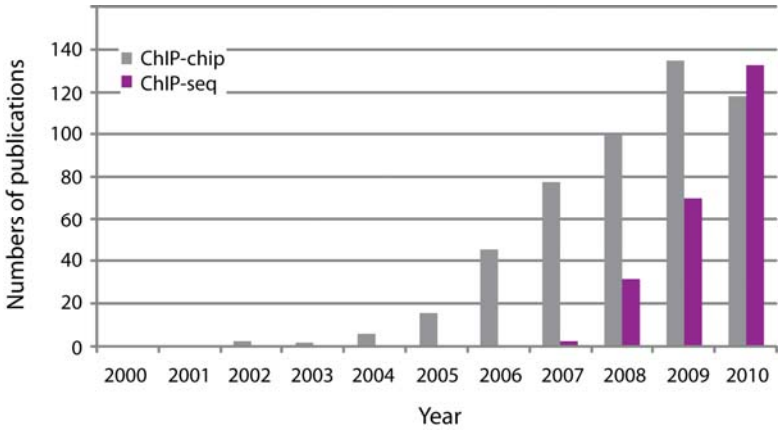

Figure 4: ChIP-seq is gaining ground. The absolute numbers of publications containing the term "ChIP-chip" (gray bars) or "ChIP-seq" (violet bars) either in the title or abstract of publications listed on Pubmed (http://www.ncbi.nlm.nih.gov/pubmed) in November 2011. Although first ChIP-chip studies were already published in 2000 and 2001, the term "ChIP-chip" was neither part of the title nor of the abstract of these publications and thus were not considered in this analysis.

Apart from ChIP-seq other innovative approaches became available during the last years (or will become available soon) which allow the systems-wide analysis of protein-DNA and protein-nascent RNA interactions as well as long-range chromatin interactions (Table 2). A spectacular improvement of the ChIP-seq method, called ChIP-exo, will become available soon. The protocol was developed in the laboratory of Franklin Pugh (Pennsylvania State University, Pennsylvania) and includes exonuclease digestion of DNA that is not directly contacted by the protein of interest (107). This increases the spatial resolution dramatically and also improves the signal-to-noise ratio and thus the sensitivity of the former ChIP-seq approach. This method will allow the mapping of transcription factor binding site at single nucleotide resolution.

Another powerful technique that allows the mapping of protein-nascent RNA interactions at single nucleotide resolution is NET-seq. This approach was developed in the laboratory of Jonathan Weissman (University of California, San Franscisco) and the protocol was published earlier this year (49). NET-seq is based on deep sequencing of 3' ends of nascent transcripts that are associated with the protein of interest. The technique was used to analyze Pol II transcription in yeast and provided strong evidence for pervasive polymerase pausing throughout the body of transcripts (49).

A very innovative approach that allows the analysis of chromatin interactions across an entire genome is Hi-C. This approach was developed in the laboratory of Job Dekker (University of Massachusetts Medical School, Worcester) and published in 2009 (183). This method couples proximity-based ligation and massively parallel sequencing, and was used to investigate the three-dimensional architecture of the human genome (183). Other state-of-the-art genome-wide approaches are listed in Table 2.

Table 2: State-of-the-art genomic tools to analyze protein-DNA, protein-nascent RNA and long-range chromatin interactions. This toolbox can be used to dissect Pol II gene transcription from a genome-wide perspective. The name of the method, its main goal and the reference are given. Please see also Table 1.

Method	Goal	Reference
4C-seq	Genome-wide interactions of one selected locus	(68, 97, 329)
ChIP-exo	Mapping of protein-genome interactions at single-nucleotide resolution	Not yet published (Pugh laboratory)
ChIP-seq	High-resolution mapping of protein-genome interactions	(263)
Competition-ChIP	Measurement of transcription factor binding dynamics	Not yet published (Lieb laboratory)
DamID	Mapping of protein-genome interactions	(66, 229)
DNase-ChIP	Mapping of nucleosome-depleted DNA regions	(59)
DNase-seq	High-resolution mapping of nucleosome-depleted regions	(284, 285)
FAIRE-seq	Alternative to DNase-seq to map open chromatin regions	(92, 104)
GRO-seq	Mapping of transcriptionally engaged Pol II	(56)
Hi-C	Mapping of genome-wide chromatin interactions	(183)
NET-seq	Protein-nascent transcriptome interactions at single-nucleotide resolution	(49)
Nucleosome mapping	Genome-wide mapping of nucleosome binding sites	(27, 138, 180, 335)
PAR-CLIP	High-resolution mapping of protein-transcriptome interactions	(113)
Permanganate genomic footprinting	Mapping of transcriptionally engaged Pol II in *Drosophila* and vertebrate cells	(100)
RIP-seq	Mapping of protein-transcriptome interactions	(342)
Sono-seq	Mapping of open chromatin regions	(12)

II. Materials and General methods

1. Materials

1.1 Bacterial and yeast strains

Table 3: *E. coli* strains

Strain	Description	Source
XL-1 Blue	Rec1A; endA1; gyrA96; thi-1; hsdR17; supE44; relA1; lac[F' proAB lacIqZΔM15Tn10(Tet^r)]	Stratagene
BL21-CodonPlus (DE3)RIL	B; F-; ompT; hasdS(rB, mB); dcm+; Tet^r; gal λ(DE3); endA; The [argU, ileY, leuW, Cam^r]	Stratagene

Table 4: *S. cerevisiae* strains

Strain	Description	Source
BY4741 (wild-type)	MATa; *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	Open Biosystems
Wild-type-pRS316	BY4741; pRS316 [URA3]	This work
S288C	MATα; *SUC2 gal2 mal mel flo1 flo8-1 hap1 ho bio1 bio6*	Euroscarf
Bur1-TAP	BY4741; *BUR1::TAP::HIS3MX6*	Open Biosystems
Cet1-TAP	BY4741; *CET1::TAP::HIS3MX6*	Open Biosystems
Clp1-TAP	BY4741; *CLP1::TAP::HIS3MX6*	Open Biosystems
Ctk1-TAP	BY4741; *CTK1::TAP::HIS3MX6*	Open Biosystems
Elf1-TAP	BY4741; *ELF1::TAP::HIS3MX6*	Open Biosystems
Hrp1-TAP	BY4741; *HRP1::TAP::HIS3MX6*	Open Biosystems
Kin28-TAP	BY4741; *KIN28::TAP::HIS3MX6*	Open Biosystems
Paf1-TAP	BY4741; *PAF1::TAP::HIS3MX6*	Open Biosystems
Pap1-TAP	BY4741; *PAP1::TAP::HIS3MX6*	Open Biosystems
Pcf11-TAP	BY4741; *PCF11::TAP::HIS3MX6*	Open Biosystems
Pcf11-TAP Ssu72-degron	BY4741; *PCF11::TAP::HIS3MX6*; *SSU72promoterΔ::CUP1promoter::degron::kanMX*;	This work
Rna14-TAP	BY4741; *RNA14::TAP::HIS3MX6*	Open Biosystems
Rna15-TAP	BY4741; *RNA15::TAP::HIS3MX6*	Open Biosystems
Rpb3-TAP	BY4741; *RPB3::TAP::HIS3MX6*	Open Biosystems
Spn1-TAP	BY4741; *SPN1::TAP::HIS3MX6*	Open Biosystems
Spn1-TAP Elf1-3HA	BY4741; *SPN1::TAP::HIS3MX6*; *ELF1::3HA::KANMX6*	This work
Spt4-TAP	BY4741; *SPT4::TAP::HIS3MX6*	Open Biosystems
Spt5-TAP	BY4741; *SPT5::TAP::HIS3MX6*	Open Biosystems
Spt6-TAP	BY4741; *SPT6::TAP::HIS3MX6*	Open Biosystems
Spt6ΔC-TAP	FY119 (isogenic to S288C); MATα; *his4-912 lys2-128 leu2-1 ura3-52 trp1-63*	Stefan Dengl (70)
Spt16-TAP	BY4741; *SPT16::TAP::HIS3MX6*	Open Biosystems
TFIIB-TAP	BY4741; *SUA7::TAP::HIS3MX6*	Open Biosystems
TFIIS C-TAP	BY4741; *DST1::TAP::HIS3MX6*	Open Biosystems
TFIIS N-TAP	BY4741; *DST1::TAP*; N-terminal TAP-tag	This work
Tfg1-TAP	BY4741; *TFG1::TAP::HIS3MX6*	Open Biosystems
Spt5 ΔCTR	BY4741; *SPT5Δ931-1063::KANMX6*	This work
Spt5 ΔCTR-pRS316	BY4741; *SPT5Δ931-1063::KANMX6*; pRS316 [URA3]	This work
Bur1-TAP Spt5 ΔCTR	BY4741; *BUR1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Cet1-TAP Spt5 ΔCTR	BY4741; *CET1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Clp1-TAP Spt5 ΔCTR	BY4741; *CLP1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work

Table 4 (continued)

Ctk1-TAP Spt5 ΔCTR	BY4741; *CTK1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Elf1-TAP Spt5 ΔCTR	BY4741; *ELF1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Hrp1-TAP Spt5 ΔCTR	BY4741; *HRP1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Paf1-TAP Spt5 ΔCTR	BY4741; *PAF1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Pap1-TAP Spt5 ΔCTR	BY4741; *PAP1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Pcf11-TAP Spt5 ΔCTR	BY4741; *PCF11::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Rna14-TAP Spt5 ΔCTR	BY4741; *RNA14::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Rna15-TAP Spt5 ΔCTR	BY4741; *RNA15::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Rpb3-TAP Spt5 ΔCTR	BY4741; *RPB3::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Spn1-TAP Spt5 ΔCTR	BY4741; *SPN1::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Spt4-TAP Spt5 ΔCTR	BY4741; *SPT4::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Spt6-TAP Spt5 ΔCTR	BY4741; *SPT6::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work
Spt16-TAP Spt5 ΔCTR	BY4741; *SPT16::TAP::HIS3MX6; SPT5Δ931-1063::KANMX6*	This work

1.2 Plasmids and primers

Table 5: Plasmids

Name	Description	Source
pET15b	T7; His-Tag 3' of MCS; pBR322 origin; Amp^r;	Novagen
pET28b(+)	T7; T7-Tag; His-Tag 5' and 3' of MCS; pBR322 origin; f1 origin; Kan^r;	Novagen
pGEX-3T	GST-Tag; pBR322 origin; Amp^r;	GE Healthcare
pGEX-4T-1	GST-Tag; pBR322 origin; Amp^r;	GE Healthcare
CIp10	*Kl URA3*; ColEI origin; Amp^r;	Heidi Feldmann; (215)
pBS1539	*Kl URA3*; Amp^r; TEV cleavage site;	Euroscarf
pFA6a-3HA-kanMX6	*3HA::KanMX6*; pBR322 origin; Amp^r;	Heidi Feldmann (190)
pKL187	*URA3*; *kanMX*; *CUP1* promoter; ts degron tag;	Euroscarf
pRS316	*URA3*; *CEN6*; *ARH4*; pBR322 origin; f1 origin; Amp^r;	ATCC; (277)

Table 6: Primers used in this work

Name	DNA sequence	Project
109_Degron_Fw	GTGTAAGGAATTCACTAGTTCATAAGCATATATACTTGTTT TAATATATTATTAAGGCGCGCCAGATCTG	Ssu72-degron
103_Degron_Rev	GATGCACAAACTGTGCAAAACTTCAAGTTTGAATTGCGAT GACTAGGCATGGCACCCGCTCCAGCGCCTG	Ssu72-degron
110_ControlA_Fw	ACTAGAGGGAGACTACGTAG	Ssu72-degron
104_ControlB_Rev	TTCCAATATAGCTTTGCTAC	Ssu72-degron
111_ControlC_Fw	CTGGTGCAGGCGCTGGAGCG	Ssu72-degron
105_ControlD_Rev	CGCTCCAGCGCCTGCACCAG	Ssu72-degron
OligodT-anchor	GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT	3'RACE
Upstream primer 1	TATGAATTCTTCTGTTTTGGGTTTGGA	3'RACE
Anchor primer 2	GACTCGAGTCGACATCGA	3'RACE
Sequencing primer	TAATACGACTCACTATAGGG	For inserts,

Table 6 (continued)

1 (forward)		pET28b(+)
Sequencing primer 2 (reverse)	GGGTTATGCTAGTTATTGC	For inserts pET28b(+)
RTh1	CTATTATTGATGCTTTGAAGACCTCCAG	Readthrough assay
RTh2	TGCCCAAAATAATAGACATACCCCATAA	Readthrough assay
RTh3	CAAGAAAGAAAAAGTACCATCCAGAG	Readthrough assay
RTh4	TTCATGTAAGTGTGTATCTTGAGTGTC	Readthrough assay
RTh5	CGTTCATGTAAGTGTGTATCTTGAGTG	Readthrough assay
RTh6	TACGTTCATGTAAGTGTGTATCTTGAG	Readthrough assay
66_Elf1-3HA_Fw	GGCCAAGTTAAAAGAGGCAGAGGCGCCTTGGTAGATAGTG ACGATGAATACCCATACGATGTTCCTGAC	Spn1-TAP Elf1-3HA
68_Elf1-3HA_Rev	AAATATATATGACCTAAGTAAATATGGTTTTTTCTCAGGAC CGGATTATGGATGGCGGCGTTAGTATCG	Spn1-TAP Elf1-3HA
68_Elf1-3HA_control_Fw	AGTGACACAGATGATGGTGAC	Spn1-TAP Elf1-3HA
66_Elf1-3HA_control_Rev	ATTCGATACTGATGACGATG	Spn1-TAP Elf1-3HA
DST1_URA_Fw	TCAAGCAGCAGAACATTCACAGTGTAGTCAGTCCGCATAA GAGCATTCATCATGGGCCGACTTGGCCAAGCCTAG	TFIIS N-TAP
DST1_URA_Rev	CATTACTTTTGTTCTTTTCTAGATTCTTAACATGTACCAGTA CTTCCTTACTATCTCTAGAAGGACCACCTTTGATTG	TFIIS N-TAP
DST1_TAP_Fw	TCAAGCAGCAGAACATTCACAGTGTAGTCAGTCCGCATAA GAGCATTCATCATGGGTCGACGGATCCCCGGGTT	TFIIS N-TAP
DST1_TAP_Rev	CATTACTTTTGTTCTTTTCTAGATTCTTAACATGTACCAGTA CTTCCTTACTATCCTGATGATTCGCGTCTACTTTCGGCG	TFIIS N-TAP
DST1_cont_Fw	TAGCCTTTCTTGTATATCCCTC	TFIIS N-TAP
DST1_cont_Rev	ATTCACGGACTCACCTACAG	TFIIS N-TAP
Spt5ΔCTR_Fw	AGCTGTAAATGCGCATGGAGGCTCAGGTGGTGGCGGTGTC TAACGGATCCCCGGGTTAATTAAG	Spt5 ΔCTR
Spt5ΔCTR_Rev	TTGATTTCTTCTTGGGTGATATTGGTTCTCCTTTTGGTGACG CATAGGCCACTAGTGGATC	Spt5 ΔCTR
ΔCTR_cont1_Fw	ACAACCGTGAGGGAGGTGAAG	Spt5 ΔCTR
ΔCTR_cont2_Fw	GTGAGGGAGGTGAAGGTA	Spt5 ΔCTR
ΔCTR_cont3_Fw	ATCATGCGTCAATCGTATGT	Spt5 ΔCTR
ΔCTR_cont1_Rev	TGCTCGCAGGTCTGCAGCGAG	Spt5 ΔCTR
ΔCTR_cont2_Rev	AATTCAACGCGTCTGTGA	Spt5 ΔCTR
ΔCTR_cont3_Rev	GGAAAGTGGCAGAAAGAAAG	Spt5 ΔCTR
TFIIB_Fw	TGAAGGGGAAATCAATGGAG	Strain Validation
Openbio_Rev	AACCCGGGGATCCGTCGACC	Strain Validation
Cet1_Fw	CCCAGACAAGAGATTCCGTC	Strain Validation
Spt4_Fw	TACCAGCTGAGGTTGTGGAG	Strain Validation
Paf1_Fw	CAGATGCTGTTCATACTG	Strain Validation
Spt6_Fw	TATCCAAGTGTTATCCAG	Strain Validation
Nrd1_Fw	AAGACATGAGGCCGAAAATG	Strain Validation
Pap1_Fw	ATACAAGGGGCAGTGACGAG	Strain Validation
Clp1_Fw	AGGATTCGCTTTGATTACAG	Strain Validation
Hrp1_Fw	TACAATCACCGTAGTGGTGG	Strain Validation
Rpb3_Fw	CTCATGGGCCGTAATATTGG	Strain Validation
Kin28_Fw	TGCTTATGACTTTGAGGGGC	Strain Validation
Tfg1_Fw	ATCGCCCGTTAAAAAGGAAG	Strain Validation
Bur1_Fw	GCTAAAACTGCGATTCCACC	Strain Validation
Spn1_Fw	TTACCTGATGGGTCTCTGCC	Strain Validation
Pcf11_Fw	AGCATTTGGACTGGCATTTC	Strain Validation
Spt16_Fw	AGAGATTTGGGCTTCCAAGG	Strain Validation
Rtt103_Fw	AACTGCAAAGCACACTGGAC	Strain Validation

Table 6 (continued)

Rna14_Fw	AGGATTCGCAGATCAAAACC	Strain Validation
Rna15_Fw	ATCGGGAGTTTCACAACAGC	Strain Validation
Ctk1_Fw	TTGTGGCTGCCTCCTGGTGG	Strain Validation
Elf1_Fw	TGTGGGCAGTCGTTCCAAAC	Strain Validation

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

Table 7: qPCR primers used in this work

Name	DNA sequence
ADH1_5'_Fw	TTTCCTTCCTTCATTCACGCACA
ADH1_5'_Rev	TCAAGTAACTGGAAGGAAGGCCGTA
ADH1_TSS_Fw	TCCTTGTTTCTTTTTCTGCAC
ADH1_TSS_Rev	GAGATAGTTGATTGTATGCTTGG
ADH1_ORF_Fw	AGCCGCTCACATTCCTCAAG
ADH1_ORF_Rev	ACGGTGATACCAGCACACAAGA
ADH1_pA_Fw	AAAACGAAAATTCTTATTCTTGA
ADH1_pA_Rev	TACCTGAGAAAGCAACCTGA
ADH1_3'_Fw	CCTGTAGGTCAGGTTGCTTT
ADH1_3'_Rev	CGGTAGAGGTGTGGTCAA
ACT1_5'_Fw	AAACCAAACTCGCCTCTCT
ACT1_5'_Rev	GGAAGGAAAGGATCAAACAA
ACT1_ORF_Fw	TCAGAGCCCCAGAAGCTTTG
ACT1_ORF_Rev	TTGGTCAATACCGGCAGATTC
ACT1_3'_Fw	TTTATCCATTGGACCGTGTA
ACT1_3'_Rev	GGGCAATTGCATAAACCTAT
ILV5_Prom_Fw	ACCCAGTATTTTCCCTTTCC
ILV5_Prom_Rev	TTGTCTATATGTTTTTGTCTTGC
ILV5_ORF_Fw	CTATCAAGCCATTGTTGACC
ILV5_ORF_Rev	CTTGAAGACTGGGGAGAAAC
ILV5_pA_Fw	CCGAAACGCGAATAAATAAT
ILV5_pA_Rev	GTCCCGATGAGGACTTATACA
PDC1_Prom_Fw	TGCCCCTTTTTCTGTTAGAC
PDC1_Prom_Rev	AATAAGGTGGTGTTGAACGA
PDC1_ORF_Fw	CAAGACCAAGAACATTGTCG
PDC1_ORF_Rev	AAAGTGGCGTTTCTGATCTT
PDC1_pA_Fw	TACCATGGAAAGACCAGACA
PDC1_pA_Rev	CCCAGACTTAAGCCTAACCA
PMA1_5'_Fw	TGACTGATACATCATCCTCTT
PMA1_5'_Rev	TTGGCTGATGAGCTGAAACAGAA
PMA1_ORF_Fw	AAATCTTGGGTGTTATGCCATGT
PMA1_ORF_Rev	CCAAGTGTCTAGCTTCGCTAACAG
PMA1_3'_Fw	GGTTTCTCTGGATGGTACTTT
PMA1_3'_Rev	TGACTTGTGTGCGTTTCATA
TEF1_Prom_Fw	ACCACTTCAAAACACCCAAG
TEF1_Prom_Rev	ACGACACCCTAGAGGAAGAA
TEF1_ORF_Fw	TTGATTATTGCTGGTGGTGT
TEF1_ORF_Rev	TGTTCTCTGGTTTGACCATC
TEF1_pA_Fw	ATTTATCCCAGTCCGATTCA
TEF1_pA_Rev	CTGATGTGATTTCGACCATT
Control(YER)_Fw	TGCGTACAAAAAGTGTCAAGAGATT
Control(YER)_Rev	ATGCGCAAGAAGGTGCCTAT

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

1.3 Antibodies

Table 8: Monoclonal antibodies used in ChIP and ChIP-chip experiments

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

Table 9: Antibodies used in Western blot experiments

Antibody	Stock solution	Host	Source
3F10 (αHA-tag)	1:700	Rat	Roche (11867423001)
8WG16 (αPol II CTD)	1:2000	Mouse	Santa Cruz (sc-56767)
α-Tubulin	1:1000	Rat	Santa Cruz (sc-69971)
α-GST-tag	1:5000	Goat	GE Healthcare (RPN1236)
yN-20 (αSpt5)	1:400	Goat	Santa Cruz (sc-26355)
PAP (Peroxidase-Anti-Peroxidase)	1:2000	Rabbit	SIGMA (P1291)
α-Goat-HRP	1:3000	Donkey	Santa Cruz (sc-2020)
α-Mouse-HRP	1:3000	Goat	Bio-Rad (170-6516)
α-Rat-HRP	1:3000	Goat	SIGMA (A9037)

1.4 Media and supplements

Table 10: Growth media

Name	Description	Application
LB	1% (w/v) tryptone; 0,5% (w/v) yeast extract; 0,5% (w/v) NaCl (+ 1,5% (w/v) agar for selective media plates);	*E. coli* culture
YPD	2% (w/v) petone; 2% (w/v) glucose; 1,5% (w/v) yeast extract (+ 1,8% (w/v) agar for solid media plates);	Yeast culture
YPDCu	YPD with 0.16 mM $CuSO_4$;	Yeast culture
5-FOA plates	SC (-ura); 0.01% (w/v) uracil; 0.2% (w/v) 5-FOA; 2% (w/v) agar;	Yeast culture
SC -ura	0.69% (w/v) yeast nitrogen base; 0.077% (w/v) drop-out -ura; 2% (w/v) glucose; (+ 2% (w/v) agar for solid media plates)	Yeast culture

Table 11: Growth media additives

Supplement	Description	Working concentration
Ampicillin	Antibiotic	100 µg/ml
Chloramphenicol (in 100% EtOH)	Antibiotic	50 µg/ml
Kanamycin	Antibiotic	30 µg/ml
Tetracycline (in 70% v/v EtOH)	Antibiotic	12.5 µg/ml
IPTG	Isopropyl-β-D-thiogalactopyranosid	0.5 mM
Geneticin (G418)	Antibiotic	400 µg/ml
6-Azauracil (6-AU)	Pyrimidine analog	50-200 µg/ml
Mycophenolic acid (MPA)	Immunosuppressive drug	15-45 g/ml

1.5 Buffers and solutions

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

Table 12: General buffers, dyes and solutions

Name	Description	Application
1x PBS	2 mM KH_2PO_4; 4 mM Na_2HPO_4; 140 mM NaCl; 3 mM KCl; pH 7.4 (25°C);	Protein purification
1x SDS-PAGE loading Buffer		SDS-PAGE
2x SDS-PAGE loading Buffer	25 mM Tris-HCl, pH 7.0 at 25°C; 0.05% (w/v) Bromophenol blue; 0.5% (v/v) β-Mercaptoethanol; 7% (w/v) Dithiothreitol; 0.05% (w/v) Lauryl sulfate; 5% (v/v) Glycerol;	SDS-PAGE
6x Loading dye (Fermentas)	1.5 g/L Bromphenol blue; 1.5 g/L Xylene cyanol; 50% (v/v) Glycerol;	Agarose gels
4x Stacking gel Buffer	0.5 M Tris-HCl; 0.4% (w/v) SDS; pH 6.8 at 25°C;	SDS-PAGE
4x Separation gel Buffer	3 M Tris-HCl; 0.4% (w/v) SDS; pH 8.9 at 25°C;	SDS-PAGE
SDS electrophoresis Buffer	25 mM Tris-HCl; 0.1% (w/v) SDS; 250 mM Glycine;	SDS-PAGE
Gel staining solution	50% (v/v) Ethanol; 7% (v/v) Acetic acid; 0.125% (w/v) Coomassie Brilliant Blue R-250;	Coomassie staining
Gel destaining solution	5% (v/v) Ethanol; 7.5% (v/v) Acetic acid;	Coomassie staining
1x TBE	8.9 mM Tris-HCl; 8.9 mM Boric acid; 2 mM EDTA; pH 8.0 at 25°C;	Agarose gels
TFB-1 Buffer	30 mM KOAc; 50 mM $MnCl_2$; 100 mM RbCl; 10 mM $CaCl_2$; 15% (v/v) Glycerol; pH 5.8 at 25°C;	Chemically competent *E. coli* cells
TFB-2 Buffer	10 mM MOPS, pH 7.0 at 25°C; 10 mM RbCl; 75 mM $CaCl_2$; 15% (v/v) Glycerol;	Chemically competent *E. coli* cells
TELit	10 mM Tris-HCl, pH 8.0 at 25°C; 155 mM LiOAc; 1 mM EDTA, pH 8.0;	Chemically competent *S. cerevisiae* cells
LitSorb	10 mM Tris-HCl, pH 8.0 at 25°C; 155 mM LiOAc; 1 mM EDTA, pH 8.0; 18.2% (w/v) D-Sorbitol	Chemically competent *S. cerevisiae* cells
LitPEG	10 mM Tris-HCl, pH 8.0 at 25°C; 155 mM LiOAc; 1 mM EDTA, pH 8.0; 40% (w/v) PEG 3350;	Chemically competent *S. cerevisiae* cells
Lyticase Buffer	1 M Sorbitol; 100 mM EDTA, pH 8.0; 14.3 mM β-Mercaptoethanol;	Preparation of genomic DNA
Spheroblast wash Buffer	1 M Sorbitol; 100 mM EDTA, pH 8.0;	Preparation of genomic DNA
TE 50/100 Buffer	50 mM Tris-HCl, pH 7.5 at 4°C; 100 mM EDTA, pH 8.0;	Preparation of genomic DNA
WB transfer Buffer	25 mM Tris; 192 mM Glycine; 20% (v/v) Ethanol;	Western Blotting
WB blocking Buffer	2% (w/v) milk powder in 1x PBS;	Western Blotting
Ab lysis Buffer	50 mM Tris-HCl, pH 8.0 at 4°C; 150 mM NaCl; 1% (w/v) NP-40; PI;	Antibody validation
Ab lysis Buffer 2	50 mM HEPES-KOH, pH 7.5 at 4°C; 150 mM NaCl; 1 mM EDTA; 1% (w/v) Triton-X-100; 0.1% (w/v) sodium-deoxycholate); PI; PhoI;	Antibody validation
Laemmli Buffer	60 mM Tris-HCl, pH 6.8 at 25°C; 2% (w/v) SDS; 10% (w/v) Glycerol; 0.01% (w/v) Bromphenol blue; 10 mM EDTA; 1 mM PMSF; 100 mM DTT;	Antibody validation
5x Tyrosine kinase Buffer	300 mM HEPES, pH 7.5 at 25°C; 15 mM $MgCl_2$; 15 mM $MnCl_2$; 6 mM DTT; 15 mM Na_3VO_4; 12.5 µM PEG 20,000;	Antibody validation
5x Serine kinase Buffer	250 mM HEPES, pH 7.9 at 25°C; 50 mM $MgCl_2$; 500 mM KCl; 5 mM DTT; 1 mM EGTA; 500 µM EDTA;	Antibody validation

Table 13: Nrd1-, Pcf11-, and Rtt103-CID purification buffers

Name	Description
Buffer A	50 mM Tris-HCl, pH 8.0 at 4°C; 300 mM NaCl; 10 mM β-Mercaptoethanol; PI;
Buffer B	50 mM Tris-HCl, pH 8.0 at 4°C; 150 mM NaCl, 5 mM DTT;
Buffer C	50 mM MES, pH 6.5 at 4°C; 50 mM NaCl, 1 mM DTT;
Buffer D	50 mM HEPES, pH 7.5 at 4°C; 100 mM NaCl; 1 mM DTT;
Buffer E	50 mM Tris-HCl, pH 8.0 at 4°C; 50 mM NaCl; 10 mM Imidazole; 10 mM β-Mercaptoethanol;
Buffer F	50 mM Tris-HCl, pH 8.0 at 4°C; 30 mM NaCl; 10 mM Imidazole; 1 mM DTT;
Buffer G	50 mM Tris-HCl, pH 8.0 at 4°C; 30 mM NaCl; 1 mM DTT;
Buffer H	50 mM HEPES, pH 7.0 at 4°C; 1 100 mM NaCl; 5 mM DTT;
Buffer I	50 mM Tris-HCl, pH 8.0 at 4°C; 300 mM NaCl; 10 mM β-Mercaptoethanol; 10% (v/v) Glycerol; PI;
Buffer J	50 mM Tris-HCl, pH 8.0 at 4°C; 50 mM NaCl; 1 mM EDTA; 1 mM DTT;
Buffer K	50 mM Tris-HCl, pH 8.0 at 4°C; 200 mM NaCl; 1 mM EDTA; 1 mM DTT;
Buffer L	50 mM Tris-HCl, pH 8.0 at 4°C; 500 mM NaCl; 5 mM DTT; PI;
Buffer M	50 mM Tris-HCl, pH 8.0 at 4°C; 300 mM NaCl; 5 mM DTT;
Buffer N	50 mM Tris-HCl, pH 8.0 at 4°C; 150 mM NaCl; 2 mM DTT;
Buffer O	20 mM HEPES, pH 8.0 at 4°C; 25 mM KCl; 2 mM DTT;

Table 14: Buffers for Fluorescence anisotropy measurements, Malachite green assay, GST pull-down assay and for Tandem Affinity Purification (TAP)

Name	Description
FluA Buffer	20 mM HEPES, pH 8.0 at 20°C; 150 mM NaCl; 5 mM DTT;
FluB Buffer	FluA, but with 10 mM NaCl;
Mal Buffer	50 mM Tris-HCl, pH 6.5 at 25°C; 200 mM NaCl; 1 mM EDTA; 1 mM DTT;
GST lysis Buffer	50 mM Tris-HCl, pH 8.0 at 4°C; 150 mM NaCl; 1 mM DTT; PI; PhoI;
TAP lysis Buffer	50 mM Tris-HCl, pH 7.5 at 4°C; 1.5 mM $MgCl_2$; 100 mM NaCl; 0.15% NP-40; 1 mM DTT; PI;
TAP wash Buffer	Identical to TAP lysis Buffer, but with 0.5 mM DTT and without PI;
Calmodulin Buffer 1	50 mM Tris-HCl, pH 7.5 at 4°C; 4 mM $CaCl_2$; 1.5 mM $MgCl_2$; 100 mM NaCl; 0.15% NP-40; 1 mM DTT;
Calmodulin Buffer 2	50 mM Tris-HCl, pH 7.5 at 4°C; 2 mM $CaCl_2$; 1.5 mM $MgCl_2$; 100 mM NaCl; 0.15% NP-40; 1 mM DTT;
TAP elution Buffer	10 mM Tris-HCl, pH 8.0 at 20°C; 5 mM EGTA;

Table 15: Buffers used in ChIP, RNase ChIP and ChIP-chip experiments.

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

*Phosphatase inhibitors (PhoI) were only applied in ChIP and ChIP-chip experiments of Pol II phospho-isoforms and in Spt5 CTR-GST pull-down assays;

Table 16: Protease and phosphatase inhibitor mixes.

Name	Description
Protease-inhibitor mix (PI)	1 mM Leupetin, 2 mM Pepstatin A, 100 mM Phenylmethylsulfonyl fluoride, 280 mM Benzamidine;
Phosphatase-inhibitor mix (PhoI)	1 mM NaN_3, 1 mM NaF, 0.4 mM Na_3VO_4;

2. General methods

2.1 Chromatin immunoprecipitation with quantitative PCR (ChIP-qPCR)

2.1.1 ChIP with TAP-tagged proteins

TAP-tagged yeast strains were usually obtained from Open Biosystems (Huntsville, USA; Yeast TAP-Tagged Collection) and were isogenic to BY4741 wild-type strain. All TAP-strains used in this work are listed in Table 4. Before these yeast strains were applied to ChIP analysis, the strains were validated. Firstly, gene-specific PCR was performed to confirm that the DNA coding for the TAP-tag was at the correct genomic position. The DNA sequences of all control primers are given in Table 6. Secondly, Western blotting with anti-TAP (PAP, Sigma) antibody was performed to verify whether the tagged protein of interest was properly expressed. Western blot experiments were conducted as described in in section II 2.7. Thirdly, the growth of the various tagged yeast strains compared to non-tagged wild-type strain was monitored to rule out any influence of the epitope tag on yeast growth. This was done by serial dilutions of the various yeast strains on YPD plates at 30°C for two days. Only yeast strains which passed all quality controls were used for ChIP analysis.

For ChIP-qPCR experiments yeast cultures were grown in 40 ml YPD medium at 30°C to mid-log phase (OD_{600} ~ 0.8), treated with 1% formaldehyde for 20 min at 20°C, and cross-linking was quenched with 5 ml 3 M glycine for 10 min. Subsequent steps were performed at 4°C with pre-cooled buffers (Table 15) containing protease inhibitors (PIs; Table 16). Cells were collected by centrifugation, washed twice with 1x TBS and twice with FA lysis Buffer. Cell pellets were flash-frozen in liquid nitrogen and stored at -80°C. Pellets were thawed, resuspended in 1 ml FA lysis Buffer, and disrupted by bead beating (Retsch) in the presence of 1 ml silica-zirconia beads for 30 min at 4°C. Lysis efficiency was typically >80% as determined by spectrophotometer measurements (BioPhotometer, Eppendorf). Chromatin was solubilized and fragmented via sonication with a BioruptorTM UCD-200 (Diagenode Inc.). Sonication was performed at intensity setting “high” (cycles of 0.5 min on and 0.5 min off) for 35 min with sample cooling on ice after 5, 15 and 25 min. 30 µl and 100 µl of fragmented chromatin samples were saved as input and for control of the average chromatin fragment size (described in section II 2.1.3), respectively. 700 µl of sample (IP sample) was immunoprecipitated with 20 µl IgG Sepharose beads at 4°C for 1 h. The IgG beads were directed against the Protein A content of the C-terminal TAP-tag. Immunoprecipitated chromatin was washed three times with FA lysis Buffer, twice with FA lysis Buffer 2, twice

with ChIP wash Buffer and once with TE Buffer. Immunoprecipitated chromatin was eluted for 10 min at 65 °C with ChIP elution Buffer. Eluted chromatin was digested with Proteinase K at 37°C for 2 h and the reversal of crosslinks was performed at 65°C over-night. DNA was purified with the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions. Thereby, DNA was eluted with 50 µl ddH_2O and was further analyzed by quantitative PCR.
All buffers and reagents used for ChIP are listed in Table 15 and 16, respectively. The RNase-ChIP assay that allows the analysis of RNA-dependent binding of the protein of interest is described in section V 2.2.

2.1.2 ChIP of Pol II phospho-isoforms

ChIP analysis of the various Pol II phospho-isoforms was performed as described in the previous section, but with the following modifications. Firstly, ChIP experiments were conducted in the presence of phosphatase inhibitors (PhoIs; Table 16). Secondly, for chromatin immunoprecipitation a set of monoclonal antibodies with strong specificity and affinity for particular phosphorylated states of the Pol II CTD were applied (Table 8). These antibodies were generated and validated in the laboratories of Dirk Eick and Elisabeth Kremmer (Helmholtz Zentrum München). The amount of antibody that was used for immunoprecipitation of chromatin was optimized as is described in (199).
700 µl of chromatin sample was immunoprecipitated with the optimized amount of the respective monoclonal antibody (see Table 8) at 4°C over-night on a rotating wheel. 25 µl of Protein A and Protein G Sepharose beads (GE Healthcare) were added and incubated at 4°C for 1.5 h on a rotating wheel. Immunoprecipitated chromatin was treated as described in the previous section.

2.1.3 Control of average chromatin fragment size

100 µl of chromatin solution was used to determine the average DNA fragment size that was reached by sonication. 92 µl of TE Buffer and 8 µl of Proteinase K (20 mg/ml) was added to the chromatin sample and incubated at 37°C for 2 h and at 65°C over-night. Next, 20 µl of LiCl (4 M), 1 µl Glycogen and 120 µl Phenol was added, mixed and centrifuged at 13,000 rpm for 10 min (20°C). The supernatant was mixed with 400 µl of pre-cooled Ethanol and incubated at -20°C for 5 h. The immunoprecipitated DNA (and RNA) was pelletized by centrifugation at 13,000 rpm for 20 min (4°C). The pellet was resuspended in 20 µl TE Buffer and RNA was removed by RNAse treatment. 10 µl of RNase A/T1 mix (2 mg/ml RNase A, 5000 U/ml RNase T1) were added and incubated at 37°C for 1 h. The resulting DNA sample was electrophoretically separated on a 1.5% agarose gel (see section II 2.8). The average DNA fragment size was usually ~250 nucleotides (nt).

2.1.4 Quantitative PCR (qPCR)

For ChIP experiments, input and immunoprecipitated (IP) samples were assayed by qPCR to assess the extent of protein occupancy at different genomic regions.
However, before ChIP DNA was analyzed by qPCR, primers were designed and the *PCR efficiencies* of the corresponding primer pairs were determined. DNA primers used for qPCR experiments were 18 to 24 nt long and were designed with the OligoPerfect^{TM} Designer (Invitrogen). The length of the amplified qPCR product was between 60 and 70 nt. The PCR efficiency was determined by quantitative PCR (as described below) with the same primer pair applied to at least four different dilutions of a DNA template (usually DNA that was fragmented by sonication). Based on the resulting standard curve, the PCR efficiency was calculated with the Bio-Rad CFX Manager software version 1.1 according to the manufacturer's instructions. Only primer pairs with a PCR efficiency of ≥ 90% were used in ChIP-qPCR experiments. A list of all qPCR primers used in this work is given in Table 7.
PCR reactions for the analysis of ChIP DNA as well as for validation of qPCR primer pairs contained 1 µl DNA template, 2 µl of 10 µM primer pairs and 12.5 µl iTaq SYBR Green Supermix (Bio-Rad). Quantitative PCR was performed on a Bio-Rad CFX96 Real-Time System (Bio-Rad Laboratories) using a 3 min denaturing step at 95°C, followed by 49 cycles of 30 sec at 95°C, 30 sec at 61°C and 15 sec at 72°C. Threshold cycle (Ct) values were determined by application of the corresponding Bio-Rad CFX Manager software version 1.1 using the Ct determination mode "Regression". The *IP efficiency* (**not** PCR efficiency) and *fold enrichment* of any given region over control regions, such as an open reading frame (ORF)-free heterochromatic region on chromosome V, was determined as described in (83).

2.2 Chromatin immunoprecipitation with tiling array hybridization (ChIP-chip)

2.2.1 ChIP for ChIP-chip

For ChIP-chip experiments the standard ChIP protocol as described in sections II 2.1.1-2.1.2 had to be adapted as follows. Firstly, yeast cultures were grown in 600 ml YPD medium at 30°C to mid-log phase (OD_{600} ~ 0.8). Secondly, cell lysis via bead beating was performed for 2 h, with cooling of the sample after 30, 60 and 90 min. Thirdly, the chromatin pellet was washed two times with FA lysis Buffer before sonication. To be more precise, the cell lysate was centrifuged at 13,200 rpm for 15 min (4°C), the supernatant was discarded and the chromatin pellet was resuspened with 1 ml FA lysis Buffer. This step was repeated one time. Fourthly, the elution from the IgG Sepharose beads was performed at 65°C for 60 min. Fifthly, after purification of DNA with the QIAquick PCR Purification Kit (Qiagen), DNA was eluted with 100 µl H_2O and 5 µl RNase A (10 mg/ml) was added, and incubated at 37°C for 20 min. DNA was purified with the QIAquick PCR Purification Kit, eluted with 50 µl H_2O, concentrated via vacuum centrifugation, and amplified as described in the following section.

2.2.2 Whole genome amplification

DNA samples were amplified and re-amplified with GenomePlex© Complete Whole Genome Amplification 2 (WGA2) Kit using the Farnham Lab WGA Protocol for ChIP-chip (226) (http://www.genomecenter.ucdavis.edu/farnham/protocol.html). Briefly, 10 µl of concentrated ChIP DNA was used to generate the PCR-amplifiable OmniPlex© Library, consisting of ChIP DNA molecules flanked by universal priming sites. Library preparation was performed essentially as described in the technical bulletin of the WGA2 Kit (Sigma). The OmniPlex© Library was then amplified by PCR within a limited number of cycles. This first whole genome amplification step was also conducted according to the manufacturer's instructions. The amplified DNA was purified with the QIAquick PCR Purification Kit (Qiagen). DNA was eluted with 50 µl H_2O and the DNA quantity and quality control was performed with a ND-1000 Spectrophotometer (NanoDrop Technologies), and was usually larger than 1 µg. In addition, DNA quality was monitored by agarose gelelectrophoresis. 15 ng of purified DNA was re-amplified as described in the technical bulletin of the WGA2 Kit, but with the following modification. Re-amplification was carried out in the presence of 0.4 mM dUTP. Incorporation of dUTP was a prerequisite for the enzymatic fragmentation of DNA (described in the following section). After amplification DNA was purified with the QIAquick PCR Purification Kit. DNA was eluted with 50 µl H_2O and the DNA quantity and quality control was again assessed with a ND-1000 Spectrophotometer, and agarose gelelectrophoresis.

2.2.3 DNA labeling, fragmentation and tiling array handling

The enzymatic fragmentation, labeling, hybridization and array scanning were done according to the manufacturer's instructions (Affymetrix Chromatin Immunoprecipitation Assay Protocol P/N 702238). Enzymatic fragmentation and terminal labeling were performed by application of the GeneChip WT Double-Stranded DNA Terminal Labeling Kit (P/N 900812, Affymetrix). Briefly, re-amplified DNA was fragmented in the presence of 1.5 µl Uracil-DNA-glycosylase (10 U/µl) and 2.25 µl APE1 (100 U/µl) at 30°C for 1 h 15 min. The average fragment size was in the range of 50-70 bases as determined by automated gel electrophoresis on an Experion system (Bio-Rad Laboratories) that allowed the analysis of small amounts of DNA. The fragmented DNA was then labeled at the 3′-end by adding 2 µl and 1 µl of Terminal nucleotidyl transferase (TdT, 30 U/µl) and GeneChip DNA Labeling Reagent (5 mM), respectively. 5.5 µg of fragmented and labeled DNA were hybridized to a high-density custom-made Affymetrix tiling array (64) (PN 520055) at 45°C for 16 h with constant rotational mixing at 60 rpm in a GeneChip Hybridization Oven 640 (Affymetrix, SantaClara, CA). Washing and staining of the tiling arrays were performed using the FS450_0001 script of the Affymetrix GeneChip Fluidics Station 450. The arrays were scanned using an Affymetrix GeneChip Scanner 3000 7G. The resulting raw data image files (.DAT) were inspected for any impairment. The CEL intensity files were used for bioinformatics analysis.
At least two independent biological replicates were analyzed for each factor.

2.2.4 Bioinformatics analysis

2.2.4.1 ChIP-chip data normalization

All data normalization procedures were performed using R (259) and Bioconductor (94). For data import of the Affymetrix CEL files and the conversion into the basic Bioconductor object class for microarray data *ExpressionSet*, we used the R package *Starr* (338). Data normalization consisted of three steps. Firstly, we performed quantile normalization between replicate measurements (not between non-replicate measurements). Secondly, for each condition (including the reference measurements) we averaged the signal for each probe by calculating the geometric average over the replicate intensities. Thirdly, data from all factors were normalized using a combined mock IP (referring to ChIP-chip experiments with non-tagged wild-type strain) plus input reference normalization. Additional aspects of data normalization were described in (199).
Normalized signal was converted to occupancy values between 0-100% by setting the genome-wide 99.8% quantile to 100% occupancy and the 10% quantile to 0% occupancy. For the Pol II phospho-isoforms, ChIP enrichments were obtained by dividing ChIP intensities by the genomic input intensities.

2.2.4.2 Transcript-wise occupancy profiles

In order to calculate occupancy profiles over genes or other genomic features the normalized occupancy signal at each nucleotide of the region was calculated as the median signal of all probes overlapping this position (6.5 probes on average). Individual probe intensities were further smoothed using the sliding window smoothing procedure (window half size of 75 bp) implemented in the R package *Ringo* (303).

2.2.4.3 Gene-averaged occupancy profiles

Before the occupancy of a particular factor was averaged over different genes, the genes were selected as follows. We started with all nuclear *Saccharomyces cerevisiae* S288C protein-coding genes classified as 'verified' or 'uncharacterized' by the *Saccharomyces* Genome Database (5769 genes; http://www.yeastgenome.org/). To align gene profiles across entire transcripts, only genes with available TSS and pA assignments from RNA-seq experiments (219) were taken into account (4366 genes). Genes with TSS (pA) measurements downstream (upstream) of the annotated ATG (Stop) codon were excluded. To remove possible wrongly annotated TSSs and pAs, we only included genes with TSS (pA) annotations showing a distance less than 200 bp to the corresponding downstream (upstream) ATG (Stop) codon (3448 genes). As a result of the limited ChIP-chip resolution and the compactness of the yeast genome with its short intergenic regions (median inter-ORF length: 368 bp, median inter-transcript length: 259 bp), a gene's factor occupancy profile can have spurious contributions from flanking genes. To minimize these “spill-over” effects, we

focused on genes exhibiting a minimal ORF and transcript distance to flanking genes of 250 bp and 200 bp, respectively (1786 genes). Furthermore, we restricted our analysis to the 50% highest expressed nuclear protein-coding genes according to measurements by Dengl et al. (70) (1140 genes, ALL set). We grouped genes into four ORF length classes: Xtremely Short (XS) ranging from 256 to 511 bp, Short (S) 512 to 937 bp, Medium (M) 938 to 1537 bp, and Long (L) 1538 to 2895 bp, comprising 93, 266, 339, and 299 genes, respectively. Profiles within these groups were scaled to median gene length. We calculated gene-averaged profiles by taking the median over gene factor profiles.

2.2.4.4 Pairwise profile correlations and correlation network

Analyses were done using 4366 genes with available TSS and pA annotations (219). Pairwise Pearson correlations over factor occupancy profiles were calculated between concatenated gene profiles ranging each from TSS minus 250 bp to pA plus 250 bp. The correlation-based network was calculated using the GraphViz's Neato algorithm (90).
Singular value decomposition (SVD) analysis was performed as described in (199).

2.3 Molecular cloning and preparation of DNA

A standard **Polymerase Chain Reaction (PCR)** such as for the validation of TAP-tagged yeast strains was performed as follows. PCR reactions contained 3 µl (150 ng) genomic DNA template, 1.25 µl of 10 µM forward and reverse primer (usually 18 to 24 nt long), 5 µl 10x Taq Buffer, 6 µl of 10 mM $MgCl_2$, 2.5 µl of 2 mM dNTP-Mix, 30 µl H_2O and 1 µl Taq DNA Polymerase. PCR was performed on a T3000 Thermocycler (Biometra) using a 2 min denaturing step at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 50°C and 1 min at 72°C. Finally, the PCR reaction was incubated at 72°C for 10 min and stored at 4°C.
In cases where a new yeast strain was generated, such as an N-terminally TAP-tagged TFIIS strain, the PCR was typically performed as follows. PCR reactions contained 1 µl (100 ng) DNA template, 2 µl of 10 µM forward and reverse primer (usually 18 to 24 nt long), 10 µl 5x HF Buffer, 5 µl of 2 mM dNTP-Mix, 29 µl H_2O and 1 µl Phusion© DNA Polymerase. PCR was performed on a T3000 Thermocycler (Biometra) using a 30 sec denaturing step at 98°C, followed by 35 cycles of 7 sec at 98°C, 20 sec at 55°C and 1 min at 72°C. Finally, the PCR reaction was incubated at 72°C for 10 min and stored at 4°C.
In cases where no PCR product was obtained, the standard protocols were adjusted by changing particular parameters. Most critical parameters turned out to be the primers itself, the type of DNA polymerase and the elongation time.

For **enzymatic restriction cleavage** DNA was digested using restriction endonucleases (New England Biolabs, Fermentas) according to the manufacturer's instructions. Cleaved PCR products and cleaved plasmids were purified with the QIAquick PCR Purification Kit (Qiagen) and with the QIAquick gel extraction Kit (Qiagen), respectively.

Ligation of digested DNA into vectors was done with T4 DNA ligase (Fermentas) and its corresponding buffer in a volume of 20 µl for 1 h at room temperature or over-night at 16°C. In a typical ligation experiment a 5- to 10-fold excess of insert, relative to linearized vector was used. Ligation products were transformed into *E. coli* and yeast cells.

For **transformation of *E. coli* cells** chemically competent XL-1 Blue cells (see II 2.4 and Table 3) were used. Approximately 2 µg of DNA (usually 3 µl ligation product) were added to 50 µl of competent cells and incubated for 5 min on ice. Next, cells were heated for 30 sec at 42°C in a water bath and then put back on ice for 2 min. Transformed *E. coli* cells were recovered by incubation at 37°C for 1 h in 700 µl of LB medium. Recovered cells were centrifuged at 13,000 rpm for 30 sec (at room temperature) and 650 µl of the supernatant was removed. Cells were resuspended in the remaining volume and plated onto solid LB medium plates containing the respective antibiotic (see Table 11). Plates were incubated at 37°C over-night.

For **preparation of plasmid DNA**, 5 ml LB medium (containing the respective antibiotic) were inoculated with a single colony/clone and incubated at 37°C over-night (160 rpm). Cells were harvested by centrifugation at 4,500 rpm (Rotana 46R) for 10 min (4°C). Next, plasmid DNA was prepared with the QIAquick Miniprep Kit (Qiagen) according to the manufacturer's instructions.

For **transformation** of chemically competent ***S. cerevisiae* cells** (see section II 2.4) 10 µl PCR product (≥ 1 µg of linear DNA) and 360 µl LitPEG was added to 50 µl of competent cells. After incubation at room temperature for 30 min, 47 µl DMSO was added and in case that the yeast strain was not temperature-sensitive heated at 42°C for 15 min. Alternatively, if the yeast strain was temperature-sensitive cells were incubated at 30°C for 10 min and at 37°C for 5 min. After recovery (30°C, 1 h, 150 rpm), cells were collected by centrifugation at 2,000 rpm for 3 min (room temperature) and the supernatant was removed. Transformed yeast cells were resuspended in 50 µl sterile ddH_2O, plated on selective media plates and incubated at 30°C.

For **preparation of genomic DNA** 5 ml yeast were grown at 30°C (150 rpm) over-night. Cells were harvested by centrifugation at 2,500 rpm for 5 min (4°C), washed with 1 ml sterile ddH_2O and resuspended in 500 µl Lyticase Buffer. Next, 20 µl Lyticase (10 U/µl) were added. After incubation at 37°C for 45 min, yeast spheroblasts were collected by centrifugation at 5,000 rpm for 5 min (4°C). The supernatant was removed and spheroblasts were gently resuspended in 1 ml Spheroblast wash Buffer. Spheroblasts were centrifuged again and resuspended in 500 µl Spheroblast wash Buffer. Spheroblasts were collected again and resupended in 500 µl of TE 50/100 Buffer. 50 µl 10% SDS were added and incubated at 70°C for 30 min. Then, 250 µl 5 M Potassium acetate were added and incubated on ice for 15 min. The precipitated genomic DNA was collected by centrifugation at 13,000 rpm for 20 min (4°C), the supernatant was removed and 700 µl Isopropanol was added. Precipitated genomic DNA was collected by centrifugation at 13,000 rpm for 10 min (4°C). Precipitated DNA was washed with 70% Ethanol and resuspended in 500 µl TE 50/100 Buffer via heating at 42°C for 30 min. DNA quantity and quality control was performed with a ND-1000 Spectrophotometer (NanoDrop Technologies). Genomic DNA was stored at -20°C.

2.4 Preparation of competent cells

Chemically **competent *E. coli* cells** were prepared as follows. 200 ml LB medium (containing the respective antibiotic) were inoculated with 5 ml of an over-night culture of the desired *E. coli* strain. Cells were grown at 37°C (160 rpm) to an OD_{600} of 0.4 to 0.5 and incubated on ice for 10 min. The following steps were performed at 4°C with pre-cooled buffers. Cells were harvested by centrifugation and washed with 50 ml TFB-1 Buffer. After a second centrifugation step, cells were resuspended in 4 ml TFB-2 Buffer. Aliquots of competent *E. coli* cells were frozen in liquid nitrogen and stored at -80°C (see Table 12).

Chemically **competent *S. cerevisiae* cells** were prepared as follows. 50 ml YPD medium was inoculated (start OD_{600} of 0.2) with yeast that was grown to stationary phase over-night. Next, yeast cells were grown at 30°C (150 rpm) to an OD_{600} of 0.5 to 0.7 and harvested by centrifugation at 4,000 rpm for 10 min (room temperature). Yeast cells were washed with 25 ml sterile ddH_2O and finally resuspended in 360 µl LitSorb. 40 µl of heated (10 min at 100°C) salmon sperm DNA was added. Aliquots of competent yeast cells were stored at -80°C (see Table 12).

2.5 Protein expression in *E. coli*

Proteins were recombinantly expressed in BL21-CodonPlus (DE3)RIL *E. coli* cells (see Table 3). The plasmids that contained the genes coding for the desired protein variants were transformed into *E. coli* cells as described in section II 2.3. 10 ml LB medium containing chloramphenicol and other antibiotics (depending on the plasmid used) were inoculated with a single *E. coli* colony and grown (37°C, 160 rpm) to stationary phase over-night. The 10 ml culture was used to inoculate 1 L LB medium containing the respective antibiotics. Cells were grown (37°C, 160 rpm) to OD_{600} of 0.6 to 0.9 and cooled on ice for 30 min. Protein expression was induced by the addition of 0.5 mM IPTG and was performed at 18°C and 160 rpm over-night. Cells were harvested by centrifugation at 4,000 rpm (SLC6000 rotor, Sorvall) for 20 min. The cell pellet was resuspended in 30 ml of the corresponding lysis buffer, frozen in liquid nitrogen and stored at -80°C.

2.6 Protein purification

For **cell lysis** frozen cell suspensions (30 ml; see previous section) were thawed rapidly. All following steps were performed at 4°C with pre-cooled buffers. 20 ml of the respective lysis buffer and 0.5 ml of protease inhibitors (PIs) were added. Cells were disrupted by sonication (3x 10 min; duty cycle of 30%; output control of 40). Afterwards, the lysate was cleared by

centrifugation at 15,000 rpm (SS34 rotor, Sorvall) for 20 min. The supernatant was centrifuged a second time. The resulting cleared supernatant was used for further purification.

Nickel-NTA-Agarose in a column was used to purify hexahistidine-tagged protein variants of interest. The Nickel-NTA-Agarose was washed with 10 ml ddH_2O and equilibrated with 10 ml of the respective lysis buffer (without PI). The cleared cell lysate was applied to the equilibrated Nickel-NTA-Agarose column (bed volume: 2 ml). The flow-through was applied to the Nickel-NTA column for a second time. After binding of the protein variant, the column was washed with 10 ml of the respective washing buffers and the protein was eluted with 10 ml of the corresponding elution buffer. All purification steps were performed at 4°C and with pre-cooled buffers.
Usually, proteins were further purified by ion exchange and gel filtration chromatography.

Protein precipitation was performed with trichloroacetic acid (TCA). An equal volume of 20% TCA was added to the protein sample and incubated on ice for 30 min. Next, the sample was centrifuged at 13,000 rpm for 15 min (4°C), the supernatant was removed and the pellet was washed with 300 µl pre-cooled Acetone. The sample was centrifuged at 13,000 rpm for 5 min (4°C), the supernatant was removed and the pellet was dried. Precipitated protein was resuspended in 1x SDS-PAGE loading Buffer and heated at 65°C for 3 min. If necessary, the protein sample was neutralized with ammonia vapor.

Protein concentration was determined with a ND-1000 Spectrophotometer (NanoDrop Technologies). Thereby, the protein concentration was calculated from the absorption rate at a wavelength of 280 nm (OD_{280}). Individual extinction coefficients for the protein variant of interest were calculated with the ProtParam tool (http://web.expasy.org/protparam/) that is part of the ExPASy Bioinformatics Resource Portal.

2.7 Western Blot analysis

For Western blotting the protein sample was separated by SDS-PAGE as described in section II 2.8. Separated proteins were then transferred to a PVDF membrane (Schleicher & Schuell; pre-equilibrated with 100% Ethanol) in the presence of the WB transfer Buffer (see Table 12). Thereby, the wet blotting system from Bio-Rad was used according to the manufacturer's instructions. Transfer was performed at 100 V for 2 h on ice or at 20 V over-night (4°C). For quantitative Western blots, transfer was always performed over-night (see also section V 2.6). After transfer, the membrane was air-dried for 30 min. Next, the dried membrane was blocked for at least 1 h with WB blocking Buffer. The blot was then incubated at room temperature for at least 1.5 h with the primary antibody in WB blocking Buffer. The blot was washed four times with WB blocking Buffer at room temperature for 10 min and then incubated with the secondary antibody in WB blocking Buffer at room temperature for at least 1 h. Afterwards, the blot was washed four times with 1x PBS at room temperature for 10 min. Secondary antibodies were usually coupled to horseradish peroxidase. Final signals were detected with the Chemiluminescence Kit (Pierce) followed by exposure of the blot to high-sensitivity films (Invitrogen). Films were developed with the X-omat M35 developing machine (Kodak).

Alternatively, the chemiluminescent signals were measured with the Mini-LAS300 System (Fujifilm Life Sciences) and quantified with the ImageQuant software suite (GE Healthcare). The standard protocol was adjusted if needed. An overview of the different antibodies that were used in this work is given in Table 8 and 9.

2.8 Electrophoretic separation of DNA and proteins

Electrophoretic separation of DNA was performed in horizontal 1x TBE agarose gels that contained either Ethidium bromide (0.7 µg/ml) or SYBR Safe® (0.01 µg/ml; Invitrogen). Depending on the size of DNA molecules that were separated, the agarose concentration varied between 0.8% and 2% (w/v). Agarose gelelectrophoresis was carried out in PerfectBlue Gelsystem electrophoresis devices from Peqlab (110 V, 35 min). DNA samples were mixed with 6x loading dye (Fermentas). The sizes of the separated DNA molecules were assessed with the help of the GeneRuler™ 1kb DNA ladder or the GeneRuler™ 100bp DNA ladder (Fermentas).

The **electrophoretic separation of protein** samples was performed by SDS-PAGE with 15% to 17% acrylamide gels (acrylamide:bisacrylamide ratio = 37.5:1) (174) in Bio-Rad gel systems. Before loading onto the gel, protein samples were mixed with SDS-PAGE loading Buffer and boiled at 95°C for 2 min. Samples that required separation at a high spatial resolution, such as purified protein complexes via TAP (see section III 2.3), were separated by ready-to-use NuPAGE Novex Bis-Tris minigels (Invitrogen). MES and MOPS running buffers (Invitrogen) were used. SDS-PAGE with ready-to-use gels was carried out according to the manufacturer's instructions.
Gels were stained with Gel staining solution at room temperature for at least 20 min and destained at room temperature over-night in Gel destaining solution. Corresponding buffers and solutions are listed in Table 12.

III. Uniform transitions of the general RNA polymerase II transcription complex

1. Introduction

Gene transcription begins with the assembly of Pol II and its initiation factors on promoter DNA (309). Pol II then starts mRNA synthesis and exchanges initiation factors for elongation factors during initiation-elongation transition (2, 155, 243). Elongation factors are required for chromatin passage and RNA processing (137). During the elongation-termination transition that usually occurs at the 3' end of genes, elongation factors are exchanged by termination factors that are required for transcription termination (see also section IV 1.3; (155)).

1.1 Pol II transcription elongation

Pol II transcription elongation represents a distinct stage within the Pol II transcription cycle (274). During transcription elongation a nascent RNA molecule is synthesized by a stable elongation complex in a processive manner (198). In the course of the last years data has accumulated showing that transcription elongation is extremely complex and highly regulated (196, 274).

1.1.1 Transcript elongation occurs discontinuously and is highly regulated

Studies in higher eukaryotes have shown for selected genes (26, 103) and genome-wide (217, 221, 339) that Pol II pauses 25-50 nt downstream of the TSS at a subset of genes. In *Drosophila* cells, promoter proximal pausing was observed at one third of all genes, mainly at developmentally regulated and stimulus-responsive genes (221). ChIP-chip analyses in yeast grown to stationary phase detected Pol II occupancy at the promoter region of many inactive genes but not downstream of that region (255). Although this observation was interpreted as Pol II pausing, it is not known whether promoter proximal pausing exists in proliferating yeast. Promoter proximal pausing is currently considered as an important regulatory step after the recruitment of Pol II to the gene promoter (196). Particular elongation factors, certain DNA sequences (may be also RNA sequences) and nucleosome occupancy play a role in specifying promoter proximal pausing (220). It is currently under debate whether promoter proximal pausing is a true pausing event of Pol II during transcription elongation or whether the observed occupancy peak of Pol II downstream of the TSS rather represents an intermediate of early termination (David Bentley, University of Colorado, and Stephen Buratowski, Harvard Medical School, personal communication; (288)).
Apart from promoter proximal pausing that usually occurs shortly after the TSS, pausing events can occur throughout the body of genes. ChIP, ChIP-chip and NET-seq studies in yeast could show that Pol II pauses at arrest sites (171), upstream of nucleosomes (49) and at the 3' end of introns (5, 39). Single cell experiments revealed that the transcript elongation rate

along a gene can vary considerably (62). This is in accord with the observation that transcription occurs in bursts with pulses of high polymerase density (48). All these observations led to the view that transcription elongation is a discontinuous process, interrupted by periods of regulated pauses and arrests. Given the reasonable assumption that several Pol II elongation complexes can transcribe the same gene at a time, pausing would lead to rear-end collisions between the leading elongation complex and the subsequent complexes (267).

1.1.2 Transcription elongation through chromatin

Gene transcription in living cells occurs in a chromatin context (170). Chromatin consists of repeating subunits, the nucleosomes (see also section I 2.2; (191)). Linear nucleosomal arrays represent an extremely strong barrier to Pol II passage (230). One possibility of how Pol II elongation can proceed is nucleosome disassembly in front of Pol II. This mechanism was observed at highly transcribed genes with a high density of Pol II molecules (165, 273). Additionally, a genome-wide analysis of nucleosome occupancy in yeast revealed that the transcription rate is inversely proportional to the histone density in the coding region of genes (178). Although, transcriptionally very active yeast genes were depleted of histones in the coding region, histones immediately re-associated with the DNA at a very high rate when genes were turned off (273). Taken together, during gene transcription the chromatin structure is reversibly altered.

To achieve transcription in a chromatin environment, cells have evolved different classes of factors that together with Pol II are able to modify chromatin structure. Among those factors are chromatin remodelers, histone chaperones, histone modifying enzymes and other transcription factors (13). Whereas chromatin remodelers use the energy of ATP hydrolysis to move, destabilize, eject or restructure nucleosomes (50), histone chaperones assemble or disassemble nucleosomes without using the energy of ATP (13). For example Spt16/Pob3 (also called FACT complex in higher eukaryotes) and Spt6 (see following section) are histone chaperones in yeast that facilitate Pol II transcription elongation *in vivo* by destabilizing nucleosome structure and by reassembling nucleosomes after Pol II passage (24).
Histone modifying enzymes predominantly modify the flexible histone tails, but also the nucleosomal core (88). Although the best characterized modifications at the moment are histone methylation and acetylation, many other post-translational modifications were known including phosphorylation, ubiquitylation and crotonylation (300). According to the histone code model (292) the post-translational histone modifications are rather used for binding of chromatin-related proteins than directly affecting chromatin compaction by its physicochemical properties. The biological function of the various histone modifications as well as the crosstalk between the different marks is currently far from being understood.
Apart from these distinct subsets of factors that alter the chromatin structure other proteins associate with Pol II during transcription elongation to achieve its passage through the nucleosomal DNA template. Among these transcription factors are the Paf1 complex, Spt4/5 (see also section V 1.1), TFIIS, Elf1 and Spn1 (167). Whereas several studies have shed some light on the biological role of the Paf1 complex (142), Spt4/5 (184), and TFIIS (45, 151),

almost nothing is known about Elf1 and Spn1. Elf1 is a non-essential nuclear protein that genetically interacts with Paf1, Spt4/5, Spt6 and Spt16, and localizes to actively transcribed regions (246). Spn1 represents an essential nuclear protein that directly interacts with Spt6 (167). Spn1 may regulates the binding of Spt6 to nucleosomes (200).

Spt6 is a highly conserved nuclear protein and is essential for viability in yeast. Spt6 is required for Pol II transcription elongation and normal chromatin structure. Consistent with its role in transcription elongation, Spt6 co-localizes to chromosomal regions in *S. cerevisiae* and *Drosophila melanogaster* actively transcribed by Pol II. Spt6 stimulates the elongation rate of Pol II in vivo and interacts with several other transcription elongation factors, including Spt4/5, Elf1 and Spn1. In addition, Spt6 contains a C-terminal tandem SH2 domain that binds phosphorylated Pol II CTD (73, 185, 294) (Figure 5). Deletion of that domain in yeast cells leads to a slow-growth phenotype and is lethal in the presence of 6-azauracil (6-AU), indicating a role in transcription elongation (294). Spt6 may also bind nascent RNA via its S1 domain (Figure 5).

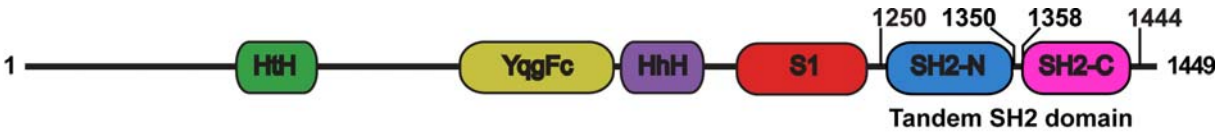

Figure 5: Domain architecture of yeast Spt6. HtH, helix-turn-helix domain, binds double-stranded DNA; YqgFc, predicted to be resolvase or ribonuclease, but in Spt6 catalytic residues are exchanged, thus probably not active; HhH, triple-helix-domain, binding of double stranded DNA; S1, RNA binding domain; SH2-N, SH2-C, tandem SH2 domain; numbers for amino acid residues are indicated; Spt6 is a modular protein with different interaction faces: nucleosome/DNA, RNA and Pol II interaction faces (70). (modified from (294))

Consistent with its role in influencing the chromatin structure, Spt6 directly interacts with histone H3 and can assemble nucleosomes *in vitro* (32). In agreement with this study, spt6 mutations lead to transcription from cryptic promoters within the coding region of genes, suggesting that Spt6 is required to re-establish correct chromatin structure after Pol II passage (145), especially at highly transcribed genes (133). Spt6 seems also to be involved in the positioning of the +1 nucleosome (133). Recently, it could be shown that Spt6 is required in heterochromatic silencing in *Schizosaccaromyces pombe* (152).

1.2 Modification of the Pol II CTD during gene transcription

In the course of the Pol II transcription cycle the CTD is dynamically phosphorylated and dephosphorylated by site-specific CTD kinases and CTD phosphatases (161). During transcription initiation Ser5 residues of the CTD consensus repeats are phosphorylated (S5P) by Kin28 and CDK7, subunits of the yeast and mammalian TFIIH complex, respectively (4, 162, 272). Srb10/CDK8 that is part of the Mediator complex can also phosphorylate Ser5 residues of the Pol II CTD (34, 121). But in contrast to Kin28, Srb10 phosphorylates the CTD

prior to the formation of the initiation complex what interferes with the recruitment of Pol II to the gene promoter and results in inhibition of transcription (121). This modification facilitates the association of the mRNA capping enzyme (98), the termination factor Nrd1 (308) and of the histone methyltransferase Set1 (222).

During transcription initiation Kin28 also phosphorylates the Pol II CTD at Ser7 residues in yeast (157). Its function, if any, is currently not known. However, in mammalian cells CTD Ser7-phosphorylation (S7P) by CDK7 occurs further downstream of the TSS, during transcription elongation (41). S7P recruits the Integrator complex and thus is important for 3' end processing of snRNAs (80, 81). Shortly downstream of the TSS, S5P and S7P marks were removed in yeast (157, 162, 272). S5P levels can be modulated by the CTD phosphatase Ssu72 (328). It is currently under debate whether Rtr1 is a bona fide phosphatase that may also contribute to reduce S5P levels (213). The S7P CTD phosphatase is not known.

In the course of transcription elongation CTD Ser2-phosphorylation (S2P) levels increase (162). The CTD Ser2-kinases in yeast are Bur1 and Ctk1 (253). Bur1 is recruited by CTD Ser5-phosphorylation (253). However, the mechanism of Ctk1 recruitment remains to be determined. In higher eukaryotes P-TEFb/CDK9 phosphorylates CTD on Ser2 residues (239). CTD Ser2-phosphorylation facilitates the association of elongation factor Spt6 (294) and recruits termination factors Pcf11 and Rtt103 to transcribing Pol II (193, 201). Levels of Ser2-phosphorylation are modulated by the CTD phosphatase Fcp1 (46).

Apart from phosphorylation of serine residues, the CTD in mammals can also be phosphorylated at tyrosine-1 residues (Y1P) (see section IV 1.2; (19)). The biological function of this modification and the corresponding CTD Y1P phosphatase is not known. It is also an open question whether Tyr1-phosphorylation exists in lower eukaryotes such as yeast.

As this work was in preparation, a study appeared showing that Pol II CTD is also phosphorylated on residue Thr4 by the Ser2 kinase P-TEFb/CDK9 (128). The study revealed that in chicken cells Thr4 phosphorylation seems to be specifically required for 3' end processing of histone mRNAs (128). Although Thr4 phosphorylation was also detected in yeast and human it remains to be determined whether it plays a similar role, as observed in chicken cells, or not. The CTD posphatase that targets phosphorylated Thr4 residues as well as factors that may bind to this mark is currently not known.

Apart from phosphorylation, the Pol II CTD can also be modified by acetylation (Melanie Ott, University of California, San Francisco, personal communication), glycosylation (147) and methylation (279). Whereas the biological function of CTD acetylation and glycosylation is currently not known, methylation of the arginine residue R1810 of the CTD in mammals by the coactivator-associated arginine methyltransferase 1 (CARM1) facilitates the expression of snRNA and snoRNA genes (279). Similar to the Ser7-phosphorylation in mammals, CTD methylation seems to have a gene-class specific role. Although R1810 is conserved in mammals, *Drosophila melanogaster* and *Caenorhabditis elegans*, the arginine residue is absent in yeast. It is not known whether CTD methylation occurs in yeast.

Additionally, the peptidylprolyl-*cis/trans* isomerase Ess1 catalyzes the *cis-trans* isomerization of peptide bonds N-terminal of proline residues (Pro-3 and Pro-6 of CTD), resulting in a conformational change of the Pol II CTD (275). Ess1 and its mammalian homologue, called Pin1, was shown to directly bind to phosphorylated CTD and to CTD that is doubly phosphorylated at Ser2 and Ser5 residues, respectively (212, 312). Inhibition of Pin1 by a

specific inhibitor blocked Pol II transcription (40). Recently, it was shown that Ess1 plays a role in snRNA transcription termination (280).

1.3 Aims and scope of this work

Until recently, most models of Pol II transcription were based on experiments performed for a very limited number of genes. Additionally, these case studies were often conducted *in vitro* in the absence of regulatory constraints that prevail in the living cell. Therefore for a long time the picture of Pol II gene transcription was gene biased and rather incomplete. However, during the last couple of years new methods were developed that allowed the analysis of Pol II transcription on a systems-wide level *in vivo* (see section I 3). Microarray and sequencing based techniques have shown that more than 75% of the genome in proliferating yeast is expressed (64, 219). Complementary to these studies, ChIP-chip analyses have shown that large parts of the genome in proliferating yeast is occupied by Pol II (287, 309).
Recently, it could be shown that initiation factors such as TFIIB are present at all active gene promoters in yeast. However, it was unknown whether all Pol II elongation factors are recruited to all active genes or only to subsets of genes. Similarly, it was also not known whether 5'- and 3'-RNA processing factors are recruited to all active genes. Therefore the main goal of this work was to investigate whether a general Pol II transcription complex exists that is required for the transcription of all active genes in yeast. To achieve this aim, key components of the Pol II transcription machinery should be subjected to systematic ChIP-chip analysis to reveal whether these factors co-localize at active genes. Statistical analyses of the data should provide additional evidence for the existence of a general Pol II transcription complex. This genome-wide localization analysis should be combined with classical biochemical approaches such as tandem affinity purifications (TAPs) to look whether factors that co-localize on chromatin also interact *in vivo*.
Our genome-wide study also bears the potential to answer another interesting question, namely whether the composition of the Pol II transcription complex changes during transcription. Case studies already indicated that parts of the initiation factors are exchanged by elongation factors, and that in turn some elongation factors are exchanged by termination factors at the 5'- and 3'-region of genes, respectively (155, 243). However, it is not known whether these 5'- and 3'-tranistions occur at all active genes in a uniform manner. Furthermore, the exact locations of the transitions within the transcribed region are also unknown.
More than a decade ago the laboratories of Stephen Buratowski (Harvard Medical School) and David Bentley (University of Colorado) could show for selected genes that the Pol II CTD is dynamically phosphorylated and dephosphorylated during transcription in yeast (162, 272). It is unknown whether this CTD phosphorylation cycle occurs on a genome-wide level. This question should be tackled by ChIP-chip analyses of Pol II and its various CTD phosphoisoforms. In addition, all the occupancy profiles should be used to investigate how factor recruitment correlates with distinct Pol II CTD phosphorylation patterns.
In higher eukaryotes, Pol II often pauses during early transcription elongation near the TSS (217, 339). Although Pol II is stalled at the promoter region of inactive genes in yeast grown

to stationary phase (255), it is unknown whether similar type of pausing occurs in proliferating yeast. ChIP-chip analyses of Pol II and its different CTD phosphorylated forms may provide evidence for pausing during transcription elongation.

This work was performed in cooperation with Michael Lidschreiber and Matthias Siebert.

2. Specific procedures

Table 17: Proteins analyzed by ChIP-chip (other proteins that were investigated by ChIP-chip are listed in Table 19 and 22)

Protein	Process involved	ChIP protocol
Rpb3 (Pol II)	Gene transcription	For TAP-tagged proteins
S5P (Pol II)	Pol II transcription early elongation/ 5'-RNA processing	For Pol II phospho-isoforms
S7P (Pol II)	Currently not known in yeast	For Pol II phospho-isoforms
S2P (Pol II)	Pol II transcription elongation/3'-RNA processing	For Pol II phospho-isoforms
TFIIB	Pol II transcription initiation	For TAP-tagged proteins
Kin28 (TFIIH)	Pol II transcription initiation	For TAP-tagged proteins
Tfg1 (TFIIF)	Pol II transcription initiation	For TAP-tagged proteins
Cet1 (mRNA capping enzyme)	5'-RNA processing (mRNA Capping)	For TAP-tagged proteins
TFIIS	Pol II transcription initiation/elongation	For TAP-tagged proteins
Bur1	Pol II transcription elongation	For TAP-tagged proteins
Ctk1 (CTDK-I complex)	Pol II transcription elongation	For TAP-tagged proteins
Elf1	Currently not known	For TAP-tagged proteins
Paf1 (Paf1 complex)	Pol II transcription elongation	For TAP-tagged proteins
Pcf11 (CFIA complex)	Pol II transcription termination/ 3'-RNA processing	For TAP-tagged proteins
Spn1 (also called Iws1)	Pol II transcription elongation	For TAP-tagged proteins
Spt4	Pol II transcription elongation	For TAP-tagged proteins
Spt5	Pol II transcription elongation	For TAP-tagged proteins
Spt6	Pol II transcription elongation	For TAP-tagged proteins
Spt6ΔC	(C-terminal truncation)	For TAP-tagged proteins
Spt16 (FACT complex)	Pol II transcription elongation	For TAP-tagged proteins

2.1 Molecular cloning and phenotyping of epitope-tagged yeast strains

The TAP-tag has a molecular weight of 25 kDa and thus leads to a markedly extension of the epitope-tagged protein of interest. It has been shown that the TAP-tag can interfere with the biological function of the tagged protein (31, 143). Although, this seems to be true for only a small subset of proteins, a C-terminal TAP-tag on **TFIIS** led to a slow-growth phenotype in the presence of 6-azauracil. This indicated that a C-terminal TAP-tag could interfere with the biological function of TFIIS. To test whether an N-terminally TAP-tagged version of TFIIS is compatible with yeast cell growth and thus with the biological function of TFIIS, an N-terminally TAP-tagged TFIIS strain was generated essentially as described in (31). Briefly, the *URA3* gene (of *Kluyveromyces lactis*) was originally amplified from the plasmid CIp10

(215) (primers: DST1_URA_Fw; DST1_URA_Rev; Table 6) and transformed into BY4741 wild-type yeast cells. The amplified DNA was inserted after the start codon of *DST1*, the gene coding for TFIIS. Positive transformants were identified on SC -ura plates. In a second step the *URA3* gene was replaced by the TAP-tag, originally amplified from the plasmid pBS1539 (251) (primers: DST1_TAP_Fw; DST1_TAP_Rev; Table 6). Transformants were plated on 5-FOA to select for loss of the inserted *URA3* gene (counter selection). Finally, the strain expressing an N-terminally TAP-tagged version of TFIIS was validated by control PCR (primers: DST1_cont_Fw; DST1_cont_Rev; Table 6), DNA sequencing and Western Blotting (see section II 2.1.1 and II 2.7). This strain was further characterized by phenotyping experiments (described in following section). By the application of this cloning strategy the tagged gene was still under the control of the endogenous *DST1* promoter and no additional sequences other than the TAP-tag encoding nucleotides were inserted into the genome.

To test whether **Elf1** and Spn1 could interact *in vivo*, a yeast strain was generated that expresses a C-terminally 3HA-tagged version of the Pol II transcription elongation factor Elf1 and a C-terminally TAP-tagged version of Spn1.
The plasmid pFA6a-3HA-kanMX6 (190) was used as a PCR template to amplify the DNA coding for the 3HA-Tag and the *kanMX6* marker cassette (primers: 66_Elf1-3HA_Fw, 68_Elf1-3HA_Rev; Table 6). The PCR product was transformed into the Spn1-TAP strain (BY4741 background) and replaced the original stop codon of *ELF1* (new stop codon at the 3'-end of the DNA region that codes for the 3HA-tag). Positive transformants were identified on G418 selective media plates. The Spn1-TAP Elf1-3HA yeast strain was validated by control PCR (primers: 68_Elf1-3HA_control_Fw, 66_Elf1-3HA_control_Rev; Table 6), DNA sequencing, Western blotting and phenotyping as described in section II 2.1.1 and II 2.7.

To test whether a TAP-tag fused either C-terminally or N-terminally to the Pol II transcription elongation factor TFIIS may interferes with yeast cell growth, **serial dilutions** of the C-terminally and N-terminally TAP-tagged TFIIS strains were grown at different temperatures or in the presence of the nucleotide depleting drug 6-azauracil (6-AU). Yeast cells were grown at 30°C to stationary phase over-night. Yeast cell cultures were diluted to an OD_{600} ~5 and spotted in 1:10 serial dilutions on YPD plates or on selective media plates containing 25, 50 or 100 µg/ml 6-AU. YPD plates were incubated at 30°C (= standard condition), 37°C or at 14°C for several days and inspected daily. 6-AU selective media plates were incubated at 30°C for several days and inspected daily. Wild-type yeast (no TAP-tag) and TAP-tagged TFIIS strains were grown on the same solid media plate. Biological duplicate measurements were performed.

2.2 ChIP-chip analysis of the Pol II transcription machinery

An overview of all factors that were investigated by ChIP-chip in this study is given in Table 17. ChIP-chip analysis of the Pol II transcription machinery in proliferating yeast was performed as described in section II 2.1, II 2.2 and according to (9, 83). However, the standard protocol had to be optimized for the genome-wide localization analyses of the Spt6ΔC mutant, the Pol II transcription elongation factor TFIIS and for the different Pol II

phospho-isoforms. For all factors at least two independent biological replicate measurements were conducted. Before ChIP-chip analyses of the various factors, the corresponding yeast strains were validated as described in section II 2.1.1.

2.2.1 Genome-wide occupancy profiling of Spt6ΔC and TFIIS

For ChIP-chip analysis the Spt6ΔC-TAP yeast strain (70) (Table 4) was used. This strain expressed a C-terminally TAP-tagged version of a C-terminally truncated form of the Pol II elongation factor Spt6 (lacking the 202 C-terminal residues). Since the lysis efficiency of Spt6ΔC-TAP cells was two-fold lower than for all the other yeast strains used for ChIP-chip analysis, the cell number was doubled. Therefore, 1.2 L (instead of 600 ml) of yeast was grown to mid-log phase (OD_{600} ~ 0.8). All the other steps were performed as described in section II 2.2.
For the genome-wide localization analysis of the Pol II transcription elongation factor TFIIS, the TFIIS N-TAP strain was used. This yeast strain expressed an N-terminally TAP-tagged form of TFIIS (see also III 2.1). This strain was used, since a C-terminal TAP-tag interfered with yeast cell growth and thus with the biological function of TFIIS. The initial ChIP-signal of TFIIS obtained with the standard protocol was low. The fold enrichment over an unoccupied genomic region was in the range of 3- to 4-fold. By doubling the amount of cells, the ChIP-signal of TFIIS could be improved. Therefore, 1.2 L (instead of 600 ml) of yeast was grown to mid-log phase (OD_{600} ~ 0.8). All the other steps were performed as described in section II 2.2.

2.2.2 ChIP-chip analysis of Pol II and its different phosphorylated forms

ChIP-chip analysis off Pol II was performed with a yeast strain expressing a C-terminally TAP-tagged version of the core subunit Rpb3. This allowed the genome-wide localization analysis of Pol II independent of the CTD phosphorylation state. ChIP-chip experiments were performed essentially as described in II 2.1.1 and II 2.2.
For ChIP-chip analyses of the Pol II phospho-isoforms a set of monoclonal antibodies with strong specificity and affinity for phosphorylated serine residues S5P (3E8), S2P (3E10) and S7P (4E12) were used (41). Since the monoclonal antibodies were generated in rats, it remained to be determined whether these antibodies work in ChIP experiments performed in yeast. Furthermore, it was reported that the amount of antibody used influences the occupancy profiles obtained for Pol II phospho-isoforms (157). Therefore, we carried out ChIP-qPCR experiments with different amounts of antibodies before ChIP-chip analyses as described in (199). Based on these experiments the amount of antibody for chromatin immunoprecipitation could be optimized (see Table 8). ChIP-chip analyses of the different Pol II phospho-isoforms were performed essentially as described in section II 2.1.2 and II 2.2.

2.3 Tandem affinity purification of Elf1 and Spn1

Tandem affinity purification (TAP) of Elf1 and Spn1 was performed essentially as described (251). Briefly, for Elf1 TAP an Elf1-TAP yeast strain was used. Spn1 TAP was performed with a yeast strain that expressed a C-terminally TAP-tagged version of Spn1 as well as a C-terminally 3HA-tagged version of Elf1 (Spn1-TAP Elf1-3HA; see section III 2.1). 4 L of yeast were grown from OD_{600} of 0.1 to ~3 at 30°C. Cells corresponding to a 1 L yeast culture were harvested, washed with 500 ml of sterile H_2O (4°C), resuspended in 25 ml TAP lysis Buffer, flash-frozen in liquid nitrogen and stored at -80°C. Two pellets were pooled (corresponding to a 2 L yeast culture), the same volume of glass beads were added and lysed by bead beating for 12 min at 500 rpm and 4°C (Pulverisette). The cell lysate was ultra-centrifuged for 1.5 h at 27,000 rpm (SW28 rotor, Beckman Coulter) and 4°C. 0.5 ml (slurry) of IgG beads pre-equilibrated in TAP lysis Buffer were added to the lysate and incubated for 1.5 h at 4°C. Beads were transferred into a column (Bio-Rad) and washed with 10 ml TAP wash Buffer (4°C). Beads were resuspended in 150 µl TAP wash Buffer and incubated with 2 µl (8 µg) TEV protease for 1.5 h at 20°C. After incubation 150 µl of TAP wash Buffer was added. The resuspension was diluted with an equal volume of Calmodulin Buffer 1 and 0.5 ml (slurry) Calmodulin beads, pre-equilibrated in Calmodulin Buffer 2, was added and incubated for 1 h at 4°C. Beads were washed with 5 ml Calmodulin Buffer 2. Next, 600 µl of TAP elution Buffer was added and proteins were eluted during a 20 min incubation step at 37°C. Proteins in the final eluate were precipitated with Trichloroacetic acid (TCA) as described in section II 2.6. All buffers are listed in Table 14.
Proteins associated with the purified TAP-tagged proteins were identified either by mass spectrometry (Zentrallabor für Proteinanalytik, Ludwig-Maximilians-Universität München) or by Western blot analysis using monoclonal antibodies directed against an epitope tag of the protein of interest (α-3HA-tag, 3F10; α-TAP-tag, PAP;) or, if available, against the protein of interest itself (α-Rpb1, 8WG16; α-Rpb3, 1Y26). For further details of the antibodies please see Table 9. Western blot analysis was performed as described in section II 2.7.

3. Results

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

We determined genome-wide occupancy profiles by ChIP in exponentially growing *Saccharomyces cerevisiae* strains expressing tandem affinity purification (TAP)-tagged proteins (see Table 4 and 17). Chromatin immunoprecipitation was performed as described in section II 2.1, II 2.2 and III 2.2. Enriched DNA fragments of an average size of 250 nt were analyzed with tiling microarrays that cover the yeast genome at 4 nt resolution (64). For data normalization, we developed a procedure that corrects for non-specific antibody binding by using both mock immunoprecipitation and input measurements (see section II 2.2.4.1). Data from two or three highly reproducible replicates were averaged. The profile for the Pol II subunit Rpb3 (Figure 6) matched our previous profiles, which were obtained with different strains, experimental protocols, and array platforms (136), but showed more details. Pol II was observed at protein-coding, snRNA, and snoRNA genes, and at regions producing cryptic unstable and unannotated transcripts (330), but was lacking at genes transcribed by Pol I and Pol III.

Out of 4366 yeast genes with annotated TSS and pA sites (219), 2465 (56%) showed Pol II peak occupancies above 20%, consistent with transcription of most of the genome (64). To average Pol II profiles over genes, genes with the 50% highest mRNA levels (70) and a minimum distance to neighbouring genes were sorted into four major length classes, scaled to adjust length differences, and aligned with their TSS and pA sites (see section II 2.2.4). The pA site marks the point of RNA 3′-cleavage and polyadenylation, but transcription continues beyond this site until termination. Consistent with this, the gene-averaged Rpb3 profile revealed Pol II occupancy through the transcribed region into the region flanking the pA site on the 3′-side (Figure 6 b,c). Although this observation already indicates that the gene-averaged occupancy profile of Pol II is biologically meaningful, we also looked whether one can find this occupancy pattern also at single genes. As shown in Figure 6 a, the gene-averaged profile of Pol II is also true on a single-gene level.

Taken together, ChIP-chip analysis in yeast revealed a uniform Pol II occupancy pattern on most protein coding genes, as well as Pol II binding at snRNA genes, snoRNA genes and at regions coding for CUTs and SUTs.

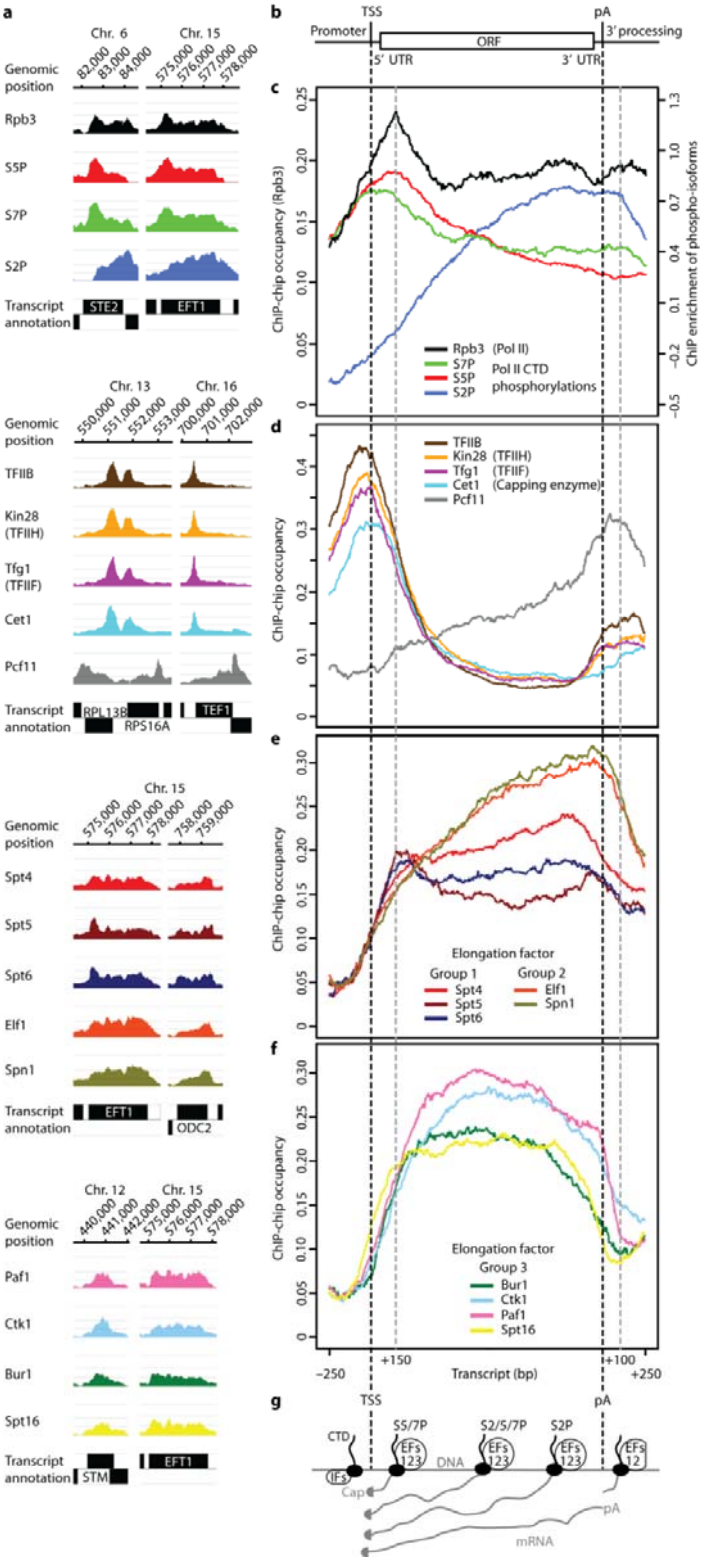

Figure 6: Genome-wide occupancy profiling of the Pol II machinery. (a) Factor occupancy on selected genes. Colored profiles represent normalized factor occupancies smoothed by a 150 nt window running median. Black boxes indicate transcripts (64) on the Watson (top) and Crick strands (bottom). **(b)** DNA frame with promoter, 5′ untranslated region (UTR), open reading frame (ORF) and 3′ UTR. Dashed black lines indicate the TSS and pA site. Dashed gray lines mark the positions 150 nt downstream of the TSS and 100 nt downstream of the pA site. **(c)** Gene-averaged profiles for the median gene length class (1,238 ± 300 nt, 339 genes) of Pol II and its phosphorylated forms. Profiles of other length classes are generally similar (exemplarily shown for Pol II initiation and 5' capping factors Figure 7). Occupancies and signal intensities are given for Rpb3 and phosphorylated Pol II on the left and right *y* axes, respectively. **(d)** Gene-averaged profiles as in **(c)** for initiation (TFIIB, TFIIF, TFIIH), 5′ capping (Cet1) and termination (Pcf11) factors. **(e,f)** Gene-averaged profiles as in **(c)** for elongation factors of groups 1 (Spt4, Spt5, Spt6), 2 (Elf1, Spn1) and 3 (Spt16, Paf1, Ctk1, Bur1). **(g)** Cartoon representation of Pol II (black dots) and its CTD (black lines) transcribing along DNA (horizontal gray line) from left to right, to produce mRNA (gray lines). IFs, initiation factors; EFs 123, elongation factors of groups 1, 2 and 3; S2/5/7P, phosphorylation of serines 2, 5 and 7.

3.2 Initiation/termination factors flank the transcribed region

Gene-averaged profiles for the initiation factors TFIIB, -F, and -H show a single strong peak 50-30 nucleotides upstream of the TSS (Figure 6 b,d and Figure 7) and was independent of the gene length (Figure 7). This indicates the presence of initiation complexes at promoters and is consistent with a scanning mechanism for TSS location in yeast (163, 168). TFIIF was only observed at promoters and not within transcribed regions, indicating that its reported elongation-stimulatory activity *in vitro* (261) is restricted *in vivo* to early RNA elongation and to downstream sites of transient association. Weaker peaks for initiation factors were observed in the 3′-region and this may indicate gene looping at selected genes. Occupancy of the capping enzyme subunit Cet1 peaked just downstream of the TSS, consistent with capping when the nascent RNA appears on the Pol II surface. The symmetric peaks of averaged occupancy of initiation factor and capping enzyme indicated that these factors are restricted to defined locations just upstream and downstream, respectively, of the TSS. Occupancy for the 3′-processing and termination factor Pcf11 peaked downstream of the pA site, consistent with transcription and completion of mRNA 3′-end formation downstream of the pA site. The location of Pcf11 peak occupancy levels were independent of the gene length (Figure 6 and 6/7). Thus, representative initiation and termination factors show peak occupancies outside the transcribed region, and are apparently not present during mRNA chain elongation.

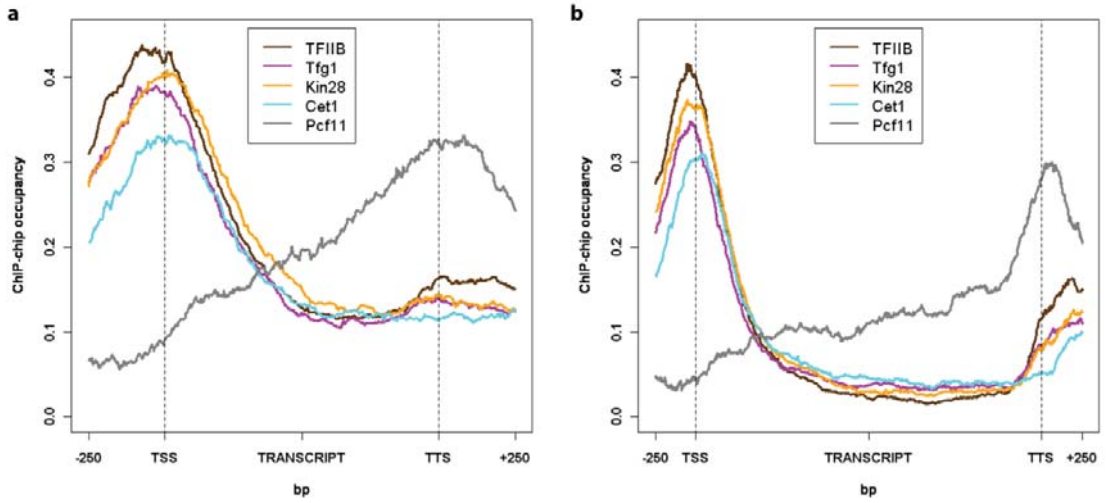

Figure 7: Gene-averaged median profiles of Pol II initiation and termination factors for (a) "Short" and (b) "Long" genes. The "Short" gene length class includes genes with a length ranging from 512 to 937 bp (in total 266 genes, see methods). The "Long" gene length class contains genes with a length ranging from 1538 to 2895 bp (in total 299 genes, see methods). The gene-averaged median profiles of Pol II initiation factors for "Medium" sized genes are shown in Figure 6.

3.3 Elongation factors enter during a single 5′-transition

Elongation factor profiles did not correlate with profiles of initiation or termination factors (Figure 8). Elongation factors were absent at the promoter, but their occupancies sharply increased downstream of the TSS within a narrow window of ~50 nt, indicating coordinated

elongation complex assembly during a single 5′-transition (Figure 6, e-f). Spt16/FACT was an exception, as it entered already ~30 nucleotides further upstream. Elongation factors showed characteristic distributions over the transcribed region. Three distinct profile shapes were observed and used to group the factors (Figure 8 b). Group 1 includes Spt4, Spt5, and Spt6, group 2 includes Spn1/Iws1 and Elf1, and group 3 includes Spt16/FACT, Paf1, and the CTD kinases Bur1 and Ctk1.

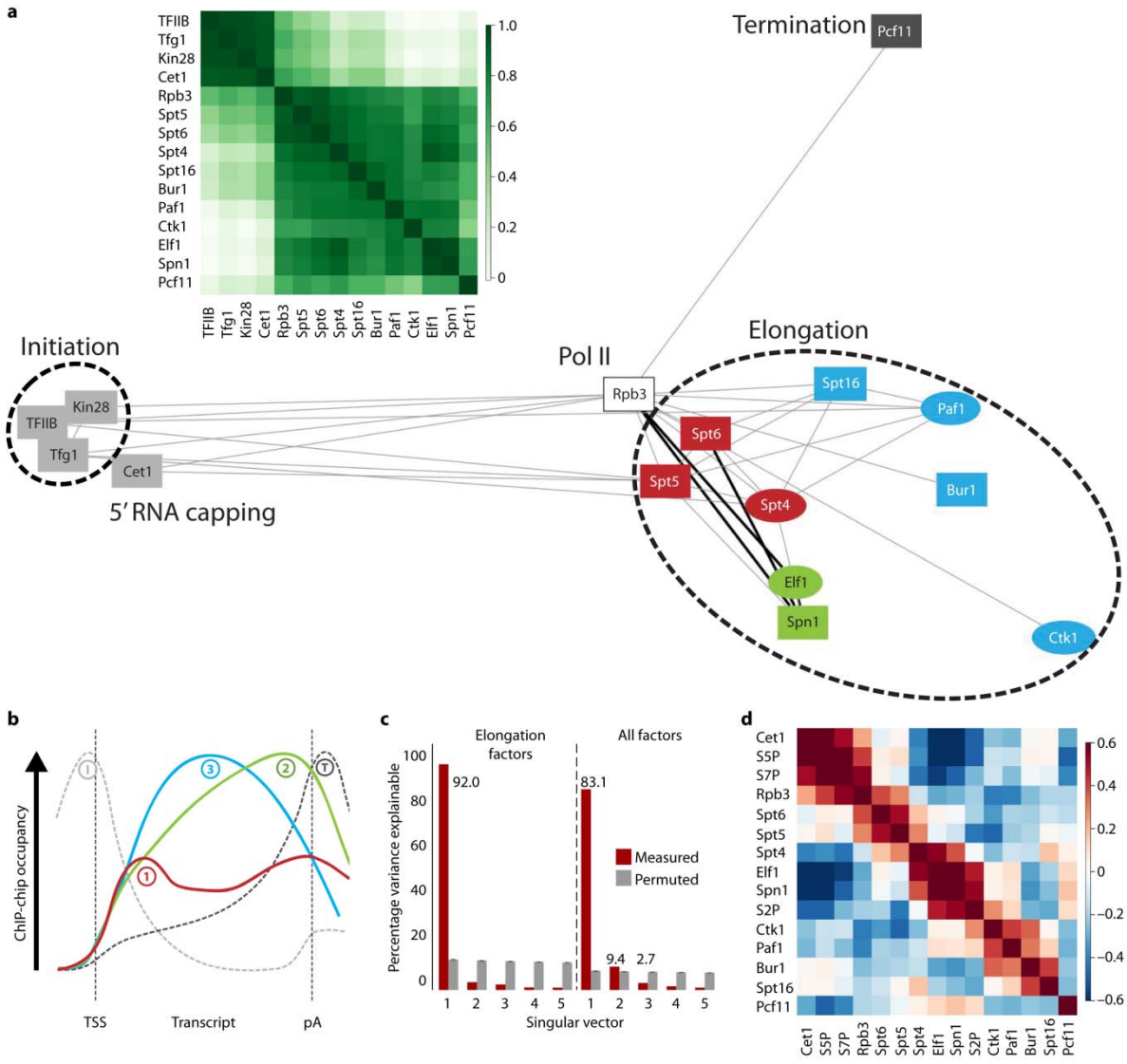

Figure 8: Statistical analysis indicates a general elongation complex. (a) Correlation analysis of genome-wide occupancy profiles. Initiation factors TFIIB, Tfg1 (TFIIF), Kin28 (TFIIH) and the capping enzyme Cet1 have similar profiles that are distinct from those of the nine elongation factors and the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

3.4 Spn1 and Elf1 interact within a Pol II complex

The close resemblance of the gene-averaged profiles of the poorly characterized factors Spn1 and Elf1 suggested that these factors may interact *in vivo* (Figure 6 e and Figure 9 a). To test this, we generated a yeast strain that expressed a C-terminally TAP-tagged version of Spn1 as well as a C-terminally HA-tagged version of Elf1 (see III 2.1). We performed tandem affinity purification (TAP) of Spn1 (see III 2.3). Spn1 TAP led to the co-purification of Elf1, Pol II (Rpb3) and Spt6 (Figure 9 b,c), consistent with an interaction between Spn1 and Elf1. We also performed an Elf1 TAP, which led to co-purification of the two Pol II subunits Rpb1 and Rpb3 (Figure 9 c). Since there were no specific antibodies available that recognize yeast Spn1, we could not test for co-purification of Spn1. These results suggested that Spn1 and Elf1 interact within a Pol II complex, and the profiling data suggests that their function and requirement during the transcription cycle is distinct from that of other elongation factors.

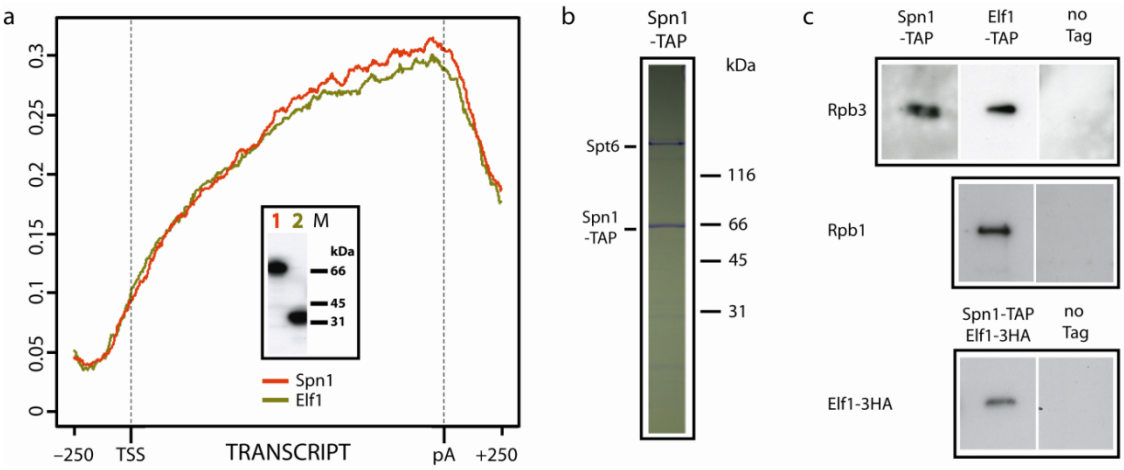

Figure 9: Elf1 co-purifies with Spn1 and Pol II. (a) Gene-averaged occupancy profiles of Elf1 (green) and Spn1 (orange). Western blotting analysis was performed to validate the TAP strains (inset). The results for the Spn1- and Elf1-TAP strains are indicated on lane 1 and 2, respectively. The molecular weight marker is given on lane M. **(b)** Spn1 TAP: the purified protein sample was analyzed by SDS-PAGE and Coomassie staining. Spt6 could be identified by mass spectrometry. The molecular weight marker (kDA) is given on the right. Spn1 copurifies with Spt6. **(c)** Elf1 and Spn1 TAPs: the protein samples were analyzed by SDS-PAGE and Western blotting. The TAP-strains as well as the control strain (with no TAP tag) are given above the panels. The proteins identified by Western blotting are indicated on the left. Elf1 co-purifies with two Pol II subunits Rpb1 and Rpb3. Spn1 copurifies with Rpb3 and Elf1 bearing a C-terminally 3HA tag. For the latter experiment, a yeast strain was generated that expressed both, a C-terminally 3HA-tagged version of Elf1 as well as a C-terminally TAP-tagged version of Spn1.

3.5 Elongation factors exit during a two-step 3′-transition

Around the pA site, two steps of a 3′-transition could be distinguished. The two-step transition was most easily seen at genes with high factor occupancies, such as ribosomal protein genes (Figure 10). Whereas group 3 factor occupancies sharply decreased upstream of the pA site, group 1+2 factors apparently exited further downstream, suggesting they are present during RNA 3′-end formation and/or transcription termination. Spn1 and Elf1 peaked just upstream

of the pA site and dropped only to around 80% of their peak occupancies at 100 nt downstream of the pA site (Figure 6 e and Figure 9 a).

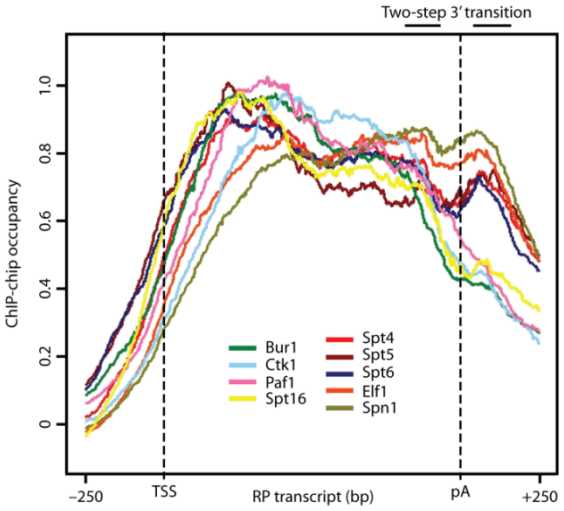

Figure 10: Two-step 3‘ transition observed at ribosomal protein (RP) genes. Shown are averaged elongation factor profiles on selected ribosomal protein genes. Dashed lines indicate the TSS and pA sites. The regions of the two-step 3′ transition are indicated. In contrast to the averaged profiles in Figure 6 and 7, the very high occupancies did not allow us to align profiles with their promoter minima along the *y* axis.

3.6 A general Pol II elongation complex for chromatin transcription

High correlations between elongation factor profiles (Figure 8 and 11) suggested that all elongation factors co-occupy active genes.

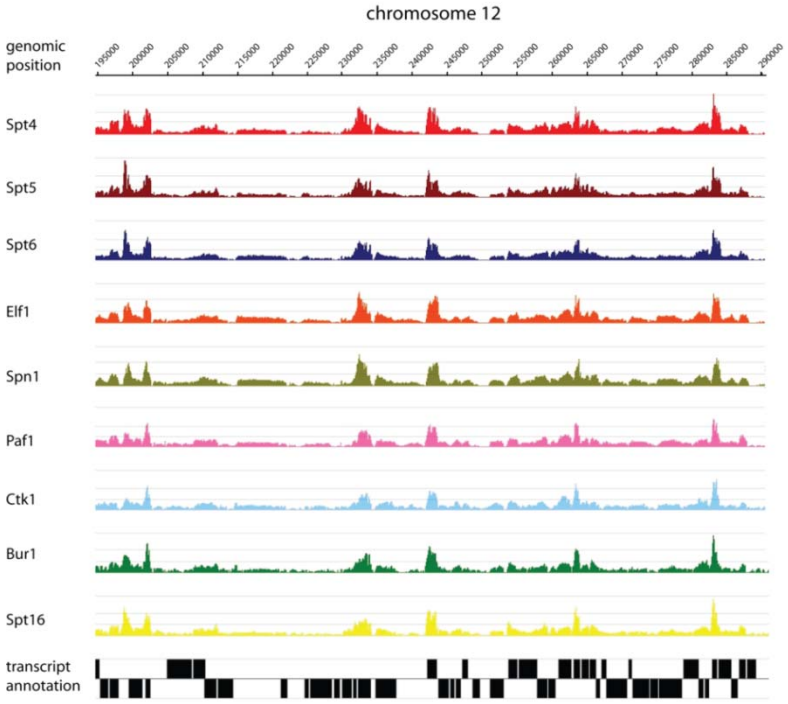

Figure 11: Pol II elongation factors co-localize genome-wide. Elongation factor occupancies are shown for a representative region of yeast chromosome 12. The exact genomic position is given above factor occupancies. The location of transcripts (64) are indicated as black boxes and are indicated below. The transcripts are shown for the Watson (upper line) and Crick strand (lower line). Each bar in the occupancy profile represents the normalized signal from a 150 bp sliding window. Note the high degree of covariation among the 9 occupancy profiles.

To investigate this, we measured co-variation in the data sets by singular value decomposition (SVD). We calculated peak occupancies for nine elongation factors within 4366 genes. After subtracting the row mean of the 9×4366 matrix from each element, the resulting matrix was subjected to SVD. The first singular value, which describes strict co-variation, explained 92% of the total variance (Figure 8 c, left panel). Thus elongation factor occupancies co-varied over all genes, consistent with a general composition of the elongation complex.
The apparent elongation complex composition and coordinated assembly during a 5′-transition were independent of gene length, expression, function, transcript type, size of the nucleosome-depleted promoter region and the presence of introns (Figure 12 and not shown). Although differences in the composition of elongation complexes in individual cells can not be ruled out, these results strongly indicate a general initiation-elongation transition and a general elongation complex composition on Pol II genes.

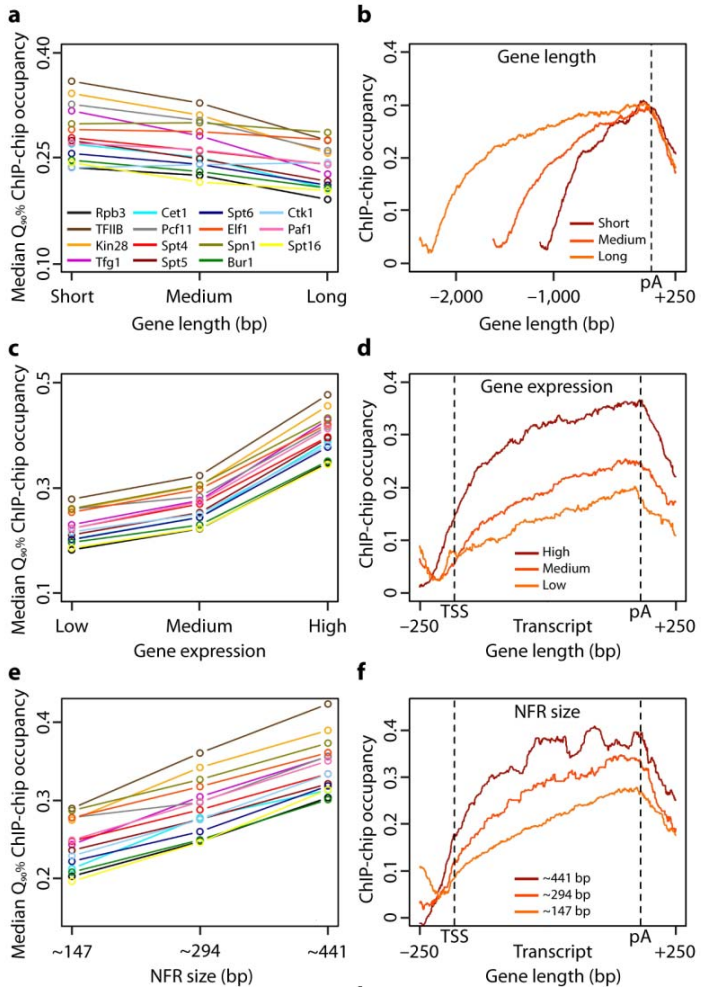

Figure 12: Transcription complex composition and transitions are independent of gene length, expression, and NFR size. (a, c, e) Medians of the peak factor occupancies covary between different length classes **(a)**, expression level classes **(c)**, nucleosome-free promoter region (NFR) size classes **(e)** (123, 335), gene function (gene functions include 213 G1 cell cycle genes, 206 amino acid (aa) synthesis genes, and 421 protein synthesis genes, obtained as transcription modules from Ihmels *et al.* (130)) **(b, d, f)** Gene-averaged profiles of the representative elongation factor Elf1 have shapes and transition points that are independent of gene length **(b)**, expression level **(d)** and NFR size **(f)**. The same holds for all other profiled factors and also for genes grouped by transcript type and functional class (data not shown).

3.7 CTD phosphorylation profiles depend on TSS location

To investigate how the observed profiles and transitions correlate with CTD phosphorylation, we determined occupancy profiles for Pol II phosphorylated at CTD residues Ser7, Ser5, and Ser2, using site-specific antibodies (41) (see Table 8). The averaged profiles reveal broad peaks of Ser7 and Ser5 phosphorylation at around 20 and 120 nt downstream of the TSS (Figure 6 a,b,c). Levels of Ser7/Ser5 phosphorylation decrease over the transcribed region, whereas Ser2 phosphorylation increases, saturates at 600-1000 nt downstream of the TSS, and sharply decrease 100-200 nt downstream of the pA site on all genes (Figure 6 b,c). The point where full Ser2, Ser5 and Ser7 phosphorylation was reached did not depend on pA site location, but rather on TSS location (Figure 13). The point where Ser2 phosphorylation decreased depended on the pA site location (Figure 13, right panel). We note however that changes in ChIP efficiency that are due to the accessibility of the phosphorylated CTD to antibodies may not be ruled out. Taken together, CTD phosphorylation patterns that were observed on individual genes (162, 272) occur globally and depend on TSS location.

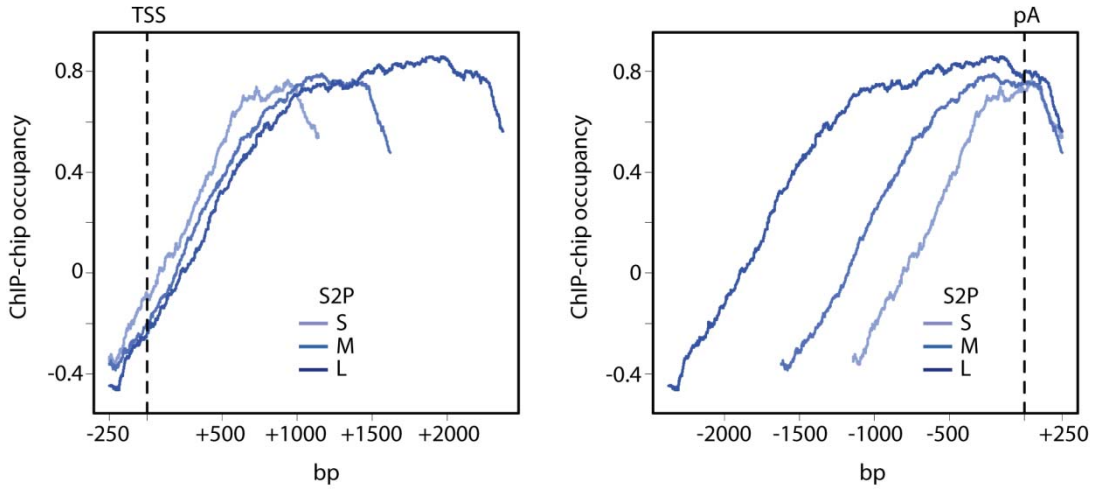

Figure 13: Occupancy levels of Pol II phospho-isoforms depend on the TSS location (exemplified for S2P). Gene-averaged median profiles of the Ser2 phosphorylated form of Pol II are shown for three gene length classes (see section II 2.2.4.3) aligned at TSS (left) and pA sites (right), respectively.

3.8 Recruitment of CTD kinases explains CTD phosphorylations

The Ser7/Ser5 peaks just downstream of the TSS are consistent with the presence of the Ser7/Ser5 kinase Kin28 (4) just upstream of the TSS (Figure 14 a). The early peak of Ser7 phosphorylation is further consistent with dependence of Ser7 phosphorylation on Mediator (30), which binds at promoters. The subsequent increase in Ser2 phosphorylation is consistent with entry of the Ser2 kinases Bur1 and Ctk1 (187, 216) during the 5′-transition, and with their continued presence in transcribed regions (Figure 14 a). Thus specific CTD phosphorylations can be explained by the recruitment of corresponding specific CTD kinases

at defined distances from the TSS. Unfortunately we could not obtain satisfactory profiles for CTD phosphatases to study their recruitment with respect to the observed decreases in CTD phosphorylations.

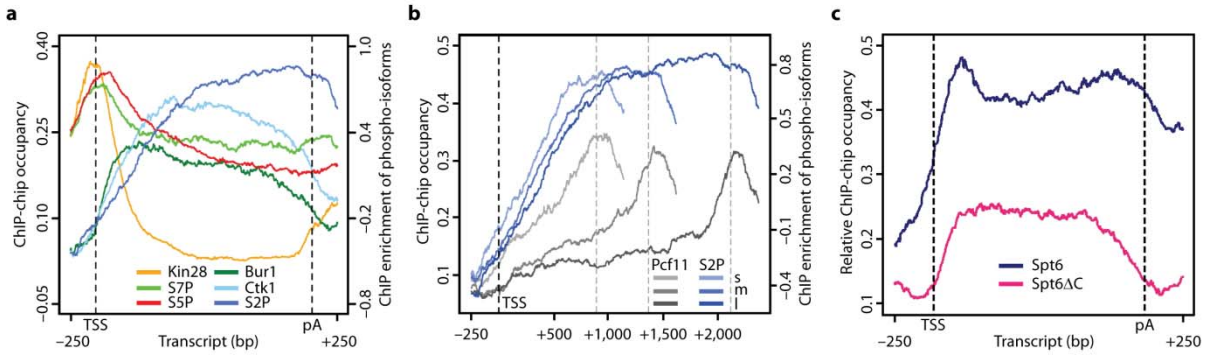

Figure 14: Pol II phosphorylation and factor occupancy. (a) Gene-averaged profiles (for gene length class long (L) only) of CTD phosphorylations, CTD Ser5 kinase Kin28 and Ser2 kinases Bur1 and Ctk1. **(b)** Profiles for Ser2-phosphorylated Pol II and Pcf11 aligned at the TSS and pA site, respectively, and averaged for length classes short (S), medium (M) and long (L). Occupancy and signal intensity for Pcf11 and S2P are plotted on the left and right y axes, respectively. **(c)** Gene-averaged profiles of Spt6 and the C-terminal deletion variant Spt6ΔC (lacking the 202 C-terminal residues) (70). As 100% occupancy levels are not expected for Spt6ΔC, the y axis shows ChIP enrichments obtained by normalization with input measurements as well as mock IPs without scaling to 100% occupancy.

3.9 CTD phosphorylation and factor recruitment

To unravel relations between CTD phosphorylations and factor occupancies, we subjected all profiles to SVD and correlated residual profiles lacking contributions of the first SVD term that described 85.6% of the covariation of factor occupancies. Ser7 and Ser5 phosphorylation correlated with capping enzyme occupancy (Figure 8 d), as expected from binding of capping enzyme to Ser5-phosphorylated CTD *in vitro* (264). Since Ser5 phosphorylation levels peaked over 100 nt downstream from the Cet1 peak, the capping enzyme may bind already when the first Ser5 residues are phosphorylated. Cet1 occupancy drops very sharply further downstream, whereas Ser5 phosphorylation levels remain high, suggesting an active mechanism to release the capping enzyme from the CTD. Ser2 phosphorylation correlated strongly with Spn1 and Elf1 occupancy (Figure 8 d), suggesting these factors are stabilized within the elongation complex by Ser2 phosphorylation.

3.10 Possible Pol II CTD masking and CTD-independent recruitment

CTD Ser2 phosphorylation did not correlate with occupancy of Pcf11 and Spt6, although these factors bind to the Ser2-phosphorylated CTD *in vitro* (15). Pcf11 was recruited mainly at the pA site (Figure 14 b), consistent with the known role of Pcf11 in RNA 3′-processing. This may be explained if the Ser2-phosphorylated CTD only becomes accessible to Pcf11 after pA site passage and 3′-RNA cleavage. Alternatively, Pcf11 crosslinking may be increased by cooperative interactions of factors and RNA around the pA site and/or

conformational changes in the elongation complex. Spt6 entered early, during the 5′-transition, suggesting a CTD Ser2 phosphorylation-independent recruitment mechanism. To investigate this, we determined the occupancy profile of Spt6 lacking its CTD-binding C-terminal domain (Spt6ΔC) with the use of a yeast strain that expresses only a truncated Spt6 variant lacking the last 202 residues (70). Deletion of the Spt6 CTD-binding domain strongly decreased Spt6 recruitment, but did not abolish its entry during the 5′-transition (Figure 14 c and Figure 15). Thus Spt6 is apparently recruited in a CTD-independent manner during the 5′-transition, but full recruitment requires the CTD-binding domain. The CTD-binding domain is required for retaining Spt6 until the pA site (Figure 14 c), consistent with its preference for binding the Ser2-phosphorylated CTD.

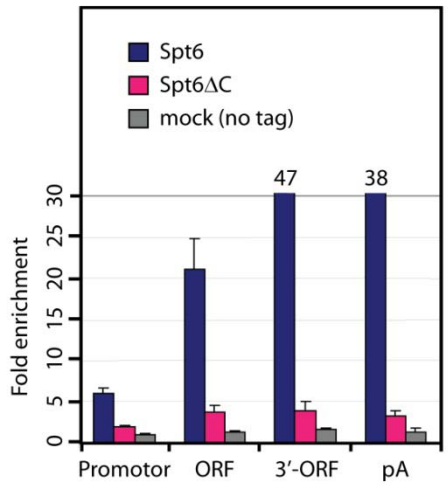

Figure 15: Deletion of the Spt6 C-terminal Pol II CTD-binding domain results in a reduction of fold enrichment at the gene encoding *ADH1*. A TAP-tagged Spt6 wild-type strain and a Spt6 variant strain lacking the 202 C-terminal residues (Spt6ΔC) as well as a BY4741 strain (without tag) were analyzed by ChIP of the *ADH1* gene. Precipitated DNA was used for quantitative PCR amplification with primers directed against the promoter, central and 3′ regions of the ORF, and against the pA site. The fold enrichments for the different positions of the *ADH1* gene over an ORF-free region on chromosome V are given on the y-axis. Error bars show standard deviations for three independent experiments.

These results indicate that binding of a factor to the phosphorylated CTD *in vitro* cannot predict factor recruitment *in vivo*. This suggests that the CTD may be transiently masked and its accessibility may be regulated, and that CTD-independent and CTD phosphorylation type-independent recruitment contributes to factor recruitment.

3.11 No evidence for promoter-proximally stalled Pol II

In higher eukaryotes, Pol II is often stalled early during elongation near the promoter (56, 221, 339), and can be released by activators (257). Our data do not provide evidence for the presence of such promoter-proximally stalled Pol II in growing yeast. Genes with stalled Pol II would show Ser5- but not Ser2-phosphorylation, or at least more Ser5- than Ser2-phosphorylation. However, we do not observe genes that show a peak for Ser5 phosphorylation and at the same time lack a peak for Ser2 phosphorylation (not shown). SVD analysis of initiation and elongation factor profiles revealed a high covariance of 83.1% (Figure 8 c), suggesting that initiation complexes are generally efficiently converted to elongation complexes. Although Rpb3 occupancy peaks around 150 nt downstream of the TSS (Figure 6 c), this does not indicate polymerase stalling, since stalling generally occurs closer to the TSS. Instead, this Pol II peak may be explained by the 5′-transition that is may be

slow due to capping, phosphorylation, and assembly events, leading to an apparent accumulation of Pol II. Alternatively or additionally, the peak may reflect transient pausing of Pol II between the +1 and +2 nucleosomes, which are positioned around 40 and 210 nt downstream of the TSS, respectively (139). Our data obtained in exponentially growing yeast also do not show evidence for polymerase peaks upstream of the TSS as observed in yeast during stationary growth (255).

3.12 General elongation complexes are productive

To investigate whether the general elongation complexes are active on most genes, we correlated averaged Rpb3 and elongation factor occupancies with mRNA levels (70). The mRNA level should be proportional to the mRNA synthesis rate of a single elongation complex times its occupancy, divided by the mRNA decay rate (Figure 16 caption). For constant mRNA synthesis and decay rates we therefore expect a linear dependence between occupancy and mRNA level, corresponding to a slope of 1 in a log-log plot. Indeed we find a robust correlation of 0.65 between the log occupancy and the log mRNA levels (Figure 16 a). However, the slope is only ~0.5. A correlation of 0.71 is obtained when we use the distance-filtered gene set. This shows that increased mRNA levels are not only due to increased elongation complex occupancy, but also due to an increased ratio of mRNA synthesis over decay rates. The same dependence leads to a high correlation of 0.79 between the observed average occupancy and the expected occupancy calculated from the mRNA level (Figure 16 b). These correlations indicate that most general elongation complexes are active in producing mRNA and that gene occupancy with the general elongation complex is a good predictor for gene expression level.

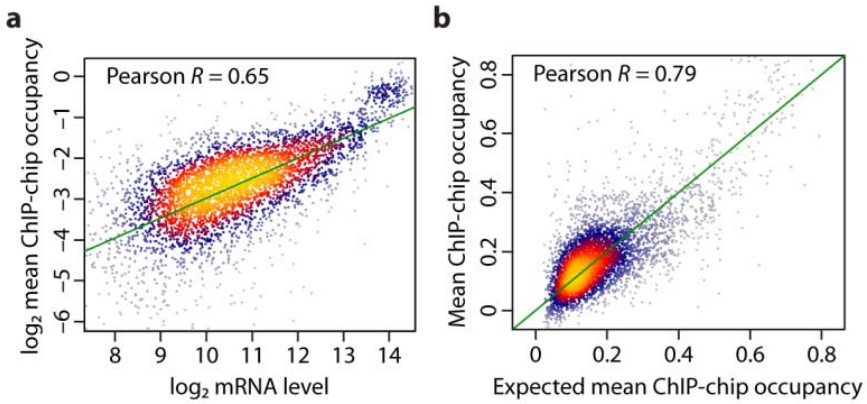

Figure 16: Elongation complex occupancy predicts mRNA expression. (a) The logarithm of the average elongation factor and Rpb3 transcript occupancy is highly correlated with the logarithm of mRNA levels. For constant mRNA synthesis rates (elongation complex speeds) v and decay rates r, we expect a linear relationship between elongation factor occupancy o and mRNA levels c, because at equilibrium the rates of mRNA synthesis and decay are equal, and thus $ov = cr$. This would result in a linear dependence between log o and log c with slope 1: log o = log c + log (r/v). The actual slope of 0.49 (green) implies that the ratio of Pol II speed to decay rate increases slightly with increasing mRNA level ($v/r \propto c^{0.51}$). For ribosomal protein genes, v/r is about three-fold higher than average. **(b)** The averaged transcript occupancy of Rpb3 and the nine elongation factors is highly correlated with the occupancy expected on the basis of the relationship between occupancies and observed mRNA levels in **(a)**.

3.13 Unpublished result: Genome-wide occupancy profiling of TFIIS

Several studies have shown that TFIIS plays a role during initiation and elongation of Pol II transcription (86, 111, 151, 153, 247, 325). However, until now it is not known whether TFIIS is recruited at the promoter of genes and is retained during transcription elongation or whether TFIIS is recruited transiently to elongating Pol II. To address this question *in vivo*, we determined the genome-wide occupancy pattern of TFIIS by ChIP-chip analysis. Since it was shown that the C-terminal region of TFIIS interacts with the Pol II active site and thus is important for triggering the nascent RNA-cleavage activity of Pol II, a 25 kDa C-terminal TAP-tag would interfere with the biological function of TFIIS (151). Therefore, we generated a new yeast strain that expressed an N-terminally TAP-tagged version of TFIIS (see section III 2.1). Phenotyping experiments in the presence of the nucleotide-depleting drug 6-azauracil (6-AU) were conducted for wild-type (wt) yeast (no TAP-tag) and for strains expressing either an N-terminally (N-TAP) or C-terminally TAP-tagged (C-TAP) version of TFIIS. These experiments revealed a slow-growth phenotype for C-TAP but not for the wt and N-TAP TFIIS strain. This indicated that a C-terminal TAP-tag, but not an N-terminal TAP-tag, indeed interferes with the biological function of TFIIS *in vivo*. However, no differences in yeast growth were observed at different temperatures.

To determine the occupancy pattern of TFIIS on a genome-wide level, we performed ChIP-chip analysis for the N-terminally TAP-tagged version of TFIIS. This analysis revealed TFIIS occupancy at the promoter and at the transcribed region of most protein coding genes in yeast (Figure 17 a,b). This suggests that TFIIS, in contrast to the other Pol II elongation factors investigated in this work, is present during transcription initiation and elongation. Furthermore, we also observed TFIIS occupancy at genes that are transcribed by RNA polymerase (Pol) III, such as tRNA genes (Figure 17 c). The latter observation is in accord with a recent study showing that TFIIS is also part of the Pol III transcription machinery (96). Taken together, ChIP-chip experiments revealed that TFIIS localizes to the promoter and to the transcribed region of protein coding genes, what agrees with the view that TFIIS is recruited at the promoter region and stays associated with Pol II during transcription elongation.

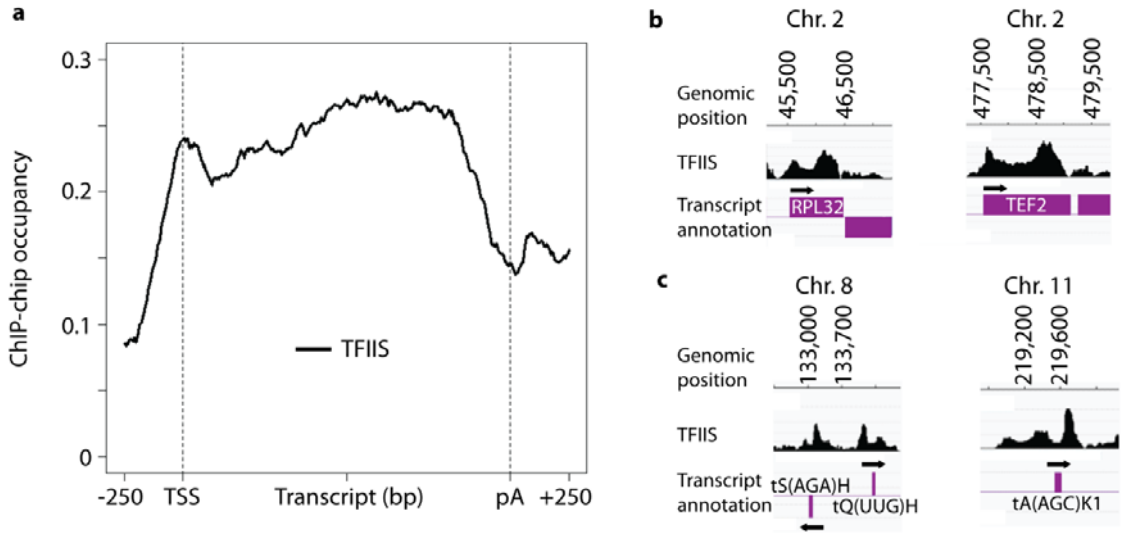

Figure 17: ChIP-chip occupancy of TFIIS at protein coding and tRNA genes. (a) Gene-averaged profile for the median gene length class (1,238 ± 300 nt, 339 genes) of TFIIS. Profiles of other length classes are generally similar (data not shown). **(b)** TFIIS occupancy on selected protein coding genes. The profiles represent normalized TFIIS occupancies smoothed by a 150-nt window running median. Violet boxes indicate transcripts (64) on the Watson (top) and Crick strands (bottom). The direction of transcription is indicated by black arrows. **(c)** TFIIS occupancy on selected tRNA genes that are transcribed by Pol III. Please note the strong TFIIS ChIP-signal at the beginning and at the end of tRNA genes. The profiles are shown as described in **(b)**.

4. Discussion

4.1 The initiation-elongation transition of the Pol II transcription cycle

Published biochemical and genetic data suggest that the 5' transition corresponds to a coordinated conversion of a general initiation complex to a general elongation complex. The conversion includes initiation factor dissociation, which liberates the Pol II clamp domain (43) for binding Spt5 (124). Spt5 could coordinate entry of group 1 factors because it binds Spt4 *in vitro* (112) and associates with Spt6 *in vivo* (184). Group 1 factors could recruit group 2 factors, since Spt6 binds Spn1 (167, 184). Consistently group 1 factors interact genetically with Elf1 (246), and, as shown here, Spn1 and Elf1 interact within a Pol II complex. Recruitment of group 3 factors may commence with CTD Ser5 phosphorylation, which recruits Bur1 (253), which in turn phosphorylates Spt5, thereby recruiting Paf1 (176, 187, 343). Spt16/FACT enters already around 30 nt upstream from other elongation factors (Figure 6 f), perhaps due to binding the +1 nucleosome (293), and consistent with its role as a histone chaperone (23). Initiation factors are not present when the 5' transition is completed around 150 nt downstream of TSS, consistent with a role of Ctk1 in promoting dissociation of initiation factors (2).

4.2 A two-step elongation-termination transition at the 3'-end of Pol II genes

The general two-step 3' transition we observed is consistent with ChIP data obtained at individual genes. The early exit of group 3 factors Paf1, Ctk1, and Bur1 and the continued presence of Spt4/5 and Spt6 have previously been observed (148, 155). Our results, however, show that the reported Spt16 occupancy downstream of the pA site (155) does not occur globally. Also, our observation of peak levels of the *bona fide* 3'-processing factor Pcf11 downstream of the pA challenges the idea of an early loading of 3'-processing factors in the 5'-region of genes (105). We observed continued presence of group 1 and group 2 factors downstream of the pA site, consistent with their reported roles in mRNA 3'-processing (144), mRNA export (333), and in re-establishing chromatin structure after Pol II passage (145).

4.3 The role of Pol II CTD phosphorylation during 5' and 3' transitions

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

4.4 TFIIS co-lovalizes with Pol II at promoter and coding regions of genes

Several studies have shown that TFIIS plays an important role during Pol II transcription initiation and elongation (86, 111, 153, 247, 325). ChIP analysis in yeast revealed that after heat shock, TFIIS localizes to the promoter and coding regions of two heat shock responsive genes *HSP82* and *SSA4* (243). Another ChIP study showed TFIIS binding to the upstream activating sequence (UAS), promoter and coding regions of *GAL1* after galactose induction (247). Since the spatial resolution of factor binding of classical ChIP experiments is very

limited it is still not known (i) whether TFIIS is transiently recruited to Pol II arrest sites during transcription elongation or (ii) whether it is stably associated with the Pol II elongation complex after its initial recruitment. ChIP-chip analysis with our optimized protocol revealed that TFIIS crosslinks to the promoter and transcribed region of genes, what provides strong evidence for the second model of TFIIS recruitment. According to this view, TFIIS is recruited during transcription initiation at the promoter of genes via a direct interaction with the Pol II core (151, 252). This is also in accord with *in vitro* studies showing that TFIIS is part of the Pol II initiation machinery (111, 153, 247). TFIIS stays associated during the initiation-elongation transition and during elongation but dissociates from the Pol II elongation complex upstream of the pA site of genes. Our ChIP-chip data support the idea that TFIIS has not to be recruited from a distant location to arrested Pol II but is rather closely associated with Pol II. Nevertheless, structural changes of the transcription complex have to occur that allow TFFIIS to insert its C-terminal domain III into the active site of Pol II to induce RNA cleavage activity and reactivation of arrested Pol II (45, 151).

5. Conclusions and Outlook

In this work we have established an improved ChIP-chip protocol that allowed us to determine the genome-wide binding behaviors of central components of the Pol II transcription machinery in yeast at a high resolution. Our data support the following view of a productive chromatin transcription cycle. The initiation complex forms ~30–50 nt upstream of the TSS and contains unphosphorylated Pol II and initiation factors such as TFIIB, TFIIF and TFIIH. The complex then scans for the TSS downstream, begins RNA synthesis and triggers RNA 5′ capping, where maximal CTD Ser7 phosphorylation levels occur. Next, the complex is converted into a general elongation complex by exchanging initiation factors for elongation factors. This initiation-elongation transition is completed around 150 nt downstream of the TSS, where CTD Ser5 phosphorylation levels peak. During subsequent elongation, CTD Ser2 phosphorylation increases until it reaches peak levels 600–1,000 nt downstream of the TSS. During a two-step 3′ transition, a group of elongation factors exits upstream of the pA site, whereas another group persists downstream, where it is joined by factors such as Pcf11, resulting in mRNA 3′ processing and transcription termination.

This genome-wide study of Pol II transcription in yeast gave rise to several new questions. As detailed below, two of the main questions were already addressed in follow-up projects and helped to elucidate new aspects of Pol II gene transcription (see section IV “Topic II” and section V “Topic III” of this work). Other questions that have arisen from that study are currently under investigation or will be investigated in future.

One interesting finding of our study was that the Pol II CTD modifications S2P, S5P and S7P occur at all active genes in proliferating yeast, indicating the importance of CTD phosphorylation cycle. However, as this project was completed it was not known whether the CTD in yeast could also be phosphorylated at highly conserved Tyr1 and Thr4 residues, and thus whether a more complicated “CTD code” exists. Based on these considerations a follow-up project was launched to investigate whether Tyr1-phosphorylation exists in yeast and if so, what the biological function of this modification is (section IV “Topic II” of this work).

Furthermore, it is currently not known whether the CTD can be glycosylated in yeast and if so, what the biological function of this new type of CTD modification is. This is currently under investigation.

This study also provided evidence for a mechanism that prevents the premature recruitment of transcription termination factors, such as Pcf11, to elongating Pol II. The mechanism of how the accessibility of the CTD is regulated was not known. This question has already been addressed by a follow-up study (section IV "Topic II" of this work).

Our study revealed that the transcription elongation factor Spt5 co-localizes with Pol II throughout the transcribed region and past the pA-site of genes. This already illustrated that Spt5 is in an excellent position to recruit other factors during transcription. As our study was in progress, it was additionally shown by two other groups that the C-terminal repetitive region (CTR) of Spt5 recruits the Paf1 complex to elongating Pol II in yeast (187, 343). All these observations prompted us to ask whether the Spt5 CTR may act as a general platform to recruit other transcription factors, in a way similar to the Pol II CTD. This question has already been addressed (section V "Topic III" of this work).

The transcription elongation factor Spt6 contains a C-terminal tandem SH2 domain that was shown to bind to Ser2-phosphorylated CTD peptides in vitro (294). This observation led to the suggestion that Spt6 is recruited to Pol II via a direct interaction of its tandem SH2 domain with the Ser2 phosphorylated CTD. However, our ChIP-chip analyses have shown that in the absence of the tandem SH2 domain, Spt6 is still recruited to active genes. What are the alternative recruitment mechanisms of Spt6 that obviously act *in vivo*?

Our study revealed an interaction between the transcription elongation factors Elf1 and Spn1. Consistently, the genome-wide occupancy profiles of both factors are almost identical and show maximum levels upstream of the pA site of genes. What is the biological function of both proteins?

We have also profiled the occupancy of CTD kinases. An important result was that the recruitment of CTD kinases could explain the appearance of CTD phosporylations. However, CTD phosphorylation by CTD kinases represents only one part of the CTD phosphorylation cycle. Another integral part is the dephosphorylation of the CTD by CTD phosphatases. Currently, no genome-wide occupancies of CTD phosphatases are available. Therefore, it is an open question whether the recruitment of CTD phosphatases can explain the decrease in CTD phosphorylation levels. Another interesting question in this context is which phosphatase targets Ser7-phosphorylated CTD?

Our study provides strong evidence for the existence of general Pol II elongation complex that is required for the expression of all genes in proliferating yeast. Although the elongation factors investigated here co-localize to active genes *in vivo*, it is currently not known whether the elongation complex is established in a step-wise manner or whether a preformed complex of factors is recruited at the 5' region of genes. For most of these factors it is also not known whether they interact with the tail-like CTD or the core of Pol II, or with both parts.

In the present ChIP-chip study we not only averaged over thousands of yeast cells but also over different cell cycle stages. Although single-cell ChIP-chip experiments are currently not available and will be extremely difficult to realize, the second issue can be tackled easily. Yeast cells can be arrested at a certain stage of the cell cycle from which the whole culture can be released simultaneously. ChIP-chip using synchronized yeast cells may provide new insights in cell-cycle dependent mechanisms of transcription regulation.

In our genome-wide study presented here, we used ChIP-chip to investigate the DNA binding of components of the Pol II transcription machinery. However, a very fascinating question is whether and if so, how the Pol II transcription machinery interacts with nascent RNA during the transcription cycle. This dynamic binding interface is at the moment almost uncharacterized. New innovative techniques such as NET-seq and PAR-CLIP now open up new ways to tackle this issue.

IV. CTD tyrosine phosphorylation impairs termination factor recruitment to transcribing RNA polymerase II

1. Introduction

1.1 Read-out of the Pol II CTD code

During the transcription cycle the C-terminal repeat domain (CTD) of the largest subunit of Pol II gets phosphorylated and dephosphorylated by CTD kinases and phosphatases (37). This leads to different phosphorylation patterns that predominate at each stage of the transcription. These different phosphorylation states of the CTD were suggested to form a code (36) (for details see section I 2.3.2 and III 1.2). However, a biological meaning of that CTD code only arises when there are at the same time proteins that can "read" the code. Proteins read that code via special domains that recognize a specific CTD phosphorylation pattern (202). CTD binding domains usually make contacts with CTD residues that are part of two consecutive heptapeptide repeats (202). Additionally, it was shown for yeast that insertion of alanine residues between diheptads had little phenotypic effect, while inserting alanines between consecutive heptades turned out to be lethal for yeast cells (289). This led to the suggestion that the functional unit of the Pol II CTD is contained within pairs of heptapeptides (289). CTD-binding proteins mainly comprise factors that are involved in mRNA processing and histone-modification (37). The list of factors that associate with the CTD is currently growing and therefore it seems likely that proteins involved in other processes may also contact the CTD (240).

1.1.1 CTD recognition by Nrd1, Pcf11 and Rtt103

Proteins that were shown to directly bind to particular CTD phosphoisoforms are the Pol II transcription termination factors Nrd1, Pcf11 and Rtt103 (193, 201, 308). All three factors contact the CTD via a conserved CTD interacting domain (CID) (193, 201, 266, 308). The structures of these CIDs are very similar and consist of eight α-helices in a right-handed superhelical arrangement (193). The Nrd1-CID was shown to preferentially bind to Ser5 phosphorylated CTD peptides *in vitro* (308). In case of Pcf11 the CID is essential for cell viability in yeast (266). Pcf11-CID binds to Ser2 phosphorylated CTD mimicking peptides *in vitro* and a co-structure of the Pcf11-CID bound to a S2P CTD peptide reveals the details of its binding interface (201). The S2P CTD adopts a β-turn conformation, when bound to Pcf11-CID. Interestingly, the phosphate group of the CTD residue Ser2 points away from the CID surface and no direct CID contacts were observed (201). Similarly to the Pcf11 and contrary to Nrd1, the Rtt103-CID preferentially associates with S2P CTD peptides *in vitro* (193). The co-structure of the Rtt103-CID bound to a S2P CTD peptide helps to explain the higher affinity of Rtt103-CID for S2P CTD as compared to the Pcf11-CID. In contrast to Pcf11, several residues of the Rtt103-CID directly contact the Ser2 phosphate group of the CTD peptide. Interestingly, the S2P CTD peptide is also in a β-turn conformation, when

bound to the Rtt103-CID (193). Recently, another interesting aspect of CTD binding via CIDs could be revealed. According to this study, CIDs achieve high affinity and specificity for the Pol II CTD by cooperatively binding to neighboring CTD repeats (193). This was shown to be true for the CIDs of Pcf11 and Rtt103 (193).

1.1.2 CTD recognition by Spt6

In yeast there are three proteins that bear a canonical CID (Nrd1, Pcf11, Rtt103) and one protein is predicted to contain a CID (Ctk3; *Saccharomyces* Genome Database (SGD), http://www.yeastgenome.org/). However, more than 100 proteins were known to interact with the Pol II CTD in yeast (240). This implies that other domains must exist that allow proteins to bind the CTD. Structural studies begin to show that CTD binding proteins use a variety of different domains that allow them to read the different phosphorylated states of the CTD (202). One example is the CTD binding of Spt6. Spt6 represents an essential nuclear protein in yeast (296) that is involved in Pol II transcription elongation in a chromatin environment and possesses histone chaperone activity (295). The C-terminal region of mammalian Spt6 that contains a SH2 domain binds to Ser2 phosphorylated CTD *in vitro* (333). Recently, sequence and structural analyses have shown that the C-terminal region of yeast Spt6 possesses a "tandem SH2 domain" that consists of two SH2 domains forming an extended interface (52, 73, 185, 294). The tandem SH2 domain binds preferentially to Ser2 phosphorylated CTD *in vitro* (52, 294). Since no co-structure of the Spt6 tandem SH2-CTD is currently available, the molecular details of this interaction remain to be determined.

1.2 Tyrosine-phosphorylation of the mammalian Pol II CTD

The CTD is modified by phosphorylation at Ser2, Ser5 and Ser7 residues in yeast and mammals (for details see section I 2.3.2 and III 1.3), (37, 41, 157)). In mammals the Pol II CTD can also be phosphorylated and dephosphorylated at at tyrosine residues (19). Kinases that can target tyrosine residues of the mammalian Pol II CTD are c-Abl (19) and Arg (18), but not the c-Src tyrosine kinase (19). The Pol II CTD specificity of c-Abl is conferred by its N-terminal SH2 domain (77) as well as by a C-terminally located CTD binding domain (17). Furthermore, c-Abl can interact with the Pol II CTD *in vitro* and with Pol II *in vivo* (17). Tyrosine phosphorylation of the Pol II CTD can enhance transcription of certain genes (17). Additionally, genotoxic stress such as treatment of mouse cells with the DNA damaging reagent methyl methanesulphonate (MMS) or via an exposure of cells to ionizing radiation leads to an increase of tyrosine phosphorylation of the Pol II CTD by the c-Abl kinase, which in turn is activated via phosphorylation by the ATM kinase (21, 188). Although, the mammalian Pol II CTD represents a substrate for the c-Abl tyrosine kinase, *Drosophila* and yeast CTD are not efficiently phosphorylated by c-Abl *in vitro* (20). The biological role of CTD tyrosine phosphorylation is not known. It is also not known whether CTD tyrosine phosphorylation is restricted to mammals or also exists in lower eukaryotes such as yeast.

1.3 Pol II transcription termination

Accurate termination is important since stopping transcription too late or too early would disrupt proper gene expression. Pol II does not terminate transcription at a specific DNA sequence at the 3' end of genes. Transcription termination sites are rather located in a region up to 1 kb downstream of the poly(A) site (218). Different models of how Pol II may terminate transcription were proposed.

1.3.1 Antitermination versus torpedo model

Based on the mechanism of how Pol II transcription termination may occur at protein coding genes, two major models have been proposed. According to the "anti-terminator" model (also called "allosteric" model) transcription through the poly(A) signal changes the properties of the elongating Pol II complex. This may occurs by dissociation of positive elongation factors or by the recruitment of termination factors (189). The model is supported by chromatin immunoprecipitation (ChIP) experiments showing that some elongation factors are exchanged for termination factors during an elongation-termination transition at the 3' end of genes (155, 199). However, the "torpedo" model proposes that endonucleolytic cleavage of the nascent RNA transcript at the poly(A) site transmits a signal to Pol II that leads to the destabilization of the elongation complex and thus to transcription termination (54). This model is supported by experiments showing that the 5' to 3' exoribonuclease Rat1/Xrn2 degrades RNA from the 5' end that is created by transcript cleavage at the poly(A) site (156, 322). Therefore, Rat1 may act as a torpedo that removes Pol II from DNA. It was also shown that apart from Rat1 other proteins are involved in this termination pathway including Rai1, Rtt103 and yet uncharacterized factors (69, 156). Since both models have experimental support, it seems likely that Pol II transcription termination can occur by more than one mechanism.

1.3.2 Early versus late transcription termination pathway

Apart from the mechanism of transcription termination there is accumulating evidence for gene class specific Pol II termination pathways. In yeast, at least two termination pathways can be distinguished: the "early" and the "late" termination pathway (37, 158, 265). The early termination pathway functions at snRNA, snoRNA genes and at regions coding for cryptic unstable transcripts (CUTs) (37, 287). Termination of these non-coding RNAs requires the Sen1/Nrd1/Nab3 termination complex (287, 308) and occurs within the first few hundred nucleotides of elongation at a region with high levels of Ser5 phosphorylated Pol II CTD (37). The late termination pathway functions at most protein-coding genes. Messenger RNA transcripts terminate downstream of the coding regions in a process that is coupled to RNA cleavage and polyadenylation (63). An obvious question is how Pol II discriminates between both termination pathways? There is no simple answer to that question, but sequences in the nascent RNA seem to play a role. Nrd1 and Nab3 were shown to bind to specific RNA sequences. Those binding sites for Nrd1 and Nab3 may predominantly occur at genes coding for snRNAs, snoRNAs and CUTs. However, it is very likely that other processes also help Pol II to distinguish between both termination pathways, including histone modifications (301).

1.4 Aims and scope of this work

More than ten years ago the laboratories of Steve Buratowski (Harvard Medical School) and David Bentley (University of Colorado) could show that the CTD is mainly phosphorylated at Ser5 and Ser2 residues and that the CTD phosphorylation pattern changes during gene transcription (162, 272). In both studies ChIP experiments with monoclonal antibodies that recognize S5P or S2P CTD were conducted to reveal S5P and S2P peak occupancies at the 5' and 3' regions of genes, respectively (162, 272). Similar observations were made in other organisms including mammals. In the following years plenty of studies have elucidated the binding of particular factors to certain phosphorylated forms of the Pol II CTD and thus have assigned distinct functions to the different phospo-modifications. Structural studies additionally revealed the details of some of these binding interfaces at atomic resolution (202).

In 2007 and 2009, with the availability of a new monoclonal antibody that was generated and validated in the laboratories of Elisabeth Kremmer and Dirk Eick (Helmholtz Zentrum Munich), it could be shown by ChIP experiments that the CTD is also phosphorylated at Ser7 residues in mammals and yeast (41, 157). Similar to S5P, S7P was predominantly detected at the 5' region of genes (157). Recently, we and others have shown that S2P, S5P and S7P occur globally in yeast ((154, 199, 302); section III "Topic I" of this work).

Although it is widely accepted that phosphorylation at the different serine residues of the Pol II CTD plays an important role in eukaryotic gene transcription, it is currently not known whether this is also true for CTD tyrosine phosphorylation. Almost two decades ago, one laboratory described CTD tyrosine phosphorylation in mammals, but its function remained unknown (19). Tyr1 represents the most conserved residue within the CTD consensus repeat and exchanging Tyr1 by Phe residues is lethal for yeast cells (321). This already illustrates how important Tyr1 is and to be more precise, how important its hydroxyl-group is. With this published data in hand we asked whether CTD tyrosine phosphorylation exists in yeast and if so, what the biological role of this type of CTD modification is. We addressed these questions in cooperation with Dirk Eick, Martin Heidemann and Elisabeth Kremmer (Helmholtz Zentrum Munich).

Recently, we observed that CTD Ser2 phosphorylation did not correlate with the occupancy of the Pol II termination factor Pcf11 (199), although this factor binds to Ser2 phosphorylated CTD *in vitro* (15). The maximum occupancy levels of S2P Pol II were reached long before Pcf11 recruitment occurred (see section III 3.10). This observation led us to suggest that accessibility of the CTD is may be transiently masked and its accessibility regulated (section III "Topic I" of this work, (199)). Therefore, we were also interested in the influence of tyrosine phosphorylation on the regulation of the CTD accessibility and CTD masking.

2. Specific procedures

Table 18: List of plasmids used in this study

Plasmid	Insert	Type	Tag	Restriction sites	Source
AM001	*S. cerevisiae* Nrd1-CID (6-151)	pET28b(+)	Ct His	NcoI, NotI	This work
AM002	*S. cerevisiae* Rtt103-CID (1-131)	pET28b(+)	Ct His	NcoI, NotI	This work
213 (Cramer laboratory)	*S. cerevisiae* Pcf11-CID (1-140)	pET28b(+)	Ct His	NcoI, NotI	Anton Meinhart, (201)
MS001	*C. glabrata* Spt6 tandem SH2 domain (1250-1444)	pET28b(+)	Nt His	NdeI, NotI	Mai Sun, (294)
134 (Cramer laboratory)	Human Ssu72 (full length)	pET28b(+)	Ct His	NcoI, NotI	Anton Meinhart, (203)
KS001	*D. melanogaster* Ssu72 (C13D/D144N; full length)	pET15b	Nt His	NdeI, BamHI	Kathrin Schwinghammer

Table 19: Proteins analyzed by ChIP-chip in this study

Protein	Process involved	ChIP protocol
Y1P (Pol II)	Under investigation	For Pol II phospho-isoforms
Nrd1	Pol II transcription termination	For TAP-tagged proteins
Rtt103	Pol II transcription termination	For TAP-tagged proteins

2.1 Generation and characterization of the monoclonal antibody 3D12

The rat monoclonal antibody 3D12 (subclass IgG1) was generated as previously described (41). For immunization we used the CTD-specific phosphopeptide YSPTSPKme2YPSPTSPSC (Peptide Specialty Laboratories GmbH, Heidelberg, Germany) coupled to ovalbumin.

For characterization of specificity, antibodies were analyzed in ELISA experiments using CTD-like peptides with different modification patterns (Peptide Specialty Laboratories GmbH, Heidelberg, Germany) coupled to 96-well maleimide plates (Thermo Fisher Scientific Inc., Rockford, IL USA) as antigen. Peptides were incubated with the monoclonal antibody and biotinylated, subclass-specific antibodies respectively. After incubation with horseradish peroxidase (HRP)-coupled avidin, H_2O_2 and TMB (3,3',5,5'-tetramentylbenzidine) was added. Absorbance of each well was measured at 650 nm after color change and quantitated with an ELISA reader.

Immunoprecipitation for antibody validation was carried out as follows. HeLa cells were washed twice with cold 1x PBS and lysed in 100 µl Ab lysis Buffer per $3x10^6$ cells for 30 min on ice. All samples were sonicated on ice using a BRANSON Sonifier 250 (15 sec on, 15 sec off, 50% duty) and centrifuged at 16,400 rpm (FA-45-24-11 rotor) for 20 min at 4°C. The supernatant was incubated with antibody-coupled protein G-Sepharose beads (2.5 µg of

antibodies for 4 h at 4 °C, followed by 3 washes with 1 ml Ab lysis Buffer) rotating over-night. Beads were washed five times with 1 ml Ab lysis Buffer before continuative experiments.
For IPs from yeast, logarithmically grown cells (OD_{600} = 0.8) were pelleted, washed and lysed in 1 ml Ab lysis Buffer 2. After centrifugation (13,000 rpm, FA-45-24-11 rotor for 30 min at 4°C) the supernatant was added to the antibody-precoupled beads respectively. After incubation over-night, the beads were washed 5 times with Ab lysis Buffer 2 and proteins were boiled off Sepharose beads in Laemmli Buffer for SDS-PAGE.

For the *in vitro* kinase assay, immuno-purified Pol II from whole cell extracts with an antibody recognizing only unphosphorylated CTD (1C7) was used as substrate. 10 µl of the substrate coupled Sepharose G beads were incubated with 40 µl Tyrosine kinase Buffer or Serine kinase Buffer, 200 µM ATP, 1 µg BSA and 200 ng of the recombinant kinase (ProQinase GmbH, Freiburg, Germany) at 30°C for 30 minutes. 10 µl of 6x Laemmli Buffer was added and samples were boiled for 7 minutes at 95°C followed by SDS-PAGE and western analysis.

Corresponding buffers are listed in Table 12.

2.2 ChIP-chip of the Tyr1 Pol II phospho-isoform

ChIP analysis of the Tyr1 Pol II phospho-isoform was conducted as decribed in section II 2.1.2, but with the following modifications. Chromatin immunoprecipitation was performed with 3D12, a monoclonal antibody with strong specificity and affinity for Tyr1-phosphorylated form of the Pol II CTD. Since it was reported that the amount of antibody used influences the occupancy profiles obtained for Pol II phospho-isoforms (157), we carried out ChIP experiments with different amounts of 3D12 antibody before ChIP-chip analyses. The antibody titration ChIP experiments revealed a similar occupancy pattern of the Tyr1 Pol II phospho-isoform at different amounts of antibody tested. These experiments also revealed that 50 µl of 3D12 antibody resulted in high fold enrichments (over control region) by an only minimal dilution of chromatin during immunoprecipitation. This optimized amount of 3D12 was used in all consequent ChIP and ChIP-chip experiments. All the other steps were performed essentially as described in section II 2.1.2 and II 2.2. A list of all factors that were analyzed by ChIP-chip in this study is shown in Table 19.

2.3 Ssu72 depletion in yeast cells

2.3.1 Molecular cloning and validation of a Ssu72-degron yeast strain

Since Ssu72 is essential for viability in yeast, a deletion of *SSU72* gene is not possible. Therefore, Ssu72 had to be inactivated conditionally. We applied the head-inducible degron system to deplete Ssu72 in living yeast cells (268). This strategy was already shown to work successfully for Ssu72 in yeast (119). According to this system a degron cassette is inserted at

the beginning of the open reading frame of interest. This leads to a protein, with the degron tag fused to the N-terminus. The heat-inducible degron tag targets the protein of interest for degradation when the temperature is raised from 24°C to 38°C (268). Therefore, we amplified the degron cassette, consisting of the *kanMX* marker gene, the *CUP1* promoter and the degron tag from the pKL187 plasmid (primers: 109_Degron_Fw; 103_Degron_Rev; Table 6) and used the PCR product to transform wild-type yeast (Pcf11-TAP; BY4741 background). Positive clones were selected on YPDCu+G418 plates (see Table 10).
Yeast strain was validated by PCR to control whether the degron cassette has inserted at the corrected genomic position (primers: 110_ControlA_Fw; 104_ControlB_Rev; 111_ControlC_Fw; 105_ControlD_Rev; Table 6). Furthermore, serial dilutions of wild-type and Ssu72-degron strain were grown on YPDCu+G418 plates at 24°C and 38°C, respectively. Whereas growth of wild-type yeast was observed at 38°C, no growth was detected for the Ssu72-degron strain, consistent with the observation that Ssu72 is essential for yeast cell viability.

2.3.2 ChIP of Pcf11 and the Tyr1 Pol II phospho-isoform under Ssu72 depletion

ChIP experiments were performed as described in section II 2.1, but with the following modifications. Wild-type (Pcf11-TAP strain) and Ssu72-degron yeast cells were grown in 40 ml of YPDCu medium at 24°C from an OD_{600} of 0.1 to 0.2. The cultures were then transferred to 38°C and were grown to an OD_{600} of 0.7. This took 2.5 h and 3.5 h for wild-type and Ssu72 depleted cells, respectively. Cells were harvested by centrifugation and treated essentially as described in section II 2.1.

2.4 Protein sample preparations

The DNA sequence coding for the *S. cerevisiae* **Nrd1-CID** (amino acid residues 6-151; (308)) was synthesized by Mr Gene GmbH (Regensburg) and cloned into a pET28b(+) vector (AM001; Table 18), resulting in an N-terminally hexahistidine tagged protein version. Proteins were expressed in *E. coli* BL21 (DE3) RIL cells in 1 L LB medium for 16 h at 18°C. Cells were harvested, washed with 1x PBS buffer, and lysed by sonication in the presence of Buffer A. The lysate was centrifuged two times at 15,000 rpm (SS34 rotor, Sorvall) for 20 min at 4°C. The supernatant was loaded on a Ni-NTA column (Qiagen). The Ni-NTA column was equilibrated with Buffer B. The column was washed with Buffer B containing 10-50 mM Imidazole, and protein was eluted with Buffer B containing 100-300 mM Imidazole. The eluate fractions were pooled and applied to a cation exchange chromatography (Mono S, GE Healthcare). The column was equilibrated with Buffer C, and Nrd1 CID was eluted with a linear gradient of 20 column volumes from 50 mM to 1 M NaCl. After concentration, the sample was applied to a Superose 12 size exclusion column (GE Healthcare) equilibrated with Buffer D. Peak fractions were pooled and concentrated to 4-5 mg/ml as determined by ND-1000 Spectrophotometer (NanoDrop Technologies).

The plasmid 213 (Patrick Cramer laboratory; pET28b(+) based; originally generated by Anton Meinhart; Table 18), coding for N-terminally hexahistidine tagged **Pcf11-CID** (amino acid residues 1-140 (201); *S. cerevisiae*), was used for recombinant expression. Proteins were expressed in *E. coli* BL21 (DE3) RIL cells in 1 L LB medium for 16 h at 18°C. Cells were harvested, washed with 1x PBS buffer, and lysed by sonication in the presence of Buffer A. The lysate was centrifuged two times at 15,000 rpm (SS34 rotor, Sorvall) for 20 min at 4°C. The supernatant was loaded on a Ni-NTA column (Qiagen). The Ni-NTA column was equilibrated with Buffer E. The column was washed with Buffer E containing 10-30 mM Imidazole, and protein was eluted with buffer E containing 50-100 mM Imidazole. The eluate was dialysed against 2 L of Buffer F at 4°C over night. Pcf11 CID was purified by anion exchange chromatography (Mono Q, GE Healthcare). The column was equilibrated with Buffer G and Pcf11 CID was eluted with a linear gradient of 20 column volumes from 30 mM to 600 mM NaCl. After concentration, the sample was applied to a Superose 12 size exclusion column (GE Healthcare) equilibrated with buffer D. Pooled peak fractions were concentrated to 5 mg/ml as determined by ND-1000 Spectrophotometer (NanoDrop Technologies).

The DNA sequence coding for the *S. cerevisiae* **Rtt103-CID** (amino acid residues 1-131; (193)) was synthesized by Mr Gene GmbH (Regensburg) and cloned into a pET28b(+) vector (AM002; Table 18), resulting in an N-terminally hexahistidine tagged protein version. The expression and protein purification was performed as described for Nrd1-CID, but with the following modifications. The eluate of the Ni-NTA affinity chromatography was concentrated and applied to a Superdex 75 size exclusion column (GE Healthcare) equilibrated with Buffer D. Pooled peak fractions were concentrated to 5 mg/ml as determined by ND-1000 Spectrophotometer (NanoDrop Technologies).

The plasmid MS001 (Patrick Cramer laboratory; pET28b(+) based; generated by Mai Sun, (294); Table 18), coding for N-terminally hexahistidine tagged **Spt6 tandem SH2 domain** (amino acid residues 1250-1444 (294); *Candida glabrata*), was used for recombinant expression as described for Nrd1-CID. The recombinantly expressed Spt6 tandem SH2 domain was purified essentially as described in (294). The purified protein was concentrated to 10 mg/ml as determined by ND-1000 Spectrophotometer (NanoDrop Technologies) and stored in Buffer H.

The plasmid 134 (Patrick Cramer laboratory; pET28b(+) based; generated by Anton Meinhart; (203); Table 18) coding for N-terminally hexahistidine tagged **human Ssu72** was expressed in *E. coli* BL21 (DE3) RIL cells in 1 L LB medium for 16 h at 18°C. The recombinantly expressed Ssu72 was purified as described in (203). Briefly, cells were harvested, washed with 1x PBS buffer, and lysed by sonication in the presence of Buffer H. The lysate was centrifuged two times at 15,000 rpm (SA600 rotor, Sorvall) for 20 min at 4°C. The supernatant was loaded on a Ni-NTA column (Qiagen), pre-equilibrated with Buffer A (without PI). The column was washed with Buffer A (without PI) containing 10-30 mM Imidazole, and protein was eluted with Buffer A (without PI) containing 50-100 mM Imidazole. The eluates were diluted with 8 volumes Buffer J and loaded onto a Mono Q anion exchange column (GE Healthcare). The column was equilibrated with Buffer J and protein was eluted with a linear gradient of 25 column volumes from 50 mM to 1 M NaCl. After protein concentration, the sample was applied to a Superose 12 size exclusion column (GE

Healthcare) equilibrated with Buffer K. Pooled peak fractions were concentrated to 5 mg/ml as determined by ND-1000 Spectrophotometer (NanoDrop Technologies). Ssu72 phosphatase activity was measured essentially as described in (203).

The DNA sequence coding for a mutated version of *D. melanogaster* **Ssu72 (C13D/D144N**, according to (319)) was synthesized by Mr Gene GmbH (Regensburg) and cloned into a pET15b vector using the restriction sites NdeI and BamHI (Table 18). This resulted in an N-terminally hexahistidine tagged protein version. Proteins were expressed in *E. coli* BL21 (DE3) RIL cells in 1 L LB medium for 18 h at 20°C, harvested by centrifugation, resuspended in Buffer L and lysed by sonification. The lysate was cleared by centrifugation at 15,000 rpm (SS34 rotor, Sorvall). The supernatant was loaded onto a Ni-NTA column equilibrated with Buffer M and washed with buffer M containing 20 mM imidazole. Protein was eluted with Buffer M containing 250 mM imidazole. The N-terminal hexahistidine tag was cleaved by dialysis in Buffer N in the presence of Thrombin (1 U per mg protein). Uncleaved protein was removed by a second Ni-NTA affinity chromatography. The eluate was loaded onto a Superose 12 10/300 column (GE Healthcare) equilibrated with Buffer O. Pooled peak fractions were concentrated to 10 mg/mL, flash frozen in liquid nitrogen and stored at -80°C.

All buffers used for protein purification are listed in Table 12 and 13.

2.5 Pol II CTD phosphopeptide interaction assay

For fluorescence anisotropy measurements, we used a HORIBA Jobin-Yvon Fluoromax 3 spectrofluorometer. The Pol II CTD mimicking phospho-peptides were synthesized by Peptide Specialty Laboratories (PSL) GmbH (Heidelberg) and PANATecs GmbH (Tübingen). An overview of all CTD-peptides used in this work is given in Table 20. The CTD-peptides were N-terminally labeled with fluorescein aminocaproic acid. 500 nM of the CTD-peptide was dissolved in FluA Buffer (150 mM NaCl) and was titrated at 20°C with pure yeast Nrd1-CID, Pcf11-CID, Rtt103-CID or Ssu72 (C13D/D144N), respectively. For Pcf11-CID measurements were also performed in FluB Buffer (10 mM NaCl). For the Spt6 tandem SH2 domain measurements were performed in FluB Buffer only. However, analyses were generally conducted at 150 mM NaCl to avoid protein precipitation and to reduce the risk of artificial protein interactions. The fluorescence anisotropy was measured after 2 min when equilibrium was reached. Data were fit to 1:1 binding models using non-linear least squares algorithm (Hill fit) as implemented in the software package OriginPro 8G (OriginLab Corporation, Northampton). Corresponding buffers are listed in Table 14.

Table 20: Synthetic CTD peptides used in this study

Name	Amino acid sequence	Company
No-P	$P_6\ S_7\ Y_1\ S_2\ P_3\ T_4\ S_5\ P_6\ S_7\ Y_1\ S_2\ P_3\ T_4\ S_5\ P_6\ S_7$	PANATecs
1,1*	$P_6\ S_7$ **Y_1** $S_2\ P_3\ T_4\ S_5\ P_6\ S_7$ **Y_1** $S_2\ P_3\ T_4\ S_5\ P_6\ S_7$	PSL
2,2	$P_6\ S_7\ Y_1$ **S_2** $P_3\ T_4\ S_5\ P_6\ S_7\ Y_1$ **S_2** $P_3\ T_4\ S_5\ P_6\ S_7$	PANATecs
5,5*	$P_6\ S_7\ Y_1\ S_2\ P_3\ T_4$ **S_5** $P_6\ S_7\ Y_1\ S_2\ P_3\ T_4$ **S_5** $P_6\ S_7$	PSL
1,2,1,2*	$P_6\ S_7$ **Y_1** **S_2** $P_3\ T_4\ S_5\ P_6\ S_7$ **Y_1** **S_2** $P_3\ T_4\ S_5\ P_6\ S_7$	PSL
1,5,1,5	$P_6\ S_7$ **Y_1** $S_2\ P_3\ T_4$ **S_5** $P_6\ S_7$ **Y_1** $S_2\ P_3\ T_4$ **S_5** $P_6\ S_7$	PSL
1*	$P_3\ T_4\ S_5\ P_6\ S_7$ **Y_1** $S_2\ P_3\ T_4\ S_5$	PSL
1,2*	$P_3\ T_4\ S_5\ P_6\ S_7$ **Y_1** **S_2** $P_3\ T_4\ S_5$	PSL
7*	$Y_1\ S_2\ P_3\ T_4\ S_5\ P_6$ **S_7** $Y_1\ S_2\ P_3\ T_4\ S_5\ P_6\ S_7\ C$	PSL

■ = Phosphorylated residue
* = Peptides used in the Pol II CTD phosphatase assay

2.6 Pol II CTD phosphatase assay

Ssu72 phosphatase activity was quantified with the Malachite Green Phosphate Assay Kit (BioAssay Systems) that measures the formation of a green complex formed between Malachite Green, molybdate and free orthophosphate. The assay applicability was approved by the cleavage of the 5' phosphate group of a random oligonucleotide by the antarctic phosphatase (New England Biolabs) (data not shown). The assay was performed according to the instructions of the manufacturer. Briefly, 80 µl of each sample were incubated at 37°C and contained 50 nM human Ssu72 and 20 µM CTD peptide or orthophosphorylated aminoacids (o-phospho-L-serine; o-phospho-L-threonine; o-phospho-L-tyrosine; SIGMA-ALDRICH GmbH, Schnelldorf, Germany) in Mal Buffer (Table 14). Except of the tyrosine phosphorylated CTD peptide ("1,1", Table 20), the CTD peptides did not possess a fluorescein label. Substrate compounds used in this assay were dissolved in water. Reactions were stopped by the addition of 20 µL kit solution in a 96 well plate (286). Complex formation was measured via the absorption at 620 nm using an Infinite®M1000 (Tecan). As a control for each reaction one sample was prepared without enzyme and treated as the long-term sample.

3. Results and discussion

3.1 A monoclonal antibody that recognizes tyrosine phosphorylated Pol II CTD in yeast and mammals

Chromatin immunoprecipitation (ChIP) with antibodies that recognize different phosphorylated forms of the Pol II CTD have revealed that the CTD is phosphorylated at Ser2 (S2P), Ser5 (S5P), Ser7 (S7P) and Thr4 *in vivo* (128, 154, 157, 162, 199, 272, 302). Although it is now accepted that the Pol II CTD is phosphorylated at serine and threonine residues *in vivo*, it is currently not known whether the CTD is also phosphorylated at tyrosine residues in

yeast. To address this question we generated a monoclonal antibody (3D12) that recognizes CTD heptads that are phosphorylated at Tyr1 (Y1P).
Previous results have shown, that the functional unit of the Pol II CTD lies in between a pair of heptapeptides (186, 314). Therefore, a panel of di-heptapeptides bearing various modification patterns was used (Figure 1a) to characterize the binding epitope and specificity of our antibody. The analysis revealed that some modifications on neighboring residues prevent Y1P recognition. For example, the binding of α-Tyr1-P antibody 3D12 interferes with upstream S7P, but is not affected by downstream S2P (Figure 18 a). To show that tyrosine phosphorylation of Pol II CTD occurs *in vivo*, *Saccharomyces cerevisiae* (*S. cerevisiae*) whole cell extracts were analyzed in immunoprecipitation and immunoblotting experiments (Figure 18 b). The data reveals phosphorylation of Tyr1 *in vivo* and immunoprecipitation of RNA Pol II by the monoclonal antibody 3D12. To further confirm the Y1P specificity of monoclonal antibody 3D12, unphosphorylated Pol IIA form was purified from HeLa whole cell extracts as substrate for the mammalian CTD-tyrosine-kinase c-Abl (20). Phosphorylation of Pol IIA by c-Abl leads to a strong signal for monoclonal antibody 3D12 while addition of reaction buffer and ATP alone was not sufficient (Figure 18 c). The dramatic increase of Y1P additionally results in a shift of Pol IIA into the hyperphosphorylated Pol II0 form. Taken together, these experiments show that the antibody 3D12 specifically recognizes the Tyr1-phosphorylated CTD and that Tyr1 phosphorylation occurs in yeast.
This work was performed in cooperation with Dirk Eick, Martin Heidemann and Elisabeth Kremmer (Helmholtz Zentrum Munich).

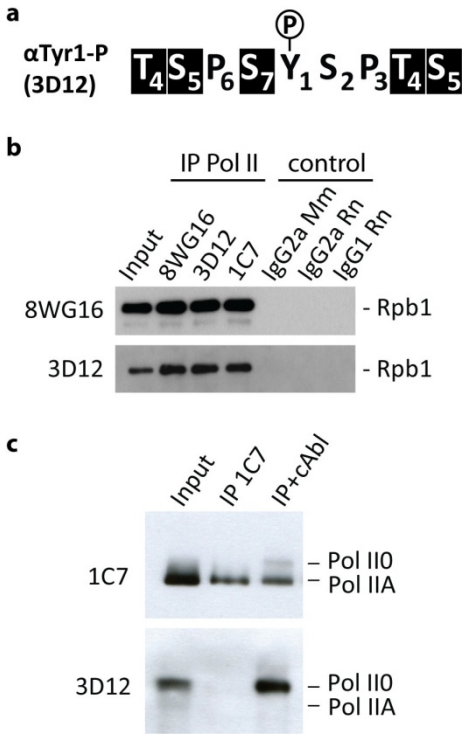

Figure 18: The Pol II CTD is phosphorylated on Tyr1-residues in proliferating yeast. (a) Specificity of the 3D12 antibody. Phosphate groups on black shaded residues interfere with the binding of the 3D12 antibody. **(b)** Western blot analysis with yeast whole cell extract reveals CTD Tyr1 phosphorylation. Whole cell extract prepared from proliferating yeast was either directly probed by 8WG16 or 3D12 antibodies (Input) or used for immunoprecipitations of Pol II and then analyzed by Western blotting ("IP Pol II"). The corresponding isotype controls are displayed ("control"). **(c)** Western blot analysis with HeLa whole cell extract also detects CTD Tyr1 phosphorylation (Input). 1C7 antibody immunoprecipitates Pol II that is unphosphorylated at its CTD. Consistently, no signal was detected with 3D12 antibody (lane "IP 1C7"). However, incubation of immunoprecipitated (1C7) Pol II with cAbl tyrosine kinase led to a very strong signal in case of 3D12 antibody (lane "IP+cAbl"). The corresponding heights for the hyperphosphorylated form (II0) and the hypophosphorylated form (IIA) of Pol II are indicated on the right.

3.2 The CTD of genome-associated Pol II is phosphorylated on Tyr1 (Y1P) residues

To investigate whether genome-associated Pol II can be phosphorylated at CTD Tyr1 residues, we carried out occupancy profiling by ChIP experiments coupled to high-resolution microarray analysis in proliferating yeast as described (199). Data from two biological replicates (R = 0.94) were averaged and showed strong signals over protein-coding gene regions. We next averaged occupancy profiles of genes of similar lengths (199). The resulting gene-averaged profiles revealed that Tyr1-phosphorylated Pol II occupies the coding region of protein-coding genes (Figure 19 a,b). Whereas at promoter regions Y1P occupancy levels were lowest, the signal increased downstream of the TSS towards the 3'-end of the coding region, with the highest levels in a region 200-600 nt upstream of the pA site (Figure 19 a,b). The profile resembled that of Ser2-phosphorylated Pol II (154, 199, 302), with two striking differences. Firstly, maximal Y1P occupancy levels were reached earlier than those of S2P. Secondly, Y1P levels started to decrease ~180 nucleotides upstream of the pA site, whereas S2P occupancy levels decreased ~200 nucleotides downstream of that site (Figure 19 a,b).

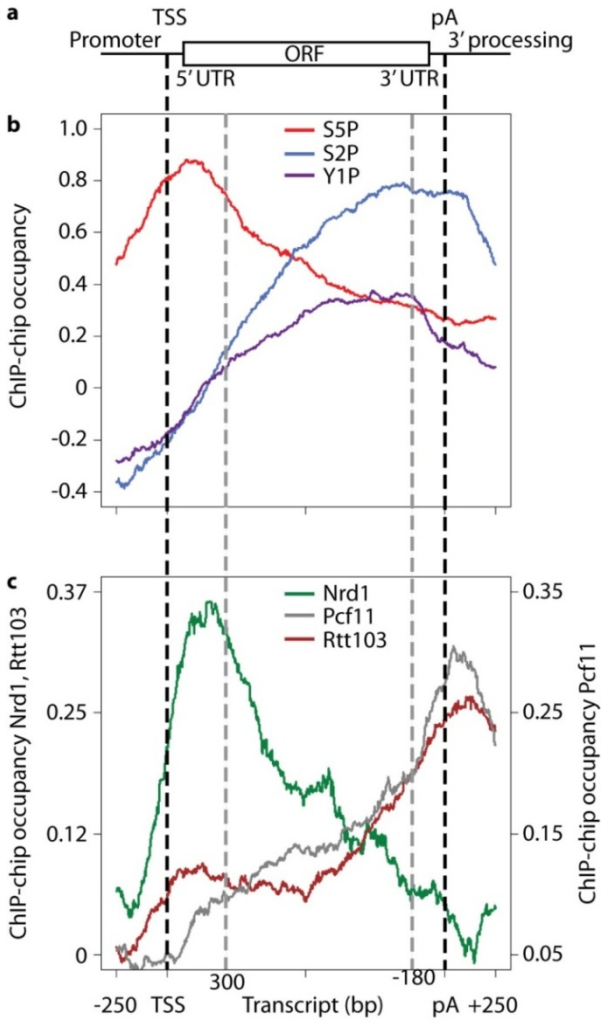

Figure 19: Genome-wide occupancy profiling of Pol II phospho-isoforms and CTD binding termination factors. (a) DNA frame with promoter, 5' untranslated region (UTR), open reading frame (ORF) and 3' UTR. Dashed black lines indicate the TSS and pA site. Dashed gray lines mark the positions 300 nt downstream of the TSS and 180 nt upstream of the pA site. **(b)** Gene-averaged profiles for the "medium" gene-length class (1,238 ± 300 nt, 339 genes) of the Ser5- ("Topic I" of this work, (199)), Ser2- ("Topic I" of this work, (199)) and Tyr1-phosphorylated form of Pol II. Profiles of other length classes are similar (data not shown). **(c)** Gene-averaged profiles as in **(b)** for Pol II termination factors Nrd1, Pcf11 ("Topic I" of this work, (199)) and Rtt103. All three termination factors contain a CTD-interacting domain (CID) that were shown to directly contact particular phosphorylated forms of the Pol II CTD (193, 201, 308) ChIP-chip occupancy of Nrd1 and Rtt103 is on the left y-axis, the occupancy of Pcf11 is on the right y-axis.

The point of Y1P level increase was dependent on the TSS, whereas the point of decrease was dependent on the pA site (Figure 20), and not on gene length or expression level (Figure 20 Figure 22 b), arguing that the Tyr1 phosphorylation mark is set and removed within a general Pol II transcription cycle.

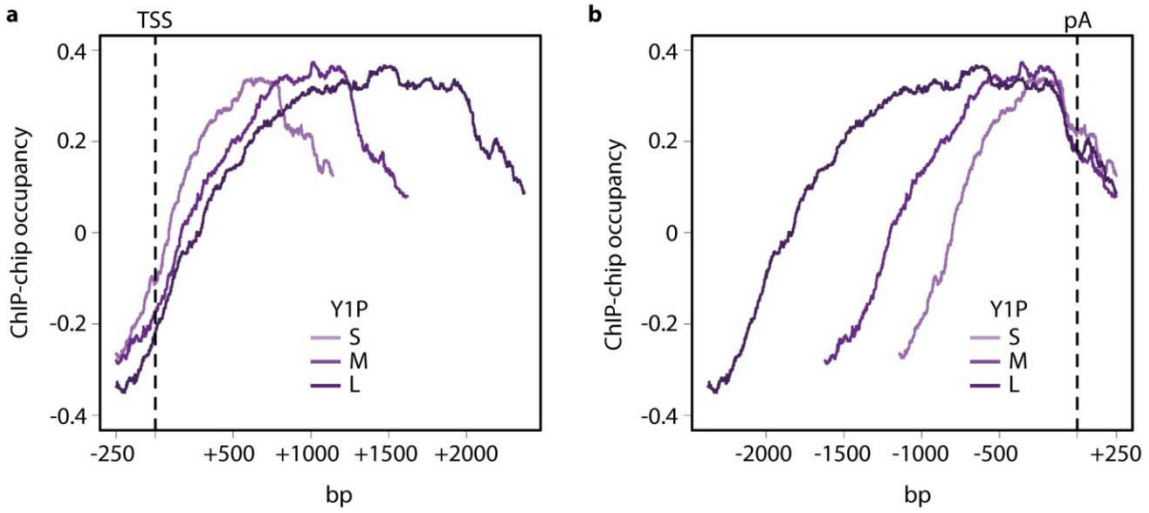

Figure 20: CTD Tyr1 phosphorylation occurs within a general Pol II transcription cycle. (a) Gene-averaged profiles of the Tyr1 Pol II phospho-isoform of three gene length classes ("small": 725 ± 213 nt, 266 genes; "medium": 1,238 ± 300 nt, 339 genes; "long": 2,217 ± 679 nt, 299 genes;) aligned at TSS. The point where Tyr1 phosphorylation levels start to increase depends on the distance to the TSS. **(b)** Gene-averaged profiles of the Tyr1 Pol II phospho-isoform as in **(a)**, but aligned at the pA site. The location where Tyr1 phosphorylation starts to decrease depends on the distance to the pA site.

3.3 Y1P is a mark for all active genes in proliferating yeast

Recently, we and others have shown that S2P and S5P represent general types of CTD modifications that were detected at all actively transcribed Pol II genes in proliferating yeast (154, 199, 302). To test whether Pol II phosphorylation at Tyr1 occurs also on all transcribed genes, we measured co-variation in the occupancy data for Tyr1-phosphorylated Pol II and the other phosphorylated forms of Pol II (199) by singular value decomposition (SVD) as described (199). The first singular value describes strict co-variation and explained 83.8% of the total variance (Figure 21). Thus, Pol II phospho-isoform occupancies co-varied over all genes, indicating that Tyr1 phosphorylation is not a gene-specific or rare modification, but rather comparable in occurrence to Ser2 and Ser5 phosphorylation, which have clearly defined functional roles.

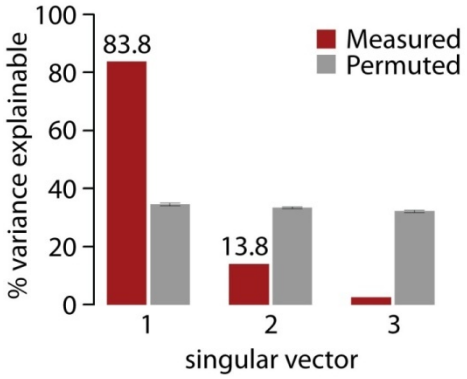

Figure 21: Tyr1 phosphorylation is a general type of CTD modification. SVD analysis of S2P, S5P and Y1P occupancies. The contributions of the first three singular vectors to the variance (red) are shown in comparison to a control with randomly permuted matrix elements (gray). SVD reveals that 83.8% of the variance is explained by covariation.

Consistent with this observation, there is a robust correlation of 0.65 between the occupancy of Tyr1-phosphorylated Pol II and mRNA expression levels (70) (Figure 22 a). Interestingly, the correlation was as high as detected between components of the general Pol II elongation complex and mRNA levels (199), also indicating that Y1P is a mark of active genes. Next, we grouped yeast genes based on their mRNA levels (70) in three different gene expression classes ("high", "medium" and "low") and determined the corresponding Y1P occupancies. As illustrated in Figure 22 b, this analysis revealed that the higher the expression level of a gene, the stronger was the corresponding Y1P occupancy signal.
Thus, Tyr1 phosphorylation similar to S2P and S5P represents a general CTD modification that is associated with active genes in proliferating yeast.

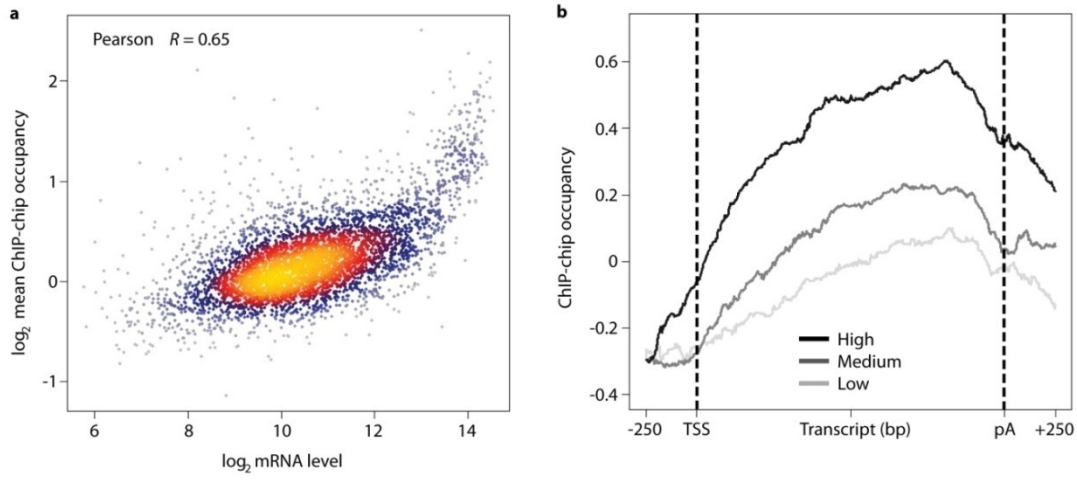

Figure 22: Tyr1 phosphorylation of the CTD is a mark of actively transcribed genes in proliferating yeast. (a) Correlation analysis between the logarithm of the average Y1P occupancy and the logarithm of mRNA levels. **(b)** Gene-averaged profiles for genes in three different mRNA expression level classes. The genes were partitioned into three groups: low (25-50% quantile), medium (50-75% quantile) and high (>75% quantile) expression according to data from (70). From this set of genes, those with ORF lengths between 938 and 1538 were selected.

3.4 Termination factors localize to gene regions with low Y1P levels

To investigate whether and how Tyr1 phosphorylation could influence the recruitment of factors that bind the phosphorylated CTD, we determined genome-wide occupancy profiles of the three yeast proteins that contain CTD-interacting domains (CIDs), Nrd1, Pcf11, and Rtt103, which are all factors involved in transcription termination. The genome-wide occupancy profile of Rtt103 showed peak occupancy in the region downstream of the pA site, resembling the previously determined profile for Pcf11 (199), whereas Nrd1 showed peak occupancy at the beginning of the transcribed region, with peak levels ~150 nt downstream of the TSS in protein-coding genes (Figure 19 c). Comparison with the occupancies of the different phosphorylated forms of Pol II revealed an Nrd1 peak at a region with maximal levels of S5P and very low levels of Y1P and S2P. Consistently, correlation analysis could detect a correlation between Nrd1 and S5P occupancy levels (R = 0.6) but not for S2P (R = 0.1) and only a weak correlation for Y1P (R = 0.4). This agrees with the previous observation that Nrd1 can bind to Ser5-phosphorylated CTD *in vitro* (308). Strong Pcf11 and Rtt103 occupancy peaks were detected at the region immediately downstream of the pA site where the maximum difference between Y1P and S2P signals occurs (Figure 19 c). Consistently, the occupancies of Pcf11 and Rtt103 correlate only weakly with Ser2 phosphorylation (R = 0.4, for both), although they are both known to bind the Ser2-phosphorylated CTD *in vitro* (193, 201). This observation suggested that the Ser2-phosphorylated CTD may be masked by an unknown mechanism within coding regions and we now hypothesized that this masking is achieved by Tyr1 phosphorylation. This would offer an explanation for why Pcf11 and Rtt103 are primarily recruited downstream of the pA site, although maximal S2P levels were reached long before the pA site.

3.5 Y1P impairs CID-CTD interactions *in vitro*

To investigate whether Tyr1 phosphorylation would prevent the direct interaction of the CID-containing termination factors with the CTD, we determined the affinity of the three CID domains for various CTD diheptad phosphopeptides (Figure 23). We purified recombinant CIDs of yeast Nrd1 (residues 6-151), Pcf11 (residues 1-140), and Rtt103 (residues 1-131), and determined their binding affinity to CTD peptides (Table 20) by fluorescence anisotropy. None of the CID domains bound to unphosphorylated CTD (Figure 23, black curves). This is in contrast to a previous report showing that Pcf11-CID can also bind to unphosphorylated CTD (15). Consistent with previous results (193, 308), the Pcf11-CID and the Rtt103-CID bind to the Ser2-phosphorylated CTD peptide with dissociation constants of 54±6 µM and 12±2 µM, respectively, whereas the Nrd1-CID binds to both Ser2- and Ser5-phosphorylated CTD peptides, with slightly higher affinity for Ser5-phosphorylated form (K_D = 85±25 µM, compared to 131±15 µM for the Ser2-phosphorylated peptide) (Figure 23). However, Nrd1-CID, Pcf11-CID and Rtt103-CID cannot bind CTD peptides that are phosphorylated on Tyr1. Furthermore, Nrd1-CID cannot bind to Ser2- and Ser5-phosphorylated CTD peptides if they are additionally phosphorylated at Tyr1 (Figure 23 c). Similarly, Pcf11- and Rtt103-CID can also not bind to Ser2-phosphorylated CTD if additionally phosphorylated at Tyr1 residues.

Thus, Tyr1 phosphorylation prevents CTD binding of all CIDs in yeast, consistent with the idea that Tyr1 phosphorylation masks the Ser2- and Ser5-phosphorylated CTD.
Although Pcf11 is highly conserved from yeast to human, Nrd1 and Rtt103 have no clear homologues in humans. Nevertheless, the human genome encodes a protein, SCAF8, that shares sequence homology to yeast Nrd1 in the CID and the RRM domains and has a similar domain arrangement (337). Similarly to Nrd1, SCAF8 associates with Pol II during transcription elongation and also plays a role in pre-mRNA processing (234). SCAF8-CID was shown to preferentially interact with Ser2-phosphorylated CTD *in vitro* (22). Interestingly, a previous report has demonstrated that CTD tyrosine-phosphorylation prevents the binding of SCAF8-CID (337). Therefore, the finding that Tyr1 phosphorylation prevents CID-CTD interactions in yeast seems also to be true in humans.

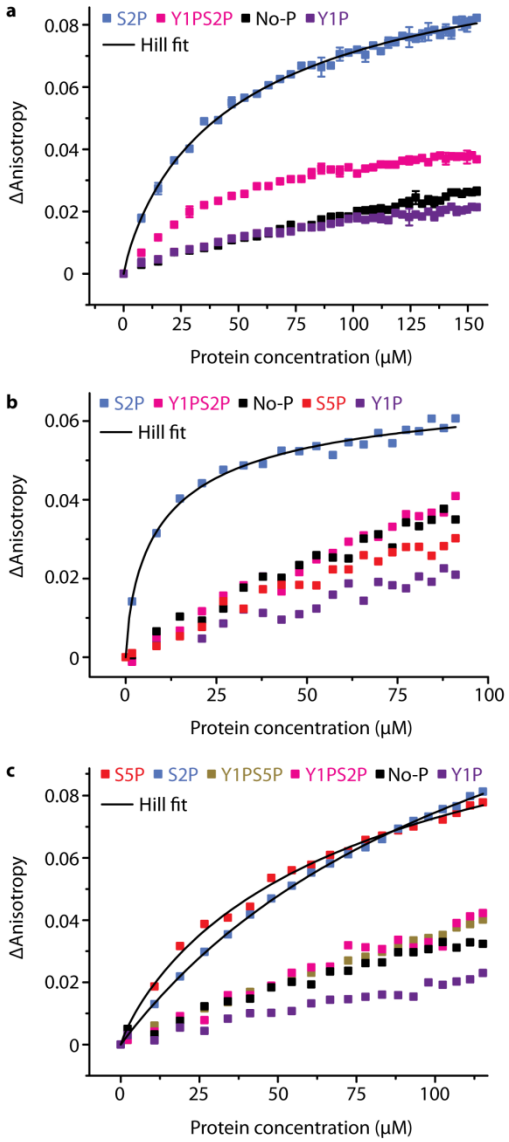

Figure 23: CTD Tyr1 phosphorylation prevents the binding of CID-containing termination factors *in vitro*. (a) Analysis of the binding affinity and specificity of the Pcf11-CID for different CTD-mimicking synthetic peptides by fluorescence anisotropy titration experiments. The Pcf11-CID from *S. cerevisiae* was recombinantly expressed in *E. coli* cells and purified to homogeneity. Purified Pcf11-CID was titrated to fluorescently labeled CTD peptides and the fluorescence anisotropy was detected in the presence of 10 mM NaCl (for details see section IV 2.5). A non-linear binding curve could be fitted in case of S2P, allowing the determination the corresponding binding affinity (Hill fit, Origin). Standard deviations refer to two replicate measurements. **(b-c)** Analysis of the binding affinities of Rtt103-CID (*S. cerevisiae*) and Nrd1-CID (*S. cerevisiae*) for different CTD peptides, respectively. Experiments were performed as in **(a)** but in the presence of 150 mM NaCl. For details about the CTD peptides used in this study, please see Table 20.

3.6 Molecular basis for the impairment of CID-CTD interaction by Y1P

High-resolution X-ray structures of Nrd1-CID (308), Pcf11-CID (201) and Rtt103-CID (193) are available. Co-crystal structures of CID-CTD complexes are available for Pcf11-CID and Rtt103-CID, both bound to CTD-S2P peptides (193, 201). In the Pcf11-CTD structure, the Tyr1 hydroxyl group forms a hydrogen bond with the Asp68 side chain of the CID, that is conserved from yeast to human (201). Thus binding of a Tyr1-phosphorylated peptide would be impaired due to electrostatic repulsion of the Tyr1 phosphate and the Asp68 side chain. Modeling a Tyr1 phosphate group onto the CTD peptide results in a steric clash (Figure 24 a,b). The same situation is expected for Nrd1 CID, which contains a highly conserved CTD-binding site, including the asparate residue (308). In the Rtt103-CTD structure, the Tyr1 hydroxyl group forms a hydrogen bond with the side chain carboxyl of Asn65 (193), which corresponds to Asp68 in Pcf11. This hydrogen bond would not be possible after Tyr1 phosphorylation, which again leads to a steric clash when modeled onto the structure (Figure 24 c,d).

Additionally, the co-structures reveal that both the Pcf11-CID and the Rtt103-CID bind to Ser2-phosphorylated CTD that is in a β-turn conformation (193, 201). The β-turn is formed by S2P-P3-T4-S5 motif within the same CTD repeat and is stabilized by an intramolecular hydrogen bond between the S2P phosphate and the side-chain hydroxyl group of Thr4 (193, 201). A phosphate group on Tyr1 instead of Ser2 interferes with the formation of that hydrogen bond and thus destabilizes the β-turn formation what in turn interferes with the CID-CTD interactions. The high conservation of CID residues that directly contact the CTD and the observation that all CIDs in yeast can bind to Ser2-phosphorylated CTD peptides illustrates the importance of the β-turn stabilization. The results described for the Pcf11- and Rtt103-CIDs are also true for human SCAF8 bound to several Ser2-phosphorylated CTD peptides (22).

Taken together, structural data indicate that a phosphate group on Tyr1 would impair formation of all CID-CTD complexes, due to charge repulsion, steric hindrance and destabilizazion of the β-turn conformation of the CTD.

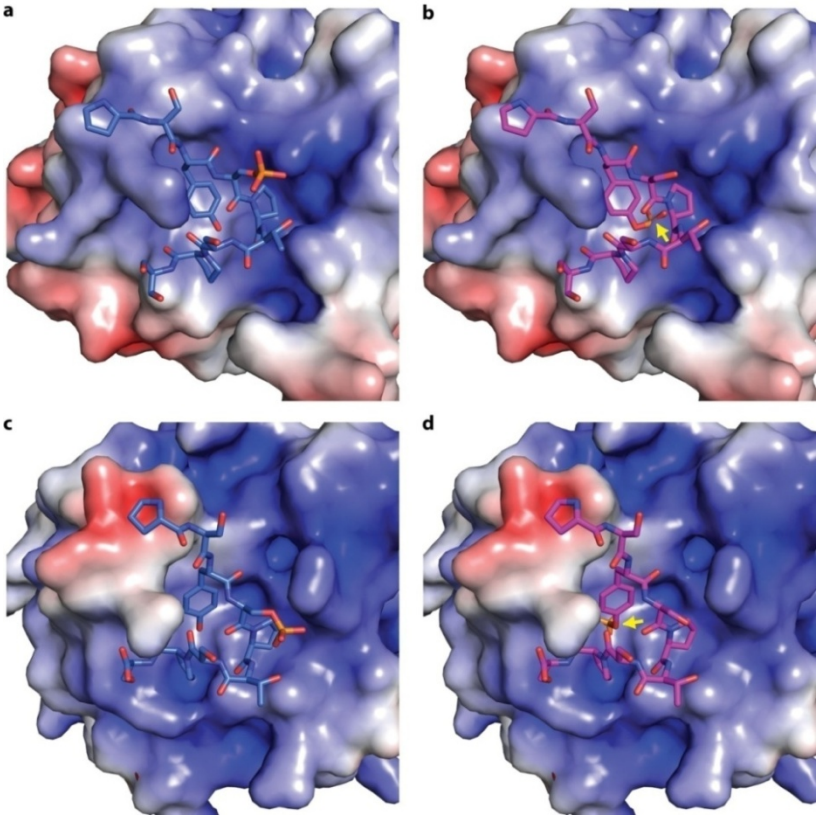

Figure 24: Tyr1 phosphorylation blocks CID-CTD interactions. (a) Surface representation of Pcf11-CID bound to Ser2-phosphorylated CTD peptide (stick model) (201). Blue, white and red colored surfaces correspond to positively charged, electrically neutral and negatively charged regions, respectively. The phosphate group (orange) on the Ser2 residue points away from the CID surface and does not contact the CID directly. **(b)** Structural model of Pcf11-CID (201) bound to Tyr1-phosphorylated CTD peptide. The Pcf11-CID is shown in the same orientation and the CTD peptide is bound in the same register as in **(a)**. The modeled phosphate group on Tyr1 resulted in a steric clash (yellow arrow) with Pcf11-CID. The surface area of Pcf11-CID where the steric clash occured is positively charged, resulting also in an electrostatic repulsion. **(c)** Surface representation of the Rtt103-CID bound to Ser2-phosphorylated CTD peptide (stick model) (193). The phosphate group on Ser2 directly contacts the Rtt103-CID. **(d)** Structural model of Rtt103-CID (193) bound to Tyr1-phosphorylated CTD peptide. Rtt103-CID is shown in the same orientation as in **(c)**. The Tyr1-phosphorylated CTD peptide is bound in the same register as in **(c)**. Modeling a phosphate group on Tyr1 resulted in a steric clash with Rtt103-CID (yellow arrow) and in electrostatic repulsion. Structural modeling and generation of Figures was performed with PyMol version 1.4.1. The surface charges were calculated with APBS.

3.7 The Pol II elongation factor Spt6 associates with Y1P

These results described in the previous sections converge on a model that Tyr1 phosphorylation prevents binding of termination factors, which use a CID domain, to associate with elongating Pol II, but also raise the questions whether Tyr1 phosphorylation would interfere with CTD interactions of factors that use other CTD-binding domains. To address this, we investigated the CTD binding to the tandem SH2 domain of Spt6 (residues 1250-1444), which binds the Ser2-phosphorylated but not the unphosphorylated CTD (52, 73, 185, 294) and contributes strongly to Spt6 occupancy on transcribed genes (199). Since Spt6 occupies genes throughout the coding region and downstream of the pA site (155, 199), we predicted that Tyr1 phosphorylation must not interfere with CTD binding of the tandem SH2 domain. Indeed, the purified recombinant tandem SH2 domain of Spt6 bound very well to CTD peptides phosphorylated at Tyr1 (K_D = 3.6±0.15), Tyr1 and Ser2 (K_D = 1.9±0.04) as well as Tyr1 and Ser5 (K_D = 1.3±0.06) (Figure 25). Interestingly, the interaction of the

tandem SH2 domain with Tyr1-phosphorylated CTD was even slightly stronger as compared to the Ser2-phosphorylated CTD (K_D = 8.4±0.19), representing the known interaction partner, and the Ser5-phosphorylated CTD (K_D = 5.2±0.09) (Figure 25). This observation agrees with previous reports showing that the tandem SH2 domain has a stronger affinity for CTD peptides phosphorylated on Tyr1 than on Ser2 or Ser5 residues (52, 185). This demonstrates that Tyr1 phosphorylation does not impair CTD binding of a well-studied elongation factor, but that it selectively prevents binding of CID-containing termination factors.

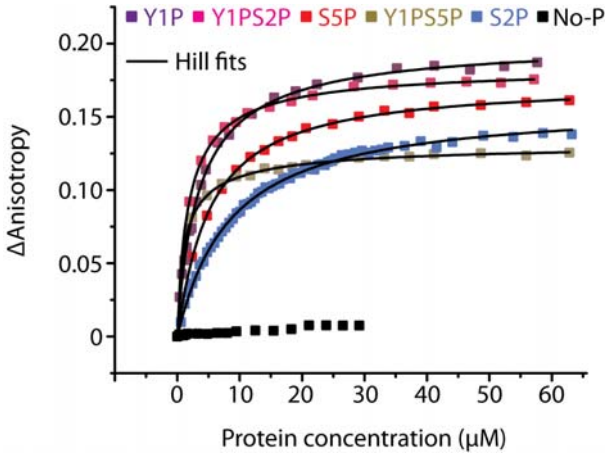

Figure 25: Spt6 binds to Tyr1-phosphorylated CTD peptides in vitro. The C-terminal tandem SH2 domain of Spt6 (*S. cerevisiae*) was recombinantly expressed in *E. coli* cells and purified to homogeneity as described (294). The binding affinity of the yeast Spt6 tandem SH2 domain for different CTD peptides (Table 20) was analyzed by fluorescence anisotropy titration experiments. The tandem SH2 domain bound to all CTD peptides tested, but with a slightly higher affinity to Tyr1-phosphorylated peptides. Fitting of non-linear binding-curves was performed with Origin (Hill fit). The color code is the same as in Figure 23.

3.8 Y1P restricts the recruitment of termination factors to narrow regions at the 5' and 3' end of genes

To illustrate the mechanism for exclusion of termination factors from a central transcribed region by Tyr1 phosphorylation, we summed up the genome-wide occupancies for Ser2- and Ser5-phosphorylated Pol II and subtracted from this sum the occupancy with Tyr1-phosphorylated Pol II. Although calculation of such a difference profile is problematic due to unknown normalization factors between data sets, we obtained a curve that contains peaks just downstream of the TSS and the pA site, and an extended depression in between (Figure 26). Whereas the peaks correspond to regions where termination factors are usually recruited (5' recruiting region, Nrd1; 3' recruiting region, Pcf11 and Rtt103), the depression indicates a central region in which the Ser2-phosphorylated CTD is masked and Pol II is shielded from termination factors. We note that additional factors can contribute to the recruitment of termination factors to elongating Pol II. For instance, Nrd1 functions in a complex with Nab3 and Sen1 (308). Nrd1 and Nab3 interact specifically with nascent RNA (60) what contribute to Nrd1 recruitment or its persistency near Pol II even when Tyr1 phosphorylation levels rise. As detailed later, this seems to be especially true at snoRNA genes. In addition, our model assumes uniform CTD phosphorylation on all repeats, which does not necessarily occur, but there is currently no data that address this issue.

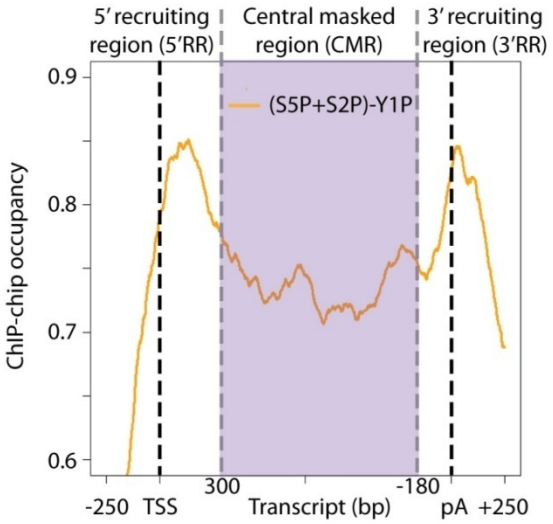

Figure 26: Tyr1 phosphorylation of the CTD restricts the recruitment of Pol II transcription termination factors to the 5' and 3' end of genes. Subtracting the genome-wide Y1P occupancy from the occupancy formed by S5P+S2P, results in a difference curve with peaks downstream of the TSS and the pA site, and an extended depression between both peaks. This observation may indicate that Tyr1 phosphorylation of the CTD establishes three distinct regions along transcribed protein-coding genes: 5' and 3' recruiting regions (5'RR and 3'RR) that are separated by an extended central masked region (CMR; violet). Whereas at the narrow 5'RR and 3'RR Pol II termination factors can associate with transcribing Pol II (5'RR: Nrd1; 3'RR: Pcf11 and Rtt103), the interaction is blocked during the CMR by Tyr1 phosphorylation of the CTD. This analysis was performed for the "medium" gene-length class, but the results of other length classes were similar (data not shown).

3.9 An extended model of the Pol II CTD code

Our results extend the previously proposed CTD code (36, 55, 79) and lead to a model for the phosphorylation-regulated recruitment of CTD-binding factors to Pol II during the transcription cycle (Figure 27). During transcription initiation the CTD is phosphorylated on Ser5 residues, which facilitates recruitment of Nrd1 at the beginning of genes. About 150 nt downstream of the TSS, peak occupancy levels for Nrd1 and Pol II (199) are reached. This observation is consistent with the idea of an early decision point, where Pol II either terminates early or continues RNA synthesis after transient pausing (37). In case Pol II continues RNA synthesis, Tyr1 and Ser2 phosphorylation levels rise, facilitating the binding of elongation factors such as Spt6. However, at the same time Tyr1 phosphorylation releases Nrd1 and impairs recruitment of termination factors Rtt103 and Pcf11. Before the pA site, Tyr1 phosphorylation levels drop, whereas Ser2 phosphorylation levels remain high, enabling recruitment of Rtt103 and Pcf11, and leading to mRNA 3'-end processing and transcription termination. Termination factors can bind cooperatively to neighboring CTD repeats (193) and can also interact with nascent RNA (60). This cooperative binding may lead to the observed sharp occupancy peaks. The key functional role of Tyr1 phosphorylation during active transcription explains why mutation of Tyr1 residues to phenylalanine, which removes the oxygen atom required for phosphorylation, is lethal (321).

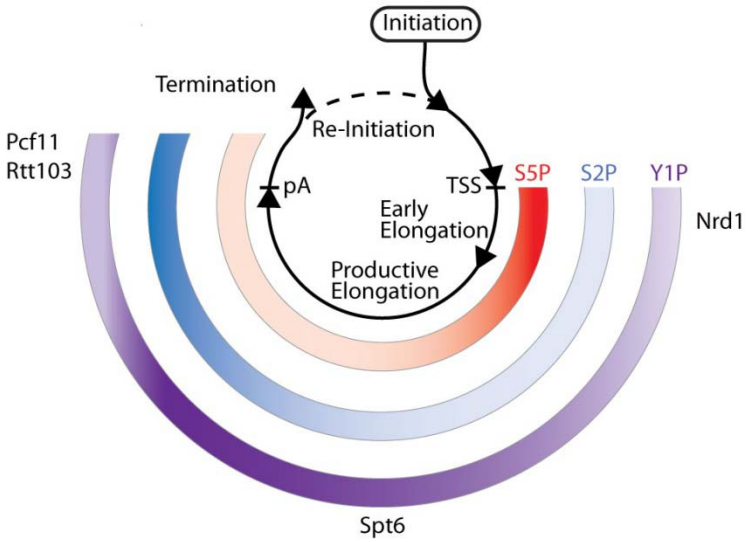

Figure 27: Extension of the Pol II CTD code. During the mRNA transcription cycle (internal black circle, different stages as well as the TSS and the pA site are indicated) the phosphorylation pattern of the CTD changes, as illustrated by red (S5P), blue (S2P) and violet (Y1P) semicircles. The color saturation corresponds to the occupancy level (the stronger the color saturation, the stronger the occupancy). The color code is the same as in Figure 2. Factors that associate with particular CTD phosphorylation patterns are indicated.

3.10 The CTD phosphatase Ssu72 targets S5P but not Y1P

Based on the presence of a protein tyrosine phosphatase (PTPase) signature motif, structural similarities to PTPases and its ability to cleave the phosphotyrosine analogue *p*-nitrophenylphosphate *in vitro*, Ssu72 was proposed to by a tyrosine phosphatase (203). Ssu72 was also shown to act as a CTD phosphatase with specificity against S5P (164, 319, 328). Therefore, we asked whether Ssu72 also targets Y1P. However, several lines of evidence indicate that this is most likely not the case.

Firstly, we performed a binding assay to determine the substrate specificity of Ssu72. We expressed and purified a catalytically inactive form of Ssu72 (C13D/D144N) from *Drosophila melanogaster* (319) and measured its affinity against various synthetic diheptad CTD peptides that represent potential substrates by fluorescence anisotropy measurements (Table 20). The data shows that Ssu72 binds to S5P (K_D = 32±8 µM) but neither to Y1P, Y1PS2P, S2P nor to the unphosphorylated CTD indicating that Ssu72 can discriminate between different Pol II phosphoisoforms (Figure 28 a).

Secondly, we measured the phosphatase activity of Ssu72 for different substrates by a malachite green assay that allows the detection of liberated orthophosphate (see section IV 2.6). We expressed and purified human Ssu72 (203) since we were not able to clone the yeast *SSU72* gene. This assay revealed a weak activity of Ssu72 for S5P but not for the other substrates tested, including Y1P and Y1PS2P (Figure 28 b).

Thirdly, we conducted ChIP experiments of Pol II (Rpb3) and of its tyrosine phosphorylated form under wild-type condition and under Ssu72 depletion. We generated an Ssu72 degron strain (see section IV 2.3.1, (8)) and used this yeast strain for ChIP experiments. Since *SSU72* is an essential gene, Ssu72 depletion was not viable for yeast cells what served as a quality control. ChIP data for the *ADH1* gene illustrates that the decrease of Y1P occupancy under Ssu72 depleting condition was substantially stronger than for Pol II (Rpb3) (Figure 28 c). This was true over the whole length of the transcribed region and indicates that Ssu72 rather plays a role in inhibiting than in promoting Y1P dephosphorylation under wildtype

conditions. Taken together, our *in vitro* and *in vivo* data indicate that Ssu72 targets S5P but most likely not Y1P.

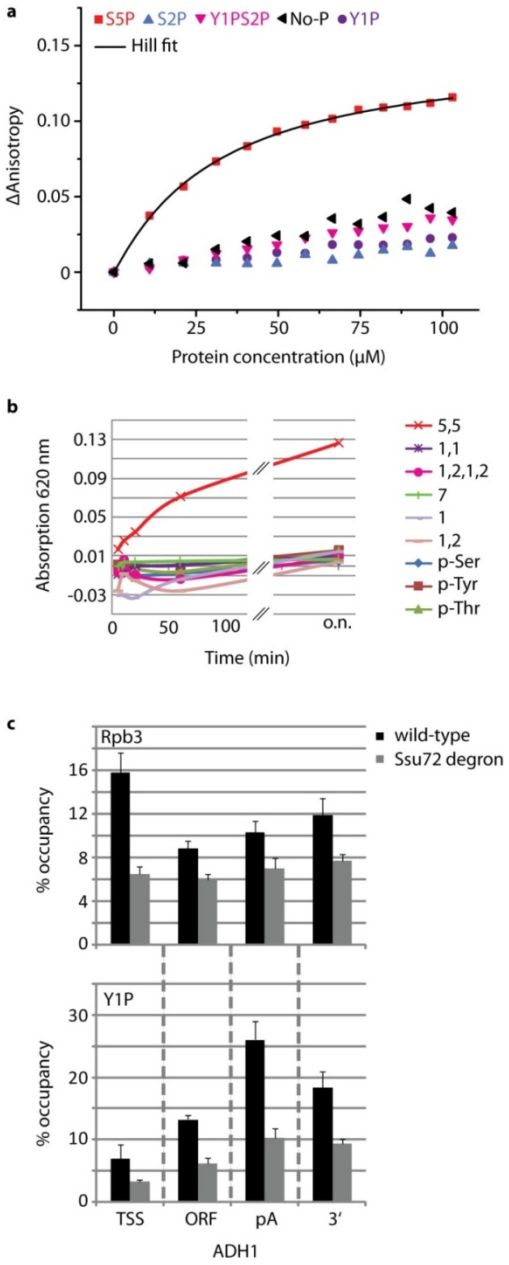

Figure 28: Ssu72 is most likely not the Y1P CTD phosphatase. (a) Analysis of the binding affinity of Ssu72 for different CTD peptides by fluorescence anisotropy titration experiments. A catalytically inactive form of Ssu72 (Ssu72 C13D/D144N, *D. melanogaster*) was recombinantly expressed in *E. coli* cells and purified to homogeneity. A binding of Ssu72 could only be detected for the S5P CTD peptide. Fitting of non-linear binding-curves was performed with Origin (Hill fit). The color code is the same as in Figure 23. **(b)** Analysis of the phosphatase activity of human Ssu72 for different CTD substrates by a malachite green assay. Consistent with the fluorescence anisotropy measurements, a weak phosphatase activity was only detected for the S5P CTD peptide. **(c)** ChIP experiments of Rpb3 (Pol II, independent of CTD phosphorylation status; upper panel) and the Tyr1-phosphorylated form of Pol II (lower panel) for wild-type (black bars) and for Ssu72 depleting conditions (gray bars; Ssu72 degron strain, see methods). Occupancies are shown for four distinct regions of the *ADH1* gene. Standard deviations refer to two biological replicates.

4. Conclusions and Outlook

In this study we used a monoclonal antibody that specifically recognizes tyrosine phosphorylated Pol II CTD to show that this type of CTD modification exists in yeast. High-resolution genome-wide occupancy profiling revealed that Tyr1 phosphorylation occurs at all transcribed genes and that its level correlates with mRNA expression. Tyr1 phosphorylation was absent at the gene promoter, increased downstream of the TSS and decreased before the pA site. Our data also indicated that Tyr1 phosphorylation prevents the recruitment of transcription termination factors to Pol II. First, the Pol II termination factors Nrd1 and Pcf11/Rtt103 showed occupancy peaks upstream or downstream of the region containing high Tyr1 phosphorylation levels, *in vivo*. Second, Tyr1 phosphorylation blocked CTD binding to the conserved CTD-interacting domains (CIDs) of Pol II termination factors *in vitro*. Third, structural modeling implied that Tyr1 phosphorylation results in an electrostatic repulsion from an aspartate residue. However, Tyr1 phosphorylation did not impair CTD binding of the elongation factor Spt6, which is present in transcribed regions. The data consistently supported the view that Tyr1 phosphorylation of the CTD shields active Pol II from termination factors within transcribed regions.

This study raises several questions that can be addressed in follow up projects.
Firstly, it is currently not known which kinase(s) is responsible for the Tyr1 phosphorylation of the CTD in yeast. One way to tackle this question *in vivo* is to inhibit the kinase of interest, for instance by the use of analog-sensitive (as) yeast strains (29), and observe whether the Tyr1 phosphorylation level is altered by ChIP experiments. The identification of the Tyr1-kinase(s) would enable plenty of follow up studies. For example, the Tyr1-phosphorylation could be inhibited by conditionally depleting the Tyr1-kinase. Under those conditions ChIP time-course experiments would reveal factors that are recruited viaTyr1-phosphorylation *in vivo*.
Secondly, the phosphatase(s) which targets phosphorylated Tyr1 residues is also not known. One way to address this question in vivo is conditionally deplete the phosphatase of interest and observe the levels of Tyr1-phosphorylation via ChIP experiments. Also *in vitro* phosphatase assays with purified phosphatases and tyrosine phosphorylated CTD as substrates could contribute to the identification of the Tyr1-phosphatase.
Thirdly, it is currently unknown whether apart from Spt6 other proteins interact with the tyrosine phosphorylated form of the CTD. One way to systematically identify potential interaction partners, including kinases and phosphatases, is by a combined approach of affinity purification (such as tandem affinity purification, TAP (251)) and mass-spectrometry (MS). Once new interaction partners were discovered, co-crystallization trials could be performed to reveal details of the respective binding interface at atomic resolution. These experiments would not only contribute to a better understanding of basic aspects of transcription regulation but also bear the potential to assign entirely new functions to the Pol II CTD.
Fourthly, our current knowledge of CTD modifications are based on studies using specific monoclonal antibodies. Since, a particular antibody recognizes a certain epitope on the CTD our current view of how the CTD is modified is rather indirect and most likely incomplete. Additionally, antibodies always bear the risk of crossreactivity as was also shown to be an

issue for some antibodies directed against certain CTD phosphoisoforms (157). Mass-spectrometry analyses of the CTD would avoid some of the drawbacks described for antibodies and would provide a more direct view on the CTD modification status. MS could help to solve the long-standing questions what modifications occur in which parts of the CTD and in which combinations in cells. Interesting questions in this context are whether Tyr1 residues are predominantly phosphorylated at the proximal or distal part of the CTD and whether phosphorylated Tyr1 co-occur with other CTD modifications.
Fivthly, apart from the observation that Tyr1 phosphorylation of the CTD also occurs in mammals, nothing is known of its biological role in other organisms. The analyses of CTD tyrosine phosphorylation in different organisms would reveal similarities and differences that result in new insights into the evolution of gene regulation in the context of the Pol II CTD.
A more longterm perspective would be to understand whether Tyr1 phosphorylation of the CTD plays a role in the regulation of the cell cycle, in certain diseased states and during development. Another key question is whether the different CTD phosphorylation marks influence each other.

V. The Spt5 C-terminal region recruits yeast 3'-RNA cleavage factor I

1. Introduction

1.1 The transcription factor Spt5

1.1.1 Basic aspects

The gene encoding the Spt5 protein was originally identified in a genetic screen in *Saccharomyces cerevisiae* (*S. cerevisiae*) as a suppressor of reporter gene insertions in 1984 (326). Spt5 was later described as an essential nuclear protein (297). Spt5 forms a tight complex with the zinc-binding protein Spt4 (118). Spt5 physically associates with RNA polymerase II *in vivo* and mutations of *SPT5* lead to a slow-growth phenotype in the presence of the nucleotide-depleting drug 6-azauracil (6-AU), arguing for a role in Pol II transcription elongation in yeast (118). The human homolog of yeast Spt4/5, also called DSIF complex, affects transcription elongation by Pol II (315). Furthermore, Spt5 co-localizes with Spt6 and Pol II at transcriptionally active loci on *Drosophila* polytene chromosomes (7, 146). Chromatin immunoprecipitation (ChIP) analyses revealed that Spt5 co-localizes with Pol II throughout the transcribed region and past the pA site (243). Recently, it was shown that Spt4/5 is present on all transcribed yeast genes and that it is a general component of the Pol II elongation complex ((199, 311); see also section III. 3).

Spt5 is the only known Pol II-associated transcription factor that is conserved in all three domains of life (320). The bacterial Spt5 homolog NusG and archaeal Spt5 consist of an N-terminal (NGN) domain and a flexibly linked C-terminal Kyrpides-Ouzounis-Woese (KOW) domain (160, 211) (Figure 29). In contrast to bacteria, archaea also have a Spt4 homolog (Figure 29). Recently, the structures of archaeal Spt4/5 bound to the RNA polymerase clamp domain could be determined by X-ray crystallography and cryoelectron microscopy (159, 198). These structures indicate that the NGN domain closes the active center cleft to lock nucleic acids and render the elongation complex stable and processive (124, 159, 198). Eukaryotic Spt5 possesses additional regions and domains. Yeast Spt5 consists of an acidic N-terminal region, followed by an NGN domain, five KOW domains, and a repetitive C-terminal region (CTR) (245, 297, 343).

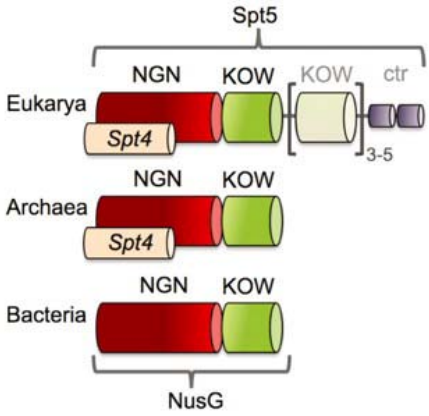

Figure 29: Domain organization of Spt4/5 in three domains of life. The NGN domain, the KOW motifs and the repeats of the C-terminal region (CTR) are indicated as red, green/white and violet barrels, respectively. Please note that Spt4 homologs only exist in eukaryotes and archaeal cells. This Figure is taken from (124).

Spt5 has recently emerged as a platform that recruits factors to elongating Pol II. Spt5 co-purifies with more than 90 yeast proteins that are involved in several distinct processes including transcription elongation, RNA processing, transcription termination and mRNA export (184). Spt4/5 interacts with the Pol II transcription factor and histone chaperone Spt6 to modulate chromatin structure in yeast (118, 298). Spt5 directly interacts with the mRNA capping enzyme in fission yeast and humans (236, 318). Mammalian Spt5 recruits the activation-induced cytidine deaminase to DNA during antibody gene diversification (235). Recently, it was shown that yeast Spt5 recruits She2 to the nascent RNA, coupling mRNA localization with Pol II transcription (276). The recruitment of factors can be mediated by the CTR of Spt5. The yeast Spt5 CTR recruits the Paf1 complex and the histone deacetylase subunit Rco1 during Pol II transcription elongation (75, 187, 343).

1.1.2 The repetitive C-terminal region (CTR) of Spt5

The CTR forms a repeat structure similar to the Pol II CTD (297). In yeast, the CTR consists of 15 hexapeptide repeats of the consensus sequence S[T/A]WGG[A/Q] (343). However, the consensus sequence of the Spt5 CTR repeats can vary from one organism to another. For example, the CTR of human Spt5 consists of pentapeptide repeats with the consensus sequence GS[R/Q]TP (331), whereas the Spt5 CTR of fission yeast consists of nonapeptide repeats with the consensus sequence TPAWNSGSK (236). In contrast to the Pol II CTD, deletion of the Spt5 CTR is not lethal in yeast (74, 187, 343), but leads to 6-AU sensitivity and a slow-growth phenotype at 16°C (187, 343). The CTR deletion is synthetically lethal with the deletion of the gene for the Pol II CTD kinase Ctk1 (187). Deletion of the CTR in fission yeast leads to a slow-growth phenotype and abnormal cell morphology (271). The slow-growth phenotype is intensified if the Pol II CTD is truncated (271), suggesting that the CTR co-operates with the CTD. Deletion of the CTR impairs embryogenesis in zebrafish and leads to a de-repression of transcription of genes in zebrafish and human cells (44). Similar to the Pol II CTD, the CTR of Spt5 can be phosphorylated by the kinases Bur1 and P-TEFb in yeast and human, respectively (187, 331, 343). CTR phosphorylation promotes transcription elongation in yeast, and is important for the co-transcriptional recruitment of the Paf1 complex and for histone modification (187, 343). In human cells, CTR phosphorylation by P-

TEFb converts Spt5 from a negative to a positive elongation factor (331). The Spt5 CTR may also play a role in the suppression of transcription-coupled nucleotide excision repair in yeast (74).

1.2 The cleavage factor (CF) I complex

Processing of mRNA 3'-ends occurs in two steps, endonucleolytic cleavage and addition of a pA tail (25). In yeast, the multisubunit complexes that carry out cleavage and polyadenylation are the cleavage factor (CF) I and the cleavage/polyadenylation factor (CPF) (195). The CFI can be separated into CFIA and CFIB (108, 150). CFIA consists of Clp1, Rna14, Rna15 and Pcf11 (6, 208, 209), whereas CFIB consists of a single subunit called Hrp1 (108, 149, 207). This machinery is highly conserved, with most of the subunits having homologs in mammalian cells (195). Rna15 forms a heterodimer with Rna14 (224). Two Rna14/15 heterodimers can in turn form tetramers *in vitro*, what led to the suggestion that CFI may functions at a dimer during 3'-RNA processing *in vivo* (14, 224). 3'-RNA processing occurs co-transcriptional (237). The CFIA subunit Pcf11 and Rna14 can bind to the phosphorylated form of the Pol II CTD (15, 266, 341). Pcf11 preferentially binds to the Ser2-phosphorylated form of the CTD ((182, 201); for details see section IV 1.1 and IV 3.5). In addition, the CFIA subunit Rna15 and the CFIB Hrp1 contain RNA recognition motifs (RRMs) that were shown to bind RNA *in vitro* (109, 181, 224, 233, 238). However, the role of nascent RNA in the recruitment of 3'-RNA processing factors *in vivo* remains to be determined.

1.3 Aims and scope of this work

Spt5 is the only Pol II-associated transcription factor that is conserved in all three domains of life, indicating its important role in gene transcription (320). Spt5 associates with Pol II during transcription elongation. Spt5 enhances the stability of the elongation complex and renders transcription elongation processive (159, 198). In yeast, Spt5 is essential for viability and co-purifies with more than 90 proteins that are involved in many different nuclear processes (184). ChIP experiments have detected Spt5 at selected genes that are actively transcribed by Pol II (243). Recently, we have shown that Spt5 co-localizes with transcribing Pol II throughout the coding region and past the pA site of all active genes in proliferating yeast (see section III Figure 6; (199)).
Interestingly, Spt5 contains a repetitive C-terminal region (CTR) that has been shown to recruit the Paf1 complex to Pol II (187, 343) and that is also involved in the recruitment of the histone deacetylase complex Rpd3 (75). All these observations have led us to assume that the Spt5 CTR, similar to the Pol II CTD, may act as a general platform that recruits factors during Pol II transcription.
Since Spt5 co-localizes well with Pol II throughout the coding region of genes, one question was whether the CTR is involved in the recruitment of particular elongation factors. An interesting finding of our genome-wide analyses was that Spt5 also co-localizes with Pol II

and transcription termination factors past the pA site until the region where transcription terminates (see section III Figure 6 e). Therefore, another question to be addressed was whether the CTR of Spt5 is also involved in the recruitment of 3'-RNA processing and termination factors. These questions should be addressed by ChIP experiments *in vivo* and by interaction studies with recombinantly expressed CTR constructs *in vitro*. In cases where valid interactions could be detected, its biological roles should be elucidated. This study bore not only the potential to assign entirely new functions to the chromatin transcription factor Spt5 but also to reveal new regulatory mechanisms of Pol II transcription. This work was performed in cooperation with Amelie Schreieck.

2. Specific procedures

Table 21: List of plasmids used in this study

Plasmid	Insert	Type	Tag	Restriction sites	Source
GST-CTR	GST-CTR (*S. cerevisiae* Spt5 CTR; 931-1063)	pET28b(+)	Nt GST	NcoI, HindIII	This work
pGEX-4T-1-GST	GST-tag	pGEX-4T-1	GST		GE Healthcare

Table 22: Proteins analyzed by ChIP-chip in this study

Protein	Process involved	ChIP protocol
Rpb3 (Pol II) in Spt5 ΔCTR cells	Gene transcription	For TAP-tagged proteins
Rna14 (CFIA complex)	3'-RNA processing	For TAP-tagged proteins
Rna15 (CFIA complex)	3'-RNA processing	For TAP-tagged proteins

Table 23: Proteins analyzed by ChIP-qPCR in this study (all factors were analyzed wild-type and Spt5 ΔCTR yeast cells, except Spt5 (only analyzed in wild-type))

Protein	Process involved	ChIP protocol
Rpb3 (Pol II)	Gene transcription	For TAP-tagged proteins
Bur1	Pol II transcription elongation	For TAP-tagged proteins
Cet1 (mRNA capping enzyme)	5'-RNA processing (mRNA Capping)	For TAP-tagged proteins
Clp1 (CFIA complex)	3'-RNA processing	For TAP-tagged proteins
Ctk1 (CTDK-I complex)	Pol II transcription elongation	For TAP-tagged proteins
Elf1	Currently not known	For TAP-tagged proteins
Hrp1 (also called CFIB)	3'-RNA processing	For TAP-tagged proteins
Paf1 (Paf1 complex)	Pol II transcription elongation	For TAP-tagged proteins
Pap1	3'-RNA processing	For TAP-tagged proteins
Pcf11 (CFIA complex)	Pol II transcription termination/ 3'-RNA processing	For TAP-tagged proteins
Rna14 (CFIA complex)	3'-RNA processing	For TAP-tagged proteins
Rna15 (CFIA complex)	3'-RNA processing	For TAP-tagged proteins
Spn1	Pol II transcription elongation	For TAP-tagged proteins
Spt4	Pol II transcription elongation	For TAP-tagged proteins
Spt5	Pol II transcription elongation	For TAP-tagged proteins
Spt6	Pol II transcription elongation	For TAP-tagged proteins
Spt16 (FACT complex)	Pol II transcription elongation	For TAP-tagged proteins

2.1 Molecular cloning and phenotyping of Spt5 ΔCTR yeast strains

For the generation of the Spt5 ΔCTR (deleted residues: 931-1063) yeast strain, the *kanMX6* marker cassette was amplified from the plasmid pFA6a-3HA-kanMX6 (primers: Spt5ΔCTR_Fw; Spt5ΔCTR_Rev; Table 6). The PCR product was used for transformation of wild-type yeast (BY4741; no epitope-tag) and of various yeast TAP-strains listed in Table 4. Positive transformants were identified on G418 selective media plates. Yeast strains were validated by control PCR (primers: ΔCTR_cont1_Fw; ΔCTR_cont2_Fw; ΔCTR_cont3_Fw; ΔCTR_cont1_Rev; ΔCTR_cont2_Rev; ΔCTR_cont3_Rev; Table 6) and DNA sequencing.Wild-type-pRS316 and Spt5 ΔCTR-pRS316 yeast strains were generated by transformation of the pRS316 plasmid into wild-type (BY4741; no epitope-tag) and Spt5 ΔCTR (no epitope-tag) cells, respectively. Positive transformants were identified on SC -ura plates.

For growth curve measurements, overnight cultures of wild-type and Spt5 ΔCTR strains were diluted with fresh YPD medium to a starting OD_{600} of 0.1. Yeast cells were grown for 18 h and the OD_{600} was determined every 90 minutes. Biological triplicate measurements were performed and analyzed by Microsoft Excel 2007. Growth of wild-type and Spt5 ΔCTR strains was also tested on YPD plates at different temperatures. Cells were grown in YPD at 30°C to stationary phase and diluted to OD_{600} ~ 1.0 with fresh medium. Equal amounts of cells were spotted on YPD plates in 20-fold serial dilutions. Plates were incubated at 16°C, 30°C or 37°C, respectively, and inspected daily. To assess potential defects in transcription elongation, serial dilutions of the wild-type-pRS316 and the Spt5 ΔCTR-pRS316 strain were spotted on SC –ura plates containing 50 µg/mL 6-azauracil (6-AU) or 15 µg/mL mycophenolic acid (MPA) at 30°C.

2.2 RNase-ChIP assay

RNase-ChIP analysis was performed as described in section II 2.1 and according to (1), but with the following modifications. Firstly, the time of formaldehyde crosslinking was reduced from 20 min to 5 min. Secondly, the prepared chromatin was separated before the immunoprecipitation (IP) step. One chromatin sample was treated with 7.5 U of RNase A and 300 U of RNase T1 (Ambion), the other sample was treated with the same volume of the RNase storage buffer. After incubating for 30 min at room temperature, the following steps were performed as described in section II 2.1. After qPCR the fold enrichment for different gene regions over an unoccupied genomic region was calculated as described in (83), but with the following modifications. The calculated fold enrichment values of the respective gene region were averaged between the different biological replicates. The fold enrichment values of chromatin samples that were not treated with the RNase-mix were averaged as well and set to 100%.

2.3 Rapid amplification of cDNA 3'-ends (3'-RACE)

For RNA isolation, overnight yeast cultures were diluted in 50 mL fresh YPD medium to OD_{600} ~ 0.1 and grown at 30°C to mid-log phase (OD_{600} ~ 0.8). The RiboPure™-Yeast Kit (Applied Biosystems) was used to isolate yeast RNA according to the manufacturer's instructions. 3'-RACE was performed for the *ACT1* gene as described (89). This gene was shown to possess several alternative pA sites and was used in 3'-end processing studies (3). The *ACT1* cDNA was synthesized as described (3) using Superscript II reverse transcriptase (Invitrogen), 0.5 µM OligodT-anchor primer containing a XhoI restriction site and 2 µg RNA template. Next, RNA was digested using RNase H (New England BioLabs Inc.) for 20 minutes at 37°C. The enzyme was inactivated by incubation at 65°C for 20 min. PCR reactions were conducted using 2 µL of the cDNA samples, 0.25 µM gene-specific Upstream primer 1 containing an EcoRI restriction site, the Anchor primer 2 and *Taq* DNA Polymerase (Fermentas, Thermo Fisher Scientific). The PCR product and the pET28b(+) vector were digested with EcoRI-HF™ as well as XhoI (New England BioLabs Inc.) and then ligated and transformed into competent *E. coli* XL1-blue cells. Plasmids from at least three different clones were sequenced (GATC Biotech) using pET28b(+) Sequencing primer 1 and 2. The sequences of the corresponding primers are listed in Table 6.

2.4 Pol II readthrough assay

The readthrough was performed for the *PMA1* gene according to (3), using the gene-specific forward primer RTh1 and five reverse primers (RTh2 to RTh6; Table 6)) positioned at different regions downstream of the transcription termination site. RNA isolation, reverse transcription and PCR were conducted as described in 3'RACE and the PCR products for wild-type and Spt5 ΔCTR mutant were analyzed by standard agarose gelelectrophoresis.

2.5 GST pull-down assay

To prepare yeast cell lysates, cell lysis was performed as for ChIP experiments but with the following modifications: (i) cell cultures were grown in 200 mL YPD medium at 30°C but not crosslinked with formaldehyde; (ii) phosphatase inhibitors (PhoI) were added to the FA lysis buffer at all steps; (iii) cell debris was removed by centrifugation. A synthetic DNA construct of the GST tag based on the pGEX3T vector fused to the DNA sequence coding for Spt5 CTR residues 931-1063 (343) was synthesized (Mr Gene GmbH, Regensburg, Germany) and cloned into vector pET28b(+) (Table 21). Expression of the GST-Spt5CTR fusion construct (pET28b(+)) and the GST tag alone (pGEX-4T-1) in *E. coli* was induced with 0.5 mM IPTG at 18 °C over night in 1 L cultures. Cells were lysed by sonification in 50 mL GST lysis Buffer for 15 minutes (Branson Sonifier 250). The cell debris was removed by centrifugation. For pull-down experiments, *E. coli* cell lysates were incubated with 100 µL Glutathione Sepharose beads (GE Healthcare), pre-equilibrated in GST lysis Buffer for 1 h at 4°C. Beads were washed 8x with GST lysis Buffer and incubated for 1 h at room temperature with 300 µL Rna14-TAP yeast cell lysate. Next, beads were washed 8x with GST lysis Buffer and proteins were eluted from the beads 8x with GST lysis Buffer containing 10 mM glutathione.

All elution fractions were pooled and the protein precipitated with 10% trichloroacetic acid (TCA). The protein pellet was washed with ice-cold acetone and resuspended in 2x SDS-PAGE loading buffer. Wash and elution fractions were analyzed by SDS-PAGE and Western blotting with antibodies against the TAP-tag of Rna14 (PAP, Sigma-Aldrich) and the GST-tag (RPN1236, GE Healthcare). To test whether the interaction depends on phosphorylation, the GST pull-down assay was performed with the following modification: 5 mM fresh ATP was added to the Rna14-TAP yeast cell lysate before incubation with Glutathione Sepharose beads. The buffers are listed in Table 14.

2.6 Quantitative Western blot analysis

Western Blot analysis was performed as described in section II 2.7, but with the following modifications. For evaluation of the GST pull-down experiments the membrane was probed with the PAP antibody against the TAP-tag of Rna14 (PAP, P1291, Sigma-Aldrich) and the GST antibody (RPN1236, GE Healthcare, Chalfont St. Giles, UK). For quantitative Western Blot analysis, quantification of the Spt5 protein levels in wild-type and Spt5 ΔCTR cells was performed with total protein samples using an antibody against the N-terminus of Spt5 (yN-20; Table 9) as well as the α-Tubulin antibody (3H3087; Table 9). The antibody signal was detected quantitatively using the Pierce® ECL Western Blotting Substrate and the LAS-3000 camera (FUJIFILM, Tokyo, Japan). Spt5 protein signals were quantified relatively to the α-Tubulin signals using the ImageQuant 5.0 software (Molecular Dynamics).

3. Results

3.1 Investigation of elongation factor recruitment by the Spt5 CTR

To investigate whether the function of the yeast Spt5 CTR in recruiting Pol II-associated factors extends to elongation factors other than Paf1 (187, 343), we carried out ChIP analysis in strains lacking the Spt5 CTR. We generated a yeast strain with a CTR deletion (described in section V 2.1). As reported previously, CTR deletion led to 6-AU sensitivity and a slow-growth phenotype at 16°C (187, 343). We also observed a slight growth defect at 30°C, but not at 37°C. In contrast to observations in *S. pombe* (271), the morphology of *S. cerevisiae* was not altered upon CTR deletion (not shown). Quantitative Western blot analysis revealed approximately two-fold higher Spt5 protein levels in cells lacking the CTR.

For ChIP analysis, we fused a tandem affinity purification (TAP) tag to the C-terminus of elongation factors in the CTR deletion background. The occupancy levels of the elongation factors were determined by ChIP at different positions of genes *ADH1, ILV5, PDC1, PMA1* and *TEF1*. We chose these genes for several reasons. Firstly, these genes encode housekeeping proteins, are highly expressed (70), and are heavily occupied by Pol II in the mid-log phase of yeast growth (199). Secondly, their DNA elements, including the transcription start site (TSS) and the pA site, are well characterized (64, 155, 219). Thirdly,

the transcription unit is long enough to distinguish between different binding levels at distinct positions of the gene. The primers used are lited in Table 7.

We performed ChIP analyses for eight Pol II elongation factors that belong to the three different groups described recently (199): (i) Spt4 and Spt6 (group 1), (ii) Elf1 and Spn1 (group 2), (iii) Bur1, Ctk1, Paf1 and Spt16 (group 3). The results are shown in Figure 30. These data revealed strong factor binding at all tested genes. A severe decrease in Paf1 occupancy to about 20% was detected at the *ADH1* gene (Figure 30 b), consistent with previous reports (187, 343) and providing a positive control. The difference in Paf1 occupancy was not due to a difference in Pol II occupancy, which was unaffected by CTR deletion (Figure 30 b). However, the other representative elongation factors tested did not show significant differences in their gene occupancies, showing that CTR deletions specifically reduce gene occupancy of Paf1.

An overview of all factors that were analyzed by ChIP and ChIP-chip in this study is given in Table 22 and 23.

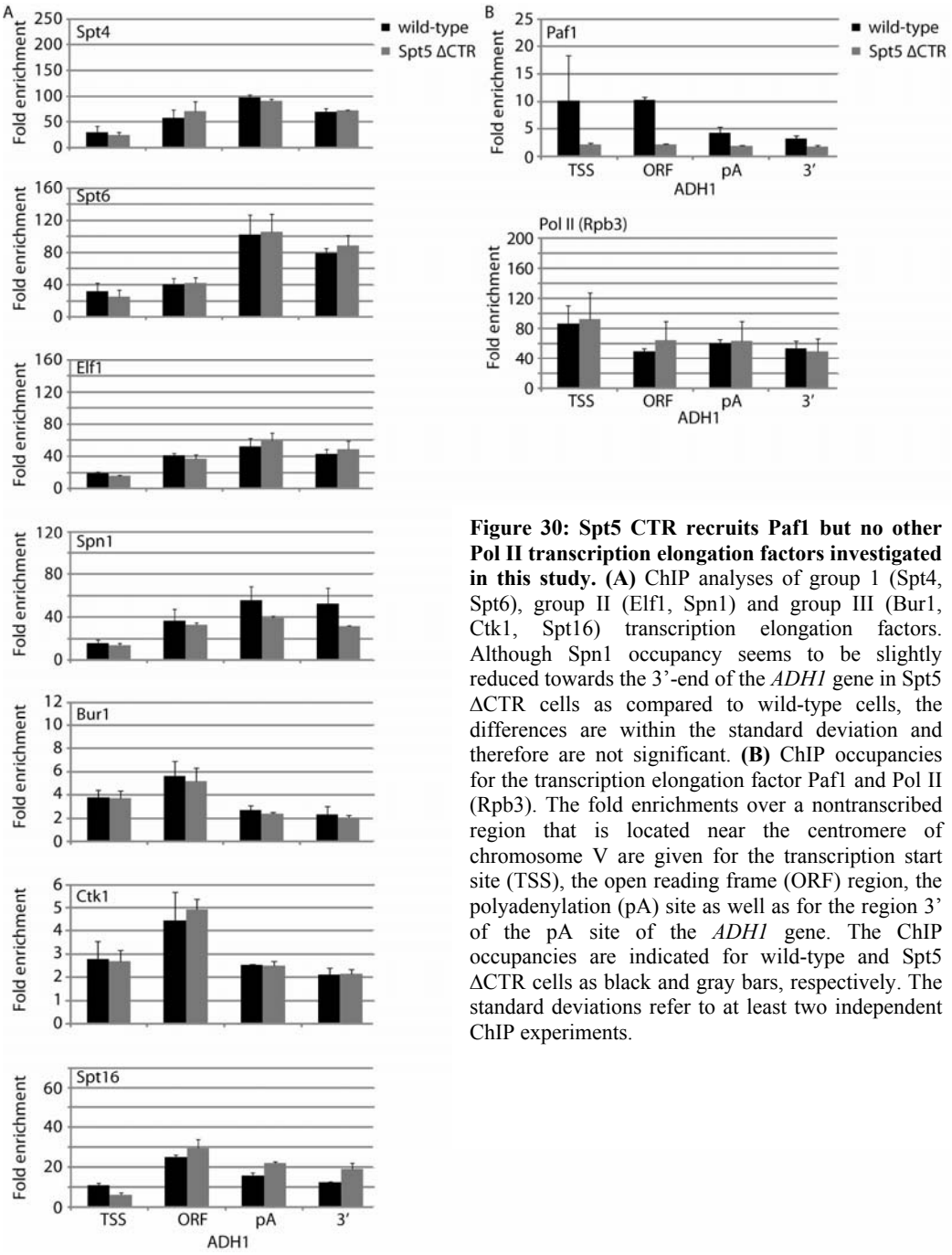

Figure 30: Spt5 CTR recruits Paf1 but no other Pol II transcription elongation factors investigated in this study. (A) ChIP analyses of group 1 (Spt4, Spt6), group II (Elf1, Spn1) and group III (Bur1, Ctk1, Spt16) transcription elongation factors. Although Spn1 occupancy seems to be slightly reduced towards the 3'-end of the *ADH1* gene in Spt5 ΔCTR cells as compared to wild-type cells, the differences are within the standard deviation and therefore are not significant. **(B)** ChIP occupancies for the transcription elongation factor Paf1 and Pol II (Rpb3). The fold enrichments over a nontranscribed region that is located near the centromere of chromosome V are given for the transcription start site (TSS), the open reading frame (ORF) region, the polyadenylation (pA) site as well as for the region 3' of the pA site of the *ADH1* gene. The ChIP occupancies are indicated for wild-type and Spt5 ΔCTR cells as black and gray bars, respectively. The standard deviations refer to at least two independent ChIP experiments.

3.2 Spt5 CTR is required for recruitment of CFI *in vivo*

Since Spt5 co-localizes with 3'-processing factors (155, 199, 311) and co-purifies with these factors (184), we tested whether it plays a role in the recruitment of 3'-processing and transcription termination factors. ChIP analysis revealed that all CFIA subunits, Clp1, Pcf11, Rna14, and Rna15, showed high occupancy at the 3'-end of protein-coding genes and the pA site (Figure 31). In Spt5 ΔCTR cells, occupancy of Clp1, Rna14, and Rna15 was reduced more than 50% at the *ADH1* gene, and also markedly lower at other tested genes (Figure 1 a,c). However, we observed no difference in the occupancy of Pcf11 between wild-type yeast and Spt5 ΔCTR cells (Figure 31 a).

Since CFIA is associated with CFIB/Hrp1, we investigated whether Hrp1 occupancy was affected by CTR deletion. Previous ChIP analyses have shown that Hrp1 crosslinks throughout the coding regions until the 3'-end of genes (155, 162). Our ChIP analysis revealed that Hrp1 shows strongest occupancy near the pA site, although it is recruited earlier than CFIA subunits (Figure 31 b). Similar to CFIA subunits, Hrp1 binding was markedly reduced in Spt5 ΔCTR cells (Figure 31 b). However, no occupancy difference could be observed for the poly(A) polymerase Pap1 (Figure 31 d), which is also required for 3'-end processing (195, 206). Taken together, the Spt5 CTR plays a crucial role in recruitment of CFI, but not of Pap1, to the 3'-end of protein-coding genes.

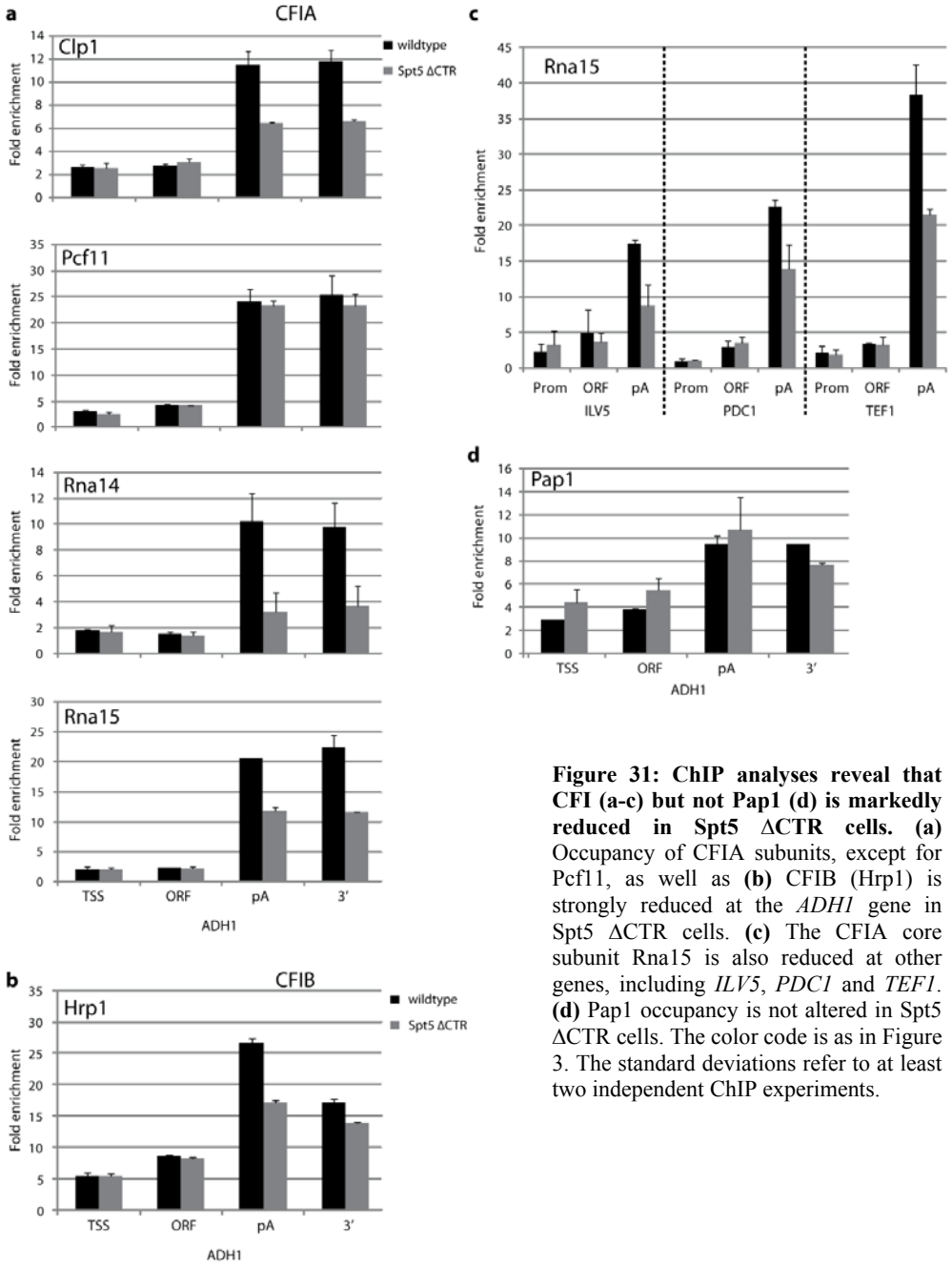

Figure 31: ChIP analyses reveal that CFI (a-c) but not Pap1 (d) is markedly reduced in Spt5 ΔCTR cells. (a) Occupancy of CFIA subunits, except for Pcf11, as well as **(b)** CFIB (Hrp1) is strongly reduced at the *ADH1* gene in Spt5 ΔCTR cells. **(c)** The CFIA core subunit Rna15 is also reduced at other genes, including *ILV5*, *PDC1* and *TEF1*. **(d)** Pap1 occupancy is not altered in Spt5 ΔCTR cells. The color code is as in Figure 3. The standard deviations refer to at least two independent ChIP experiments.

3.3 Spt5 CTR interacts with CFI *in vitro*

Since CFI recruitment to genes was impaired upon CTR deletion, we asked whether there is a physical interaction between the Spt5 CTR and CFI. We performed pull-down experiments with a recombinantly expressed glutathione S-transferase (GST)-tagged version of the Spt5 CTR and yeast cell lysates, prepared from an Rna14-TAP strain. Western blot analysis of the eluates with antibodies directed against the GST tag and the TAP tag of Rna14 revealed a co-precipitation of the GST-tagged Spt5 CTR and the CFI subunit Rna14, but not with GST alone (Figure 32). Pull-down experiments in the presence and absence of phosphatase inhibitors and an excess of kinase substrate ATP revealed no differences in the amount of Rna14 that co-precipitated with Spt5 CTR (not shown). These experiments revealed an apparently phosphorylation-independent previously unobserved interaction between the Spt5 CTR and CFI *in vitro*. Since the interaction was detected with the use of a lysate that naturally contains many non-specific competitor proteins, and since it was not observed with the GST tag alone, it must be regarded as being highly specific. It is however possible that the interaction is mediated by other proteins in the extract and thus indirect.

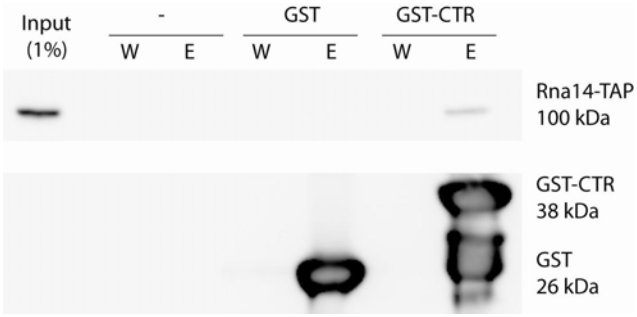

Figure 32: The Spt5 CTR interacts with Rna14 *in vitro*. GST pull-down experiments were performed with a GST-Spt5 CTR fusion protein (GST-CTR), with GST alone (GST) and without protein (-), serving as negative control. Western Blotting was performed for the last washing fractions (W), the combined elution fractions (E) of the respective sample as well as for 1% of the Rna14-TAP lysate (Input) with antibodies against the TAP tag of Rna14 and the GST tag.

3.4 RNA contributes to CFI recruitment

Since deletion of the Spt5 CTR led to a marked reduction in the occupancy of CFI subunits, but not to a complete loss, we asked which factors contribute to the residual binding of CFI. Since Rna15 and Hrp1 contain RNA recognition motifs that are known to bind RNA sequences *in vitro* (150, 181, 233, 238), we reasoned that nascent RNA may contribute to CFI recruitment. To address this, we performed an RNase-ChIP assay (1). In this assay, RNA is digested before the immunoprecipitation step, leading to a drop in factor occupancy if factor recruitment *in vivo* involves RNA.

We performed RNase-ChIP for Rna15 and Rna14 (Figure 33). Firstly, we observed a decrease in Rna15 occupancy after RNase treatment, indicating an important role of RNA in Rna15 recruitment, both in wild-type and Spt5 ΔCTR cells (Figure 33 a). Secondly, Rna15 binding most strongly depended on RNA around the pA site of the *ADH1* gene. Thirdly, the strongest reduction in Rna15 occupancy was observed when both the Spt5 CTR was deleted and when RNA was removed by RNAse treatment. The additional decrease of the Rna15 occupancy level was highly reproducible and could be observed in all four independent biological

replicates. A similar RNA dependence could be observed for Rna14, although the additional decrease of Rna14 occupancy in Spt5 ΔCTR cells was not as prominent as for Rna15 (Figure 33 b). These results indicate that RNA contributes to CFI recruitment *in vivo* and that this can explain residual recruitment of CFI in cells lacking the Spt5 CTR.

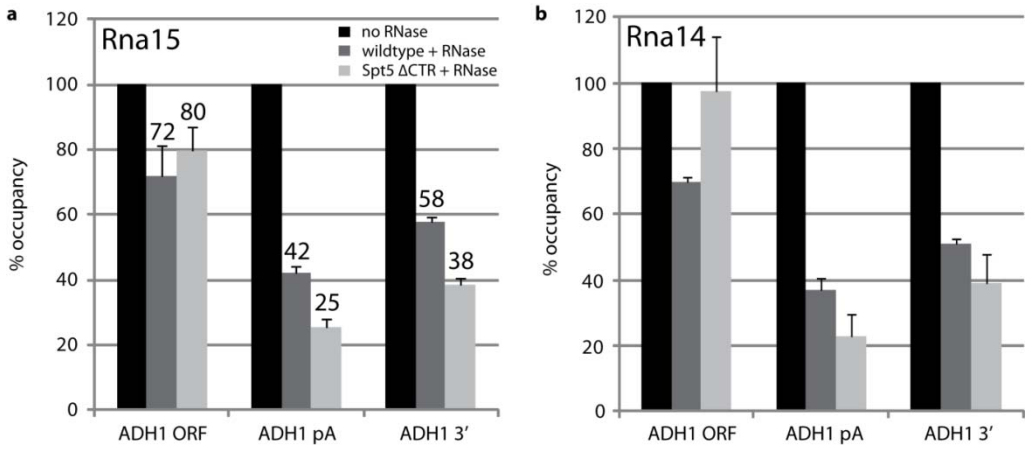

Figure 33: RNAse-ChIP assays reveal that RNA contributes to CFI recruitment in Spt5 ΔCTR cells. RNA-dependent binding of **(a)** Rna15 and **(b)** Rna14 (same color code as in **(a)**) is given for the ORF region, the pA site and a region 3' of the pA site of the *ADH1* gene. The ChIP occupancy signal without RNase treatment (black) was set to 100%. The relative ChIP signals for wild-type (dark gray) and Spt5 ΔCTR cells (light gray) after RNase treatment are indicated as well. In case of Rna15, the corresponding percentages are given above the bars. The standard deviations refer to four independent RNase ChIP experiments.

3.5 CFI co-localizes with the Ser2-phosphorylated CTD downstream of the pA site

It was shown for mammalian cells that homologs of the yeast CFI complex are already recruited at the promoter region of genes (105, 299). One study in yeast showed that some RNA 3'-processing factors, including Rna14 and Rna15, also crosslinked to the promoter and the early coding region of the *PMA1* gene, but not at other genes tested (155). However, our ChIP analysis (Figure 31) and published data in yeast (155, 182) suggested that CFI subunits crosslinked near the pA site at the 3'-end of genes.

To investigate the preferred location of CFI subunits on a genome-wide level, we performed ChIP-chip analysis for Rna14 and Rna15. This revealed CFI recruitment at all protein-coding genes that are occupied by Pol II and its elongation factors (199). The ChIP-chip profiles showed sharp occupancy peaks for Rna14 and Rna15 105 (± 10) nucleotides and 108 (± 6) nucleotides downstream of the pA site, respectively (Figure 34 a). We also detected weak Rna14 and Rna15 binding over the transcribed region, with an increase towards the 3'-end. These profiles were independent of gene length (not shown). Comparison with previous profiles (199) revealed that peak occupancies of Rna15 and Rna14 coincided with Pcf11 occupancy, which peaked 52 (± 7) nt downstream of the pA site (Figure 34 a). CFI subunit peak occupancies further occured in a region where the occupancies for Spt5 and the S2-phosphorylated Pol II were high (Figure 34 b). The sharp occupancy drop-off of CFI subunits coincided with the drop-off for the ChIP signal of the S2-phosphorylated CTD, consistent with a role of the S2-phosphorylated CTD in CFI recruitment.

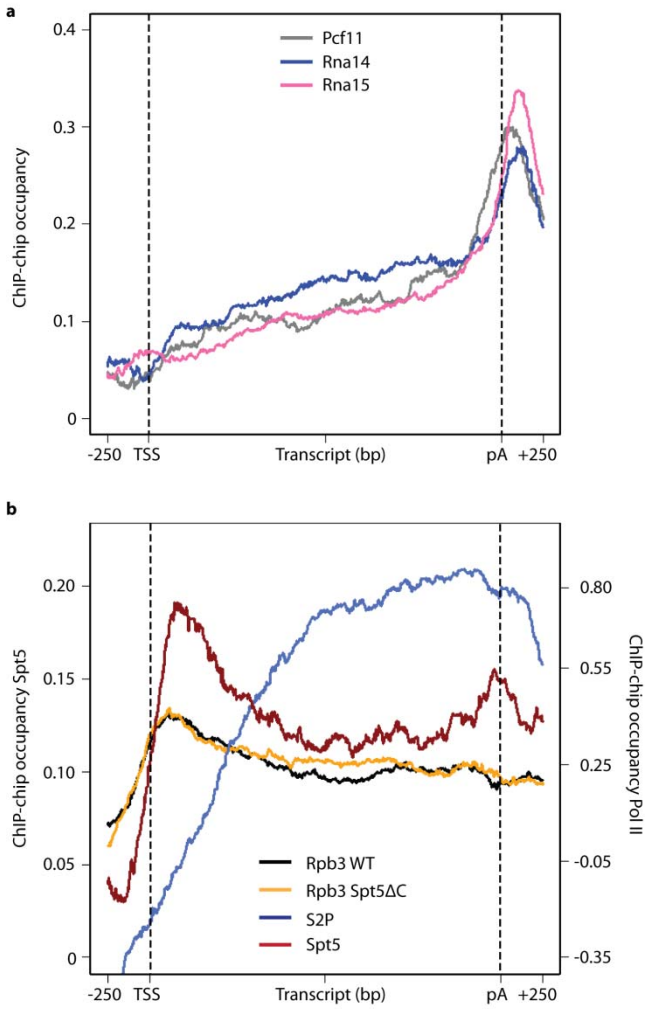

Figure 34: Genome-wide ChIP-chip occupancy profiling of CFI subunits in yeast. (a) Gene-averaged profiles for the long gene length class (2,350 ± 750 nucleotides, 299 genes, see methods) of Pcf11 (199), Rna14, and Rna15. **(b)** Gene-averaged profiles as in **(a)** for Pol II in wild-type and Spt5 ΔCTR cells, the Ser2-phosphorylated CTD form of Pol II (199) and the transcription elongation factor Spt5 (199). Occupancies and signal intensities are given for Spt5 and Pol II (including the Ser2-phosphorylated form) on the left and right y-axes, respectively. Dashed black lines indicate the TSS and pA site.

3.6 CTR deletion does not impair termination

Previous studies showed that mutations in Pcf11, Rna14, and Rna15 can lead to defects in transcript cleavage and readthrough transcription beyond the termination site (10, 28, 249). We therefore investigated whether the reduced level of CFI recruitment observed in Spt5 ΔCTR cells leads to Pol II readthrough transcription at the *ACT1*, *PMA1* and *RNA14* genes. To detect transcriptional readthrough, we chose a PCR-based method with a gene-specific forward primer and different reverse primers positioned downstream of the normal transcript termination site, as described (3) (Figure 35 a). In this assay, a prolonged transcript resulting from readthrough transcription would be detected by the generation of PCR products with a reverse primer that is located downstream of the termination site. As shown in Figure 35 b, no differences in the PCR products, and thus the length of the *PMA1* transcripts, were observed between wild-type and ΔCTR cells. Similar results were obtained for the *ACT1* and *RNA14* genes (not shown). Thus, deletion of the Spt5 CTR does not result in a termination defect that would be detected by transcriptional readthrough.

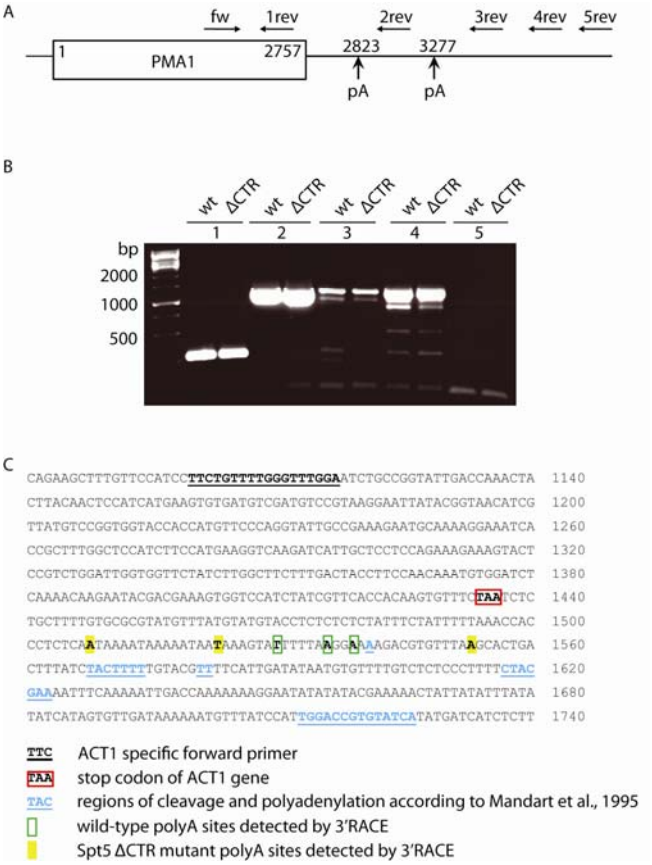

Figure 35: Spt5 CTR deletion provokes neither transcriptional readthrough of Pol II nor alternative pA site usage. (A) Schematic representation of the yeast *PMA1* locus. The ORF region and the two pA sites according to (3) are indicated by a box and vertical arrows, respectively. The forward primer (fw) and different reverse primers (1 rev to 5 rev) that were used for Pol II readthrough detection are depicted as horizontal arrows. **(B)** Agarose gelelectrophoresis of the 5 PCR products as described in **(A)** for wild-type (wt) and Spt5 ΔCTR cells (ΔCTR). No differences in the length of the PCR products could be detected between wild-type and Spt5 ΔCTR cells. The height of the marker lanes in base pairs (bp) are shown on the left. **(C)** The nucleotide sequence of the 3'-region of yeast *ACT1* is shown. Key sequence elements are labeled. The RNA cleavage and pA sites as determined by 3'RACE map to the same region for wild-type (green boxes) and Spt5 ΔCTR cells (yellow boxes).

Although we did not detect readthrough transcription at tested single genes, it may still occur at other genes. To investigate this, we measured high-resolution ChIP-chip occupancy profiles for the Pol II core subunit Rpb3 in wild-type and mutant yeast cells lacking the CTR. The high correlation between the Pol II profiles (Pearson $R = 0.89$) and the high similarity of the gene-averaged profiles (Figure 34 b) however indicated no difference in Pol II occupancy between wild-type and mutant cells. In addition, a difference profile calculated from Rpb3 occupancy in ΔCTR and wild-type cells did not reveal any clusters of altered occupancy. These results show that transcription readthrough does not occur in the ΔCTR strain.

3.7 CTR deletion does not alter pA site usage

Since mutations of 3'-end processing factors were also shown to result in the usage of alternative pA sites (194), we investigated whether CTR deletion and the resulting reduction in CFI recruitment lead to alternative pA site usage. To detect a possible change in the usage of pA sites *in vivo*, we used rapid amplification of cDNA 3'-ends (3'-RACE), which allows for a mapping of the 3'-ends of transcripts (89). We performed 3'-RACE for the *ACT1* and *PMA1* genes, which possess five and two pA sites, respectively (Figure 35 c; (155, 194)). The

experiments revealed multiple pA sites for the *ACT1* gene, which map to a distinct region at the 3'-end of the gene, 1508-1553 nucleotides from the TSS. However, no differences of the pA site pattern could be detected between wild-type and Spt5 ΔCTR cells (Figure 35 c). Similar results were obtained for the *PMA1* gene (not shown). Thus a reduced level of CFI recruitment in Spt5 ΔCTR cells does not lead to altered pA site usage *in vivo*.

4. Discussion

Two major transitions occur during the Pol II transcription cycle, the initiation-elongation transition at the 5'-end of genes, and the elongation-termination transition at the 3'-end of genes, which is coupled to 3'-RNA processing. Whereas the first transition was extensively studied (243), less is known about the second transition. Studies of the second transition revealed a role of the Ser2-phosphorylated Pol II CTD in the recruitment of 3'-processing and termination factors (3). This transition also involves the Paf1 complex (135, 214, 225), elongation factor Spt6 (144), and the transcription regulator Sin1 (122). In this work we provide evidence for a role of the Spt5 CTR in the elongation-termination transition, in particular in the recruitment of the essential mRNA 3'-processing factor CFI.

4.1 Paf1 complex-independent recruitment of CFI

There is evidence that recruitment of RNA 3'-processing factors also involves the Paf1 complex (Paf1C). Deletion of subunits of Paf1C reduces the recruitment of Pcf11 (214) and interferes with Pol II binding of Cft1 (225), a component of the yeast CPF complex. Since Paf1 occupancy levels are markedly reduced in Spt5 ΔCTR cells, it may be argued that recruitment of CFI may occur via Paf1C and the observed reduction of CFI subunit occupancy may result from a loss of Paf1C. However, several lines of evidence argue against this model and instead argue that CFI recruitment occurs via a direct interaction with the CTR. Firstly, Pcf11 occupancy is not altered in Spt5 ΔCTR cells, despite the loss of Paf1C (Figure 31). Secondly, despite extensive interactomics studies, physical interactions between Paf1C and CFI have never been observed. Thirdly, Paf1C clearly dissociates from the Pol II elongation complex upstream of the pA site (72, 155, 199), whereas CFI subunits are mainly recruited downstream of the pA site (Figure 34 a, (72, 155, 199)). These results argue for a Paf1C-independent mechanism of CFI recruitment in yeast.

4.2 Spt5 CTR is not required for normal pA site usage and transcription termination

The genes coding for the CFIA subunits Clp1, Pcf11, Rna14 and Rna15 are essential for viability in yeast (6, 99, 210). Therefore, deletion of these genes cannot be performed for functional studies. To circumvent this problem, temperature-sensitive *rna14* and *rna15*

mutants were generated to reveal that Rna14 and Rna15 are required for RNA cleavage and polyadenylation, mRNA stability, poly(A) tail length, pA site choice and for the sequestration of small nucleolar ribonucleoproteins at discrete sites in the nucleus (3). Since the Rna14 and Rna15 occupancy levels were strongly reduced in cells lacking the CTR of Spt5, we asked whether this reduction could lead to similar 3'-end processing defects *in vivo*. However, our analyses have shown that neither the pA site usage nor transcription termination at selected genes or genome-wide is altered in yeast cells lacking the CTR of Spt5. This is consistent with the non-essential nature of the Spt5 CTR in yeast and may be due to residual CFI recruitment in Spt5 ΔCTR cells that likely results from binding of the CFI subunits to RNA and from the binding of Pcf11 to the Pol II CTD. Pcf11 contains an essential CTD interaction domain (CID), which directly binds the Ser2-phosphorylated CTD (182, 201, 266), and may be responsible for normal recruitment of Pcf11 to genes upon CTR deletion. These observations indicate that CFI is recruited to a defined region downstream of the pA site by co-operative interactions with the Pol II CTD, the Spt5 CTR, and nascent RNA (Figure 36). This model is consistent with the reported binding of Rna14 and Rna15 to the phosphorylated CTD (15), and other published data. Rna15 can be cross-linked to RNA by UV light (150) and contains a RNA recognition motif (RRM) domain that binds GU-rich RNA *in vitro* (181, 233). Hrp1 has two RRMs that bind to AU-containing RNAs (238). The recruitment of CFI within a narrow region downstream of the pA site and the absence at the promoter region of genes in yeast is in contrast to observations made in mammalian cells. In mammals, 3'-RNA processing factors, including homologs of yeast CFI, are already recruited at the promoter region of genes (61, 105, 299). It is currently not known what role, if any, 3'-RNA processing factors may play at the 5'-end of mammalian genes.

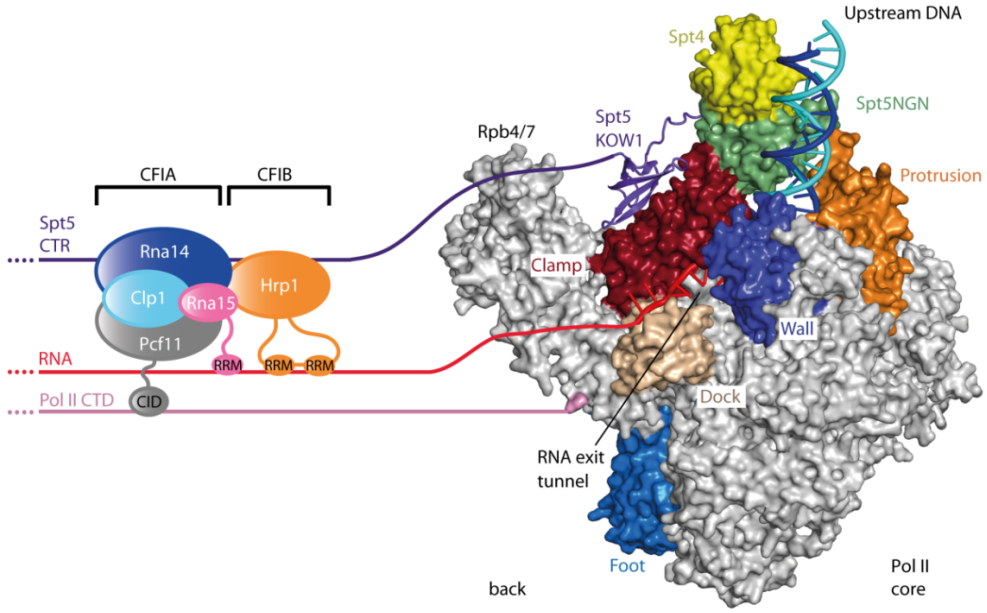

Figure 36: Model of CFI recruitment in yeast. The complete yeast Pol II elongation complex with bound Spt4/5 is viewed from the back (198). Pol II and Spt4/5 are shown as molecular surfaces with key domains highlighted in color and labeled. Exiting RNA, the C-terminal KOW domains and the CTR of Spt5, and the Pol II CTD extend towards the same side around the Rpb4/7 subcomplex, providing the main interfaces for CFI recruitment. Rna14 may directly contact the Spt5 CTR, whereas the RNA is bound by the C-terminal RRM domain of Rna15 and by two internal RRM domains of Hrp1. The Pol II CTD is mainly bound by the N-terminal CID domain of Pcf11. CFI subunits are drawn to scale. Important protein domains are illustrated as extensions from the respective protein core.

4.3 A role of Spt5 in coupling transcription and translation

Finally, our results have implications for understanding the evolution of transcription-coupled events. Spt5 represents the only known RNA polymerase-associated factor that is conserved in all three domains of life (320), and its bacterial homologue is called NusG (118). All Spt5 homologues contain two conserved domains, the NusG N-terminal domain (NGN) and a C-terminal KOW domain (124). Whereas the NGN domain binds to the polymerase clamp domain and closes the active center cleft to render transcription processive, the KOW domain extends from the polymerase surface towards exiting RNA (159, 198, 211) (Figure 36). In bacteria, the KOW domain interacts with the ribosome, thus coupling transcription to mRNA translation (38, 248). In eukaryotes, transcription and translation take place in different cellular compartments, and any coupling between these processes likely occurs via the mRNA that exits the nucleus (127, 197). Our data indicate that the CTR of Spt5 contributes to the coordination of transcription with 3'-RNA processing, which in turn is coupled to mRNA export (140, 192, 254). Since the CTR occurs only in eukaryotic Spt5 homologues, it is likely that it emerged in evolution to maintain coupling between transcription and translation after the spatial separation of these processes. Such coupling may be achieved by co-transcriptional Spt5-dependent loading of mRNA export factors onto the nascent RNA, before its maturation, nuclear export, and translation in the cytosol.

5. Conclusions and Outlook

This work reveals a new role of the Spt5 CTR in the Pol II transcription elongation-termination transition, in particular in the recruitment of the essential 3'-RNA processing factor CFI. ChIP analyses in wild-type and Spt5 ΔCTR yeast cells detected a reduction in the occupancy of Paf1 as was already described by others, but also of CFI subunits, indicating impaired CFI recruitment to the 3'-end of genes *in vivo*. A pull-down assay additionally revealed an interaction between the CFI subunit Rna14 and the Spt5 CTR *in vitro*. RNA contributes to CFI recruitment, as RNAse treatment prior to ChIP further decreases CFI ChIP signals. Genome-wide profiling by ChIP-chip revealed a sharp occupancy peak of the CFI core subunits Rna14 and Rna15 around 100 nt downstream of the pA site, which coincides with high occupancy of Spt5 and the CTD Ser2-phosphorylated form of Pol II. These observations lead to a model, according to which CFI is recruited to a defined region downstream of the pA site of genes by cooperative binding to the Spt5 CTR, nascent RNA and the Ser2-phosphorylated CTD of Pol II. Consistent with this model, the CTR is not required for pA site recognition and transcription termination *in vivo*.

Based on these findings new questions arise that can be tackled in follow-up studies. First, we confirmed that Paf1 interacts with the Spt5 CTR *in vivo*. However, Paf1 represents only one subunit of the Paf1 complex. The Paf1 complex consists of Paf1, Ctr9, Cdc73, Rtf1 and Leo1 in yeast (135). It is currently not known which subunit(s) directly contact the Spt5 CTR. This question can be addressed by in vitro binding assays with CTR peptides. As soon as the

subunit of the Paf1 complex that mediates the interaction with the Spt5 CTR is identified, the binding interface can be characterized in more detail. For example this could be done by co-crystallization trials with the respective subunit and CTR peptides.
Second, our analysis revealed a new interaction between the CFI and the CTR of Spt5. However, the details of this interaction are not known. Since CFI also represents a multiprotein complex, the same approaches, as mentioned for the Paf1 complex, can be applied to characterize the binding interface between CFI and the Spt5 CTR.
Third, our data provides strong evidence for CFI to bind to nascent RNA *in vivo*. One way to characterize the RNA binding interface on a genome-wide scale is by the application of NET-seq and PAR-CLIP. These approaches bear the potential to identify preferred RNA sequence motifs that are bound by the protein of interest.
Fourth, our study indentified CFI as a Spt5 CTR binding complex *in vivo*. It remains to be determined whether other factors also contact the CTR. An unbiased approach would be to combine an affinity-purification assay, for example with a GST-CTR construct, and mass spectrometry. The new interactions that may be discovered by this approach can in turn be investigated in a next round of follow-up projects.
Fifth, the Spt5 CTR can be phosphorylated by the Pol II CTD kinase Bur1 (187). Currently, it is not known whether the CTR is phosphorylated *in vivo* and if so, what the biological implications of this modification are. One way to characterize the post-translational modification status of the Spt5 CTR is via mass spectrometry. Another way would be to generate an antibody that specifically recognizes the phospho-CTR. This antibody can then be used in ChIP experiments to reveal possible changes in the CTR-phosphorylation pattern in the course of the transcription cycle.
The Spt5 CTR of yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) consists of 15 hexapeptide repeats. However, many of those repeats differ strongly from the consensus sequence. Furthermore, the deletion of the Spt5 CTR has almost no phenotypic effects for *S. cervisiae* cells. However, in fission yeast *Schizosaccharomyces pombe* (*S. pombe*) the deletion of the Spt5 CTR leads to a slow-growth phenotype and an abnormal cell morphology (271). In fission yeast the CTR of Spt5 comprises 18 nonamer repeats. In contrast to *S. cerevisiae*, the repeats are much more consensus-like. Additionally, the capping enzyme binds to the Spt5 CTR in *S. pombe* (236) but not in *S. cerevisiae*. Given the obvious differences of the Spt5 CTR function in both model organisms, it would be very worthwhile to extend the analyses to *S. pombe* cells. To figure out the similarities and dissimilarities of the CTR function between budding and fission yeast, would shed light on the evolution of the CTR as well as the co-evolution of CTR and its binding partners.

References

1. **Abruzzi, K. C., S. Lacadie, and M. Rosbash.** 2004. Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. EMBO J **23:**2620-2631.
2. **Ahn, S. H., M.-C. Keogh, and S. Buratowski.** 2009. Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II. EMBO J **28:**205-212.
3. **Ahn, S. H., M. Kim, and S. Buratowski.** 2004. Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3' End Processing. Molecular Cell **13:**67-76.
4. **Akhtar, M. S., M. Heidemann, J. R. Tietjen, D. W. Zhang, R. D. Chapman, D. Eick, and A. Z. Ansari.** 2009. TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. Molecular Cell **34:**387-393.
5. **Alexander, R. D., S. A. Innocente, J. D. Barrass, and J. D. Beggs.** 2010. Splicing-Dependent RNA Polymerase Pausing in Yeast. Molecular Cell **40:**582-593.
6. **Amrani, N., M. Minet, F. Wyers, M. Dufour, L. Aggerbeck, and F. Lacroute.** 1997. PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I. Mol. Cell. Biol. **17:**1102-1109.
7. **Andrulis, E. D., E. Guzmán, P. Döring, J. Werner, and J. T. Lis.** 2000. High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. Genes & Development **14:**2635-2649.
8. **Ansari, A., and M. Hampsey.** 2005. A role for the CPF 3′-end processing machinery in RNAP II-dependent gene looping. Genes & Development **19:**2969-2978.
9. **Aparicio, O., J. V. Geisberg, E. Sekinger, A. Yang, Z. Moqtaderi, and K. Struhl.** 2005. Chromatin Immunoprecipitation for Determining the Association of Proteins with Specific Genomic Sequences In Vivo, Current Protocols in Molecular Biology. John Wiley & Sons, Inc.
10. **Aranda, A., and N. Proudfoot.** 2001. Transcriptional Termination Factors for RNA Polymerase II in Yeast. Molecular Cell **7:**1003-1011.
11. **Asin-Cayuela, J., and C. M. Gustafsson.** 2007. Mitochondrial transcription and its regulation in mammalian cells. Trends in Biochemical Sciences **32:**111-117.
12. **Auerbach, R. K., G. Euskirchen, J. Rozowsky, N. Lamarre-Vincent, Z. Moqtaderi, P. Lefrançois, K. Struhl, M. Gerstein, and M. Snyder.** 2009. Mapping accessible chromatin regions using Sono-Seq. Proceedings of the National Academy of Sciences **106:**14926-14931.
13. **Avvakumov, N., A. Nourani, and J. Côté.** 2011. Histone Chaperones: Modulators of Chromatin Marks. Molecular Cell **41:**502-514.
14. **Bai, Y., T. C. Auperin, C.-Y. Chou, G.-G. Chang, J. L. Manley, and L. Tong.** 2007. Crystal Structure of Murine CstF-77: Dimeric Association and Implications for Polyadenylation of mRNA Precursors. Molecular Cell **25:**863-875.
15. **Barillà, D., B. A. Lee, and N. J. Proudfoot.** 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences **98:**445-450.
16. **Barrandon, C., B. Spiluttini, and O. Bensaude.** 2008. Non-coding RNAs regulating the transcriptional machinery. Biology of the Cell **100:**83-95.
17. **Baskaran, R., G. Chiang, and J. Wang.** 1996. Identification of a binding site in c-Ab1 tyrosine kinase for the C- terminal repeated domain of RNA polymerase II. Mol. Cell. Biol. **16:**3361-3369.
18. **Baskaran, R., G. G. Chiang, T. Mysliwiec, G. D. Kruh, and J. Y. J. Wang.** 1997. Tyrosine Phosphorylation of RNA Polymerase II Carboxyl-terminal Domain by the Abl-related Gene Product. Journal of Biological Chemistry **272:**18905-18909.
19. **Baskaran, R., M. E. Dahmus, and J. Y. Wang.** 1993. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. Proceedings of the National Academy of Sciences **90:**11167-11171.
20. **Baskaran, R., S. R. Escobar, and J. Y. J. Wang.** 1999. Nuclear c-Abl Is a COOH-Terminal Repeated Domain (CTD)-Tyrosine Kinase-specific for the Mammalian RNA Polymerase II: Possible Role in Transcription Elongation. Cell Growth Differ **10:**387-396.
21. **Baskaran, R., L. D. Wood, L. L. Whitaker, C. E. Canman, S. E. Morgan, Y. Xu, C. Barlow, D. Baltimore, A. Wynshaw-Boris, M. B. Kastan, and J. Y. J. Wang.** 1997. Ataxia telangiectasia mutant protein activates c-Abl tyrosine kinase in response to ionizing radiation. Nature **387:**516-519.
22. **Becker, R., B. Loll, and A. Meinhart.** 2008. Snapshots of the RNA Processing Factor SCAF8 Bound to Different Phosphorylated Forms of the Carboxyl-terminal Domain of RNA Polymerase II. Journal of Biological Chemistry **283:**22659-22669.

23. **Belotserkovskaya, R., S. Oh, V. A. Bondarenko, G. Orphanides, V. M. Studitsky, and D. Reinberg.** 2003. FACT Facilitates Transcription-Dependent Nucleosome Alteration. Science **301:**1090-1093.
24. **Belotserkovskaya, R., and D. Reinberg.** 2004. Facts about FACT and transcript elongation through chromatin. Current Opinion in Genetics & Development **14:**139-146.
25. **Bentley, D. L.** 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Current Opinion in Cell Biology **17:**251-256.
26. **Bentley, D. L., and M. Groudine.** 1986. A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. Nature **321:**702-706.
27. **Bernstein, B., C. Liu, E. Humphrey, E. Perlstein, and S. Schreiber.** 2004. Global nucleosome occupancy in yeast. Genome Biology **5:**R62.
28. **Birse, C. E., L. Minvielle-Sebastia, B. A. Lee, W. Keller, and N. J. Proudfoot.** 1998. Coupling Termination of Transcription to Messenger RNA Maturation in Yeast. Science **280:**298-301.
29. **Bishop, A. C., O. Buzko, and K. M. Shokat.** 2001. Magic bullets for protein kinases. Trends in Cell Biology **11:**167-172.
30. **Boeing, S., C. Rigault, M. Heidemann, D. Eick, and M. Meisterernst.** 2010. RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. J Biol Chem **285:**188-196.
31. **Booher, K. R., and P. Kaiser.** 2008. A PCR-based strategy to generate yeast strains expressing endogenous levels of amino-terminal epitope-tagged proteins. Biotechnology Journal **3:**524-529.
32. **Bortvin, A., and F. Winston.** 1996. Evidence That Spt6p Controls Chromatin Structure by a Direct Interaction with Histones. Science **272:**1473-1476.
33. **Bourbon, H.-M.** 2008. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. Nucleic Acids Research **36:**3993-4008.
34. **Bourbon, H.-M., A. Aguilera, A. Z. Ansari, F. J. Asturias, A. J. Berk, S. Bjorklund, T. K. Blackwell, T. Borggrefe, M. Carey, M. Carlson, J. W. Conaway, R. C. Conaway, S. W. Emmons, J. D. Fondell, L. P. Freedman, T. Fukasawa, C. M. Gustafsson, M. Han, X. He, P. K. Herman, A. G. Hinnebusch, S. Holmberg, F. C. Holstege, J. A. Jaehning, Y.-J. Kim, L. Kuras, A. Leutz, J. T. Lis, M. Meisterernest, A. M. Naar, K. Nasmyth, J. D. Parvin, M. Ptashne, D. Reinberg, H. Ronne, I. Sadowski, H. Sakurai, M. Sipiczki, P. W. Sternberg, D. J. Stillman, R. Strich, K. Struhl, J. Q. Svejstrup, S. Tuck, F. Winston, R. G. Roeder, and R. D. Kornberg.** 2004. A Unified Nomenclature for Protein Subunits of Mediator Complexes Linking Transcriptional Regulators to RNA Polymerase II. Molecular Cell **14:**553-557.
35. **Braunstein, M., A. B. Rose, S. G. Holmes, C. D. Allis, and J. R. Broach.** 1993. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. Genes & Development **7:**592-604.
36. **Buratowski, S.** 2003. The CTD code. Nat Struct Mol Biol **10:**679-680.
37. **Buratowski, S.** 2009. Progression through the RNA Polymerase II CTD Cycle. Molecular cell **36:**541-546.
38. **Burmann, B. M., K. Schweimer, X. Luo, M. C. Wahl, B. L. Stitt, M. E. Gottesman, and P. Rösch.** 2010. A NusE:NusG Complex Links Transcription and Translation. Science **328:**501-504.
39. **Carrillo Oesterreich, F., S. Preibisch, and K. M. Neugebauer.** 2010. Global Analysis of Nascent RNA Reveals Transcriptional Pausing in Terminal Exons. Molecular Cell **40:**571-581.
40. **Chao, S.-H., A. L. Greenleaf, and D. H. Price.** 2001. Juglone, an inhibitor of the peptidyl-prolyl isomerase Pin1, also directly blocks transcription. Nucleic Acids Research **29:**767-773.
41. **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:**1780-1782.
42. **Chapman, R. D., M. Heidemann, C. Hintermair, and D. Eick.** 2008. Molecular evolution of the RNA polymerase II CTD. Trends in Genetics **24:**289-296.
43. **Chen, H.-T., L. Warfield, and S. Hahn.** 2007. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nat Struct Mol Biol **14:**696-703.
44. **Chen, H., X. Contreras, Y. Yamaguchi, H. Handa, B. M. Peterlin, and S. Guo.** 2009. Repression of RNA Polymerase II Elongation In Vivo Is Critically Dependent on the C-Terminus of Spt5. PLoS ONE **4:**e6918.
45. **Cheung, A. C. M., and P. Cramer.** 2011. Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature **471:**249-253.
46. **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 & Development **15:**3319-3329.
47. **Choudhary, C., C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen, and M. Mann.** 2009. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. Science **325:**834-840.

48. **Chubb, J. R., T. Trcek, S. M. Shenoy, and R. H. Singer.** 2006. Transcriptional Pulsing of a Developmental Gene. Current Biology **16:**1018-1025.
49. **Churchman, L. S., and J. S. Weissman.** 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature **469:**368-373.
50. **Clapier, C. R., and B. R. Cairns.** 2009. The Biology of Chromatin Remodeling Complexes. Annual Review of Biochemistry **78:**273-304.
51. **Clayton, R. A., O. White, K. A. Ketchum, and J. C. Venter.** 1997. The first genome from the third domain of life. Nature **387:**459-462.
52. **Close, D., S. J. Johnson, M. A. Sdano, S. M. McDonald, H. Robinson, T. Formosa, and C. P. Hill.** 2011. Crystal Structures of the S. cerevisiae Spt6 Core and C-Terminal Tandem SH2 Domain. Journal of Molecular Biology **408:**697-713.
53. **Collins, S. R., K. M. Miller, N. L. Maas, A. Roguev, J. Fillingham, C. S. Chu, M. Schuldiner, M. Gebbia, J. Recht, M. Shales, H. Ding, H. Xu, J. Han, K. Ingvarsdottir, B. Cheng, B. Andrews, C. Boone, S. L. Berger, P. Hieter, Z. Zhang, G. W. Brown, C. J. Ingles, A. Emili, C. D. Allis, D. P. Toczyski, J. S. Weissman, J. F. Greenblatt, and N. J. Krogan.** 2007. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature **446:**806-810.
54. **Connelly, S., and J. L. Manley.** 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes & Development **2:**440-452.
55. **Corden, J. L.** 2007. Seven Ups the Code. Science **318:**1735-1736.
56. **Core, L. J., J. J. Waterfall, and J. T. Lis.** 2008. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. Science **322:**1845-1848.
57. **Cramer, P., K. J. Armache, S. Baumli, S. Benkert, F. Brueckner, C. Buchen, G. E. Damsma, S. Dengi, S. R. Geiger, A. J. Jasiak, A. Jawhari, S. Jennebach, T. Kamenski, H. Kettenberger, C. D. Kuhn, E. Lehmann, K. Leike, J. F. Sydow, and A. Vannini.** 2008. Structure of Eukaryotic RNA Polymerases. Annual Review of Biophysics **37:**337-352.
58. **Cramer, P., D. A. Bushnell, and R. D. Kornberg.** 2001. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. Science **292:**1863-1876.
59. **Crawford, G. E., S. Davis, P. C. Scacheri, G. Renaud, M. J. Halawi, M. R. Erdos, R. Green, P. S. Meltzer, T. G. Wolfsberg, and F. S. Collins.** 2006. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nat Meth **3:**503-509.
60. **Creamer, T. J., M. M. Darby, N. Jamonnak, P. Schaughency, H. Hao, S. J. Wheelan, and J. L. Corden.** 2011. Transcriptome-Wide Binding Sites for Components of the Saccharomyces cerevisiae Non-Poly(A) Termination Pathway: Nrd1, Nab3, and Sen1. PLoS Genet **7:**e1002329.
61. **Dantonel, J.-C., K. G. K. Murthy, J. L. Manley, and L. Tora.** 1997. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. Nature **389:**399-402.
62. **Darzacq, X., Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer.** 2007. In vivo dynamics of RNA polymerase II transcription. Nat Struct Mol Biol **14:**796-806.
63. **David, B.** 2002. The mRNA assembly line: transcription and processing machines in the same factory. Current Opinion in Cell Biology **14:**336-342.
64. **David, L., W. Huber, M. Granovskaia, J. Toedling, C. J. Palm, L. Bofkin, T. Jones, R. W. Davis, and L. M. Steinmetz.** 2006. A high-resolution map of transcription in the yeast genome. Proceedings of the National Academy of Sciences **103:**5320-5325.
65. **de Godoy, L. M. F., J. V. Olsen, J. Cox, M. L. Nielsen, N. C. Hubner, F. Frohlich, T. C. Walther, and M. Mann.** 2008. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature **455:**1251-1254.
66. **de Wit, E., F. Greil, and B. van Steensel.** 2007. High-Resolution Mapping Reveals Links of HP1 with Active and Inactive Chromatin Components. PLoS Genet **3:**e38.
67. **Dedon, P. C., J. A. Soults, D. C. Allis, and M. A. Gorovsky.** 1991. A simplified formaldehyde fixation and immunoprecipitation technique for studying protein-DNA interactions. Analytical Biochemistry **197:**83-90.
68. **Dekker, J., K. Rippe, M. Dekker, and N. Kleckner.** 2002. Capturing Chromosome Conformation. Science **295:**1306-1311.
69. **Dengl, S., and P. Cramer.** 2009. Torpedo Nuclease Rat1 Is Insufficient to Terminate RNA Polymerase II in Vitro. Journal of Biological Chemistry **284:**21270-21279.
70. **Dengl, S., A. Mayer, M. Sun, and P. Cramer.** 2009. Structure and in Vivo Requirement of the Yeast Spt6 SH2 Domain. Journal of Molecular Biology **389:**211-225.
71. **DeRisi, J. L., V. R. Iyer, and P. O. Brown.** 1997. Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. Science **278:**680-686.
72. **Dermody, J. L., and S. Buratowski.** 2010. Leo1 Subunit of the Yeast Paf1 Complex Binds RNA and Contributes to Complex Recruitment. Journal of Biological Chemistry **285:**33671-33679.

73. **Diebold, M.-L., E. Loeliger, M. Koch, F. Winston, J. Cavarelli, and C. Romier.** 2010. Noncanonical Tandem SH2 Enables Interaction of Elongation Factor Spt6 with RNA Polymerase II. Journal of Biological Chemistry **285:**38389-38398.
74. **Ding, B., D. LeJeune, and S. Li.** 2010. The C-terminal Repeat Domain of Spt5 Plays an Important Role in Suppression of Rad26-independent Transcription Coupled Repair. Journal of Biological Chemistry **285:**5317-5326.
75. **Drouin, S., L. Laramée, P.-É. Jacques, A. Forest, M. Bergeron, and F. Robert.** 2010. DSIF and RNA Polymerase II CTD Phosphorylation Coordinate the Recruitment of Rpd3S to Actively Transcribed Genes. PLoS Genet **6:**e1001173.
76. **Dutrow, N., D. A. Nix, D. Holt, B. Milash, B. Dalley, E. Westbroek, T. J. Parnell, and B. R. Cairns.** 2008. Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. Nat Genet **40:**977-986.
77. **Duyster, J., R. Baskaran, and J. Y. Wang.** 1995. Src homology 2 domain as a specificity determinant in the c-Abl-mediated tyrosine phosphorylation of the RNA polymerase II carboxyl-terminal repeated domain. Proceedings of the National Academy of Sciences **92:**1555-1559.
78. **Egelhofer, T. A., A. Minoda, S. Klugman, K. Lee, P. Kolasinska-Zwierz, A. A. Alekseyenko, M.-S. Cheung, D. S. Day, S. Gadel, A. A. Gorchakov, T. Gu, P. V. Kharchenko, S. Kuan, I. Latorre, D. Linder-Basso, Y. Luu, Q. Ngo, M. Perry, A. Rechtsteiner, N. C. Riddle, Y. B. Schwartz, G. A. Shanower, A. Vielle, J. Ahringer, S. C. R. Elgin, M. I. Kuroda, V. Pirrotta, B. Ren, S. Strome, P. J. Park, G. H. Karpen, R. D. Hawkins, and J. D. Lieb.** 2011. An assessment of histone-modification antibody quality. Nat Struct Mol Biol **18:**91-93.
79. **Egloff, S., and S. Murphy.** 2008. Cracking the RNA polymerase II CTD code. Trends in Genetics **24:**280-288.
80. **Egloff, S., D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick, and S. Murphy.** 2007. Serine-7 of the RNA Polymerase II CTD Is Specifically Required for snRNA Gene Expression. Science **318:**1777-1779.
81. **Egloff, S., S. A. Szczepaniak, M. Dienstbier, A. Taylor, S. Knight, and S. Murphy.** 2010. The Integrator Complex Recognizes a New Double Mark on the RNA Polymerase II Carboxyl-terminal Domain. Journal of Biological Chemistry **285:**20564-20569.
82. **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. Molecular Cell **11:**1549-1561.
83. **Fan, X., N. Lamarre-Vincent, Q. Wang, and K. Struhl.** 2008. Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments. Nucleic Acids Research **36:**e125.
84. **Fath, S., M. S. Kobor, A. Philippi, J. Greenblatt, and H. Tschochner.** 2004. Dephosphorylation of RNA Polymerase I by Fcp1p Is Required for Efficient rRNA Synthesis. Journal of Biological Chemistry **279:**25251-25259.
85. **Ficarro, S. B., M. L. McCleland, P. T. Stukenberg, D. J. Burke, M. M. Ross, J. Shabanowitz, D. F. Hunt, and F. M. White.** 2002. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotech **20:**301-305.
86. **Fish, R. N., and C. M. Kane.** 2002. Promoting elongation with transcript cleavage stimulatory factors. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression **1577:**287-307.
87. **Fodor, S., J. Read, M. Pirrung, L. Stryer, A. Lu, and D. Solas.** 1991. Light-directed, spatially addressable parallel chemical synthesis. Science **251:**767-773.
88. **Frederiks, F., I. J. E. Stulemeijer, H. Ovaa, and F. van Leeuwen.** 2011. A Modified Epigenetics Toolbox to Study Histone Modifications on the Nucleosome Core. ChemBioChem **12:**308-313.
89. **Frohman, M. A., M. K. Dush, and G. R. Martin.** 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proceedings of the National Academy of Sciences **85:**8998-9002.
90. **Gansner, E., and S. North.** 1998. Improved Force-Directed LayoutsGraph Drawing, p. 364-373. *In* S. Whitesides (ed.), vol. 1547. Springer Berlin / Heidelberg.
91. **García-Martínez, J., A. n. Aranda, and J. E. Pérez-Ortín.** 2004. Genomic Run-On Evaluates Transcription Rates for All Yeast Genes and Identifies Gene Regulatory Mechanisms. Molecular Cell **15:**303-313.
92. **Gaulton, K. J., T. Nammo, L. Pasquali, J. M. Simon, P. G. Giresi, M. P. Fogarty, T. M. Panhuis, P. Mieczkowski, A. Secchi, D. Bosco, T. Berney, E. Montanya, K. L. Mohlke, J. D. Lieb, and J. Ferrer.** 2010. A map of open chromatin in human pancreatic islets. Nat Genet **42:**255-259.
93. **Gavin, A.-C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A.-M. Michon, C.-M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.-A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M.**

Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature **415:**141-147.

94. **Gentleman, R., V. Carey, D. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Yang, and J. Zhang.** 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology **5:**R80.
95. **Gerber, S. A., J. Rush, O. Stemman, M. W. Kirschner, and S. P. Gygi.** 2003. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proceedings of the National Academy of Sciences **100:**6940-6945.
96. **Ghavi-Helm, Y., M. Michaut, J. Acker, J.-C. Aude, P. Thuriaux, M. Werner, and J. Soutourina.** 2008. Genome-wide location analysis reveals a role of TFIIS in RNA polymerase III transcription. Genes & Development **22:**1934-1947.
97. **Gheldof, N., M. Leleu, D. Noordermeer, J. Rougemont, and A. Reymond.** 2012. Detecting Long-Range Chromatin Interactions Using the Chromosome Conformation Capture Sequencing (4C-seq) MethodGene Regulatory Networks, p. 211-225. *In* B. Deplancke and N. Gheldof (ed.), vol. 786. Humana Press.
98. **Ghosh, A., S. Shuman, and Christopher D. Lima.** 2011. Structural Insights to How Mammalian Capping Enzyme Reads the CTD Code. Molecular Cell **43:**299-310.
99. **Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A. P. Arkin, A. Astromoff, M. El Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.-D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kotter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C.-y. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis, and M. Johnston.** 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature **418:**387-391.
100. **Gilmour, D. S., and R. Fan.** 2009. Detecting transcriptionally engaged RNA polymerase in eukaryotic cells with permanganate genomic footprinting. Methods **48:**368-374.
101. **Gilmour, D. S., and J. T. Lis.** 1984. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. Proceedings of the National Academy of Sciences **81:**4275-4279.
102. **Gilmour, D. S., and J. T. Lis.** 1985. In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. Molecular and Cellular Biology **5:**2009-2018.
103. **Gilmour, D. S., and J. T. Lis.** 1986. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Molecular and Cellular Biology **6:**3984-3989.
104. **Giresi, P. G., J. Kim, R. M. McDaniell, V. R. Iyer, and J. D. Lieb.** 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Research **17:**877-885.
105. **Glover-Cutter, K., S. Kim, J. Espinosa, and D. L. Bentley.** 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat Struct Mol Biol **15:**71-78.
106. **Gnad, F., L. M. F. de Godoy, J. Cox, N. Neuhauser, S. Ren, J. V. Olsen, and M. Mann.** 2009. High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. PROTEOMICS **9:**4642-4652.
107. **Goodrich, J., and J. Kugel.** 2010. Genome-wide insights into eukaryotic transcriptional control. Genome Biology **11:**305.
108. **Gross, S., and C. Moore.** 2001. Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. Proceedings of the National Academy of Sciences **98:**6080-6085.
109. **Gross, S., and C. L. Moore.** 2001. Rna15 Interaction with the A-Rich Yeast Polyadenylation Signal Is an Essential Step in mRNA 3'-End Formation. Mol. Cell. Biol. **21:**8045-8055.
110. **Guenther, M. G., S. S. Levine, L. A. Boyer, R. Jaenisch, and R. A. Young.** 2007. A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. Cell **130:**77-88.
111. **Guglielmi, B., J. Soutourina, C. Esnault, and M. Werner.** 2007. TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. Proceedings of the National Academy of Sciences **104:**16062-16067.

112. **Guo, M., F. Xu, J. Yamada, T. Egelhofer, Y. Gao, G. A. Hartzog, M. Teng, and L. Niu.** 2008. Core Structure of the Yeast Spt4-Spt5 Complex: A Conserved Module for Regulation of Transcription Elongation. **16:**1649-1658.
113. **Hafner, M., M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano Jr, A.-C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan, and T. Tuschl.** 2010. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell **141:**129-141.
114. **Hahn, S.** 2004. Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol **11:**394-403.
115. **Hampsey, M., B. N. Singh, A. Ansari, J.-P. Lainé, and S. Krishnamurthy.** 2011. Control of eukaryotic gene expression: Gene loops and transcriptional memory. Advances in Enzyme Regulation **51:**118-125.
116. **Hanlon, S. E., and J. D. Lieb.** 2004. Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. Current Opinion in Genetics & Development **14:**697-705.
117. **Harbison, C. T., D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J.-B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, and R. A. Young.** 2004. Transcriptional regulatory code of a eukaryotic genome. Nature **431:**99-104.
118. **Hartzog, G. A., T. Wada, H. Handa, and F. Winston.** 1998. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II inSaccharomyces cerevisiae. Genes & Development **12:**357-369.
119. **He, X., A. U. Khan, H. Cheng, D. L. Pappas, M. Hampsey, and C. L. Moore.** 2003. Functional interactions between the transcription and mRNA 3′ end processing machineries mediated by Ssu72 and Sub1. Genes & Development **17:**1030-1042.
120. **Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford, and B. Ren.** 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet **39:**311-318.
121. **Hengartner, C. J., V. E. Myer, S.-M. Liao, C. J. Wilson, S. S. Koh, and R. A. Young.** 1998. Temporal Regulation of RNA Polymerase II by Srb10 and Kin28 Cyclin-Dependent Kinases. Molecular Cell **2:**43-53.
122. **Hershkovits, G., H. Bangio, R. Cohen, and D. J. Katcoff.** 2006. Recruitment of mRNA cleavage/polyadenylation machinery by the yeast chromatin protein Sin1p/Spt2p. Proceedings of the National Academy of Sciences **103:**9808-9813.
123. **Hesselberth, J. R., X. Chen, Z. Zhang, P. J. Sabo, R. Sandstrom, A. P. Reynolds, R. E. Thurman, S. Neph, M. S. Kuehn, W. S. Noble, S. Fields, and J. A. Stamatoyannopoulos.** 2009. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Meth **6:**283-289.
124. **Hirtreiter, A., G. E. Damsma, A. C. M. Cheung, D. Klose, D. Grohmann, E. Vojnic, A. C. R. Martin, P. Cramer, and F. Werner.** 2010. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. Nucleic Acids Research **38:**4040-4051.
125. **Ho, J., E. Bishop, P. Karchenko, N. Negre, K. White, and P. Park.** 2011. ChIP-chip versus ChIP-seq: Lessons for experimental design and data analysis. BMC Genomics **12:**134.
126. **Ho, Y., A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S.-L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A. R. Willems, H. Sassi, P. A. Nielsen, K. J. Rasmussen, J. R. Andersen, L. E. Johansen, L. H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B. D. Sorensen, J. Matthiesen, R. C. Hendrickson, F. Gleeson, T. Pawson, M. F. Moran, D. Durocher, M. Mann, C. W. V. Hogue, D. Figeys, and M. Tyers.** 2002. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature **415:**180-183.
127. **Hocine, S., R. H. Singer, and D. Grünwald.** 2010. RNA Processing and Export. Cold Spring Harbor Perspectives in Biology **2**.
128. **Hsin, J.-P., A. Sheth, and J. L. Manley.** 2011. RNAP II CTD Phosphorylated on Threonine-4 Is Required for Histone mRNA 3′ End Processing. Science **334:**683-686.
129. **Huh, W.-K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman, and E. K. O'Shea.** 2003. Global analysis of protein localization in budding yeast. Nature **425:**686-691.
130. **Ihmels, J., G. Friedlander, S. Bergmann, O. Sarig, Y. Ziv, and N. Barkai.** 2002. Revealing modular organization in the yeast transcriptional network. Nat Genet **31:**370-377.
131. **Ingolia, N. T., S. Ghaemmaghami, J. R. S. Newman, and J. S. Weissman.** 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science **324:**218-223.

132. **Ito, T., K. Tashiro, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K. Yamamoto, S. Kuhara, and Y. Sakaki.** 2000. Toward a protein–protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proceedings of the National Academy of Sciences **97:**1143-1147.
133. **Ivanovska, I., P.-É. Jacques, O. J. Rando, F. Robert, and F. Winston.** 2011. Control of Chromatin Structure by Spt6: Different Consequences in Coding and Regulatory Regions. Molecular and Cellular Biology **31:**531-541.
134. **Iyer, V. R., C. E. Horak, C. S. Scafe, D. Botstein, M. Snyder, and P. O. Brown.** 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature **409:**533-538.
135. **Jaehning, J. A.** The Paf1 complex: Platform or player in RNA polymerase II transcription? Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1799:**379-388.
136. **Jasiak, A. J., H. Hartmann, E. Karakasili, M. Kalocsay, A. Flatley, E. Kremmer, K. Strässer, D. E. Martin, J. Söding, and P. Cramer.** 2008. Genome-associated RNA Polymerase II Includes the Dissociable Rpb4/7 Subcomplex. Journal of Biological Chemistry **283:**26423-26427.
137. **Jesper Q, S.** 2002. Chromatin elongation factors. Current Opinion in Genetics & Development **12:**156-161.
138. **Jiang, C., and B. F. Pugh.** 2009. A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. Genome Biology **10:**R109.
139. **Jiang, C., and B. F. Pugh.** 2009. Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet **10:**161-172.
140. **Johnson, S. A., G. Cubberley, and D. L. Bentley.** 2009. Cotranscriptional Recruitment of the mRNA Export Factor Yra1 by Direct Interaction with the 3' End Processing Factor Pcf11. Molecular Cell **33:**215-226.
141. **Jonathan R, W.** 1999. The economics of ribosome biosynthesis in yeast. Trends in Biochemical Sciences **24:**437-440.
142. **Judith A, J.** The Paf1 complex: Platform or player in RNA polymerase II transcription? Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1799:**379-388.
143. **Jung, J., Y.-J. Ahn, and L.-W. Kang.** 2008. A novel approach to investigating protein/protein interactions and their functions by TAP-tagged yeast strains and its application to examine yeast transcription machinery. Journal of Microbiology and Biotechnology **18:**631-638.
144. **Kaplan, C. D., M. J. Holland, and F. Winston.** 2005. Interaction between Transcription Elongation Factors and mRNA 3′-End Formation at the Saccharomyces cerevisiae GAL10-GAL7 Locus. Journal of Biological Chemistry **280:**913-922.
145. **Kaplan, C. D., L. Laprade, and F. Winston.** 2003. Transcription Elongation Factors Repress Transcription Initiation from Cryptic Sites. Science **301:**1096-1099.
146. **Kaplan, C. D., J. R. Morris, C.-t. Wu, and F. Winston.** 2000. Spt5 and Spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. Genes & Development **14:**2623-2634.
147. **Kelly, W. G., M. E. Dahmus, and G. W. Hart.** 1993. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. Journal of Biological Chemistry **268:**10416-10424.
148. **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:**7005-7018.
149. **Kessler, M. M., M. F. Henry, E. Shen, J. Zhao, S. Gross, P. A. Silver, and C. L. Moore.** 1997. Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3′-end formation in yeast. Genes & Development **11:**2545-2556.
150. **Kessler, M. M., J. Zhao, and C. L. Moore.** 1996. Purification of the Saccharomyces cerevisiae Cleavage/Polyadenylation Factor I. Journal of Biological Chemistry **271:**27167-27175.
151. **Kettenberger, H., K.-J. Armache, and P. Cramer.** 2003. Architecture of the RNA Polymerase II-TFIIS Complex and Implications for mRNA Cleavage. Cell **114:**347-357.
152. **Kiely, C. M., S. Marguerat, J. F. Garcia, H. D. Madhani, J. Bähler, and F. Winston.** 2011. Spt6 Is Required for Heterochromatic Silencing in the Fission Yeast Schizosaccharomyces pombe. Molecular and Cellular Biology **31:**4193-4204.
153. **Kim, B., A. I. Nesvizhskii, P. G. Rani, S. Hahn, R. Aebersold, and J. A. Ranish.** 2007. The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. Proceedings of the National Academy of Sciences **104:**16068-16073.
154. **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:**1279-1286.
155. **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[prime] ends of genes. EMBO J **23:**354-364.

156. **Kim, M., N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedea, J. F. Greenblatt, and S. Buratowski.** 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature **432:**517-522.
157. **Kim, M., H. Suh, E.-J. Cho, and S. Buratowski.** 2009. Phosphorylation of the Yeast Rpb1 C-terminal Domain at Serines 2, 5, and 7. Journal of Biological Chemistry **284:**26421-26426.
158. **Kim, M., L. Vasiljeva, O. J. Rando, A. Zhelkovsky, C. Moore, and S. Buratowski.** 2006. Distinct Pathways for snoRNA and mRNA Termination. Molecular cell **24:**723-734.
159. **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. Proceedings of the National Academy of Sciences **108:**546-550.
160. **Knowlton, J. R., M. Bubunenko, M. Andrykovitch, W. Guo, K. M. Routzahn, D. S. Waugh, D. L. Court, and X. Ji.** 2003. A Spring-Loaded State of NusG in Its Functional Cycle Is Suggested by X-ray Crystallography and Supported by Site-Directed Mutants. Biochemistry **42:**2275-2281.
161. **Kobor, M. S., and J. Greenblatt.** 2002. Regulation of transcription elongation by phosphorylation. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression **1577:**261-275.
162. **Komarnitsky, P., E.-J. Cho, and S. Buratowski.** 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & Development **14:**2452-2460.
163. **Kostrewa, D., M. E. Zeller, K. J. Armache, M. Seizl, K. Leike, M. Thomm, and P. Cramer.** 2009. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature **462:**323-330.
164. **Krishnamurthy, S., X. He, M. Reyes-Reyes, C. Moore, and M. Hampsey.** 2004. Ssu72 Is an RNA Polymerase II CTD Phosphatase. Molecular cell **14:**387-394.
165. **Kristjuhan, A., and J. Q. Svejstrup.** 2004. Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo. EMBO J **23:**4243-4252.
166. **Krogan, N. J., G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrín-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Y. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili, and J. F. Greenblatt.** 2006. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature **440:**637-643.
167. **Krogan, N. J., M. Kim, S. H. Ahn, G. Zhong, M. S. Kobor, G. Cagney, A. Emili, A. Shilatifard, S. Buratowski, and J. F. Greenblatt.** 2002. RNA Polymerase II Elongation Factors of Saccharomyces cerevisiae: a Targeted Proteomics Approach. Mol. Cell. Biol. **22:**6979-6992.
168. **Kuehner, J. N., and D. A. Brow.** 2006. Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. J Biol Chem **281:**14119-14128.
169. **Kuehner, J. N., E. L. Pearson, and C. Moore.** 2011. Unravelling the means to an end: RNA polymerase II transcription termination. Nat Rev Mol Cell Biol **12:**283-294.
170. **Kulaeva, O. I., D. A. Gaykalova, and V. M. Studitsky.** 2007. Transcription through chromatin by RNA polymerase II: Histone displacement and exchange. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **618:**116-129.
171. **Kulish, D., and K. Struhl.** 2001. TFIIS Enhances Transcriptional Elongation through an Artificial Arrest Site In Vivo. Molecular and Cellular Biology **21:**4162-4168.
172. **Kumar, A., and M. Snyder.** 2001. Emerging technologies in yeast genomics. Nat Rev Genet **2:**302-312.
173. **Kung, L. A., S.-C. Tao, J. Qian, M. G. Smith, M. Snyder, and H. Zhu.** 2009. Global analysis of the glycoproteome in Saccharomyces cerevisiae reveals new roles for protein glycosylation in eukaryotes. Mol Syst Biol **5**.
174. **Laemmli, U. K.** 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature **227:**680-685.
175. **Lantermann, A. B., T. Straub, A. Stralfors, G.-C. Yuan, K. Ekwall, and P. Korber.** 2010. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nat Struct Mol Biol **17:**251-257.
176. **Laribee, R. N., N. J. Krogan, T. Xiao, Y. Shibata, T. R. Hughes, J. F. Greenblatt, and B. D. Strahl.** 2005. BUR Kinase Selectively Regulates H3 K4 Trimethylation and H2B Ubiquitylation through Recruitment of the PAF Elongation Complex. Current Biology **15:**1487-1493.
177. **Lashkari, D. A., J. L. DeRisi, J. H. McCusker, A. F. Namath, C. Gentile, S. Y. Hwang, P. O. Brown, and R. W. Davis.** 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proceedings of the National Academy of Sciences **94:**13057-13062.
178. **Lee, C.-K., Y. Shibata, B. Rao, B. D. Strahl, and J. D. Lieb.** 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. Nature Genetics **36:**900-905.

179. **Lee, T. I., N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young.** 2002. Transcriptional Regulatory Networks in Saccharomyces cerevisiae. Science **298:**799-804.
180. **Lee, W., D. Tillo, N. Bray, R. H. Morse, R. W. Davis, T. R. Hughes, and C. Nislow.** 2007. A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet **39:**1235-1244.
181. **Leeper, T. C., X. Qu, C. Lu, C. Moore, and G. Varani.** 2010. Novel Protein-Protein Contacts Facilitate mRNA 3'-Processing Signal Recognition by Rna15 and Hrp1. Journal of Molecular Biology **401:**334-349.
182. **Licatalosi, D. D., G. Geiger, M. Minet, S. Schroeder, K. Cilli, J. B. McNeil, and D. L. Bentley.** 2002. Functional Interaction of Yeast Pre-mRNA 3' End Processing Factors with RNA Polymerase II. Molecular Cell **9:**1101-1111.
183. **Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker.** 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science **326:**289-293.
184. **Lindstrom, D. L., S. L. Squazzo, N. Muster, T. A. Burckin, K. C. Wachter, C. A. Emigh, J. A. McCleery, J. R. Yates, III, and G. A. Hartzog.** 2003. Dual Roles for Spt5 in Pre-mRNA Processing and Transcription Elongation Revealed by Identification of Spt5-Associated Proteins. Mol. Cell. Biol. **23:**1368-1378.
185. **Liu, J., J. Zhang, Q. Gong, P. Xiong, H. Huang, B. Wu, G. Lu, J. Wu, and Y. Shi.** 2011. Solution Structure of Tandem SH2 Domains from Spt6 Protein and Their Binding to the Phosphorylated RNA Polymerase II C-terminal Domain. Journal of Biological Chemistry **286:**29218-29226.
186. **Liu, P., J. M. Kenney, J. W. Stiller, and A. L. Greenleaf.** 2010. Genetic Organization, Length Conservation, and Evolution of RNA Polymerase II Carboxyl-Terminal Domain. Molecular Biology and Evolution **27:**2628-2641.
187. **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:**4852-4863.
188. **Liu, Z.-G., R. Baskaran, E. T. Lea-Chou, L. D. Wood, Y. Chen, M. Karin, and J. Y. J. Wang.** 1996. Three distinct signalling responses by murine fibroblasts to genotoxic stress. Nature **384:**273-276.
189. **Logan, J., E. Falck-Pedersen, J. E. Darnell, and T. Shenk.** 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. Proceedings of the National Academy of Sciences **84:**8306-8310.
190. **Longtine, M. S., A. McKenzie Iii, D. J. Demarini, N. G. Shah, A. Wach, A. Brachat, P. Philippsen, and J. R. Pringle.** 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast **14:**953-961.
191. **Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond.** 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389:**251-260.
192. **Luna, R., H. Gaillard, C. González-Aguilera, and A. Aguilera.** 2008. Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. Chromosoma **117:**319-331.
193. **Lunde, B. M., S. L. Reichow, M. Kim, H. Suh, T. C. Leeper, F. Yang, H. Mutschler, S. Buratowski, A. Meinhart, and G. Varani.** 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat Struct Mol Biol **17:**1195-1201.
194. **Mandart, E., and R. Parker.** 1995. Effects of mutations in the Saccharomyces cerevisiae RNA14, RNA15, and PAP1 genes on polyadenylation in vivo. Mol. Cell. Biol. **15:**6979-6986.
195. **Mandel, C., Y. Bai, and L. Tong.** 2008. Protein factors in pre-mRNA 3′-end processing. Cellular and Molecular Life Sciences **65:**1099-1122.
196. **Margaritis, T., and F. C. P. Holstege.** 2008. Poised RNA Polymerase II Gives Pause for Thought. Cell **133:**581-584.
197. **Martin, W., and E. V. Koonin.** 2006. Introns and the origin of nucleus–cytosol compartmentalization. Nature **440:**41-45.
198. **Martinez-Rucobo, F. W., S. Sainsbury, A. C. M. Cheung, and P. Cramer.** 2011. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J **30:**1302-1310.
199. **Mayer, A., M. Lidschreiber, M. Siebert, K. Leike, J. Soding, and P. Cramer.** 2010. Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol **17:**1272-1278.
200. **McDonald, S. M., D. Close, H. Xin, T. Formosa, and C. P. Hill.** 2010. Structure and Biological Importance of the Spn1-Spt6 Interaction, and Its Regulatory Role in Nucleosome Binding. Molecular Cell **40:**725-735.

201. **Meinhart, A., and P. Cramer.** 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3[prime]-RNA-processing factors. Nature **430:**223-226.
202. **Meinhart, A., T. Kamenski, S. Hoeppner, S. Baumli, and P. Cramer.** 2005. A structural perspective of CTD function. Genes & Development **19:**1401-1415.
203. **Meinhart, A., T. Silberzahn, and P. Cramer.** 2003. The mRNA Transcription/Processing Factor Ssu72 Is a Potential Tyrosine Phosphatase. Journal of Biological Chemistry **278:**15917-15921.
204. **Mercer, Tim R., S. Neph, Marcel E. Dinger, J. Crawford, Martin A. Smith, A.-Marie J. Shearwood, E. Haugen, Cameron P. Bracken, O. Rackham, John A. Stamatoyannopoulos, A. Filipovska, and John S. Mattick.** 2011. The Human Mitochondrial Transcriptome. Cell **146:**645-658.
205. **Miller, C., B. Schwalb, K. Maier, D. Schulz, S. Dumcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dolken, D. E. Martin, A. Tresch, and P. Cramer.** 2011. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol **7**.
206. **Millevoi, S., and S. Vagner.** 2010. Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. Nucleic Acids Research **38:**2757-2774.
207. **Minvielle-Sebastia, L., K. Beyer, A. M. Krecic, R. E. Hector, M. S. Swanson, and W. Keller.** 1998. Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. EMBO J **17:**7454-7468.
208. **Minvielle-Sebastia, L., P. Preker, and W. Keller.** 1994. RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. Science **266:**1702-1705.
209. **Minvielle-Sebastia, L., P. J. Preker, T. Wiederkehr, Y. Strahm, and W. Keller.** 1997. The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3′-end formation. Proceedings of the National Academy of Sciences **94:**7897-7902.
210. **Minvielle-Sebastia, L., B. Winsor, N. Bonneaud, and F. Lacroute.** 1991. Mutations in the yeast RNA14 and RNA15 genes result in an abnormal mRNA decay rate; sequence analysis reveals an RNA-binding domain in the RNA15 protein. Mol. Cell. Biol. **11:**3075-3087.
211. **Mooney, R. A., K. Schweimer, P. Rösch, M. Gottesman, and R. Landick.** 2009. Two Structurally Independent Domains of E. coli NusG Create Regulatory Plasticity via Distinct Interactions with RNA Polymerase and Regulators. Journal of Molecular Biology **391:**341-358.
212. **Morris, D. P., H. P. Phatnani, and A. L. Greenleaf.** 1999. Phospho-Carboxyl-Terminal Domain Binding and the Role of a Prolyl Isomerase in Pre-mRNA 3′-End Formation. Journal of Biological Chemistry **274:**31583-31587.
213. **Mosley, A. L., S. G. Pattenden, M. Carey, S. Venkatesh, J. M. Gilmore, L. Florens, J. L. Workman, and M. P. Washburn.** 2009. Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation. Molecular Cell **34:**168-178.
214. **Mueller, C. L., S. E. Porter, M. G. Hoffman, and J. A. Jaehning.** 2004. The Paf1 Complex Has Functions Independent of Actively Transcribing RNA Polymerase II. Molecular Cell **14:**447-456.
215. **Murad, A. M. A., P. R. Lee, I. D. Broadbent, C. J. Barelle, and A. J. P. Brown.** 2000. CIp10, an efficient and convenient integrating vector for Candida albicans. Yeast **16:**325-327.
216. **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:**4089-4096.
217. **Muse, G. W., D. A. Gilchrist, S. Nechaev, R. Shah, J. S. Parker, S. F. Grissom, J. Zeitlinger, and K. Adelman.** 2007. RNA polymerase is poised for activation across the genome. Nat Genet **39:**1507-1511.
218. **N.J, P.** 1989. How RNA polymerase II terminates transcription in higher eukaryotes. Trends in Biochemical Sciences **14:**105-110.
219. **Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder.** 2008. The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. Science **320:**1344-1349.
220. **Nechaev, S., and K. Adelman.** 2011. Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1809:**34-45.
221. **Nechaev, S., D. C. Fargo, G. dos Santos, L. Liu, Y. Gao, and K. Adelman.** 2010. Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila. Science **327:**335-338.
222. **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. Molecular Cell **11:**709-719.
223. **Niedringhaus, T. P., D. Milanova, M. B. Kerby, M. P. Snyder, and A. E. Barron.** 2011. Landscape of Next-Generation Sequencing Technologies. Analytical Chemistry **83:**4327-4341.
224. **Noble, C. G., P. A. Walker, L. J. Calder, and I. A. Taylor.** 2004. Rna14–Rna15 assembly mediates the RNA-binding capability of Saccharomyces cerevisiae cleavage factor IA. Nucleic Acids Research **32:**3364-3375.

225. **Nordick, K., M. G. Hoffman, J. L. Betz, and J. A. Jaehning.** 2008. Direct Interactions between the Paf1 Complex and a Cleavage and Polyadenylation Factor Are Revealed by Dissociation of Paf1 from RNA Polymerase II. Eukaryotic Cell **7:**1158-1167.
226. **O'Geen, H., C. M. Nicolet, K. Blahnik, R. Green, and P. J. Farnham.** 2006. Comparison of sample preparation methods for ChIP-chip assays. BioTechniques **41:**577-580.
227. **Oler, A. J., R. K. Alla, D. N. Roberts, A. Wong, P. C. Hollenhorst, K. J. Chandler, P. A. Cassiday, C. A. Nelson, C. H. Hagedorn, B. J. Graves, and B. R. Cairns.** 2010. Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. Nat Struct Mol Biol **17:**620-628.
228. **Olins, D. E., and A. L. Olins.** 2003. Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol **4:**809-814.
229. **Orian, A., B. van Steensel, J. Delrow, H. J. Bussemaker, L. Li, T. Sawado, E. Williams, L. W. M. Loo, S. M. Cowley, C. Yost, S. Pierce, B. A. Edgar, S. M. Parkhurst, and R. N. Eisenman.** 2003. Genomic binding by the Drosophila Myc, Max, Mad/Mnt transcription factor network. Genes & Development **17:**1101-1114.
230. **Orphanides, G., G. LeRoy, C.-H. Chang, D. S. Luse, and D. Reinberg.** 1998. FACT, a Factor that Facilitates Transcript Elongation through Nucleosomes. Cell **92:**105-116.
231. **Ozsolak, F., A. R. Platt, D. R. Jones, J. G. Reifenberger, L. E. Sass, P. McInerney, J. F. Thompson, J. Bowers, M. Jarosz, and P. M. Milos.** 2009. Direct RNA sequencing. Nature **461:**814-818.
232. **Pan, X., P. Ye, D. S. Yuan, X. Wang, J. S. Bader, and J. D. Boeke.** 2006. A DNA Integrity Network in the Yeast Saccharomyces cerevisiae. Cell **124:**1069-1081.
233. **Pancevac, C., D. C. Goldstone, A. Ramos, and I. A. Taylor.** 2010. Structure of the Rna15 RRM–RNA complex reveals the molecular basis of GU specificity in transcriptional 3′-end processing factors. Nucleic Acids Research **38:**3119-3132.
234. **Patturajan, M., X. Wei, R. Berezney, and J. L. Corden.** 1998. A Nuclear Matrix Protein Interacts with the Phosphorylated C-Terminal Domain of RNA Polymerase II. Molecular and Cellular Biology **18:**2406-2415.
235. **Pavri, R., A. Gazumyan, M. Jankovic, M. Di Virgilio, I. Klein, C. Ansarah-Sobrinho, W. Resch, A. Yamane, B. R. San-Martin, V. Barreto, T. J. Nieland, D. E. Root, R. Casellas, and M. C. Nussenzweig.** 2010. Activation-Induced Cytidine Deaminase Targets DNA at Sites of RNA Polymerase II Stalling by Interaction with Spt5. Cell **143:**122-133.
236. **Pei, Y., and S. Shuman.** 2002. Interactions between Fission Yeast mRNA Capping Enzymes and Elongation Factor Spt5. Journal of Biological Chemistry **277:**19639-19648.
237. **Perales, R., and D. Bentley.** 2009. "Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions. Molecular Cell **36:**178-191.
238. **Perez-Canadillas, J. M.** 2006. Grabbing the message: structural basis of mRNA 3' UTR recognition by Hrp1. EMBO J **25:**3167-3178.
239. **Peterlin, B. M., and D. H. Price.** 2006. Controlling the Elongation Phase of Transcription with P-TEFb. Molecular Cell **23:**297-305.
240. **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:**15702-15719.
241. **Picotti, P., B. Bodenmiller, L. N. Mueller, B. Domon, and R. Aebersold.** 2009. Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics. Cell **138:**795-806.
242. **Pikaard, C. S., J. R. Haag, T. Ream, and A. T. Wierzbicki.** 2008. Roles of RNA polymerase IV in gene silencing. Trends in plant science **13:**390-397.
243. **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. Molecular Cell **9:**799-809.
244. **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. Herbolsheimer, J. Zeitlinger, F. Lewitter, D. K. Gifford, and R. A. Young.** 2005. Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast. Cell **122:**517-527.
245. **Ponting, C. P.** 2002. Novel domains and orthologues of eukaryotic transcription elongation factors. Nucleic Acids Research **30:**3643-3652.
246. **Prather, D., N. J. Krogan, A. Emili, J. F. Greenblatt, and F. Winston.** 2005. Identification and Characterization of Elf1, a Conserved Transcription Elongation Factor in Saccharomyces cerevisiae. Mol. Cell. Biol. **25:**10122-10135.
247. **Prather, D. M., E. Larschan, and F. Winston.** 2005. Evidence that the Elongation Factor TFIIS Plays a Role in Transcription Initiation at GAL1 in Saccharomyces cerevisiae. Mol. Cell. Biol. **25:**2650-2659.
248. **Proshkin, S., A. R. Rahmouni, A. Mironov, and E. Nudler.** 2010. Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. Science **328:**504-508.

249. **Proudfoot, N.** 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. Current Opinion in Cell Biology **16:**272-278.
250. **Ptacek, J., G. Devgan, G. Michaud, H. Zhu, X. Zhu, J. Fasolo, H. Guo, G. Jona, A. Breitkreutz, R. Sopko, R. R. McCartney, M. C. Schmidt, N. Rachidi, S.-J. Lee, A. S. Mah, L. Meng, M. J. R. Stark, D. F. Stern, C. De Virgilio, M. Tyers, B. Andrews, M. Gerstein, B. Schweitzer, P. F. Predki, and M. Snyder.** 2005. Global analysis of protein phosphorylation in yeast. Nature **438:**679-684.
251. **Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, and B. Séraphin.** 2001. The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification. Methods **24:**218-229.
252. **Qian, X., C. Jeon, H. Yoon, K. Agarwal, and M. A. Weiss.** 1993. Structure of a new nucleic-acid-binding motif in eukaryotic transcriptional elongation factor TFIIS. Nature **365:**277-279.
253. **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. Molecular Cell **33:**752-762.
254. **Qu, X., S. Lykke-Andersen, T. Nasser, C. Saguez, E. Bertrand, T. H. Jensen, and C. Moore.** 2009. Assembly of an Export-Competent mRNP Is Needed for Efficient Release of the 3'-End Processing Complex after Polyadenylation. Mol. Cell. Biol. **29:**5327-5338.
255. **Radonjic, M., J.-C. Andrau, P. Lijnzaad, P. Kemmeren, T. T. J. P. Kockelkorn, D. van Leenen, N. L. van Berkum, and F. C. P. Holstege.** 2005. Genome-Wide Analyses Reveal RNA Polymerase II Located Upstream of Genes Poised for Rapid Response upon S. cerevisiae Stationary Phase Exit. Molecular Cell **18:**171-183.
256. **Raha, D., Z. Wang, Z. Moqtaderi, L. Wu, G. Zhong, M. Gerstein, K. Struhl, and M. Snyder.** 2010. Close association of RNA polymerase II and many transcription factors with Pol III genes. Proceedings of the National Academy of Sciences **107:**3639-3644.
257. **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:**432-445.
258. **Rando, O. J., and K. Ahmad.** 2007. Rules and regulation in the primary structure of chromatin. Current Opinion in Cell Biology **19:**250-256.
259. **RDevelopmentCoreTeam.** 2009. R: a language and environment for statistical computing, Vienna, Austria.
260. **Ren, B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell, and R. A. Young.** 2000. Genome-Wide Location and Function of DNA Binding Proteins. Science **290:**2306-2309.
261. **Renner, D. B., Y. Yamaguchi, T. Wada, H. Handa, and D. H. Price.** 2001. A highly purified RNA polymerase II elongation control system. J Biol Chem **276:**42601-42609.
262. **Ringel, R., M. Sologub, Y. I. Morozov, D. Litonin, P. Cramer, and D. Temiakov.** 2011. Structure of human mitochondrial RNA polymerase. Nature **478:**269-273.
263. **Robertson, G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O. L. Griffith, A. He, M. Marra, M. Snyder, and S. Jones.** 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Meth **4:**651-657.
264. **Rodriguez, C. R., E.-J. Cho, M.-C. Keogh, C. L. Moore, A. L. Greenleaf, and S. Buratowski.** 2000. Kin28, the TFIIH-Associated Carboxy-Terminal Domain Kinase, Facilitates the Recruitment of mRNA Processing Machinery to RNA Polymerase II. Mol. Cell. Biol. **20:**104-112.
265. **Rondon, A. G., H. E. Mischo, and N. J. Proudfoot.** 2008. Terminating transcription in yeast: whether to be a 'nerd' or a 'rat'. Nat Struct Mol Biol **15:**775-776.
266. **Sadowski, M., B. Dichtl, W. Hubner, and W. Keller.** 2003. Independent functions of yeast Pcf11p in pre-mRNA 3[prime] end processing and in transcription termination. EMBO J **22:**2167-2177.
267. **Saeki, H., and J. Q. Svejstrup.** 2009. Stability, Flexibility, and Dynamic Interactions of Colliding RNA Polymerase II Elongation Complexes. Molecular Cell **35:**191-205.
268. **Sanchez-Diaz, A., M. Kanemaki, V. Marchesi, and K. Labib.** 2004. Rapid Depletion of Budding Yeast Proteins by Fusion to a Heat-Inducible Degron. Sci. STKE **2004:**pl8-.
269. **Schena, M., D. Shalon, R. W. Davis, and P. O. Brown.** 1995. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. Science **270:**467-470.
270. **Schneider, D. A., S. L. French, Y. N. Osheim, A. O. Bailey, L. Vu, J. Dodd, J. R. Yates, A. L. Beyer, and M. Nomura.** 2006. RNA polymerase II elongation factors Spt4p and Spt5p play roles in transcription elongation by RNA polymerase I and rRNA processing. Proceedings of the National Academy of Sciences **103:**12707-12712.
271. **Schneider, S., Y. Pei, S. Shuman, and B. Schwer.** 2010. Separable Functions of the Fission Yeast Spt5 Carboxyl-Terminal Domain (CTD) in Capping Enzyme Binding and Transcription Elongation Overlap with Those of the RNA Polymerase II CTD. Mol. Cell. Biol. **30:**2353-2364.
272. **Schroeder, S. C., B. Schwer, S. Shuman, and D. Bentley.** 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes & Development **14:**2435-2440.

273. **Schwabish, M. A., and K. Struhl.** 2004. Evidence for Eviction and Rapid Deposition of Histones upon Transcriptional Elongation by RNA Polymerase II. Molecular and Cellular Biology **24:**10111-10117.
274. **Selth, L. A., S. Sigurdsson, and J. Q. Svejstrup.** 2010. Transcript Elongation by RNA Polymerase II. Annual Review of Biochemistry **79:**271-293.
275. **Shaw, P. E.** 2007. Peptidyl-prolyl cis/trans isomerases and transcription: is there a twist in the tail? EMBO Rep **8:**40-45.
276. **Shen, Z., A. St-Denis, and P. Chartrand.** 2010. Cotranscriptional recruitment of She2p by RNA pol II elongation factor Spt4–Spt5/DSIF promotes mRNA localization to the yeast bud. Genes & Development **24:**1914-1926.
277. **Sikorski, R. S., and P. Hieter.** 1989. A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in Saccharomyces cerevisiae. Genetics **122:**19-27.
278. **Sikorski, Timothy W., Scott B. Ficarro, J. Holik, T. Kim, Oliver J. Rando, Jarrod A. Marto, and S. Buratowski.** 2011. Sub1 and RPA Associate with RNA Polymerase II at Different Stages of Transcription. Molecular Cell **44:**397-409.
279. **Sims, R. J., L. A. Rojas, D. Beck, R. Bonasio, R. Schüller, W. J. Drury, D. Eick, and D. Reinberg.** 2011. The C-Terminal Domain of RNA Polymerase II Is Modified by Site-Specific Methylation. Science **332:**99-103.
280. **Singh, N., Z. Ma, T. Gemmill, X. Wu, H. DeFiglio, A. Rossettini, C. Rabeler, O. Beane, R. H. Morse, M. J. Palumbo, and S. D. Hanes.** 2009. The Ess1 Prolyl Isomerase Is Required for Transcription Termination of Small Noncoding RNAs via the Nrd1 Pathway. Molecular Cell **36:**255-266.
281. **Snyder, M., and J. E. G. Gallagher.** 2009. Systems biology from a yeast omics perspective. FEBS Letters **583:**3895-3899.
282. **Solomon, M. J., P. L. Larsen, and A. Varshavsky.** 1988. Mapping proteinDNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. Cell **53:**937-947.
283. **Solomon, M. J., and A. Varshavsky.** 1985. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. Proceedings of the National Academy of Sciences **82:**6470-6474.
284. **Song, L., and G. E. Crawford.** 2010. DNase-seq: A High-Resolution Technique for Mapping Active Gene Regulatory Elements across the Genome from Mammalian Cells. Cold Spring Harbor Protocols **2010:**pdb.prot5384.
285. **Song, L., Z. Zhang, L. L. Grasfeder, A. P. Boyle, P. G. Giresi, B.-K. Lee, N. C. Sheffield, S. Gräf, M. Huss, D. Keefe, Z. Liu, D. London, R. M. McDaniell, Y. Shibata, K. A. Showers, J. M. Simon, T. Vales, T. Wang, D. Winter, Z. Zhang, N. D. Clarke, E. Birney, V. R. Iyer, G. E. Crawford, J. D. Lieb, and T. S. Furey.** 2011. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Research **21:**1757-1767.
286. **Starosta, A. L., H. Qin, A. Mikolajka, G. Y. C. Leung, K. Schwinghammer, K. C. Nicolaou, D. Y. K. Chen, B. S. Cooperman, and D. N. Wilson.** 2009. Identification of Distinct Thiopeptide-Antibiotic Precursor Lead Compounds Using Translation Machinery Assays. Chemistry & Biology **16:**1087-1096.
287. **Steinmetz, E. J., C. L. Warren, J. N. Kuehner, B. Panbehi, A. Z. Ansari, and D. A. Brow.** 2006. Genome-Wide Distribution of Yeast RNA Polymerase II and Its Control by Sen1 Helicase. Molecular cell **24:**735-746.
288. **Stephen, B.** 2009. Progression through the RNA Polymerase II CTD Cycle. Molecular Cell **36:**541-546.
289. **Stiller, J. W., and M. S. Cook.** 2004. Functional Unit of the RNA Polymerase II C-Terminal Domain Lies within Heptapeptide Pairs. Eukaryotic Cell **3:**735-740.
290. **Stiller, J. W., E. C. S. Duffield, and B. D. Hall.** 1998. Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II. Proceedings of the National Academy of Sciences **95:**11769-11774.
291. **Stiller, J. W., and B. D. Hall.** 2002. Evolution of the RNA polymerase II C-terminal domain. Proceedings of the National Academy of Sciences **99:**6091-6096.
292. **Strahl, B. D., and C. D. Allis.** 2000. The language of covalent histone modifications. Nature **403:**41-45.
293. **Stuwe, T., M. Hothorn, E. Lejeune, V. Rybin, M. Bortfeld, K. Scheffzek, and A. G. Ladurner.** 2008. The FACT Spt16 “peptidase” domain is a histone H3–H4 binding module. Proceedings of the National Academy of Sciences **105:**8884-8889.
294. **Sun, M., L. Larivière, 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). Journal of Biological Chemistry **285:**41597-41603.
295. **Svejstrup, J. Q.** 2003. Histones Face the FACT. Science **301:**1053-1055.

296. **Swanson, M. S., M. Carlson, and F. Winston.** 1990. SPT6, an essential gene that affects transcription in Saccharomyces cerevisiae, encodes a nuclear protein with an extremely acidic amino terminus. Mol. Cell. Biol. **10:**4935-4941.
297. **Swanson, M. S., E. A. Malone, and F. Winston.** 1991. SPT5, an essential gene important for normal transcription in Saccharomyces cerevisiae, encodes an acidic nuclear protein with a carboxy-terminal repeat. Mol. Cell. Biol. **11:**3009-3019.
298. **Swanson, M. S., and F. Winston.** 1992. SPT4, SPT5 and SPT6 Interactions: Effects on Transcription and Viability in Saccharomyces cerevisiae. Genetics **132:**325-336.
299. **Swinburne, I. A., C. A. Meyer, X. S. Liu, P. A. Silver, and A. S. Brodsky.** 2006. Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription. Genome Research **16:**912-921.
300. **Tan, M., H. Luo, S. Lee, F. Jin, Jeong S. Yang, E. Montellier, T. Buchou, Z. Cheng, S. Rousseaux, N. Rajagopal, Z. Lu, Z. Ye, Q. Zhu, J. Wysocka, Y. Ye, S. Khochbin, B. Ren, and Y. Zhao.** 2011. Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. Cell **146:**1016-1028.
301. **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:**3569-3583.
302. **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:**1154-1161.
303. **Toedling, J., O. Sklyar, and W. Huber.** 2007. Ringo - an R/Bioconductor package for analyzing ChIP-chip readouts. BMC Bioinformatics **8:**221.
304. **Tong, A. H. Y., M. Evangelista, A. B. Parsons, H. Xu, G. D. Bader, N. Pagé, M. Robinson, S. Raghibizadeh, C. W. V. Hogue, H. Bussey, B. Andrews, M. Tyers, and C. Boone.** 2001. Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. Science **294:**2364-2368.
305. **Tong, A. H. Y., G. Lesage, G. D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G. F. Berriz, R. L. Brost, M. Chang, Y. Chen, X. Cheng, G. Chua, H. Friesen, D. S. Goldberg, J. Haynes, C. Humphries, G. He, S. Hussein, L. Ke, N. Krogan, Z. Li, J. N. Levinson, H. Lu, P. Ménard, C. Munyana, A. B. Parsons, O. Ryan, R. Tonikian, T. Roberts, A.-M. Sdicu, J. Shapiro, B. Sheikh, B. Suter, S. L. Wong, L. V. Zhang, H. Zhu, C. G. Burd, S. Munro, C. Sander, J. Rine, J. Greenblatt, M. Peter, A. Bretscher, G. Bell, F. P. Roth, G. W. Brown, B. Andrews, H. Bussey, and C. Boone.** 2004. Global Mapping of the Yeast Genetic Interaction Network. Science **303:**808-813.
306. **Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg.** 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature **403:**623-627.
307. **van Dijk, E. L., C. L. Chen, Y. d/'Aubenton-Carafa, S. Gourvennec, M. Kwapisz, V. Roche, C. Bertrand, M. Silvain, P. Legoix-Ne, S. Loeillet, A. Nicolas, C. Thermes, and A. Morillon.** 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature **475:**114-117.
308. **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:**795-804.
309. **Venters, B. J., and B. F. Pugh.** 2009. A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. Genome Research **19:**360-371.
310. **Venters, B. J., and F. B. Pugh.** 2009. How eukaryotic genes are transcribed. Critical Reviews in Biochemistry and Molecular Biology **44:**117-141.
311. **Venters, B. J., S. Wachi, T. N. Mavrich, B. E. Andersen, P. Jena, A. J. Sinnamon, P. Jain, N. S. Rolleri, C. Jiang, C. Hemeryck-Walsh, and B. F. Pugh.** 2011. A Comprehensive Genomic Binding Map of Gene and Chromatin Regulatory Proteins in Saccharomyces. Molecular Cell **41:**480-492.
312. **Verdecia, M. A., M. E. Bowman, K. P. Lu, T. Hunter, and J. P. Noel.** 2000. Structural basis for phosphoserine-proline recognition by group IV WW domains. Nature Structural Biology **7:**639-643.
313. **Vojnic, E., A. Mourão, M. Seizl, B. Simon, L. Wenzeck, L. Larivière, S. Baumli, K. Baumgart, M. Meisterernst, M. Sattler, and P. Cramer.** 2011. Structure and VP16 binding of the Mediator Med25 activator interaction domain. Nat Struct Mol Biol **18:**404-409.
314. **Vojnic, E., B. Simon, B. D. Strahl, M. Sattler, and P. Cramer.** 2006. Structure and Carboxyl-terminal Domain (CTD) Binding of the Set2 SRI Domain That Couples Histone H3 Lys36 Methylation to Transcription. Journal of Biological Chemistry **281:**13-15.
315. **Wada, T., T. Takagi, Y. Yamaguchi, A. Ferdous, T. Imai, S. Hirose, S. Sugimoto, K. Yano, G. A. Hartzog, F. Winston, S. Buratowski, and H. Handa.** 1998. DSIF, a novel transcription elongation

factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes & Development **12:**343-356.

316. **Washburn, M. P., D. Wolters, and J. R. Yates.** 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotech **19:**242-247.
317. **Weinzierl, R. O.** 1999. Mechanisms of gene expression. Imperial College Press, London.
318. **Wen, Y., and A. J. Shatkin.** 1999. Transcription elongation factor hSPT5 stimulates mRNA capping. Genes & Development **13:**1774-1779.
319. **Werner-Allen, J. W., C.-J. Lee, P. Liu, N. I. Nicely, S. Wang, A. L. Greenleaf, and P. Zhou.** 2011. cis-Proline-mediated Ser(P)5 Dephosphorylation by the RNA Polymerase II C-terminal Domain Phosphatase Ssu72. Journal of Biological Chemistry **286:**5717-5726.
320. **Werner, F., and D. Grohmann.** 2011. Evolution of multisubunit RNA polymerases in the three domains of life. Nat Rev Micro **9:**85-98.
321. **West, M. L., and J. L. Corden.** 1995. Construction and Analysis of Yeast RNA Polymerase II CTD Deletion and Substitution Mutations. Genetics **140:**1223-1233.
322. **West, S., N. Gromak, and N. J. Proudfoot.** 2004. Human 5' to 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature **432:**522-525.
323. **White, R. J.** 2011. Transcription by RNA polymerase III: more complex than we thought. Nat Rev Genet **12:**459-463.
324. **Wilhelm J, A.** 2009. Next-generation DNA sequencing techniques. New Biotechnology **25:**195-203.
325. **Wind, M., and D. Reines.** 2000. Transcription elongation factor SII. BioEssays **22:**327-336.
326. **Winston, F., D. T. Chaleff, B. Valent, and G. R. Fink.** 1984. MUTATIONS AFFECTING TY-MEDIATED EXPRESSION OF THE HIS4 GENE OF SACCHAROMYCES CEREVISIAE. Genetics **107:**179-197.
327. **Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, A. M. Chu, C. Connelly, K. Davis, F. Dietrich, S. W. Dow, M. El Bakkoury, F. Foury, S. H. Friend, E. Gentalen, G. Giaever, J. H. Hegemann, T. Jones, M. Laub, H. Liao, N. Liebundguth, D. J. Lockhart, A. Lucau-Danila, M. Lussier, N. M'Rabet, P. Menard, M. Mittmann, C. Pai, C. Rebischung, J. L. Revuelta, L. Riles, C. J. Roberts, P. Ross-MacDonald, B. Scherens, M. Snyder, S. Sookhai-Mahadeo, R. K. Storms, S. Véronneau, M. Voet, G. Volckaert, T. R. Ward, R. Wysocki, G. S. Yen, K. Yu, K. Zimmermann, P. Philippsen, M. Johnston, and R. W. Davis.** 1999. Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science **285:**901-906.
328. **Xiang, K., T. Nagaike, S. Xiang, T. Kilic, M. M. Beh, J. L. Manley, and L. Tong.** 2010. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. Nature **467:**729-733.
329. **Xu, Z., G. Wei, I. Chepelev, K. Zhao, and G. Felsenfeld.** 2011. Mapping of INS promoter interactions reveals its role in long-range regulation of SYT8 transcription. Nat Struct Mol Biol **18:**372-378.
330. **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:**1033-1037.
331. **Yamada, T., Y. Yamaguchi, N. Inukai, S. Okamoto, T. Mura, and H. Handa.** 2006. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol Cell **21:**227-237.
332. **Yang, Y. H., and T. Speed.** 2002. Design issues for cDNA microarray experiments. Nat Rev Genet **3:**579-588.
333. **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 & Development **21:**160-174.
334. **Yu, H., P. Braun, M. A. Yıldırım, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.-F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A.-S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. P. Roth, A.-L. Barabási, J. Tavernier, D. E. Hill, and M. Vidal.** 2008. High-Quality Binary Protein Interaction Map of the Yeast Interactome Network. Science **322:**104-110.
335. **Yuan, G.-C., Y.-J. Liu, M. F. Dion, M. D. Slack, L. F. Wu, S. J. Altschuler, and O. J. Rando.** 2005. Genome-Scale Identification of Nucleosome Positions in S. cerevisiae. Science **309:**626-630.
336. **Yudkovsky, N., J. A. Ranish, and S. Hahn.** 2000. A transcription reinitiation intermediate that is stabilized by activator. Nature **408:**225-229.
337. **Yuryev, A., M. Patturajan, Y. Litingtung, R. V. Joshi, C. Gentile, M. Gebara, and J. L. Corden.** 1996. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proceedings of the National Academy of Sciences **93:**6975-6980.

338. **Zacher, B., P. Kuan, and A. Tresch.** 2010. Starr: Simple Tiling ARRay analysis of Affymetrix ChIP-chip data. BMC Bioinformatics **11:**194.
339. **Zeitlinger, J., A. Stark, M. Kellis, J.-W. Hong, S. Nechaev, K. Adelman, M. Levine, and R. A. Young.** 2007. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet **39:**1512-1516.
340. **Zhang, Y., M. L. Sikes, A. L. Beyer, and D. A. Schneider.** 2009. The Paf1 complex is required for efficient transcription elongation by RNA polymerase I. Proceedings of the National Academy of Sciences **106:**2153-2158.
341. **Zhang, Z., J. Fu, and D. S. Gilmour.** 2005. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor, Pcf11. Genes & Development **19:**1572-1580.
342. **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. Molecular Cell **40:**939-953.
343. **Zhou, K., W. H. W. Kuo, J. Fillingham, and J. F. Greenblatt.** 2009. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. Proceedings of the National Academy of Sciences **106:**6956-6961.

Curriculum Vitae

Andreas Mayer

Personal Data

Date of birth: 18.08.1981
Place of birth: Augsburg, Germany
Nationality: German

Publications

- *__Mayer A__, *Schreieck A, Lidschreiber M, Leike K, Martin DE, Cramer P (2011). The Spt5 C-terminal region recruits yeast 3'-RNA cleavage factor I. Mol Cell Biol, accepted. (* equally contributed)
- Blattner C, Jennebach S, Herzog F, **Mayer A**, Cheung ACM, Witte G, Lorenzen K, Hopfner KP, Heck AJR, Aebersold U, Cramer P (2011). Architecture of the RNA polymerase I-Rrn3 initiation complex and implications for cell growth regulation. Genes Dev 25: 2093-2105.
- *Müller M, *Heym RG, **Mayer A**, Kramer K, Schmid M, Cramer P, Urlaub H, Jansen RP, Niessing D (2011). A cytoplasmic complex mediates specific mRNA recognition and localization in yeast. PLoS Biology 9: e1000611. (* equally contributed)
- *Miller C, *Schwalb B, *Maier K, Schulz D, Dümcke S, Zacher B, **Mayer A**, Sydow J, Marcinowski L, Dölken L, Martin DE, Tresch A, Cramer P (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol 7, article number 458. (* equally contributed)
- Clausing E, **Mayer A**, Chanarat S, Muller, B, Germann SM, Cramer P, Lisby M, Strasser K (2010). The transcription elongation factor Bur1-Bur2 interacts with replication protein A and maintains genome stability during replication stress. J Biol Chem 285: 41665-41674.
- Sun M, Lariviere L, Dengl S, **Mayer A**, Cramer P (2010). A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II CTD. J Biol Chem 285: 41597-41603.
- ***Mayer A**, *Lidschreiber M, *Siebert M, Leike K, Söding J, Cramer P (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol 17: 1272-1278. (* equally contributed)
- *Röther S, *Burkert C, *Brünger KM, ***Mayer A**, Kieser A, Strässer K (2010). Nucleocytoplasmic shuttling of the La motif-containing protein Sro9 might link its nuclear and cytoplasmic functions. RNA 16: 1393-1401. (* equally contributed)
- Dengl S, **Mayer A**, Sun M, Cramer P (2009). Structure and in vivo requirement of the yeast Spt6 SH2 domain. J Mol Biol 389: 211-225.
- Hepperger C, **Mayer A**, Merz J, Vanderwall DK, Dietzel S (2009). Parental genomes mix in mule and human cell nuclei. Chromosoma 118: 335-347.