Regulation of RNA polymerase II transcription by Spt4 in *Saccharomyces cerevisiae*

Ülkü Uzun

A thesis submitted for the degree of Doctor of Philosophy

Supervisor: Prof. Jane Mellor

Department of Biochemistry, University of Oxford

St Hugh's College, Trinity Term, 2020

Acknowledgement

I was surrounded by many brilliant people who helped and supported me and made this thesis possible. To begin with, I would like to thank my supervisor Prof. Jane Mellor for giving me the chance to work on this project and for her endless encouragement, patience, and guidance throughout my time in her lab. I am also grateful to Prof. Andre Furger, Dr. Andrew Angel, Prof. Rob Klose, and Prof. Catherine Pears for their precious time and critical evaluations on my research project.

My deep gratitude to all of the current and former members of the Mellor lab for their great friendship, support as well as valuable suggestions throughout my research. In particular, Dr. Harry Fischl for teaching me how to perform and analyse NET/TEF-seq and helping me with the MNase-seq analysis. Dr. Tom Brown, for modelling and making my data more meaningful. Without Drs. Fischl and Brown's input my work would not have been completed.

I especially would like to thank Silvia Raineri for keeping me going every time I run into her when I feel low in the lab and in my life, Tom Brown for sharing his precious home-made beers with me, Harry Fischl for bringing amusing mysteries into our lives, Anna Lamstaes for putting a positive spin to literally everything, Phil Lorenz for our time travels to the sequencing facility, Meredith Wouters for being a social butterfly and reminding us that we should party more. To Struan Murray, Shidong Xi, Anitha Nair, Françoise Howe, Walaa Khushaim, Jack Feltham and Charlie George for creating an enjoyable working environment and helping with the problems on the way. Beside the Mellor Lab, I would like to thank my friends in the department, Sibyl Bertrand, Hannah Walters, Javier Viñals Camallonga, Florian Zobel, and Alastair Louey for being around to share both my worries and happy moments. And thanks to, Amanda Williams, our beloved Zoology sequencing facility queen for turning sequencing days into therapy sessions.

Warm thanks to my past and current housemates at 19 WW. Especially, Martin, Luca, Andrew, Byron, Rosa, Safa, Pauline, Arzhia, Aleks and Brigita for making me feel home after long days in the lab and enjoy my time in Oxford even during the chaotic Covid-19 times.

Very special thanks to my mum, dad, and brother for providing me with constant love, keeping me sane and strong, and even trying to troubleshoot my experiments. You basically troubleshoot my life every time I need it. And, I cannot thank enough to David for being such a supportive and loving partner, my best friend, my part-time mentor, and my spider catcher. You improved my work in so many ways and made me a stronger and happier person.

Finally, I would like to give a special acknowledgement to the CRUK Oxford Centre and the Department of Biochemistry for their funding that allowed me to pursue this DPhil research.

Abstract

Transcription of RNA polymerase II (RNAPII) through chromatin is coordinated by numerous transcription elongation factors (TEFs). Spt4 and Spt5 are two of the most highly conserved TEFs in organisms with nucleosomes. Spt4 forms a complex with Spt5, known as the DSIF complex in humans, which was discovered several decades ago and extensively studied, yet its exact function is still not well understood. Early experiments showed that *spt4* and *spt5* mutants share similar phenotypes with histone chaperone mutants, s*pt6* and s*pt16*. Recently, *in vitro* studies showed that the Spt4/5 complex reduces RNAPII stalling upon transcribing into nucleosomes and cryo-EM images indicated that Spt4/5 occupy a space between the DNA exit site on RNAPII and a nucleosome. These observations implicate the Spt4/5 complex in the regulation of the interplay between transcription and chromatin, but it has not been addressed *in vivo*.

To understand the precise role of Spt4 in budding yeast, various high throughput sequencing techniques and mass spectrometry were used. The position of the Spt4 on RNAPII was mapped at single nucleotide resolution using TEF-seq. With the same technique, the position of Spt5 on RNAPII was investigated in wild type (WT) and *spt4* knock-out (*spt4Δ*) cells. RNAPII dynamics were examined using NET-seq in *spt4Δ* cells or cells in which Spt4 has been anchored away to the cytoplasm (Spt4 AA). For comparison, RNAPII was also mapped in cells in which Spt5 has been anchored away to the cytoplasm (Spt5 AA). Additionally, nucleosome positions were studied using MNase-seq and the composition of the transcription complex was investigated using affinity purification coupled mass spectrometry in WT and *spt4Δ* cells.

The data show that the interaction between the Spt4/5 complex and RNAPII periodically changes as RNAPII transitions through nucleosomes. The dynamic interaction of Spt5 with RNAPII is dependent on Spt4. In *spt4Δ* and Spt4 AA cells, RNAPII distribution is altered in a

similar way compared to their respective controls. After transcribing into nucleosomes, RNAPII accumulates upstream of nucleosome dyads, especially around the +2 nucleosome. In Spt5 AA cells, the distribution and amount of RNAPII on genes are severely affected.

Nucleosome positions are altered in *spt4Δ* cells. Although the position of the +1 nucleosome remains unchanged, nucleosome spacing from this point is increased, and the level of increase in nucleosome spacing correlates with the level of RNAPII accumulation.

The composition of the transcription complex is changed in *spt4Δ* compared to WT cells. The levels of some transcription factors related to chromatin regulation are lowered, while the abundance of TFIIH subunits are significantly increased on the *spt4Δ* transcription complex.

Overall, this work suggests that Spt4 regulates nucleosome positioning by a mechanism related to transcription and promotes RNAPII movement through nucleosome barriers. The findings presented here add to the current understanding of the function of the Spt4/5 complex in transcription.

Declaration of Authorship

I hereby confirm that except where stated explicitly to the contrary in the text and acknowledgements, this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Word count: 44283

Table of Contents

Abbreviations

AA	Anchor away
AID	Auxin inducible degron
ATP	Adenosine tri-phosphate
CBC	Cap binding complex
CF	Cleavage factors
CFP	Cleavage and polyadenylation factors
ChIP	Chromatin immunoprecipitation
CTD	C-terminal domain (of RNAPII)
CTR	C-terminal repeat
DRB	5,6-dichloro-1-β-D-ribofuranosylbenzimidazole
DSIF	DRB-sensitivity inducing factor
FACT	Facilitates chromatin Transcription
GFP	Green fluorescent protein
GTF	General transcription factor
IF	Immunofluorescence
IP	Immunoprecipitation
MNase	Micrococcal nuclease
MS	Mass spectrometry
NDR	Nucleosome depleted region
NELF	Negative elongation factor
NET-seq	Native elongating transcript sequencing
nt	Nucleotide
OD_{600}	Optical density measured at wavelength 600 nm
Paf1	Polymerase-associated factor 1

Paf1C	Paf1 complex
PAS	Polyadenylation site
PCG	Protein coding genes
PIC	Pre-initiation complex
P-TEFb	Positive elongation factor
PTM	Post-translational modification
RNAPII	RNA polymerase II
sd	Standard deviation
Ser2/5/7P	Serine 2, 5 or 7 phosphorylation/phosphorylated RNAPII-CTD
SHL	Superhelix location
SPT	Suppressor of Ty
TAF	TBP-associated factor
TBP	TATA box binding protein
TC-NER	Transcription coupled nucleotide repair
TEC	Transcription elongation complex
TEF	Transcription elongation factor
TEF-seq	Transcription elongation factor sequencing
TF	Transcription factor
TREX	Transcription export protein complex
TTS	Transcription termination site
TU	Transcriptional unit
WT	Wild type

1 Introduction

1.1 Gene expression and transcription

Gene expression is the process whereby the information stored in DNA is converted into a functional form. This process starts with transcription; synthesis of RNA from the DNA template. RNA is synthesised by DNA-dependent RNA polymerase (RNAP) and eukaryotic cells have three forms of RNAP. RNAPI synthesises ribosomal RNA (rRNA) which forms ribosomes with ribosomal proteins. RNAPIII synthesises small non-coding RNA such as transfer RNA (tRNA) which carries amino acids to ribosome during translation. RNAPII transcribes all protein-coding genes (PCG), generating messenger RNA (mRNA) and some non-coding genes generating ncRNAs in the cell. From thereon, the term transcription will be used to describe transcription by RNAPII as it is the main focus of this thesis.

Transcription is a complex, dynamic, and tightly regulated process and has been studied extensively as it has a direct impact on protein expression. Defects in this process may negatively affect cell functioning and may lead to health issues in mammalian cells. Although the major aspects and general rules of the transcription are well described, we are still far from seeing the full picture of this complicated process. Studies in transcription are of great importance to understand the fundamentals of gene expression, which in turn will help us advance in disease treatment and biotechnology.

1.2 RNAPII structure

RNAPII is composed of 12 subunits (Rpb1-12, Rpb1 also known as Rpo21). The two largest subunits Rpb1 and Rbp2 perform the main reactions of transcription: nucleotide selection, RNA synthesis, translocation, and cleavage of RNA (Cramer, Bushnell, and Kornberg 2001). Rpb1 and Rpb2 surround two sides of the DNA template and form the active site cleft which consists of the conserved bridge helix and trigger loop. DNA is unwound downstream of the

active site which serves a single strand DNA template for transcription. During transcription, the 3'-end of RNA in the cleft aligns with the active site of RNAPII. The last 8-9 nt RNA at the 3'-end together with the unwound template DNA create a DNA-RNA duplex which is called the transcription bubble (**Figure 1.1**). The transcription bubble forms a 90-130° angle with downstream DNA which helps with the maintenance of the open DNA structure. As transcription progresses, newly synthesised RNA dissociates from DNA and leaves RNAPII through the RNA exit channel. Exiting upstream DNA is rewound (Bernecky, Plitzko, and Cramer 2017; Martinez-Rucobo and Cramer 2013).

Numerous transcription factors (TFs) are recruited to RNAPII at different stages of transcription. The unstructured carboxyl-terminal domain (CTD) of Rbp1 controls the recruitment of TFs by creating a binding platform. The CTD contains 26 repeats of $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$ in yeast (52 repeats in mammals) which can be reversibly phosphorylated by kinases at each of these residues except prolines (Jeronimo, Collin, and Robert 2016). At initiation, RNAPII is recruited as a hypo-phosphorylated form. As RNAPII progresses through different stages of transcription, the CTD is dynamically phosphorylated and dephosphorylated on different residues. The most studied and major phosphorylation sites are Ser5 and Ser2 residues of the CTD (Schüller et al. 2016). Ser5 phosphorylation (Ser5-P) peaks close to the transcription start sites (TSS) and is mostly associated with initiation and early transcription stages. Ser2 phosphorylation (Ser2-P) starts later in elongation and peaks towards the end of transcription (Zaborowska, Egloff, and Murphy 2016). Similar to Ser5-P, Ser7 phosphorylation is observed at initiation sites. Although most of these phosphorylation patterns are conserved in mammals and yeast, Tyr1 phosphorylation seems to be dominant in early transcription in mammalian cells, whereas it peaks towards the end of genes in yeast. Thr4 phosphorylation correlates with elongating RNAPII and shows a less distinctive pattern compared to other phosphorylation marks in yeast (Zaborowska et al. 2016).

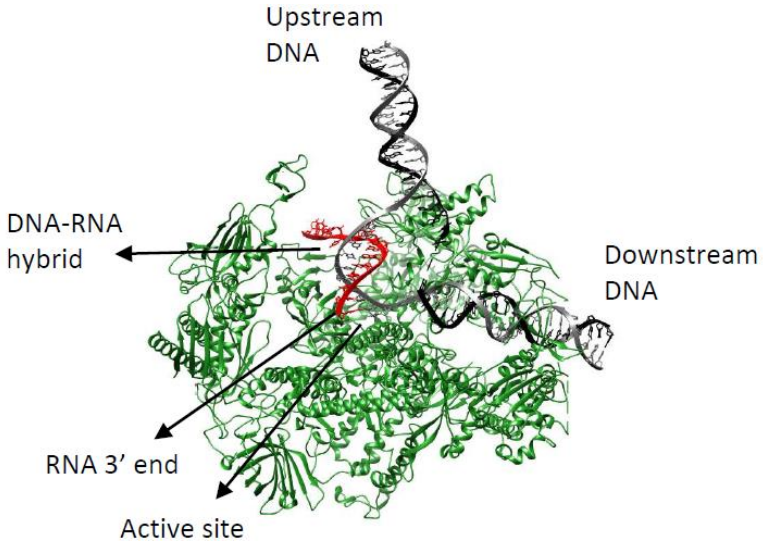


Figure 1.1) RNAPII structure.

RNAPII, template DNA, non-template DNA and RNA are coloured in green, grey, black and red, respectively (PDB: 5XON) (Ehara et al. 2017). The Rpb2 subunit is removed for clarity.

Transcription is generally evaluated in three main parts: initiation, elongation, and termination based on the major reactions taking place and factors involved. Initiation defines the recruitment of RNAPII to promoter and transcription of initial ~10 nt long RNA. Elongation is the stage where RNAPII actively transcribes. Termination defines the release of transcribed RNA from RNAPII and release of RNAPII from template DNA. These three main stages of transcription repeat for every successful transcription event and are tightly regulated to ensure the effective progress of RNAPII on the DNA template.

1.3 Transcription initiation

Transcription initiation includes the formation of pre-initiation complex (PIC), DNA opening, synthesis of first a few nt RNA and clearance of promoters from RNAPII (**Figure 1.2**). The PIC is formed by general transcription factors (GTFs), TFIIB, TFIID, TFIIE, TFIIF, TFIIH at a promoter. The promoter is located 40-120 nt upstream of the TSS (+1) in yeast and contain an A/T rich region which is important for the recruitment of GTFs. In classic transcription models, this A/T rich region is defined as TATA box bearing a conserved TATA(A/T)A(A/T)(A/G) sequence (Faitar, Brodie, and Ponticelli 2001; Pugh and Tjian 1991). In yeast, most promoters have one

or two mismatches from the consensus TATA-box sequence (Rhee and Pugh 2012), usually defined as TATA-less or TATA-like promoters although they are regulated in a similar way to TATA-box containing promoters in terms of transcription initiation (Baptista et al. 2017; Sainsbury, Bernecky, and Cramer 2015).

The formation of the PIC starts with binding of TATA-box binding protein (TBP; one of the TFIID subunit) to promoters. TBP binding is essential as it bends DNA and prepares a recognition site for other TFs and RNAPII (Sainsbury et al. 2015). TFIID includes 13 TBP-associated factors (TAFs) that can bind to promoters. Although their specific functions are not clear, TAFs are thought to adjust the strength of promoters and communicate with transcriptional activators (Joo et al. 2017). At some promoters, TFIID binding is accompanied by the binding of TFIIA which further stabilise the TBP-DNA complex (Sainsbury et al. 2015). TFIIB (Sua7 in yeast) is another essential factor that binds to TBP and interacts with promoter around the A/T rich region (known as TFIIB recognition element; BRE) (Ly et al. 2020). TFIIB binding completes the promoter recognition and the formation of closed promoter complex assembly, which then recruits RNAPII-TFIIF complex via direct interaction between TFIIB and RNAPII.

After the recruitment of RNAPII-TFIIF, promoter opening is initiated by the binding of TFIIE and TFIIH. TFIIE directly binds to RNAPII and mediates the interaction between TFIIH and RNAPII. TFIIH has a 7-subunit core (Ssl2 (ATPase), Rad3 (ATPase), Tfb1, Tfb2, Tfb4, Tfb3, Ssl1) and a 3-subunit kinase modules (Kin28 (kinase, CDK7 in humans), Ccl1, Tfb3) (Kokic et al. 2019). The Ssl2 subunit of TFIIH promotes ATP-dependent DNA opening by acting as a translocase and feeding DNA into RNAPII (Grünberg and Hahn 2013). The open DNA structure allows promoter scanning and start site selection, helps the positioning of the single-stranded DNA template and initiates the transcription of a few nt of RNA. This transcribing RNAPII together with GTFs is described as the initially transcribing complex (ITC). Once a 8-15 nt long

RNA has been transcribed by ITC, RNAPII breaks its contact with TFIIB and transcribes away from the promoter (Han and He 2016; Ly et al. 2020). As upstream DNA exits the transcription complex, the transcription bubble starts reannealing and causes GTFs to dissociate (Fujiwara et al. 2019). This event is defined as promoter escape and it is assisted by Ser5-P of the RNAPII-CTD by kinase subunit of TFIIH (Kin28) which also helps the recruitment of RNA processing and transcription elongation factors (Jeronimo and Robert 2014; Wong, Jin, and Struhl 2014).

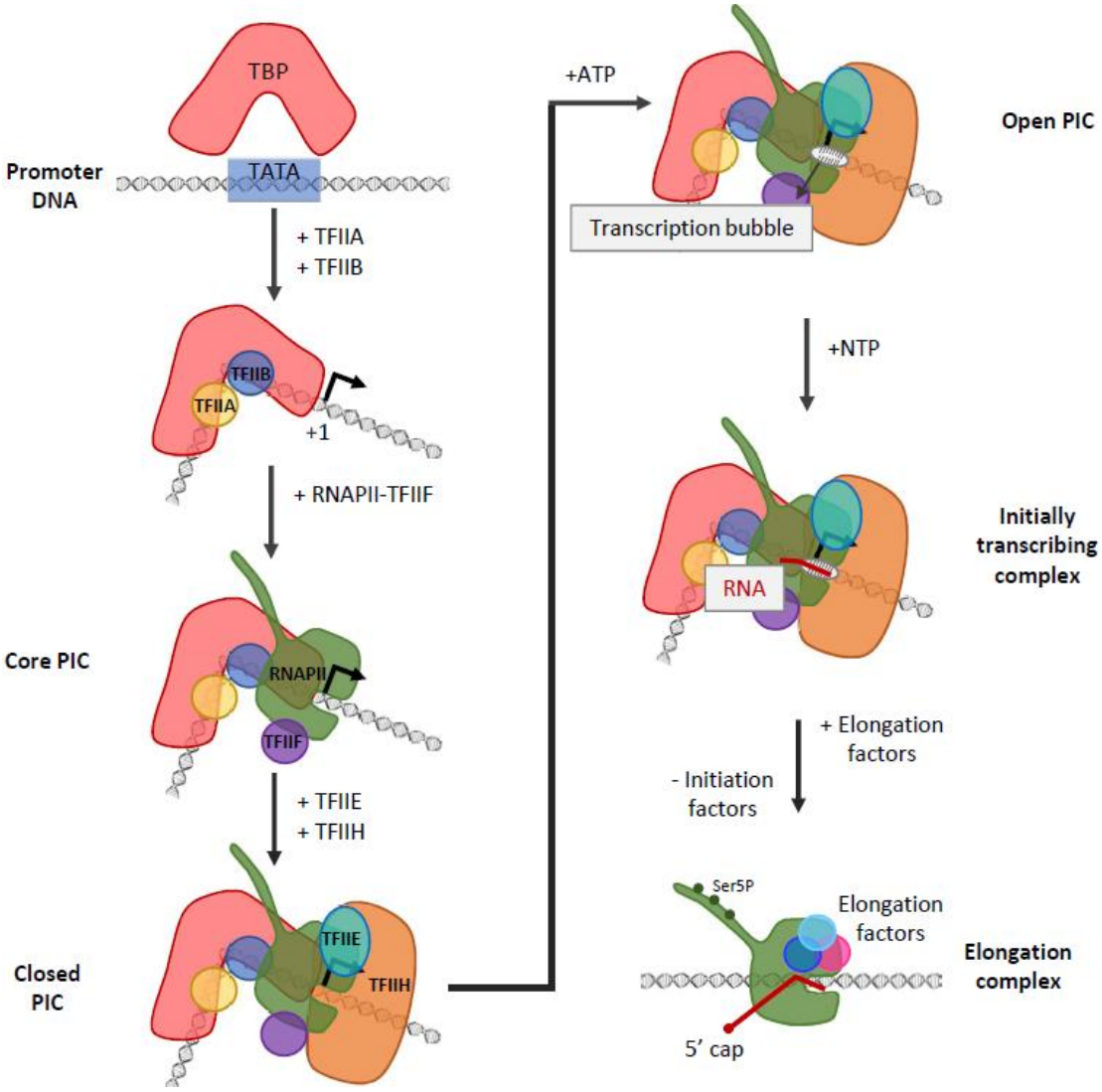


Figure 1.2) PIC assembly

Formation of the PIC at promoters is regulated, based on the cell's needs and coordinated by transcription activators and repressors that bind to upstream of promoters. Activators are positive regulators of transcription and possess a DNA binding domain providing specificity

for promoters to be activated and a protein interaction domain to bind co-activators. Co-activators such as Mediator, Spt-Ada-Gcn5 acetyltransferase (SAGA) and chromatin remodellers can directly interact with the PIC and/or modify the chromatin environment to upregulate transcription. Similarly, transcription can be negatively regulated by promoter-specific repressors and co-repressors.

1.4 Transcription elongation

Transcription elongation defines the stage of RNA synthesis after the promoter escape until RNAPII reaches to the termination site. During elongation, RNAPII forms the transcription elongation complex (TEC) with transcription elongation factors (TEFs). Elongation relies on the timely association and dissociation of TEFs that mainly ensure mRNA processing and efficient movement of RNAPII through chromatin. Based on its relative position, the composition of the TEC dynamically changes. The order of events and the involvement of TEFs are controlled by the CTD of RNAPII, TEFs that can recruit other factors, and native RNA.

1.4.1 RNA processing factors

Newly synthesised RNA is co-transcriptionally processed as RNAPII keeps adding new nucleotides to the growing RNA. With the association of proteins, mRNA forms messenger ribonucleoprotein complexes (mRNP). After synthesis of the first 25-30 nt, the native mRNA goes through 7-methyl-guanosine (m7G) capping at the 5'-end (Moteki and Price 2002). Capping co-transcriptionally occurs in three steps and each step is performed by a different enzyme in yeast. First, 5' triphosphatase Cet1 hydrolyses the γ-phosphate from the 5'-end of the native RNA and leaves diphosphate-end. Second, the RNA guanylyltransferase Ceg1 transfers GMP from GTP to diphosphate-end and creates 5' G-capped nucleoside. And third, the methyltransferase Abd1 methylases the 5' G-cap at the N7 position and completes m7G capping (Ramanathan, Robb, and Chan 2016). Ceg1 and Abd1 directly interact with Ser5-P on the RNAPII-CTD and Cet1 is recruited in a complex with Ceg1. Ser7-P on CTD by Kin28 also

seems to be important for the association between Ceg1 and RNAPII (Bharati et al. 2016). 5' capping protects mRNA from degradation and is important for downstream processing of mRNA, export to the cytoplasm, and translation initiation (Kachaev et al. 2020).

Next, the nuclear cap-binding complex (CBC consists of Cbc2 and Sto1) is recruited to 5'capped RNA. Although the function of CBC is not completely understood, this complex associates with RNA from capping stage to until after nuclear export, and is thought to regulate downstream RNA processing (Hossain et al. 2013). In budding yeast, about 5% of genes contain introns. In these genes, pre-mRNA is also co-transcriptionally processed by a large spliceosome complex (Parenteau et al. 2019). As the pre-mRNP grows longer, the THO complex (Hpr1, Tho2, Mft1, Thp2, and Tex1) is recruited. The components of the THO complex interact with CBC and RNAPII, ensuring efficient maturation of mRNP, and maintaining genome stability by preventing interaction between single-stranded DNA and native RNA (Heath, Viphakone, and Wilson 2016). Additionally, the THO complex couples transcription process to export by mediating interactions with transcription export protein complexes (TREX; consisting of Sub2 and Yra1). Overall, RNA processing factors co-transcriptionally engage with RNAPII and each other to ensures proper capping, maturation, and export of mRNA.

1.4.2 Transcription elongation factors

1.4.2.1 Discovery of SPT genes

A group of transcription related proteins in yeast were identified by genetic screening experiments performed after insertion of transposable element (Ty) to the 5'-end regulatory region of *HIS4* or *LYS2*. These Ty insertions in *HIS4* or *LYS2* disrupt the expression of the genes and create *his4* $^{-}$ or *lys2* $^{-}$ phenotypes, respectively. The genes that when mutated suppress the *his4* $^{-}$ or *lys2* $^{-}$ phenotypes, allowing growth on medium lacking histidine or lysine, respectively, are known as Suppressor of Ty (*SPT*) genes (Fassler and Winston 1988; Swanson

and Winston 1992; Winston et al. 1984, 1987). The proteins encoded by *SPT* genes are listed in **Table 1.1**.

Table 1.1) List and functions of Spt proteins. Ref. Functions and names taken from SGD.

Protein	Function and alternative name if know
Spt1	Subunit of nucleosome assembly complex; Hir2
Spt2	Important for RNA polyadenylation; Exa5, Sin1
Spt3	Subunit of SAGA complex
Spt4	Subunit of the Spt4/5 complex
Spt5	Subunit of the Spt4/5 complex
Spt6	Nucleosome interacting protein; Cre2, Ssn70
Spt7	Subunit of the SAGA complex; Git2
Spt8	Subunit of the SAGA complex
Spt10	H3 acetylase; Cre1, Sud2
Spt11	H2A; Hta1
Spt12	H2B; Htb1
Spt13	Subunit of the RNA polymerase II mediator complex; Med15
Spt14	UDP-glycosyltransferase subunit of the GPI-GnT complex; GPI3
Spt15	TATA-binding protein; TBP
Spt16	Subunit of the heterodimeric FACT complex (Spt16-Pob3); Ssf1

Following their discovery, a group of *SPT* genes were found to share a characteristic phenotype. Genes encoding SNF proteins (sucrose non-fermenting), components of an ATP-dependent chromatin remodelling protein complex, specifically control the expression of *SUC2* gene needed for cells to process sucrose. SNF mutants (*snf2, snf5*, and *snf6*) create the *SUC2*$^{-}$ phenotype and this phenotype is suppressed once combined with a mutation in one of the following genes: *SPT4, SPT5, SPT6, SPT11, SPT12,* and *SPT16* (Malone et al. 1991; Swanson and Winston 1992). As *SPT11 (HTA1)* and *SPT12 (HTB1)* encode histone H2A and H2B, respectively, the *SPT* genes suppressing SNF mutant phenotype were called histone group *SPT* genes (Renner et al. 2001). Consistently, Spt6 and Spt16 were later shown to be histone chaperones (see below). Although there are a large number of studies attempting to precisely

define the function of the TEFs Spt4 and Spt5 in transcription, their exact contribution is not yet clear.

1.4.2.2 Spt4 and Spt5

Spt4 is a transcription elongation factor that is conserved across archaea and eukaryotes, organisms containing histone proteins and chromatin (Malone, Fassler, and Winston 1993; Ponting 2002). Confirming the functional homology between humans and yeast, *spt4* knockouts in yeast can be complemented by the human counterpart of SPT4 (Hartzog et al. 1996). Spt4 consists of 4 zinc finger (ZF) motifs (**Figure 1.3A**). Although ZF motifs generally interact with nucleic acids, structural and biochemical studies do not support a direct interaction between Spt4 and DNA or RNA (Crickard, Fu, and Reese 2016; Ehara et al. 2019).

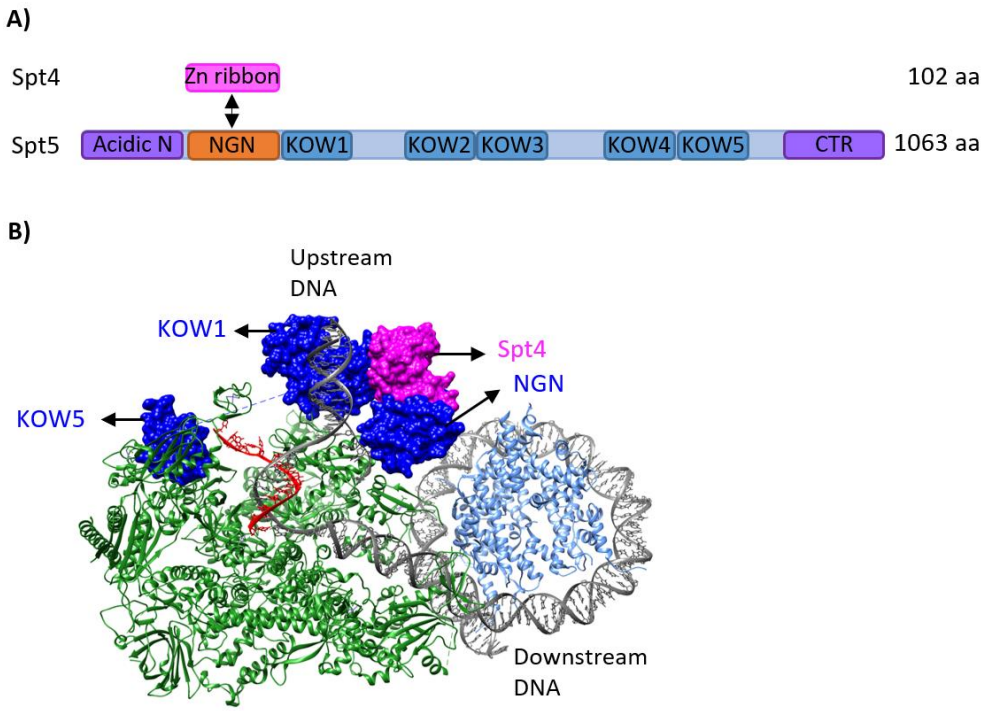


Figure 1.3) Spt4 and Spt5 structure.

A) Domain organisation of yeast Spt4 and Spt5 proteins.

B) RNAPII, Spt4, Spt5 domains (NGN, KOW1 and KOW5), DNA, RNA and histones are coloured in green, magenta, blue, grey, red and light blue, respectively (PDB: 6J51). Rpb2 subunit is removed for clarity (Ehara *et al.*, 2019).

The importance of Spt4 in transcription was shown by various genetic screening experiments over the years. Although *spt4* knockout cells are viable with a slow-growth phenotype (D. A. Schneider et al. 2006), *spt4* knockout mutants were shown to be synthetically lethal once combined with mutations in THO components (*hpr1* or *tho2*), or *elf1* (Prather et al. 2005; Rondón et al. 2003), or the Paf1C components (Squazzo et al. 2002). *dst1 spt4* double mutants are extremely sensitive to high temperatures (Quan and Hartzog 2010). Additionally, *spt4* knockout mutants are sensitive to transcriptional stress upon depletion of NTPs and this phenotype is alleviated once combined with a mutation in *ISW1*, a gene encoding a ATP-dependent chromatin remodeller known to maintain chromatin structure (Morillon et al. 2003; Ocampo et al. 2019).

Spt5 is an essential protein conserved in all domains of life. The bacterial homologue of Spt5, NusG has multiple regulatory functions. One of the most prominent functions of NusG is promoting transcription by reducing the backtracking of RNA polymerase (RNAP). NusG exerts this function by preventing backward motion of RNAP once it translocates forward following the addition of a nucleotide and thus, limits the size of the transcription bubble (Turtola and Belogurov 2016). Additionally, NusG controls transcription-coupled processes such as translation by interacting with ribosomal protein S10 (NusE) and transcription termination by interacting with bacterial transcription termination factor Rho (Artsimovitch 2018; Kang et al. 2019).

Yeast Spt5 consists of an acidic N-terminal domain, a NusG N-terminal (NGN) domain, 5 Kyprides, Ouzounis, Woese (KOW) domains and unstructured C-terminus repeat (CTR) (Ponting 2002) (**Figure 1.3A**). Like the CTD of RNAPII, the CTR of Spt5 is subject to phosphorylation by Bur1 (CDK9 in mammals, part of P-TEFb, see below) and forms a binding platform for other transcription factors. For example, unphosphorylated Spt5-CTR seems to play a role in the recruitment of capping enzymes (Doamekpor et al. 2014), while

phosphorylated CTR recruits Paf1C (Booth et al. 2018). The NGN domain of Spt5 interacts with non-template DNA (Crickard et al. 2016). Additionally, Spt5 was shown to crosslink with RNA (Baejen et al. 2017).

Spt4 binds the NGN domain of Spt5 and they form a heterodimeric complex called Spt4/5 (**Figure 1.3**). Structural studies showed that in the TEC, Spt4/5 is positioned on top of the RNAPII active cleft where upstream DNA leaves the elongating complex. The same site is occupied by TFIIF and TFIIE in the initiation complex, implying that the binding of the Spt4/5 complex is mutually exclusive with the binding of these initiation factors (Ehara et al. 2017). Additionally, Cryo-EM structures including a nucleosome also demonstrate that Spt4/5 fills a space between upstream DNA and a nucleosome downstream of the active cleft (**Figure 1.3B**). Consistently, *in vitro* studies show that the Spt4/5 complex reduces transcriptional stalling upstream of a nucleosome and stabilises the transcription complex with a partially unwrapped nucleosome (Crickard et al. 2017; Ehara et al. 2019). *In vivo* Spt4/5 was implicated in increasing RNAPII processivity (Mason and Struhl 2005).

In mammalian cells, the homologue of this complex was discovered by using a drug, 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) which interferes with transcription elongation (not initiation) by inhibiting certain kinase activities. A fraction of human nuclear extract could transcribe *in vitro* even upon addition of DRB. Once this fraction of the nuclear extract was complemented with another fraction including the Spt4/5 homologues and other factors, DRB-sensitivity was restored. From the latter fraction two complexes were purified: (1) the mammalian counterpart of the Spt4/5 complex, named as DRB sensitivity-inducing factor (DSIF) and (2) negative elongation factor (NELF) (Wada et al. 1998; Yamaguchi et al. 1999). Notably, DSIF or NELF alone was not able to induce DRB-sensitivity (Yamaguchi et al. 1999). The kinase complex that is responsible for the DRB sensitivity was called the positive transcription elongation factor B (P-TEFb). P-TEFb consists of a kinase (CDK9) and two cyclins

and phosphorylates both NELF and DSIF (Zorio and Bentley 2001). Phosphorylated NELF dissociates from the transcription complex and phosphorylated DSIF complex is considered to be switched from a negative to a positive elongation factor (Stadelmayer et al. 2014). In yeast, so far, no NELF homologue has been found and thus the negative effect of the Spt4/5 complex not detected.

1.4.2.3 Spt6 and Spn1

Spt6 is a transcription elongation factor as it binds to Tyr1, Ser2, and Ser5 phosphorylated CTD via its C-terminus. On its N-terminus, Spt6 has an unstructured acidic domain that interacts with H2A/H2B and H3, with the highest affinity toward H3 (Bortvin and Winston 1996). The N-terminal domain of Spt6 also interacts with another factor, Spn1 (also known as Iws1). As Spn1 and histones compete for the same binding site on Spt6, Spn1 is thought to regulate Spt6-histone interactions (McCullough et al. 2015).

Mutations in *SPT6* increase transcription initiation at cryptic intragenic sites, disrupt nucleosome positioning and occupancy, and reduce transcription-coupled histone recycling (Doris et al. 2018; Jeronimo, Poitras, and Robert 2019). Moreover, Spt6 was shown to help histone assembly at highly transcribed genes after RNAPII passage (Ivanovska et al. 2011). Due to its roles in histones binding, chromatin maintenance, and nucleosome re-assembly, Spt6 is considered as a histone chaperone.

1.4.2.4 Spt16 and Pob3 (FACT)

FAcilitates Chromatin Transcription (FACT) is a heterodimeric complex containing Spt16 and Pob3 and binds to histones with help of a structural protein Nhp6. The Spt16 subunit of FACT can bind to all histones, but specifically interacts with the H2A/H2B dimer at their DNA binding sites and helps co-transcriptional nucleosome disassembly. Similar to Spt6, FACT also regulates nucleosome reassembly after the RNAPII passage (Gurova et al. 2018). Consistent

with the latter function, *SPT16* mutants lead to impaired nucleosome positioning, lower nucleosome occupancy, and poor histone recycling (Feng et al. 2016; Jeronimo et al. 2019).

1.4.2.5 Paf1 complex

Polymerase-associated factor 1 (Paf1) is a conserved complex (Paf1C) in eukaryotes and consists of five subunits in yeast: Paf1, Leo1, Ctr9, Cdc73, and Rtf1. Yeast Paf1C is recruited to RNAPII downstream of the +2 nucleosome via its interaction with the phosphorylated RNAPII-CTD and phosphorylated Spt5-CTR (Fischl et al. 2017; Van Oss, Cucinotta, and Arndt 2017). Paf1C affects the recruitment of Ctk1 (CDK12) and hence the Ser2-P of the RNAPII-CTD (Dronamraju and Strahl 2014). Paf1C regulates Rad6/Ber1-dependent H2B K123 mono-ubiquitination (H2Bub1) required for the establishment of other histone marks (Van Oss et al. 2017). Recently, Paf1C has been shown to fine tune the nuclear export of native transcripts by regulating their interaction with RNA processing and nuclear export factors (Fischl et al. 2017).

1.4.2.6 Other Elongation factors

Elf1 (ELOF1 in humans) is a small zinc finger protein that physically interacts with RNAPII and genetically interacts with Spt4, Spt5, Spt6, Spt16, and Dst1 (Prather et al. 2005). Recently, together with Spt4/5, Elf1 was also shown to reduce RNAPII arrest upstream of a nucleosome during *in vitro* transcription (Ehara et al. 2019).

Dst1 (TFIIS) is a conserved elongation factor that directly binds to RNAPII (Ehara et al. 2017). Occasionally, a transcriptional barrier, such as nucleosomes or another RNAPII, induces backtracking of RNAPII. In these cases, the 3'-end of the native RNA misaligns with the active site of RNAPII. To create a new free 3'-end, the RNA needs to be cleaved. Dst1 takes a part in the rescue of backtracked RNAII by triggering the cleavage activity of RNAPII (Zatreanu et al. 2019).

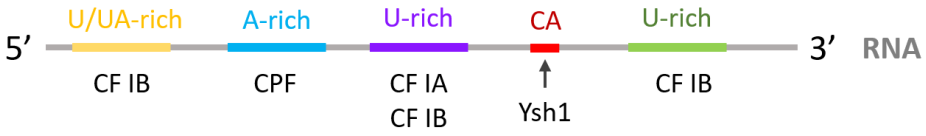
Set2 is a histone methyl transferase that interacts with Ser2 and Ser5 phosphorylated RNAPII. Set2 deposits methylation marks on lysine 36 residues of histone H3. H3K36me is important for the activation of Rpd3S a histone deacetylase that co-transcriptionally reverses acetylation marks after RNAPII passage to restore repressive chromatin state and prevent cryptic transcription (Mcdaniel et al. 2018).

1.5 Transcription termination

Transcription is completed once the native RNA is released from RNAPII and RNAPII is released from the DNA. In yeast, there are different mechanisms for canonical and early transcription termination. In canonical termination, protein-coding transcripts and some stable non-coding transcripts are terminated and processed by the cleavage and polyadenylation factor (CPF) and cleavage factor (CF) dependent pathway (CPF-CF pathway). Following this pathway, mRNAs are exported to the cytoplasm to be translated. In early termination, transcription of small non-coding RNAs including snRNAs and snoRNAs, are terminated by the Nrd1, Nab3 and Sen1 dependent pathway (NNS pathway). After the termination by the NNS pathway, snRNAs and snoRNAs are processed into stable transcripts and the rest of the ncRNAs are subject to nuclear degradation by the nuclear exosome (Bresson et al. 2017; Porrua and Libri 2015).

The CPF-CF pathway involves more than 50 proteins and mainly is coordinated by polyadenylation signal (PAS) at the 3' end of the genes (**Figure 1.4**). In mammals, the PAS is a well conserved AAUAAA sequence, whereas in budding yeast it is less well defined. The budding yeast PAS consists of U-rich elements and it bears A rich efficiency elements ~30 nt upstream of the PAS (Mischo and Proudfoot 2013) (**Figure 1.4**). After RNAPII has transcribed the PAS, termination factors (the CPF-CF pathway components) recognise the PAS and interact with the RNA or RNAPII. The native transcript is then cleaved by an endonuclease subunit of CPF, Ysh1 and polyadenylated by poly(A) polymerase, Pap1, which adds ~70 adenines (poly(A)-tail) to the 3' end of the cleaved transcript. The poly(A)-tail together with

some poly(A) binding proteins such as Pab1, stabilise the transcript (Kumar et al. 2019). After the RNA cleavage, RNAPII continues transcription downstream of the PAS until it falls off the DNA template. The RNAPII release site is termed as transcription termination site (TTS) and it generally occurs around 150 bp downstream of the PAS in yeast and a few kb in mammalian cells (Baejen et al. 2017; Proudfoot 2016). Two non-mutually exclusive models are proposed to describe transcription termination. The allosteric model suggests that the recruitment of termination factors together with dissociation of elongation factors change RNAPII conformation, destabilise RNAPII and, finally induce its release from the DNA. The torpedo model proposes that the termination is triggered by the RNA cleavage at the PAS. The cleaved RNA leaves behinds a free RNA 5'-end attached to RNAPII which is a substrate for a nuclear 5'-3' exonuclease, Rat1 (XRN2 in mammals). Rat1 then degrades the RNA until it catches up with RNAPII. This creates a kinetic race between transcribing RNAPII and degrading Rat1. The reduced speed of RNAPII is, therefore, believed to play a role at Rat1 catching up RNAPII and finally triggers RNAPII fall of the DNA (Proudfoot 2016). Several lines of evidence show that both the allosteric and the torpedo models take place in the cell and termination could happen in a combinatory fashion, recently named as the unified model (Eaton et al. 2020).



CPF		CFIA
Pap1 (Poly(A) polymerase)	Mpe1	Rna14
Ctf1/Yhh1	Pta1	Rna15
Pfs2	Ref2	Clp1
Fip1	Pti1	Pcf11 (Binds RNAPII)
Yth1	Swd2/Cps35	
Cft2	Glc7 (Phosphatase)	**CFIB**
Ysh1/Brr5 (Endonuclease)	Ssu72 (Phosphatase)	Hrp1

Figure 1.4) Polyadenylation signals and RNA recognition of the 3'-end processing factors in yeast.
CPF: cleavage and polyadenylation factors, CF: cleavage factors.

In the NNS pathway, Nrd1 specifically binds to Ser5 phosphorylated and Tyr1 unphosphorylated RNAPII-CTD. This combination of the CTD phosphorylation confines the Nrd1 and RNAPII interaction into early transcription. This interaction helps Nrd1 and Nab3 recognise their specific sequences on ncRNAs (GUAA/G and UCUUG, respectively). Subsequently, Sen1 (a helicase) is recruited to the RNA bound by Nrd1-Nab3. Although the mechanism is not clear, the recruitment of Sen1 triggers the dissociation of RNAPII from the DNA. Notably, the NNS pathway might not require cleavage of the native transcripts, as experiments do not indicate a cleavage event. Following its recruitment, NNS interact with Trf4/5-Air1/2-Mtr4-polyadenylation (TRAMP) complex, which polyadenylates the ncRNAs. The polyadenylated ncRNAs are then targeted by the nuclear exosome to be either processed into stable RNAs or completely degraded (Han et al. 2020).

1.6 Transcription and chromatin structure

The eukaryotic DNA is stored in a compact nucleoprotein structure called chromatin. Chromatin is composed of nucleosomes which include 147 bp of DNA wrapped around a histone octamer; 2 sets of each canonical histones: H2A, H2B, H3 and H4 (Kornberg 1974) (**Figure 1.5**). 147 bp nucleosomal DNA makes ~1.65 turns around the histone octamer and interacts with the histone tails approximately every 10 bp. The major grooves of DNA facing the histone octamers are defined as superhelix location (SHL). Once the midpoint of the nucleosomal DNA is taken as a reference, SHLs are numbered from SHL (0) to SHL (-/+7) to describe up- and downstream the nucleosomal DNA (**Figure 1.5**).This nucleosome structure repeats on DNA every 160-200 bp and limits the accessibility of DNA (Zhou, Gaullier, and Luger 2019).

Chromatin structure is not static and continuously undergoes dynamic regulations. During transcription or other processes requiring direct contact with DNA such as replication and DNA repair, nucleosomes need to be removed to give access to DNA and then re-positioned

after the completion of the process. Three different groups of proteins or complexes control the nucleosome organisation. First, ATP-dependent chromatin remodelling complexes (remodellers, in short) involve enzymes to assemble, evict and slide nucleosomes, and exchange histones within the nucleosomes. Histone exchange can replace a canonical histone with a variant such as H3.3 or H2A.Z (Htz1) having specific functions on chromatin (Li, Carey, and Workman 2007). Second, histone chaperones associate with free histones in the cell and help trafficking histone exchange. Third, histone modifying enzymes deposit post-translational modifications (PTMs) such as acetylation, methylation and ubiquitylation on histone tails. The PTMs on histone tails regulate the chromatin environment by altering the DNA-histone stability and creating a platform to recruit remodellers and histone chaperones. Thus remodellers, histone chaperones and histone modifiers coordinate accessibility of DNA in a process-specific manner and ensure the maintenance of the chromatin structure (Clapier et al. 2017; Venkatesh and Workman 2015).

During transcription initiation, promoters are kept free from the nucleosomes for the recruitment of the PIC and these free regions are known as nucleosome depleted regions (NDRs). Remodellers such as RSC, a member of SWI/SNF family, are recruited to promoters to slide and evict nucleosomes (Lorch, Maier-Davis, and Kornberg 2014). Together with histone chaperones such as Asf1, remodellers establish and maintain NDRs. NDRs are demarcated by well-positioned upstream and downstream nucleosomes, designated as -1 and +1 nucleosomes, respectively. Downstream of the +1 nucleosome, nucleosomes are less strictly positioned towards the end of TUs (Ocampo et al. 2019).

Over the gene bodies, nucleosomes set physically barriers for transcription and they need to be rearranged to allow RNAPII passage. Depending on the level of transcription, all or some histones are removed from the nucleosomes by histone chaperone complexes such as FACT. In highly active genes H2A/H2B histones are more readily exchanged on chromatin than

H3/H4 histones (Cole et al. 2014). After the passage of RNAPII, the disrupted chromatin structure must be reset. While histone chaperones FACT and Spt6 co-transcriptionally reassemble the nucleosomes, remodellers such as CHD and ISWI ensure proper spacing between the nucleosomes (Venkatesh and Workman 2015).

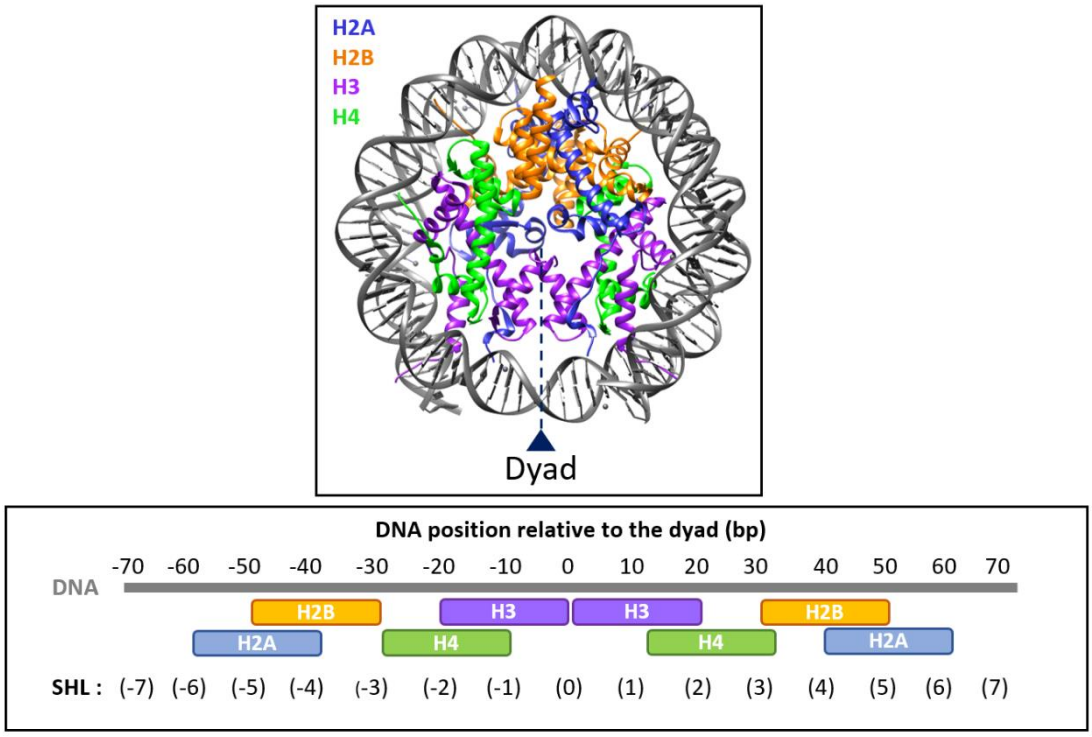


Figure 1.5) Nucleosome structure.

DNA, H2A, H2B, H3 and H4 are coloured in grey, blue, orange, purple and green respectively (PDB: 3AFA) (Tachiwana et al. 2010). Major groove of the nucleosomal DNA faces to the histone octamer approx. at every 10 bp, known as superhelix location (SHL). Approximated histone-DNA interaction regions are depicted in the bottom panel.

1.7 Transcription and health

Transcription is the first step in gene expression. All cell functions, such as metabolic processes, response to the environmental inputs, maintenance of genome integrity, cell division, survival, and apoptosis, rely on the efficient regulation of transcription. To ensure normal cell functioning, as outlined above, each stage of transcription from initiation to termination is tightly regulated. There are hundreds of proteins, if not thousands, involved in

transcription. Although a lot is known about this complicated process, we are far from understanding the full mechanism of transcription.

Even though different organisms may have different proteins or some structural alterations in the proteins involved in transcription, most steps and basic reactions are well conserved from yeast to human. Spt4 is one of the most conserved transcription elongation factors, having homologues in all organisms with nucleosomes, and the binding partner of Spt4, Spt5, has homologues in all domains of life. Consequently, studies in Spt4 (and Spt5) have great potential and impact on understanding the basics of transcription in almost all organisms, especially humans.

One example of the importance of understanding the contribution of Spt4/5 to transcription comes from studies performed on the proto-oncogene MYC that encodes MYC protein which acts as a transcription activator. At normal levels, MYC is involved in cell growth and cell cycle regulation and overexpression of this protein leads to uncontrolled cell division. Importantly, MYC is one of the most commonly overexpressed transcription activators in different cancer types (Lu et al. 2015) and thus, it is a potential therapeutic target.

There are different ways in which MYC is implicated in transcription regulation, for example, by regulating the recruitment of CDK9 and Paf1C to transcription sites (Jaenicke et al. 2016; Lu et al. 2015). A recent study showed that MYC also has direct interaction with SPT5 and regulates the amount of SPT5 transferred to RNAPII during elongation. Overexpression of MYC reduces SPT5-RNAPII interaction by sequestering SPT5, downregulates some genes, and causes cancer-like transcriptomes (Baluapuri et al. 2019). The detailed mechanism of MYC-dependent transcription regulation is open to further investigations. However, these preliminary results prove how promising the studies in transcription regulation are, and how they can improve our current understanding of health and disease.

1.8 Aims

The main aim of this thesis is to investigate the contribution of Spt4 to transcription in *S.cerevisiae* by transient or long term ablation of Spt4. To this end, genome-wide sequencing techniques were used to define the positions of Spt4, its binding partner Spt5, and RNAPII at single nucleotide resolution, and the direct role of Spt4 in maintaining these profiles investigated. The composition of the transcription complex was examined using a proteomics approach and nucleosome positions were mapped with the classic MNase-seq technique. More details on the specific aims are given in the overview to each chapter.

This thesis reveals that the interaction between the Spt4/5 complex and RNAPII periodically changes as RNAPII transitions through nucleosomes. In the absence of Spt4, RNAPII accumulates upstream of nucleosome dyads, especially around the +2 nucleosome. In cells lacking Spt4, nucleosome positions are altered; although the position of the +1 nucleosome remains unchanged, nucleosome spacing from this point is increased, and the level of increase in nucleosome spacing correlates with the level of RNAPII accumulation. Additionally, the proteomics of the transcription complex is changed in *spt4Δ* compared to WT cells. The levels of some transcription factors related to chromatin regulation are lowered, while the abundance of TFIIH subunits is significantly increased, on the *spt4Δ* transcription complex. Overall, this work suggests that Spt4 regulates nucleosome positioning by a mechanism related to transcription and promotes RNAPII movement through the nucleosome barriers.

2 Methods and materials

2.1 Yeast culture

Strains were streaked onto 2% agar YPD plates from glycerol stocks and were grown (48h, 30°C). Cells were pre-cultured in YPD (overnight, 30°C). The overnight culture was used to inoculate appropriate volume of YPD culture at OD_{600} 0.2, which was grown (30°C, 160 rpm) to OD_{600} 0.6-0.7 for all experiments unless stated otherwise.

2.2 Yeast strains

All strains used in this study, and the plasmids used to construct new strains for this study, are listed in **Table 2.1** and **Table 2.2**. Genetic manipulation of cells was performed by using the homologous recombination method (Longtine et al. 1998). For the deletion mutants, PCR products were amplified containing a selection marker with promoter and terminator sequences flanked at both ends by 40 bp of sequence homologous to sequences either side of the region to be deleted. For C-terminus tagged strains, PCR products were amplified with a 40 bp sequence homologous to the first 40 bp upstream of the stop codon of the gene to be tagged followed by a tag sequence, selection marker and 40bp of sequence homologous to a region downstream of the gene to be tagged.

Competent cells were prepared by growing to log phase, pelleting (1,000 g, 3 min) and resuspending in 45 µl 0.1M LiAc/TE (overnight, 4°C). 100 µl competent cells were mixed with 10 µl PCR product, 10 µl calf thymus DNA (Sigma), 700 µl 0.1 M LiAc/TE in 40% PEG and incubated (30 min, 30°C) and then heat-shocked (20 min, 42°C). Cells were pelleted (7000 rpm, 5 min), resuspended in H_20 and plated onto appropriate selection media. When the antibiotic resistance selection marker was used, cells were incubated (4 h, 30°C) in YPD before plating. DNA was extracted from transformants, screened by PCR, using primers flanking the site of insertion, and confirmed by sequencing and western blot where possible.

Table 2.1) Yeast strains used in this study

Strain	Source	Genotype
BY4741	Euroscarf	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*
BY4741 *spt4::KanMX6 (SPT4 KO)*	Euroscarf	*spt4::KanMX6*
BY4741 Rpb3-FLAG (WT)	H. Fischl	*RPB3-3xFLAG-His3MX6*
BY4741 *SPT4 KO* Rpb3-FLAG (*spt4Δ*)	This study	*RPB3-3xFLAG-His3MX6; spt4::KanMX6*
BY4741 Spt4-FLAG	H. Fischl	*SPT4-3xFLAG-His3MX6*
BY4741 Spt5-FLAG	This study	*SPT5-3xFLAG-His3MX6*
BY4741 *SPT4 KO* Spt5-FLAG	This study	*SPT5-3xFLAG-His3MX6; spt4::KanMX6*
BY4741 Bur1-FLAG	Harry Fischl	*BUR1-3xFLAG-His3MX6*
BY4741 *SPT4 KO* Bur1-FLAG	This study	*BUR1-3xFLAG-His3MX6; spt4::KanMX6*
BY4741 Sua7-FLAG	This study	*SUA7-3xFLAG-His3MX6*
BY4741 *SPT4 KO* Sua7-FLAG	This study	*SUA7-3xFLAG-His3MX6; spt4::KanMX6*
BY4741 Kin28-FLAG	H. Fischl	*KIN28-3xFLAG-His3MX6*
BY4741 *SPT4 KO* Kin28-FLAG	This study	*KIN28-3xFLAG-His3MX6; spt4::KanMX6*
BY4741 Rad26-FLAG	This study	*RAD26-3xFLAG-His3MX6*
BY4741 *SPT4 KO* Rad26-FLAG	This study	*RAD26-3xFLAG-His3MX6; spt4::KanMX6*
S.pombe Rpb9-FLAG	L. Vasilieva	*h+; leu1-32; ura4Δ18; ade16-M216; his3Δ1; RPB9-3xFLAG-KanMX4*
AA Spt4-FRB-GFP Rpb3-FLAG	This study	*tor1-1; Δfpr1; RPL13-2xFKBP12-NATMX6; met15; LYS2; his3-1; leu2; ura3; MATa; SPT4-FRB-eGFP-HygMX*
AA Spt5-FRB-GFP Rpb3-FLAG	This study	*tor1-1; Δfpr1; RPL13-2xFKBP12-NATMX6; met15; LYS2; his3-1; leu2; ura3; MATa; SPT5-FRB-eGFP-HygMX*
AA Rpb3-FLAG (No FRB)	This study	*tor1-1; Δfpr1; RPL13-2xFKBP12-NATMX6; met15; LYS2; his3-1; leu2; ura3; MATa*
BY4741 *dst1::kanMX6*	Euroscarf	*dst1::kanMX6*
BY4741 *dst1::kanMX6* Rpb3-FLAG	This study	*RPB3-3xFLAG-His3MX6; dst1::kanMX6*
BY4741 osTIR1 (No mAID)	This study	*OsTIR1::URA3*
BY4741 osTIR1 Spt4-mAID	This study	*OsTIR1::URA3; SPT4-mAID::KANMX6*

Table 2.2) Plasmids used in this study

Plasmid	Source	Description
pFA6a-3-FLAG-His3MX6	H. Fischl	PCR template for FLAG-tagging
pFA6a-FRB-yEGFP-hygromycin	F. Holstege	PCR template for FRB-tagging
pNHK53 (BYP6744)	NBSP	*ADH1* promoter-OsTIR1 (from *Oryza sativa*)
pMK68 (BYP7425)	NBSP	PCR template for codon optimised mini-degron IAA17 tagging (from *Arabidopsis thaliana*)

2.3 Non-phenol genomic DNA extraction

500 µl overnight culture was pelleted (15,800 g, 1 min) and resuspended in 100 µl of 200 mM LiOAc/1 % SDS and incubated (5 min, 65°C). Then 300 µl 100 % EtOH added and centrifuged (15,800 g, 3 min). The supernatant was discarded, and the pellet was washed with 100 µl 70 % EtOH, followed by centrifugation (15,800 g, 1 min). The supernatant was discarded, and the pellet was suspended in 100 µl TE and incubated (5 min, 65°C).

2.4 Doubling time measurements and calculations

Overnight cultures were diluted to OD_{600} 0.10-0.15 in 250 µl YPD and grown in 100 well plates for the Bioscreen (22 h, 30°C), with readings at OD_{600} taken every 20 min with shaking (200 rpm). For the anchor away testing, YPD is supplemented with DMSO or 1 mg/ml rapamycin dissolved in DMSO and for the auxin-inducible degron testing YPD is supplemented with EtOH or 500 µM IAA in EtOH. A minimum of four technical replicates were performed for each condition and strain. OD_{600} measurements were analysed in R. Reads were blanked by subtracting medium-only reads. Doubling times were calculated by choosing the exponential growth phase (OD_{600} 0.2 to 0.7) and using the following equation:

Doubling time = log (2)*time/ [log (max (OD_{600}) – log (min (OD_{600}))]

2.5 Western blotting

Cells were grown until mid-log phase and 10 ml of OD_{600} 0.5 equivalent of cells were pelleted (1,000 g, 3 min). The pellet was resuspended in 1 ml of 0.1N NaOH and incubated (5 min, RT). Then lysed cells were pelleted (15,800 g, 1min) and resuspended in 250 µl loading buffer (50 mM Tris-Cl pH 6.8, 10 % glycerol, 2 % SDS, 0.05 % bromophenol blue, 100 mM DTT, 2 M Urea). Protein samples were run on (various percentages usually 5-15 %) SDS polyacrylamide gels by gel electrophoresis and transferred to nitrocellulose membranes. Membranes were incubated in 5% BSA in TBST (20 mM Tris-Cl (pH 7.5), 150 mM NaCl, 0.1 % Tween-20) for 2 h, then primary antibody in 2.5% BSA in TBST for 1.5 h (RT) or overnight (4°C), washed three times in 20 ml TBST for 10 min, then incubated with rabbit, mouse or rat HRP-conjugated secondary antibody (Sigma) diluted 1:4000 in 2.5% BSA in TBST for 45 min and washed again. Primary antibodies and their dilutions are detailed in **Table 2.3**. Antibody binding was visualised using chemiluminescent reagents (Thermo Scientific Pierce ECL) and x-ray film.

For methyl methanesulfonate (MMS) treatment, samples were grown to OD_{600} 0.5-0.6 and MMS was added to media at 0.05 % final concentration. Cells were grown 45 min in MMS and collected as above except that the 0.1N NaOH solution and loading buffer was supplemented with 1x PhosSTOP (Roche) to inhibit dephosphorylation during sample preparation.

Table 2.3 Primary antibodies used in this study.

Antibody	Epitope	Species reactivity	Dilution for WB
M2 (Sigma)	FLAG	Mouse	1:8000
3E8 (Millipore)	RNAPII-CTD-Ser5P	Rat	1:2000
Ab290 (Abcam)	GFP	Rabbit	1:2000

2.6 Silver staining

Protein samples were separated on 4-20% precast gradient SDS polyacrylamide gel (Biorad) by gel electrophoresis at 100 V for 2 h. Silver staining was performed with Pierce Silver Stain Kit (Thermo Fisher) according to the manufacturer's instructions. Briefly, the gel was washed

with MQ water (10 min) and fixed in 30 % ethanol 10 % acetic acid (30 min). After washing with 10% ethanol (10 min) and MQ water (10 min), it was sensitized (1 min), stained (30 min) and developed (2-3 min) with the buffers provided in the kit. The developed gel was fixed with 5 % acetic acid (10 min).

2.7 NET and TEF-seq

2.7.1 Cell growth

2 L of cells were grown in YPD to OD_{600} 0.65 (30°C, 160 rpm shaking), collected by filtering and flash frozen in liquid nitrogen. 1.28 g of frozen *S.cerevisiae* pellet was combined with 0.32 g of frozen *S.pombe* pellet. The combined pellet was ground with mixer mill (6 cycles, 3 min, 15 hz) in a metal chamber with a metal ball and the chamber was submerged into liquid nitrogen between the milling runs.

2.7.2 Immunoprecipitation

IPs were carried out in the cold room, all buffers used were ice-cold and all centrifugations were at 4°C. 1 g of grindate was resuspended in 5.66 ml of Lysis Buffer A (20 mM HEPES (pH 7.4), 110 mM KOAc, 0.5 % Triton-X-100, 0.1 % Tween 20, 10 mM MnCl2, 1x proteinase inhibitors (Roche; complete, EDTA-free), 50 U/ml SUPERase.In RNase inhibitors (Invitrogen), 132 U/ml DNase I (Promega)) by continuous pipetting up and down for several minutes. The lysate was incubated in ice for 20 min and then centrifuged (16,000 g, 10 min). The supernatant was taken and 400 µl of M2 agarose beads pre-washed twice with 10 ml Lysis Buffer A (without SUPERase.In and DNase I) was added to the supernatant. IPs were performed on a rotating wheel for 2.5 h and then washed 4 times for 2 min with 10 ml Wash Buffer A (20 mM HEPES (pH 7.4), 110 mM KOAc, 0.5 % Triton-X-100, 0.1 % Tween 20, 1 mM EDTA). Excess wash buffer was removed by centrifugation (1,000 g, 2 min). Samples were eluted twice with 300 µl 1 mg/ml of 3xFLAG peptide (Sigma) (prepared in Lysis Buffer A without SUPERase.In and DNase I) for 30 min by mild rotation. Eluates were collected by

centrifugation (1,000 g, 2 min) and combined. RNAPII bound RNA was isolated with Qiagen miRNA kit according to the manufacturer's instructions, RNA was eluted in 31 µl of elution buffer. 1 µl of the sample was used to measure RNA amount in Nanodrop. During the IP, 20 µl of samples were taken from the input, unbound (the first flow through after 2.5 h IP incubation) and eluate samples, and mixed with 20 µl of 2x SDS buffer (100 mM Tris-Cl pH 6.8, 20 % glycerol, 4 % SDS, 0.1 % bromophenol blue, 200 mM DTT) for western blot controls.

2.7.3 Library preparation

Adapter ligation and fragmentation: A minimum of 2.5 µg of immunoprecipitated RNA was diluted in 30 µl H_2O, split into 3 tubes and denaturated (2 min, 80°C) and placed on ice (2min). RNA was ligated with 5'end adenylated and 3'end blocked adapter (**Table 2.4**) by adding 10 µl of ligation mix (50 ng/µl cloning linker 1, 12 % PEG 8000, 1 x T4 RNA ligase2 truncated ligation buffer, 10 U/µl T4 RNA ligase2 (truncated) (NEB)) to each tube (3 h, 37°C). Then the reaction was stopped by adding 0.7 µl of 0.5 M EDTA. Adaptor ligated RNA was fragmented by adding 20 µl of Alkaline Fragmentation Buffer (AFB; 100 mM NaCO3 (pH 9.2), 2 mM EDTA) (35-40 min, 95°C). Exact incubation time was determined for each batch of AFB. Then 0.56 ml RNA precipitation buffer (500 µl H_20, 60 µl 3M NaOAc (pH 5.5), 2 µl 15 mg/ml GlycoBlue (Ambion)) and 0.75 ml isopropanol was added, and samples were incubated at -20°C (>30 min). RNA was collected by centrifuge (20,000 g, 30 min,4°C) and washed with cold 80% EtOH. RNA in three tubes was resuspended in the same 10 µl of 10 mM Tris-HCl pH 7.0.

RNA size selection: Adapter ligated, and fragmented RNA was mixed with 10 µl gel loading bufferII (Invitrogen), denaturated (2 min, 80°C) and placed on ice (3 min). Denaturated RNA was run on 10% TBE-Urea gel (Biorad) (200 V, 35 min) in 1 x TBE buffer (diluted from RNase-free 10 X TBE (Ambion)). The gel was stained with SybrGold (Invitrogen) (5 min, RT) and RNA corresponding to 40-90nt was excised. For physical disruption, the gel slices were spun through 0.5 ml tubes with holes at the bottoms nested in 1.5 ml tubes. The disrupted gel

slurry was incubated in 600 µl water (10 min, 70°C, 1400rpm shaking). RNA was cleared from the gel by transferring the mix into a Costar-spin column (Corning) and centrifuging (20,000 g, 3 min, RT). 50 µl 3 M Sodium Acetate (pH 5.5), 2 µl Glycoblue and 0.75 ml of isopropanol was added to RNA mix and incubated at -20°C (>30 min). RNA was collected by centrifugation (20,000 g, 30 min, 4°C), washed with 0.75 ml cold 70% EtOH and resuspended in 10 µl of 10 mM Tris pH 7.0.

Reverse transcription (RT): Size selected RNA was mixed with 4.6 µl of RT mix (3.28 µl 5 x First-Strand buffer, 0.82 µl dNTPs (10 mM each), 0.5 µl 100 µM RT primer (**Table 2.4**)) and denaturated (2 min, 80°C). Then 1.32 µl Superase.In/DTT and 0.82 µl SuperScriptIII added and incubated (30 min, 48°C). 1.8 µl 1 M NaOH added (20 min, 98°C) to degrade RNA. 1.8 µl 1M HCl added after the incubation to neutralise the cDNA.

cDNA size selection: cDNA was mixed with 20 µl gel loading buffer II (Invitrogen), denaturated (3 min, 95°C) and placed on ice (3 min). Denaturated cDNA was run on 10% TBE-Urea gel (Biorad) (200 V, 50 min) in 1xTBE buffer. The gel was stained with SybrGold (Invitrogen) (5 min, RT) and cDNA corresponding to 80-130 nt was excised. For physical disruption, the gel slices were spun through 0.5 ml tubes with holes at the bottoms nested in 1.5 ml tubes. The disrupted gel slurry was incubated in 400 µl water (10 min, 70°C, 1400 rpm shaking). cDNA was cleared from the gel by transferring the mix into a Costar-spin column (Corning) and centrifuging (20,000 g, 3 min, RT). 25 µl 3 M NaCl, 2 µl Glycoblue and 0.75 ml of isopropanol was added to the cDNA mix. Samples were incubated at -20°C (>30 min). cDNA was collected by centrifuge (20,000 g, 30 min, 4°C), washed with 0.75 ml cold 80% EtOH and resuspended in 15 µl of 10 mM Tris-HCl (pH 8.0).

Circularisation: 4 µl circularisation mix (2 µl 10 x CircLigase buffer, 1 µl 1 mM ATP, 1 µl 50 mM $MnCl_2$) and 1 µl of CircLigase (Epicentre) was added to the size selected cDNA and incubated (60 min, 60°C). Then the enzyme was heat inactivated (10 min, 80°C).

Amplification and barcoding: Circularised cDNA was amplified and barcoded (**Table 2.4**) by adding 15 µl of PCR master mix (8 µl HF master mix (NEB), 0.8 µl 10 µM reverse barcoding primer, 0.8 µl 10 µM forward barcoding primer, 5.4 µl water) per 1 µl template (1 cycle: 30 sec 98°C;; 3-to-7 cycles: 10 sec 98°C; 10 sec 60°C; 5 sec 72°C;; 1 cycle: Hold 4°C). Tubes were taken at the end of 3-4-5-7 cycles. PCR products were mixed with 3µl loading dye (NEB) and run on 8% TBE gel (Invitrogen) (90 V, 95 min) in 1xTBE buffer. The gel was stained with SybrGold (Invitrogen) (5 min, RT) and DNA corresponding to 120-170 nt was excised. For physical disruption, the gel slices were spun through 0.5 ml tubes with holes at the bottoms nested in 1.5 ml tubes. Then 0.67 ml DNA soaking buffer (0.3 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA) was added to the gel slurry and tubes were incubated overnight on a rotating wheel.

Table 2.4) Oligos used for NET-seq library prep and control.

Oligonucleotide	**Sequence**
Linker-1	5'AppCTGTAGGCACCATCAAT/3ddC 3'
RNA control oligo	5' agu cac uua gcg aug uac acu gac ugu g3'
RT primer	5'/5Phos/ATCTCGTATGCCGTCTTCTGCTTG/iSp18/CACTCA/iSp18/TCCGA CGATCATTGATGGTGCCTACAG 3'
Barcoding reverse	5'CAAGCAGAAGACGGCATACGA 3'
Barcode-1	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTG AACTCCAGTCACATGCCATCCGACGATCATTGATGG 3'
Barcode-2	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTG AACTCCAGTCACTGCATCTCCGACGATCATTGATGG 3'
Barcode-4	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTG AACTCCAGTCACTTAGGCTCCGACGATCATTGATGG 3'
Barcode-5	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTG AACTCCAGTCACTGACCATCCGACGATCATTGATGG 3'
Barcode-6	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTG AACTCCAGTCACGCCAATTCCGACGATCATTGATGG 3'
Barcode-7	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTG AACTCCAGTCACCAGATCTCCGACGATCATTGATGG 3'
Barcode-8	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTG AACTCCAGTCACACTTGATCCGACGATCATTGATGG 3'

Barcode-9	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTG AACTCCAGTCACGATCAGTCCGACGATCATTGATGG 3'
Barcode-10	5'AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTG AACTCCAGTCACTAGCTTTCCGACGATCATTGATGG 3'
Sequencing primer	5'TCCGACGATCATTGATGGTGCCTACAG 3'

Oligos ordered from Integrated DNA Technologies. iSp18: carbon spacer in the RT primer. Indexing sequences in barcodes are underlined.

2.7.4 Sequencing and data analysis

Barcoded libraries were pooled and sequenced on Illumina NextSeq 500 (50cycle, single-end) with custom reading primer (**Table 2.4**). Reads were analysed using online tools on usegalaxy.org and RStudio/Bioconductor as outlined in **Table 2.5**.

Table 2.5) NET-seq data analysis

Steps	Package	Conditions	Platform
1) Grooming	FASTQ Groomer (Blankenberg et al. 2010)	Sanger & Illumina 1.8+	usegalaxy.org
2) FASTQ quality control	FASTQ Summary Statistics		usegalaxy.org
3) Adaptor trimming	Clip	Custom: ATCTCGTATGCCGTCTTC Discard < 15 nt reads	usegalaxy.org
4) Aligning libraries	Map with Bowtie for Illumina (Langmead 2011)	Cer3 & Pombe combined genome (-n 1, -e 70, -l 28, -k 1, -m 1)	usegalaxy.org
5) Convert data to BAM	SAM-to-BAM (Li 2011)		usegalaxy.org
6) Identifying uniquely aligned reads	Bioconductor		RStudio
7) Narrowing reads to the 3'-end	Bioconductor		RStudio
8) No tag normalisation	Bioconductor		RStudio
9) *S.pombe* spike-in normalisation	Bioconductor		RStudio
10) Annotation to TIF-seq	Bioconductor		RStudio
11) Visualisation			IGV (Thorvaldsdóttir, Robinson, and Mesirov 2013)
13) Plotting (metagene)	Bioconductor, rafalib (Irizarry 2016)		RStudio

14) Plotting (heatmaps)	Othercolor	MATLAB

No tag normalisation: NET-seq was performed on strains without a FLAG-tag to detect background signal during the IP. The *SCR1* gene is transcribed by RNAPIII and gives a high non-specific signal in the NET-seq and TEF-seq IPs, and this locus was used for no-tag normalisation. The reads on chrV [442007:442458] were split into 10 nt bins and FLAG-tag over no tag sample ratio is calculated for each bin. The mean *SCR1* ratio then multiplied by the no tag data and subtracted from the FLAG-tag samples.

FLAG-tag - [Mean *SCR1* ratio (FLAG-tag/no tag)] x no tag

Spike in normalisation: NET-seq data were aligned to the combined genome of Cer3 and Pombe. After the removal of non-uniquely aligned reads and no-tag background signal, counts table was created for *S.pombe* genes by using RStudio/Bioconductor. Then estimateSizeFactors function in the DEseq2/RStudio package was applied to calculate the relative amounts of *S.pombe* reads (i.e. normalisation ratio) in each sample (Love, Huber, and Anders 2014). NET-seq data were calibrated by dividing *S.cerevisiae* reads by the normalisation ratios.

Metagene plots: Protein-coding genes (PCGs) >750 nt were taken and genes with negative values due to no tag normalisation were discarded. To avoid genes with wrong TSS annotation, genes having 1.5x more reads upstream of the TSS (-150 to 0 nt) than in the downstream (+1 to 150 nt) were also discarded.

PCGs were plotted relative to the TSS in a window of TSS-250 nt to TSS+750 nt or relative to the PAS in a window of PAS-250 nt to PAS+250 nt. The mean number of counts for each nt position was calculated excluding top and bottom 1% of reads to avoid random spikes introduced during sequencing. The mean number of counts then was split into 10 nt bins for the metagene plots.

Heatmaps: Matrices prepared with Rpb3 or factor intensities at single nucleotide resolution ranked based on the gene length or the read intensities in RStudio. Then, the matrices were converted to MATLAB matrices using the R.matlab/RStudio package (Bengtsson 2018). Heatmaps were plotted in MATLAB.

2.8 RNA-seq

2.8.1 Sample collection and library preparation

Samples were taken from the same grindate that was prepared for the NET-seq samples as described above. 0.1 g grindate was resuspended in 250 µl of Lysis Buffer A (20 mM HEPES (pH 7.4), 110 mM KOAc, 0.5 % Triton-X-100, 0.1 % Tween 20, 10 mM MnCl2, 1x proteinase inhibitors (Roche; complete, EDTA-free), 50 U/ml SUPERase.In RNase inhibitors (Invitrogen), 132 U/ml DNase I (Promega)) by continuous pipetting up and down for several minutes. The lysate was incubated in ice for 20 min and then centrifuged (16,000 g, 10 min). Total RNA was extracted from supernatant using Qiagen miRNA kit according to the manufacturer's instructions. RNA concentrations were measured by nanodrop. Ribosomal RNAs were depleted from the total RNA using Invitrogen RiboMinus Transcriptome Isolation Kit (for yeast) and libraries prepared with the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina according to the manufacturer's instructions.

2.8.2 Sequencing and data analysis

Barcoded libraries were pooled and sequenced on Illumina NextSeq 500 (75 cycle, paired). Only forward reads were used for data processing. Reads were analysed using online tools on usegalaxy.org and RStudio/Bioconductor as outlined in **Table 2.6**.

Table 2.6) RNA-seq data analysis

Steps	Package	Conditions	Platform
1) Trimming	Trim Galore!	Single end	usegalaxy.org
1) Second trimming to remove poly(A)	Trim Galore!	Single end Adaptor sequences to be trimmed off: AAAAAAAAAA	usegalaxy.org
2) Aligning reads	TopHat (Kim et al. 2013)	Single end Cer3 & Pombe combined genome	usegalaxy.org
3) Removing PCR duplicates	RmDup		usegalaxy.org
4) Filtering low quality reads	Filter SAM or BAM	MAPQ > 49	usegalaxy.org
5) Create counts table	Bioconductor	Two separate count tables for Cer3 and Pombe genomes	RStudio
6) Spike-in normalisation and differential expression analysis	Bioconductor and DEseq2 (Love et al. 2014)	Normalised Cer3 reads to size factors calculated from Pombe counts table	RStudio

2.9 ChIP-seq

2.9.1 Immunoprecipitation

50 mL of cells were grown in YPD to OD_{600} 0.6 (30°C, 160 rpm), collected by centrifuge (1,000 g, 4min) and resuspended in 45 ml of 1x PBS. Samples were crosslinked by addition of 1.25 ml of 37% formaldehyde (1% final) at RT for 5 min with shaking at 85 rpm on a rocker. Then the reaction was quenched by the addition of 2.5 ml of 2.5 M glycine (0.125 M final) at RT for 5 min with shaking at 85 rpm. Cells were pelleted (1,000 g, 4 min, 4°C) and washed twice with 10 ml of 1x PBS (1,000 g, 2 min, 4°C). Pellet was resuspended in 500 µl cold FA-150 buffer (10 mM HEPES pH 7.9, 150 mM NaCl, 0.1% SDS, 0.1% sodium deoxycholate, 1% Triton X-100) and spiked with 16.7% *S.pombe*. The cell suspension was lysed with 1ml of glass beads using the MagnaLyser (Roche; 6 x 45 s runs, 2500 g, 4°C) with cooling on ice for 2 min between the runs. The lysate was taken into 15 ml tubes by piercing the bottom of the tubes with a heated needle and spinning (300 g, 1 min). Beads were washed and spun again after addition of 500 µl FA-150 buffer. 1 ml of more lysis buffer added to 1 ml lysate to dilute samples. The lysate was sheared 30-40 min with a bioruptor sonicator 30 sec ON/30 sec OFF at high setting. The sheared lysate was cleared by centrifuge (10,000 g, 15 min, 4°C) and the supernatant was

used for IP. 500 µl sample was incubated with ~100 µg (25 µl) of the FLAG (M2) in 1.5 ml siliconized Eppendorf tubes for 15–20 h rotating at 4°C. When the IP was performed for ChIP-qPCR, 50 µl sample was diluted to 200 µl with FA-150 buffer and incubated with 5 µl (~20 µg) of the GFP antibody in 1.5 ml siliconized Eppendorf tubes for 15–20 h rotating at 4°C. Bound chromatin was immunoprecipitated for 90 min at 22°C with 50 µl protein A or G-Dynabeads pre-blocked with bovine serum albumin and sonicated salmon sperm DNA. Beads and attached chromatin were pelleted by centrifugation (640 g , 1 min) and washed with 400 µl of TSE-150 buffer (20 mM Tris–Cl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100) for 3 min, 400 µl of TSE-500 buffer (20 mM Tris-Cl pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100) for 3 min, 400 µl of LiCl buffer (0.25 M LiCl, 10 mM Tris-Cl pH 8.0, 1 mM EDTA, 1% deoxycholate, 1% NP-40) for 15 min and twice with 400 µl of TE. After washing, chromatin was eluted from the beads for 30 min at 65°C with 100 µl of elution buffer (0.1 M $NaHCO_3$, 1% SDS). Addition of 7 µl of 5 M NaCl and incubation for 3 h at 65°C reversed the cross-links before treatment of samples with 1 µl of 10 mg/ml RNase A for 1 h at 37°C and 2 µl of 20 mg/ml proteinase K overnight at 65°C. DNA was eluted with Zymoclean ChIP concentrator kit according to the manufacturer's instructions. 150 µl of the DNA was taken for Input control and isolated following by de-crosslinking and proteinase K digestion as above. DNA concentrations were measured by qubit and libraries prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina according to the manufacturer's instructions. qPCR was performed using a Corbett Rotorgene and Sybr green mix (Bioline).

2.9.2 Sequencing and data analysis

Barcoded libraries were pooled and sequenced on Illumina NextSeq 500 (75 cycle, paired). Reads were analysed using online tools on usegalaxy.org and RStudio/Bioconductor as outlined in **Table 2.7**.

Table 2.7) ChIP-seq data analysis

Steps	Package	Conditions	Platform
1) Trimming	Trim Galore!	Paired end	usegalaxy.org
2) Aligning spike-ins	Bowtie2	Paired end Pombe genome	usegalaxy.org
3) Removing PCR duplicates in spike-ins	RmDup		usegalaxy.org
4) Filtering low quality reads in spike-ins	Filter SAM or BAM	MAPQ >20	usegalaxy.org
5) Counting spike-in reads	Convert, merge, randomize	Count BAM	usegalaxy.org
6) Aligning libraries (from step 1)	Bowtie2	Paired end Cer3 genome	usegalaxy.org
7) Removing PCR duplicates in libraries	RmDup		usegalaxy.org
8) Filtering low quality reads in libraries	Filter SAM or BAM	MAPQ >20	usegalaxy.org
9) Normalising libraries to spike-ins	Downsample SAM/BAM	% of reads calculated based on Pombe reads	usegalaxy.org
10) Peak calling	MACS2 callpeak	Input reads processed as above used as control. Effective genome size = 12.5K, Scores in bedgraph	usegalaxy.org
11) Background removal	MACS2 bdgcmp	Subtraction from treatment	usegalaxy.org
12) Creating IGV files	Wig/bedgraph-to-BigWig		usegalaxy.org
13) Visualisation			IGV (Thorvaldsdóttir et al. 2013)
14) Metagene analysis	Bioconductor and DEseq2 (Love et al. 2014)		RStudio

2.10 Anchor Away

For NET-seq, 2.3 L of cells were grown in YPD to OD_{600} 0.3 (30°C, 160 rpm) and DMSO or 1 mg/ml rapamycin dissolved in DMSO added. Cells were grown to OD_{600} 0.65 (140 min for Spt4 depletion, 180 min for Spt5 depletion and 120 min for DMSO control) and harvested as described above for NET-seq. ChIP-seq, microscopy and WB samples were collected at 0, 60, 140 min (or at 0, 60, 120 and 180 min for Spt5) after rapamycin treatment.

For ChIP-seq, 45 ml of cells collected and cross-linked with 1% FA (10 min, RT). Then the reaction was quenched for 5 min with the addition of 2.5 ml of 2.5 M glycine. Cells were pelleted (1,000 g, 2.5 min, 4°C) and washed twice with 10 ml cold PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4). ChIP-qPCR was performed as described above.

For fluorescence microscopy, 13 ml of cells collected and kept in a falcon tube wrapped with aluminium foil to limit light exposure as much as possible. Then the harvested cells were cross-linked with 4% PFA (40 min, RT), pelleted (1,000 g, 2.5 min, 4°C) and washed twice with 5 ml of cold buffer B (1.2 M sorbitol, 100 mM $KHPO_4$ pH 7.5). The pellet was resuspended in residual buffer B after the second centrifuge, and 200 µl of the suspension placed on poly-L-lysine coated coverslips and incubated (30 min, 4°C). Then coverslips were washed by dipping into MQ water twice and mounted on slides with a drop of ProLong Diamond Antifade Mountant with DAPI (Vector Shield). Slides were left at RT overnight in the dark and corners of the coverslips sealed with transparent nail polish. Slides were imaged with DeltaVision CORE wide-field fluorescence deconvolution microscope using a 100x/1.4 objective lens, T%32 filter, with exposure times of 0.05 s for DAPI and 1 s for FITC channels, respectively.

For western blotting (WB), 10 ml of cells were pelleted (1,000 g, 2.5 min, at 4°C) and washed once with water**.**

2.11 MNase-seq

2.11.1 Prep of yeast nuclei

1 L of cells were grown in YPD to OD_{600} 0.6 (30°C, 160 rpm), collected by filtering and resuspended in 45 ml of cold water. Then cells were pelleted by centrifuge (1,000 g, 5min, 4°C). After discarding the water, the weight of cells (wet weight) was noted and the following volumes were used per 1 g of wet cells. 2 ml of pre-incubation solution (2.8 mM EDTA, pH 8, 0.7 M 2-mercaptoethanol) was added to wet cells and incubated (30°C, 30 min). Then samples were pelleted (1,000 g, 5 min, 4°C) and the pellet was washed with 40 ml of 1 M sorbitol. After

centrifuging (1,000 g, 5 min, 4°C) and discarding sorbitol, the pellet was resuspended in 5 ml of sorbitol/B-ME (1 M sorbitol, 5 mM 2-mercaptoethanol) solution and 200 µl of 2% of zymolase solution at 30°C for 30 min with shaking. The lysate was pelleted (3,000 g, 8 min, 4°C) and the pellet was washed with 40 ml of 1 M sorbitol. Nuclei were resuspended in 7 ml of Ficoll solution (18% Ficoll, 20 mM KH_2PO_4, 1 mM $MgCl_2$, 0.25 mM EGTA, 0.25 mM EDTA) then collected by centrifuge (20,000 g, 30 min, 4°C).

2.11.2 MNase digestion

The nuclei obtained from 0.5 g equivalent of cells were resuspended in 3 ml of freshly prepared SDB (1 M sorbitol, 50 mM NaCl, 10 mM Tris-Cl pH 7.5, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.075% NP-40, 1 mM 2-mercaptoethanol), split into 6 tubes. 6 reactions were set up with 20-40-80-160-320U of 10U/µl MNase (prepared in 200mM Tris pH 7.5, 50 mM NaCl, 50% glycerol; Nuclease S7 Roche) and incubated (37°C, exactly 10 min). Reactions were quenched with 50 µl of pre-warmed stopping buffer (5% SDS, 50 mM EDTA, at 65°C). 50 µl of 20 mg/ml proteinase K (Roche) added to MNase treated samples and incubated (overnight, 65°C). Samples were treated with 1µl of 10 mg/ml RNase A (1 h, 37°C) and then DNA was eluted with Zymoclean ChIP concentrator kit according to the manufacturer's instructions. Isolated DNA was run on a 1.5% agarose-TBE gel and right digestion (80U) was chosen based on the enriched amount of mono-nucleosome bands; faint di-nucleoseome bands are still visible without over digestion. The mononucleosomal DNA band was gel extracted. DNA concentrations were measured by qubit and libraries prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina according to the manufacturer's instructions.

2.11.3 MNase sequencing and data analysis

Barcoded libraries were pooled and sequenced on Illumina NextSeq 500 (75 cycle, paired). Reads were analysed using online tools on usegalaxy.org, a peak calling software DANPOS2

(Chen et al. 2013) and RStudio/Bioconductor as outlined in **Table 2.8**. DANPOS2 analysis was performed by Dr. Harry Fischl.

Table 2.8) MNase-seq data analysis

Steps	Package	Conditions	Platform
1) Trimming	Trim Galore!	Paired end	usegalaxy.org
2) Aligning reads	Bowtie2	Paired end Cer3 genome	usegalaxy.org
3) Removing PCR duplicates	RmDup		usegalaxy.org
4) Filtering low quality reads	Filter SAM or BAM	MAPQ >20	usegalaxy.org
5) Peak-calling for MNase-seq			DANPOS2 (Chen et al. 2013)
6) Visualisation			IGV (Thorvaldsdóttir et al. 2013)
7) Metagene analysis	Bioconductor		RStudio

2.12 Proteomics

2.12.1 Transcription complex purification

The IP was performed as in the NET-seq IP with the following modifications. 1 L of cells were grown instead of 2 L and the working volumes were halved accordingly. Lysis and wash buffers were supplemented with 1 x PhosSTOP (Roche). After the first four washes as in NET-seq, the fifth wash was performed for 20 min in wash buffer A. Next, four more washes were performed with wash buffer A-150 (20 mM HEPES (pH 7.4), 110 mM KOAc, 0.5 % Triton-X-100, 0.1 % Tween 20, 1 mM EDTA, 150 mM NaCl) for 1 min each for the first three and 30 min for the final wash.

2.12.2 Mass Spectrometry and analyses

200 µl of the eluate was submitted to Advanced Proteomics Facility at University of Oxford, Department of Biochemistry. Samples were Trypsin FASP digested with detergent. Raw MS data files were analysed with MaxQuant with (< 1% FDR) and searched against *S.cerevisiae* database. For protein quantification, LFQ intensities were used. Proteins with less than 2

peptides in FLAG IP experiments were discarded. Data were imputed and p-values were calculated with DEP package /RStudio (Zhang et al. 2018).

Significant proteins were decided by log_2FC(*spt4Δ*/WT) > 2 and p.adj < 0.01.

3 Construction and characterisation of yeast strains used in this thesis

3.1 Overview

Various genetically manipulated *S.cerevisiae* strains were used in this work to address the functions of Spt4 and Spt5 in transcription. In this chapter, these strains are fully characterised. The specific aims of this chapter are (1) to characterise the Spt4 complete knock-out strains, (2) to characterise FLAG-tagged strains with and without the Spt4 complete knock-out, and (3) to characterise strains with conditional degradation or depletion of Spt4 and Spt5 using either the auxin inducible degron (AID) or the anchor away (AA) system.

Spt5 is an essential protein that can only be depleted or deleted using a conditional system, while the s*pt4* deletion mutant is viable with a slow-growth phenotype (D A Schneider et al. 2006). The list of strains used and characterised is shown in Table 3.1.

Table 3.1 Strains used or made in this thesis

Strain	Source	Experiment performed
BY4741	Euroscarf	NET-seq, Mass Spectrometry (MS)
BY4741 *spt4::KanMX6 (SPT4 KO)*	Euroscarf	NET-seq, MS
BY4741 Rpb3-FLAG (WT)	H. Fischl	NET-seq, MNase-seq, MS, ChIP-seq, RNA-seq
BY4741 *SPT4 KO* Rpb3-FLAG (*spt4Δ*)	This study	NET-seq, MNase-seq, MS, ChIP-seq, RNA-seq
BY4741 Spt4-FLAG	H. Fischl	TEF-seq
BY4741 Spt5-FLAG	This study	TEF-seq
BY4741 *SPT4 KO* Spt5-FLAG	This study	TEF-seq
BY4741 Bur1-FLAG	Harry Fischl	ChIP-seq
BY4741 *SPT4 KO* Bur1-FLAG	This study	ChIP-seq
BY4741 Sua7-FLAG	This study	ChIP-seq
BY4741 *SPT4 KO* Sua7-FLAG	This study	ChIP-seq
BY4741 Kin28-FLAG	H. Fischl	ChIP-seq
BY4741 *SPT4 KO* Kin28-FLAG	This study	ChIP-seq

S.pombe Rpb9-FLAG	The Vasilieva Lab	NET-seq, ChIP-seq, RNA-seq
AA Spt4-FRB-GFP Rpb3-FLAG	This study	Anchor Away, NET-seq, IF, ChIP-qPCR
AA Spt5-FRB-GFP Rpb3-FLAG	This study	Anchor Away, NET-seq, IF
AA Rpb3-FLAG (No FRB)	This study	Anchor Away, NET-seq
BY4741 *dst1::kanMX6*	Euroscarf	NET-seq
BY4741 *dst1::kanMX6* Rpb3-FLAG	This study	NET-seq
BY4741 osTIR1 (No mAID)	This study	NA
BY4741 osTIR1 Spt4-mAID	This study	NA
BY4741 Rad26-FLAG	This study	NA
BY4741 *SPT4 KO* Rad26-FLAG	This study	NA
BY4741 Rad14-FLAG	This study	NA
BY4741 *SPT4 KO* Rad14-FLAG	This study	NA
BY4741 Elf1-FLAG	This study	NA
BY4741 *SPT4 KO* Elf1-FLAG	This study	NA
BY4741 Ssl2-FLAG	This study	NA
BY4741 *SPT4 KO* Ssl2-FLAG	This study	NA
BY4741 Rsc9-FLAG	This study	NA
BY4741 *SPT4 KO* Rsc9-FLAG	This study	NA
BY4741 *pADH1::SPT4*	This study	NA
BY4741 *pADH1::3xHA-SPT4*	This study	NA
BY4741 *pADH1::3xHA-SPT4* Rpb3-FLAG	This study	NA
BY4741 *ISW1 KO* Rpb3-FLAG	This study	NA
BY4741 *SPT4 KO ISW1 KO* Rpb3-FLAG	This study	NA
BY4741 *FKH1 KO* Rpb3-FLAG	This study	NA
BY4741 *SPT4 KO FKH1 KO* Rpb3-FLAG	This study	NA
BY4741 *DST1 KO ISW1 KO* Rpb3-FLAG	This study	NA
BY4741 *DST1 KO FKH1 KO* Rpb3-FLAG	This study	NA

Most of the previous studies focusing on Spt4 were performed in *SPT4* deletion mutants. To be able to compare the results in this work with the previous data, the function of Spt4 was addressed by performing experiments on *SPT4* knock-out (*spt4Δ*) cells and comparing them to the isogenic WT strain BY4741. However, to assess more directly the effect of loss of Spt4 and Spt5 in real time, two conditional depletion systems were tested, the anchor away (AA) system and the auxin induced degron (AID) system.

3.2 Results

3.2.1 Spt4 knock out cells have a slow growth phenotype

SPT4 knock out cells were supplied from Euroscarf (BY4147 *spt4::KanMX6*). Complete deletion of *SPT4* was verified by PCR with primers targeting endogenous *SPT4* sequence (**Figure 3.1**). A *SPT4* knock-out strain was previously reported to grow ~ 20% slower than its parental strain (D A Schneider et al. 2006). The slow-growth phenotype of the *SPT4* knock-out strain used here was tested by calculating the doubling time from the growth curve obtained by taking optical densities (OD_{600}) at every 20 min as described in Methods. The quantitative analysis showed that the *SPT4* knock-out strain grows significantly slower (13%, $p < 0.001$) than the isogenic WT strain BY4147 (**Figure 3.3**). This strain was used as the *SPT4* knock out background to create further strains for this work. Both WT and *SPT4* knock out cells were engineered to create strains expressing C-terminal endogenously FLAG-tagged proteins (Rpb3, Spt4, Spt5, Sua7, Kin28, and Bur1).

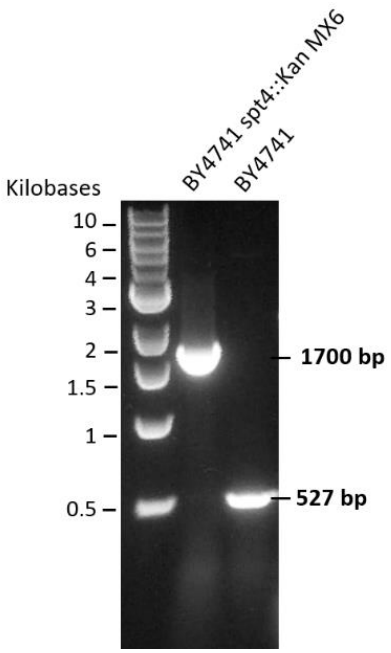


Figure 3.1) *SPT4* deletion was confirmed by PCR

PCR products amplified by primers targeting *SPT4* gene in BY4147 and BY4147 *spt4::KanMX6* . Samples were run subject to electrophoresis on a 1% TBE agarose gel.

3.2.2 Construction and characterisation of FLAG-tagged strains

Strains expressing endogenously C-terminal FLAG-tagged proteins (Rpb3, Spt4, Spt5, Sua7, Kin28, and Bur1) are used for the IP step in the NET/TEF/ChIP-seq protocols in BY4147 and BY4147 *SPT4* KO backgrounds (**Table 3.1**). The FLAG-tagging of the strains was confirmed by western blotting using antibodies raised against the FLAG epitope (**Figure 3.2**).

The doubling time of the FLAG-tagged strains was assayed to test if the tagging interferes with the function of the proteins. There was a trend towards slower growth upon Bur1 FLAG-tagging in WT (3%, $p<0.001$) and Rpb3, Sua7, and Kin28 FLAG-tagging in Spt4 knock-out background (5%, 3%, and 4%, respectively, all $p<0.001$) (**Figure 3.3**). Given that Rpb3, Sua7, Bur1 and Kin28 are essential proteins, a 3-5% reduction in the doubling time after the addition of the FLAG-tag was not considered a major growth deficiency and experiments were performed using the appropriate strains as summarised at **Table 3.1**.

The BY4147 *Rpb3-3xFLAG-HisMX6* cells were designated wild type (WT) and BY4147 *Rpb3-3xFLAG-HisMX6 spt4::KanMX6* cells were designated *SPT4* knock-out (*spt4Δ)* cells. This allows direct comparisons to be made between experiments such as those assessing (i) nascent transcription (NET-seq), (ii) steady-state transcript levels (RNA-seq), (iii) pull-down of transcriptionally engaged RNAPII and characterisation of the associated factors by mass spectrometry and (iv) assessing the position of nucleosomes by MNase-seq.

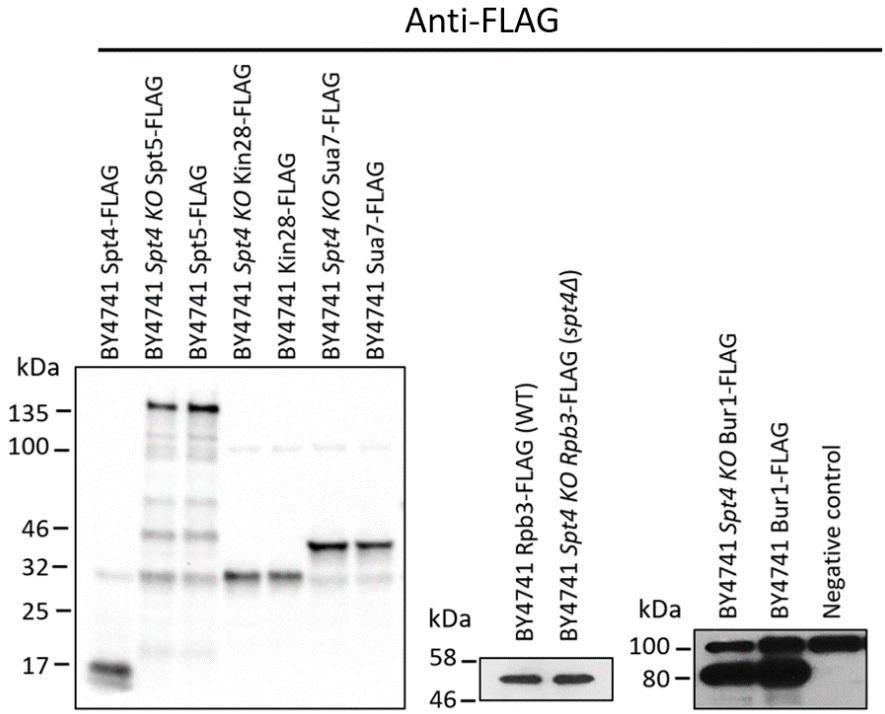


Figure 3.2) Western blot of the FLAG-tagged strains

FLAG-tagged strains were run in 4-20% gradient and 10% SDS-PAGE and blotted against the FLAG epitope with anti-FLAG (M2) antibody. The blot on the right shows Bur1-FLAG proteins at around 80 kDa and the upper band around 100 kDa is non-specific. See Appendix 1 for the whole blots.

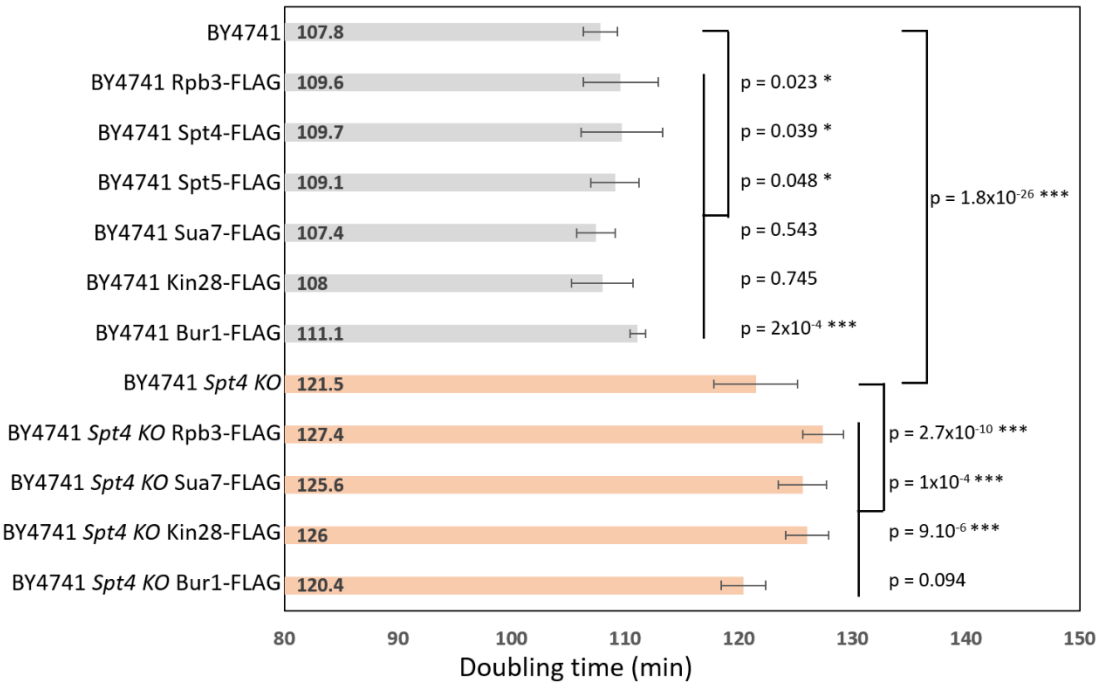


Figure 3.3) Doubling times of the strains used in this study

Experiments are performed at least in two biological and four technical replicates. OD_{600} of the strains were noted at every 20 min by generating growth curve using the Bioscreen for 22 h at 30°C. OD_{600} measurements were analysed in R. Reads were blanked by subtracting medium-only reads. Doubling times were calculated by choosing exponential growth phase (OD_{600} 0.2 to 0.7). Error bars indicate standard deviation of 2-5 biological replicates performed at 4 technical repeats. * p-value <0.05, ***p-value <0.001 (Student's t-test, unpaired, two-tailed).

3.2.3 The auxin-inducible degron (AID) system for conditional protein degradation

For conditional removal of Spt4, the auxin-inducible degron (AID) system was tested for its effectiveness. The AID system makes use of a plant degradation mechanism relying on the interaction between the three proteins/complexes triggered by the plant hormone, auxin. In the presence of auxin, transport inhibitor response 1 (TIR1) interacts with E3 ubiquitin ligase and the Auxin/Indole-3-acetic acid (Aux/IAA) family transcriptional repressors. TIR1-E3 ubiquitin ligase then mediates ubiquitination and subsequent degradation of the Aux/IAA family proteins by targeting them to the proteasome (**Figure 3.4**). The degradation complex and proteasome are conserved in eukaryotes, but TIR1 is a plant-specific protein. To adapt the AID system for yeast, the target protein is tagged with the Aux/IAA family protein domain

IAA17 (the AID-tag) in yeast cells expressing TIR1 (OsTIR1; cloned from *Oryza sativa*). Thus, the TIR1-E3 ligase can induce targeted ubiquitination of the AID-tagged protein upon auxin addition, and lead to its degradation by the proteasome (Yesbolatova et al. 2019).

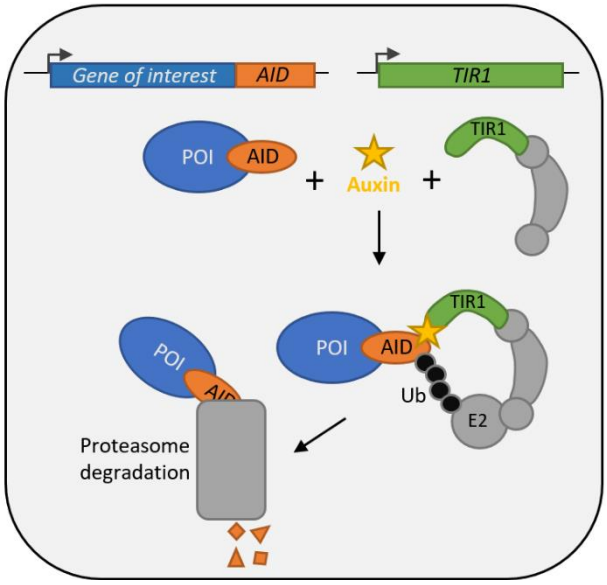


Figure 3.4) Auxin inducible degron (AID) system

The AID system allows conditional degradation of protein of interest (POI).

Spt4 is endogenously tagged with a shorter and codon optimised version of the AID-tag (mAID-tag) at its C-terminus (Nishimura and Kanemaki 2014). To induce degradation of Spt4-mAID, the growth media was supplemented with indole-3-acetic acid (IAA) which is the most abundant auxin hormone. Successful degradation of Spt4 is expected to change doubling times of the Spt4-mAID cells upon IAA addition, as *SPT4* knock out cells have a slow-growth phenotype. To test if the AID system is working, the AID background (BY4147 *osTIR*; No mAID) and Spt4-mAID strains were grown in three different conditions; YPD, YPD + EtOH (vehicle control), or YPD + auxin (500 µM IAA in EtOH) for 22 hours and their doubling times were measured as described in the Methods (**Figure 3.5**).

First, the doubling time of the No mAID control strain was analysed. In the No mAID cells, the addition of IAA (dissolved in EtOH) leads to a mild growth defect (7%, $p<0.001$), compared to the cells were grown in YDP, meaning that auxin has a small effect in the cells even without

an mAID tagged protein (**Figure 3.5**). Notably, in these cells, the vehicle control (the addition of EtOH in YPD) was also detected to create a very mild (2%), yet significant ($p<0.01$) slow growth phenotype (**Figure 3.5**). Next, the doubling time of Spt4-mAID cells was examined. Even without auxin addition, the presence of the mAID tag on Spt4 increased the doubling time of the cells by 20% ($p<0.001$) compared to No mAID cells grown in YPD (**Figure 3.5**). This is comparable with the 13% increase in the doubling time of the *SPT4* knock out cells compared to WT cells (see above, **Figure 3.3**). The non-specific slow growth phenotype observed in Spt4-mAID cells might indicate several problems in the AID system constructed to degrade Spt4. One possibility is that the mAID tagging of Spt4 interferes with the function of the protein and leads to the *SPT4*$^{-}$ slow growth phenotype without degradation of the protein. In this case, the loss of function of Spt4 would not be conditional. Another possibility is that Spt4-mAID is getting degraded without the addition of exogenous IAA/auxin. Indeed, auxin independent degradation of AID-tagged proteins has recently been reported in various organisms, including yeast (Sathyan et al. 2019). Moreover, even though IAA was first described as a plant hormone, studies reported high levels of IAA in yeast as an intermediate of the tryptophan degradation pathway, as well as low levels of IAA in the absence of tryptophan (Sun et al. 2014). Therefore, mAID-tagged Spt4 might have been degraded without the addition of exogenous IAA/auxin.

Despite the slow growth phenotype of the Spt4-mAID cells without the addition of exogenous IAA, upon IAA addition, the doubling time of Spt4-mAID cells was increased by a further 10% compared to that of Spt4-mAID cells grown in YPD (**Figure 3.5**). This additional 10% increase is similar to the 7% increase observed in the doubling time of No mAID cells upon IAA treatment. Therefore, it is not clear if the slow growth rate in Spt4-mAID cells grown in IAA is due to the degradation of Spt4-mAID or due to the secondary effects of auxin addition. Overall, the data show that (1) IAA addition leads to a non-specific growth phenotype even in the absence of mAID tagged protein, and (2) the mAID-tagging of Spt4 creates a nonspecific

slow growth phenotype even in the absence of the addition of IAA. As the changes in the doubling times of AID strains failed to provide an auxin specific and conditional phenotype, next, the AA system was tested as an alternative method of conditional removal of Spt4.

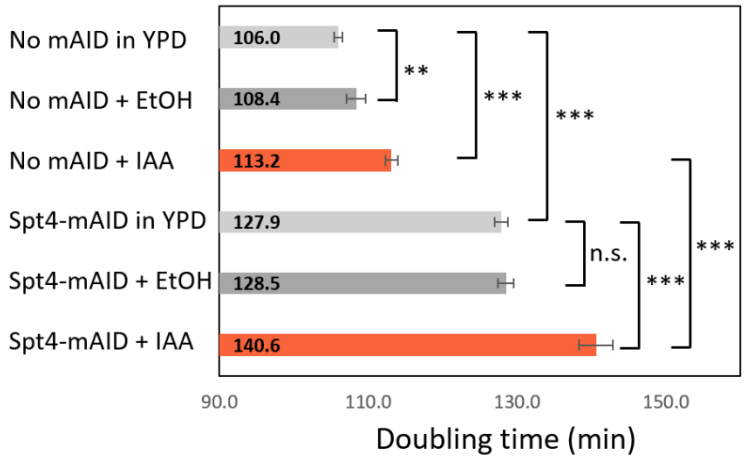


Figure 3.5) Doubling times of the degron strains

Cells were grown in YDP, YPD + EtOH and YPD + IAA (500 µM in EtOH) for 22 h. OD_{600} was recorded every 20 min using the Bioscreen and doubling times were calculated for exponential growth phase (OD_{600} 0.2 to 0.8) as described in the methods. Error bars indicate standard deviation of 3 biological replicates performed at 4 technical repeats. * p-value <0.05, ***p-value <0.001 (Student's t-test, unpaired, two-tailed).

3.2.4 Anchor away system: Depleting Spt4 and Spt5 from the nucleus

The AA system was tested as an alternative method to conditionally deplete Spt4 and Spt5. This system allows conditional removal of a target protein from the nucleus by utilising the heterodimerization of human FKBP12 protein (anchor) with FRB protein (target) upon rapamycin addition. A highly abundant protein Rpl13A (a large ribosomal subunit) shuttling between cytoplasm and nucleus is tagged with FKBP12 and the target protein is tagged with FRB in rapamycin resistant cells (*tor1* mutant). Therefore, the target protein can be rapidly exported to the cytoplasm and removed from the nucleus (Haruki, Nishikawa, and Laemmli 2008) (**Figure 3.6**).

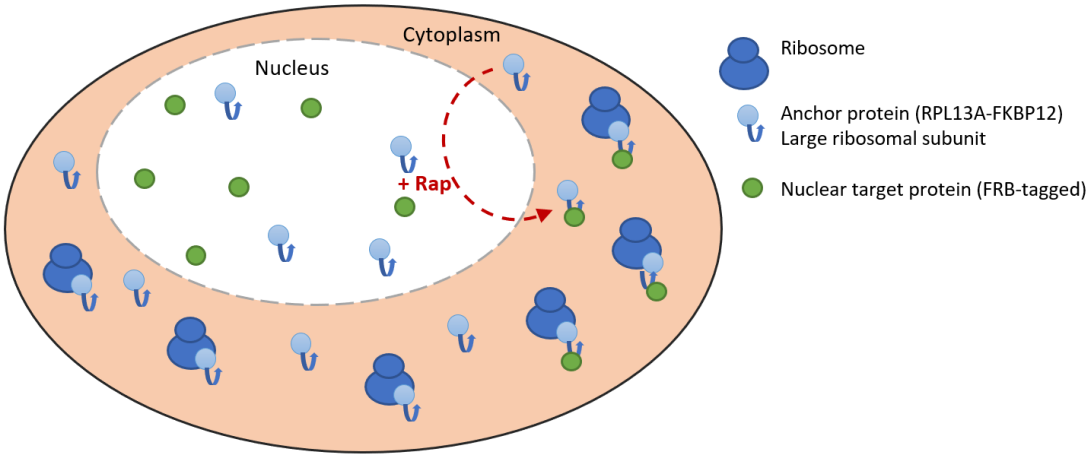


Figure 3.6) Anchor Away system

Anchor away allows conditional depletion of nuclear proteins from nucleus to cytoplasm.

In this work, both an FRB-tag, for FRB dimerization upon rapamycin addition, and a GFP-tag to track and verify the protein localisation in individual cells by immunofluorescence (IF) microscopy are fused to the C-terminal region of Spt4 and Spt5 (FRB-GFP-tag) in a strain also expressing FLAG-tagged Rpb3 to enable NET-seq to be performed. The FRB-GFP-tagging of the strains was verified by western blotting using antibodies against the GFP epitope (**Figure 3.7**). In all experiments, data obtained from the FRB-GFP-tagged Spt4 or Spt5 strains (Spt4-FRB or Spt5-FRB) was compared to the signal in the AA background strain with no FRB tag (No FRB) or DMSO (vehicle for rapamycin) treated FRB-GFP-tagged strains.

Depletion of Spt4 and Spt5 from the nucleus should change their doubling times, as an *SPT4* knock out grows slowly and an *SPT5* knock out is inviable. To test this, the doubling times of the AA background (No FRB), Spt4-FRB and Spt5-FRB strains were measured as described in the methods while the strains are growing in three different conditions; YPD, YPD + DMSO, and YPD + rapamycin (1 mg/ml in DMSO) for 22 hours (**Figure 3.8**).

No growth defects were observed upon FRB-tagging of Spt4 and Spt5, as the doubling times of the Spt4-FRB and Spt5-FRB cells were not different than that of the No-FRB cells in YPD (3%, and 2% increase in the doubling times, respectively, but not significant, **Figure 3.8**). In DMSO, the doubling times of the Spt4-FRB cells did not change, while there was a small

increase in the doubling times of the Spt5-FRB cells (3%, $p<0.05$) compared to their YPD controls (**Figure 3.8**). Additionally, the doubling times of the No-FRB cells were not affected by the addition of DMSO or rapamycin to the medium (**Figure 3.8**).

Upon rapamycin treatment, the doubling time of the Spt4-FRB cells increased (37%, $p<0.001$) compared to the controls (Spt4-FRB in DMSO or YPD or No-FRB cells), suggesting depletion of Spt4 from the nucleus (**Figure 3.8**). Importantly, however, this increase was greater than the 13% increase in the doubling time of the *SPT4* knock out cells compared to WT. Therefore, it is possible that the rapamycin addition affects the cells non-specifically in addition to the effect of Spt4 depletion. Alternatively, the growth defect could simply reflect the conditional loss of Spt4, which is compensated for in the deletion strain by the accumulation of additional secondary mutations or changes.

Once treated with rapamycin, the Spt5-FRB cells could only grow until mid-log phase and then stop dividing, reflecting the essential function of Spt5 (**Figure 3.8**). Thus, the data indicate that rapamycin addition causes slow growth in the Spt4-FRB or Spt5-FRB, but not in the No-FRB cells, and this is likely to reflect the specific depletion of the target proteins from the nucleus.

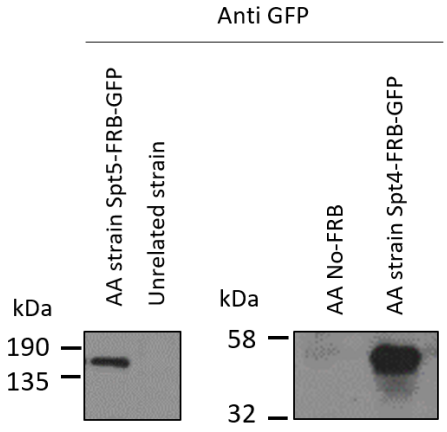


Figure 3.7) Western blot of the FRB-GFP-tagged strains

FRB-GFP-tagged strains were run in 15% SDS-PAGE and blotted against the GFP epitope with anti-GFP (Abcam290) antibody. See Appendix 1D&E for the whole blots.

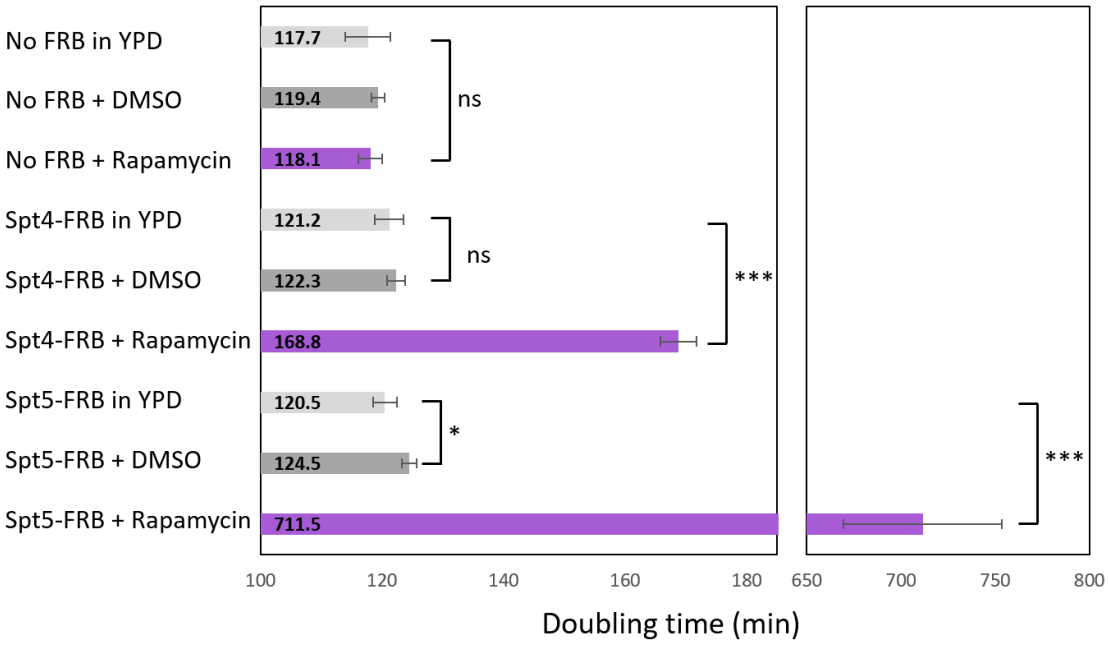


Figure 3.8) Doubling times of the anchor away strains

Cells were grown in YDP, DMSO and rapamycin (1 mg/ml in DMSO) for 22 h. OD_{600} was recorded every 20 min using the Bioscreen and doubling times were calculated for exponential growth phase (OD_{600} 0.2 to 0.8) as described in the methods. Error bars indicate standard deviation of 3 biological replicates performed at 4 technical repeats. * p-value <0.05, ***p-value <0.001 (Student's t-test, unpaired, two-tailed).

Notably, the growth assay was performed for 22 hours to test if the AA system was working which is assessed by the effect in doubling time. The growth assay proved that the AA system was working. However, here, the immediate changes in the cells were desired to be monitored and therefore, the rapamycin treatment was carried out for a shorter time. 60 min of depletion was shown to be sufficient for most of the proteins tested (Haruki et al. 2008). To demonstrate and quantify the efficiency of the AA in the Spt4-FRB strain, IF and ChIP-qPCR were performed on Spt4-FRB at time points 0, 60, 140 min after rapamycin addition. IF images show a gradual change in Spt4 localisation from time point 0 to 140 min reflected by the GFP signal. At time point 140 min, the GFP signal wass mainly in the cytoplasm with very low nuclear signal compared to the untreated cells (**Figure 3.9A**). Consistently, ChIP-qPCR analysis shows a significant reduction on the Spt4 binding at three representative genes at time points

60 and 140, compared to time point zero (**Figure 3.9B**). Thus, 140 min of treatment with rapamycin was chosen as optimal for depletion of Spt4-FRB from the nucleus.

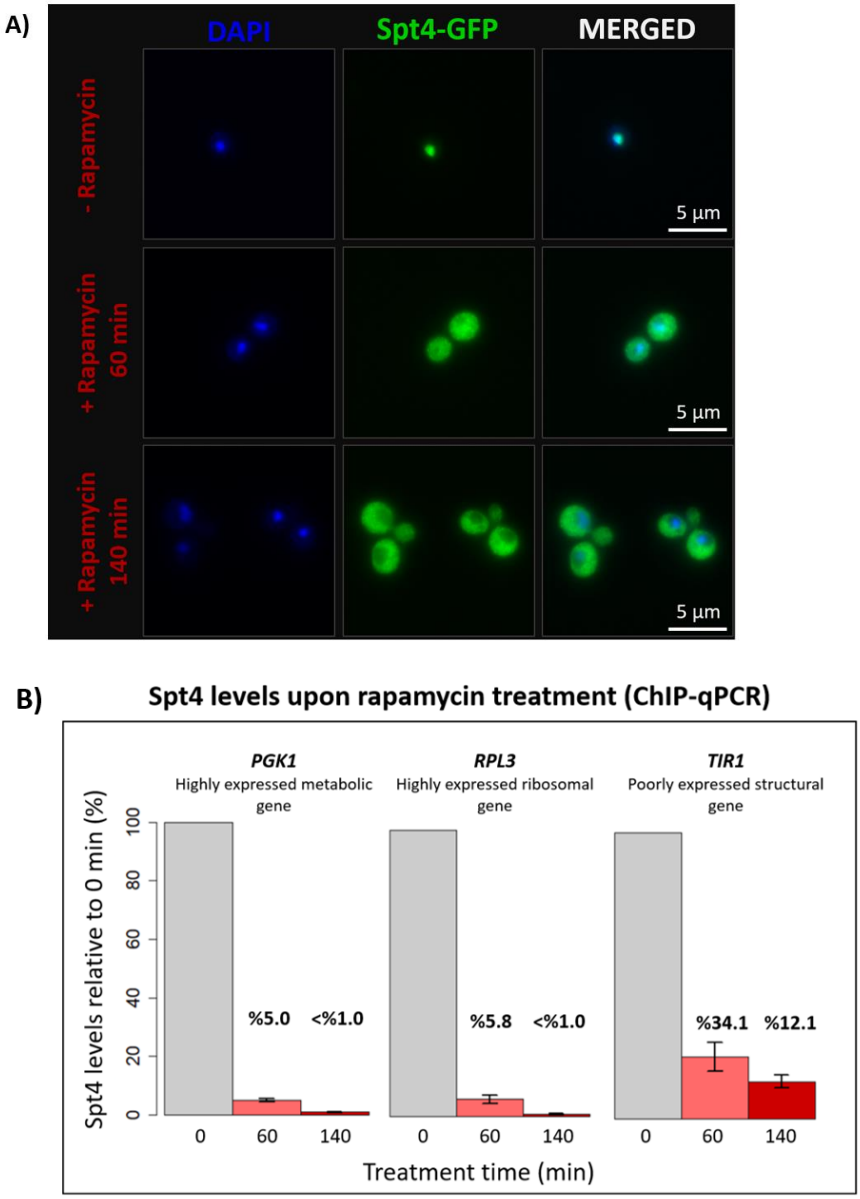


Figure 3.9) Spt4 is depleted from nucleus by anchor away system

A) IF images for Spt4-FRB samples at time points 0, 60 and 140 min upon rapamycin addition. DAPI staining is used for nucleus, GFP is expressed with Spt4 (Spt4-FRB-GFP).

B) Percentage of Spt4 levels relative to time point 0 levels at the three representative genes tested by ChIP against GFP (targeting Spt4-FRB-GFP) followed by qPCR. Error bars indicates standard deviation of the two biological replicates.

Similarly, in the Spt5-FRB-GFP strain, a gradual change in Spt5 localisation from time point 0 to 180 min by IF images was observed. Note that although the GFP signal was mainly in the cytoplasm at time point 180 min, there was still some signal in the nucleus (**Figure 3.10**). However, cells do not divide after 3 hours of treatment with rapamycin (OD_{600} does not exceed 0.7) indicating effective depletion of functional Spt5. Thus, 180 min was chosen as the optimal time for Spt5 depletion.

In conclusion, the majority of Spt4 and Spt5 proteins were depleted in Spt4-FRB and Spt5-FRB strains, respectively, after treated with rapamycin. The effects of these depletions on RNAPII distribution will be discussed in chapters 4 and 5, respectively.

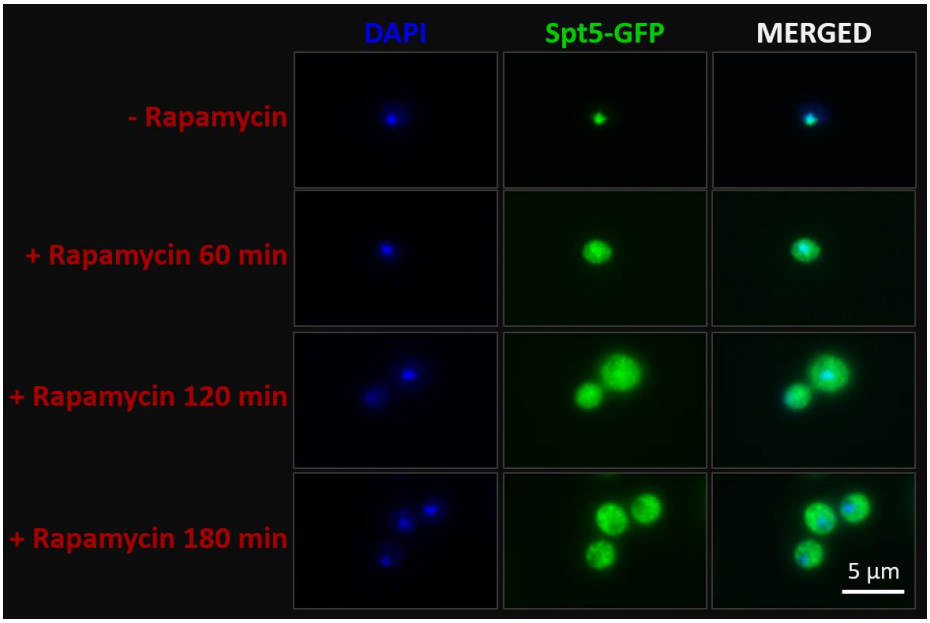


Figure 3.10) Spt5 is depleted from nucleus by anchor away system
IF images for Spt5-FRB samples at time points 0, 60, 120 and 180 min upon rapamycin addition. DAPI staining is used for nucleus, GFP is expressed with Spt5 (Spt5-FRB-GFP).

3.3 Discussion

In this chapter, the Spt4 knock out cells, and cells in which Spt4 or Spt5 are anchored away in the cytoplasm (known as Spt4/5 AA cells) were characterised to achieve a detailed study of their functions in transcription, described in the following chapters. The *SPT4* knockout strains were shown to reproduce the published slow-growth phenotype and were used to construct

a number of FLAG-tagged strains for further analysis. A range of different methods were examined to facilitate the loss of Spt4 and Spt5 in the cells. The AID system was constructed for Spt4 but shown to produce non-conditional (auxin-independent) effects on growth and was not used in these studies. The AA system was shown to deplete both Spt4 and Spt5 from the nucleus in real time and produce the expected effect on growth and was the method of choice for conditional depletion in these studies.

One drawback of AA system is that anchoring the target protein away could potentially anchor other proteins associated with the target away from the nucleus as well. In the work presented here, it was assumed that anchoring Spt4 away did not anchor Spt5 away since Spt5 depletion results in severe growth defects, whereas Spt4 depletion results in a slow-growth phenotype, but not complete growth defect. The localisation of Spt4 upon anchoring Spt5 away, on the other hand, has not been tested. In fact, it is very likely that the interaction of Spt4 with RNAPII is dependent on Spt5 as shown by multiple structural studies both in yeast and human (Ehara et al. 2017; Vos et al. 2018). In that case, the depletion of Spt5 regardless of the method being used would inevitably result in depletion of Spt4 from the transcription complex as well. This could be tested by labelling Spt4 with different fluorescent marker, such as RFP and checking the localisation of it upon Spt5 depletion.

Conditional deletion or depletion of a protein (from the nucleus, for the case studied here) are usually used to abrogate the function of essential proteins. For example, *SPT5* is an essential gene and the anchor away protocol was used to enable conditional depletion of Spt5. Although Spt4 is not an essential protein, a similar approach was taken with Spt4 to rapidly deplete this protein from the nucleus, in addition to using *spt4Δ* cells. In this way, the data collected from Spt4 AA and *spt4Δ* cells complement each other, and enable a comparison between the long and short term effects of the loss of Spt4 on the distribution of RNAPII and verify the results in different backgrounds. These data will be presented in Chapter 5.

4 Mapping Spt4 and Spt5 on genes across the genome

4.1 Overview

It is key to know the localisation of a protein to understand its function. In this chapter, the occupancy profiles of Spt4 and Spt5 on RNAPII across the genome at high resolution are determined to provide insights into their function in transcription. As a secondary question, the dependence of Spt4 to position its binding partner Spt5 was examined. To do this transcription elongation factor (TEF) associated nascent elongating transcript sequencing (TEF-seq) was used (Fischl et al. 2017). Specifically, the aims of this chapter are (1) identifying the genome-wide position of Spt4 and Spt5 in WT cells, (2) determining what genes Spt4 and Spt5 preferentially bind, and (3) mapping the genome-wide position of Spt5 in *spt4Δ*.

The positions of Spt4 and Spt5 on genes were previously shown by ChIP-based methods (Baejen et al. 2017; Mayer et al. 2010). Although giving a global view of Spt4 and Spt5 occupancy, ChIP-based experiments suffer from the lack of high resolution and strand specificity. Moreover, crosslinking in ChIP can distort mapping as a result of low affinity or indirect binding. Therefore, we sought a more sensitive technique and decided to use a native elongating transcript sequencing (NET-seq) based method.

NET-seq is a technique to map the position of RNAPII based on the position of the 3' nucleotide of the nascent transcript in the active site of the enzyme that can be analysed after immunoprecipitation of the transcription elongation complex (TEC) using FLAG-tagged Rbp3 (**Figure 4.1**). The native RNA is isolated from the immunoprecipitated complex and sequenced from the 3'-end. The most 3'-end of the RNA indicates the position of RNAPII at single-nucleotide resolution, in a strand-specific manner without requiring crosslinking (Churchman and Weissman 2011).

TEF-seq is a variation of NET-seq (Fischl et al. 2017) (**Figure 4.1**). In this technique, instead of RNAPII, the TEC is immunoprecipitated with a FLAG-tagged transcription elongation factor. This protocol, therefore, relies on the affinity between RNAPII and the tagged-TEF, and the resulting TEF position is mapped relative to the RNAPII position. The position of the 3'-end of the native transcript in the immunoprecipitated complex is mapped as for NET-seq. This gives the single nucleotide-resolution mapping of the position of the TEF on RNAPII. Some factors such as kinases and phosphatases proved hard to IP with RNAPII, but others including Spt6, Spt16, and Paf1 showed stable association during transcription (Fischl et al. 2017).

FLAG-tagged Rpb9 *S.pombe* spike-ins were included in the NET-seq and TEF-seq protocols. In this way, global changes in RNAPII or factor densities could be considered and the data could be quantitively interpreted. *S.pombe* was used as spike-in controls for a few reasons. First, *S.pombe* cells could be cultured easily without needing a different setup than for budding yeast. Second, a FLAG-tagged RNAPII subunit of *S.pombe* was already available. Lastly and most importantly, *S.pombe* and *S.cerevisiae* genomes are distinct enough to enable unique alignment of > 15 nt long RNA reads originating from the respective organisms. Therefore, 4 parts of frozen *S.cerevisiae* pellet were spiked-in with 1 part of frozen *S.pombe* by weight before the grinding steps in the NET-seq protocol. The addition of spike-ins early in the protocol was also useful to account for technical variations that could occur at all the steps from grinding to library preparations.

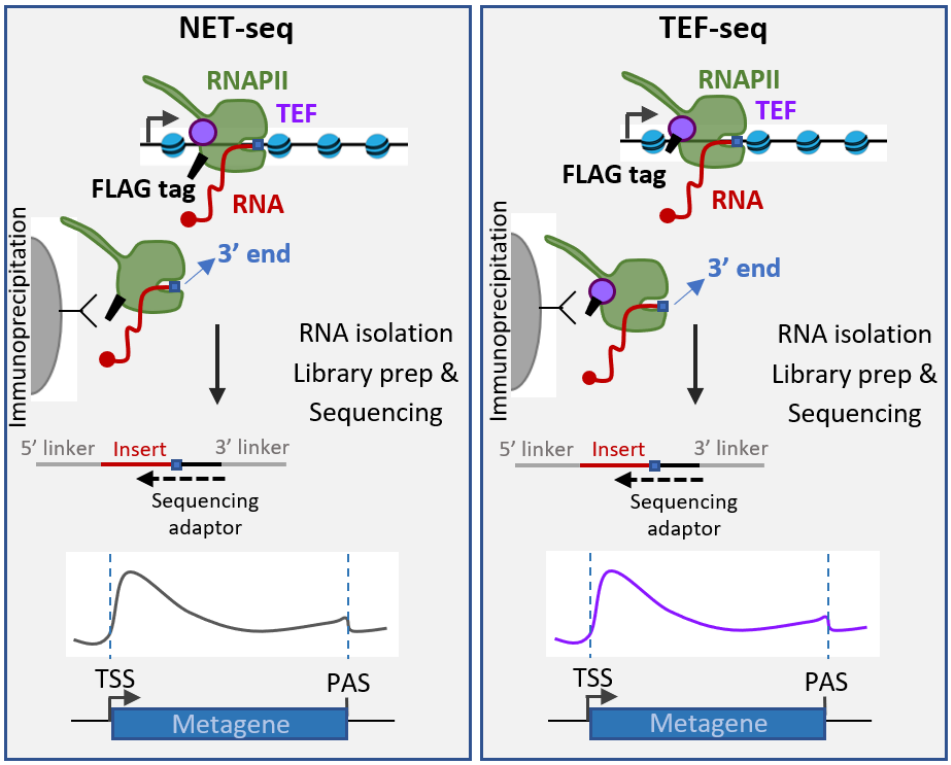


Figure 4.1) Sequencing techniques to map RNAPII and TEFs.

Native elongating transcript sequencing (NET-seq) pulls down elongation competent RNAPII and the 3'-end sequencing allows mapping of RNAPII at single nucleotide resolution. Similarly, transcription elongation factor (TEF) associated nascent elongating transcript sequencing (TEF-seq) maps TEF-associated RNAPII.

4.2 Results

4.2.1 Immunoprecipitation and data analysis of NET-seq and TEF-seq

The efficiency of the NET-seq and TEF-seq IPs were checked by western blotting using antibodies against the FLAG epitope. Strong bands at expected sizes for Rpb3, Spt4, and Spt5 in the eluates verify effective pull-down of the target proteins (**Figure 4.2**). Following the IPs in experimental duplicate, RNAs were isolated, sequenced, and aligned to the yeast genomes. Notably, to account for the IP of the undesired RNA and to remove the background signal during the data analysis, samples without FLAG-tag (no tag control) were also processed in parallel to the tagged samples.

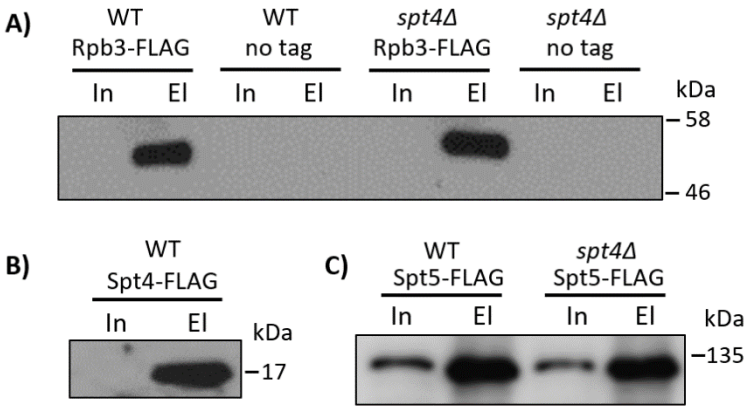


Figure 4.2) Immunoprecipitation of FLAG-tagged cells

Western blot of Input (In) and eluate (El) samples with anti-FLAG (M2) antibody run in 15 and 10 % SDS-PAGE. **A)** Rpb3-FLAG, no tag, *spt4Δ* Rpb3-FLAG, and *spt4Δ* no tag. **B)** Spt4-FLAG and **C)** Spt5-FLAG and *spt4Δ* Spt5. See Appendix 2 for the whole blots.

Sequencing reads were analysed in RStudio as described in the Methods. Briefly, multiply aligned reads were removed, and only uniquely aligned reads were kept. This filtering usually leaves ~18-24% of the total reads for the analysis (**Table 4.1**). Aligned reads were then trimmed to the 3'-ends, normalised based on *S.pombe* reads, and annotated to the *S.cerevisiae* genes using those described in (Pelechano, Wei, and Steinmetz 2013).

Table 4.1) NET-seq and TEF-seq read counts, alignment and spike-in ratios.

Sample	Total reads	Uniquely aligned reads	% of uniquely aligned reads	% of uniquely aligned *S.pombe* reads
Rpb3 #1	65825472	15542590	23.6	18.3
Rpb3 #2	63731533	15459921	24.3	18.0
Spt4 #1	53144178	9281278	17.5	27.6
Spt4 #2	73834296	15982773	21.7	24.5
Spt5 #1	64786577	13268751	20.8	23.9
Spt5 #2	47288902	8663001	18.3	23.1
No tag	65690548	12368537	18.83	69.5
***spt4Δ* Rpb3 #1**	66269665	14833491	22.4	12.1
***spt4Δ* Rpb3 #2**	87285765	20887292	23.9	15.1
***spt4Δ* Spt5 #1**	57895501	12712909	22.0	24.8
***spt4Δ* Spt5 #2**	43787211	8269081	18.9	22.3
***spt4Δ* no tag**	60251158	10589267	17.6	75.0

See Appendix table 1 for the extended list of all NET-seq and TEF-seq performed in this study.

4.2.2 NET-seq and TEF-seq are reproducible

The reproducibility of the experimental repeats was tested by counting reads over the transcription units (taken from the TSS to the PAS) and calculating Spearman's correlation coefficient. Correlation plots show that Rpb3, Spt4, Spt5, *spt4Δ* Rpb3, and *spt4Δ* Spt5 repeats are reproducible (Spearman's ρ=0.983, 0.976, 0.977, 0.982, and 0.984 respectively, **Figure 4.3**). The NET-seq data presented here is also consistent with the previously NET-seq published data (Fischl et al. 2017) (**Figure 4.3**).

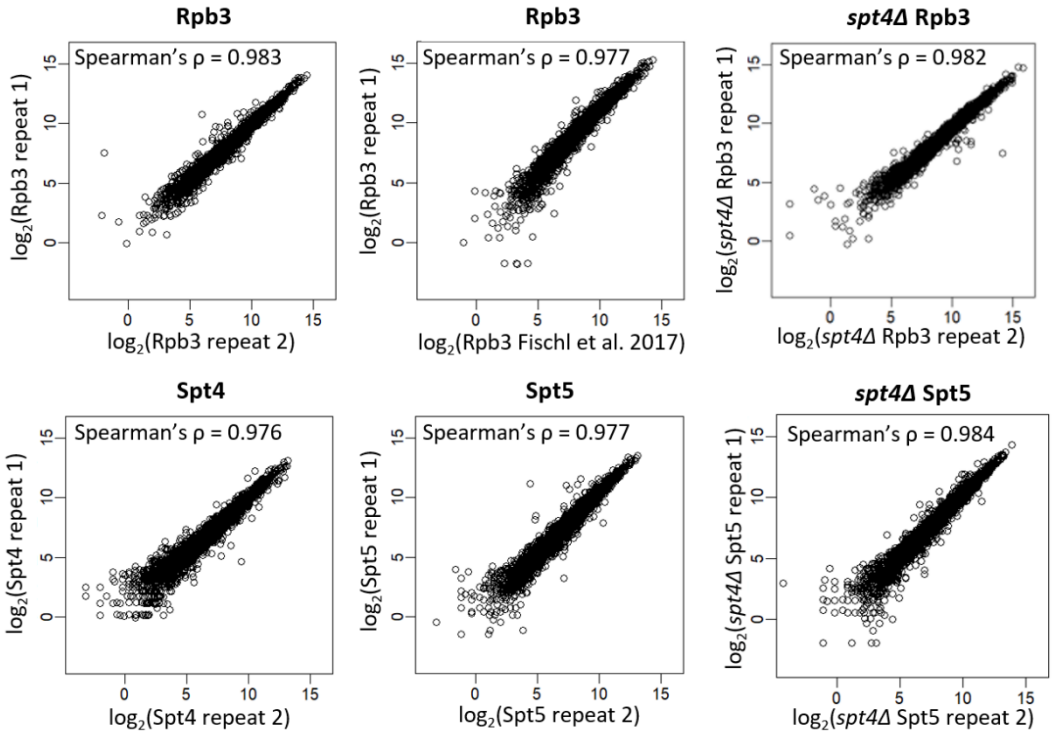


Figure 4.3) Correlations between NET-seq or TEF-seq repeats from this study and with published NET-seq data from Fischl et al. (2017)

Reads are counted from the TSS to the PAS for each gene. Log_2 transformed gene counts are correlated and Spearman's ρ calculated for each pair.

4.2.3 Where are Spt4 and Spt5 on genes across the genome?

First, the positions of Spt4 and Spt5 in WT conditions were analysed. Integrative genomics viewer (IGV; **Figure 4.4**) tracks of Rpb3, Spt4, and Spt5 display the reproducibility between the repeats at representative genes (Thorvaldsdóttir et al. 2013). In addition, reads from the intronic region of *YDR381W* show the successful detection of the native RNA (**Figure 4.4**). Unsurprisingly, the Spt4 and Spt5 signals are very similar to each other in terms of both patterns and levels since Spt4 and Spt5 form a very strong complex and are expected to be co-localised (Hartzog et al. 1998). The Spt4 and Spt5 profiles also match the Rpb3 profile; the factor signals are detected as early as the Rpb3 signal and follow the Rpb3 profile over the gene bodies, suggesting engagement of these factors with RNAPII throughout transcription (**Figure 4.4**).

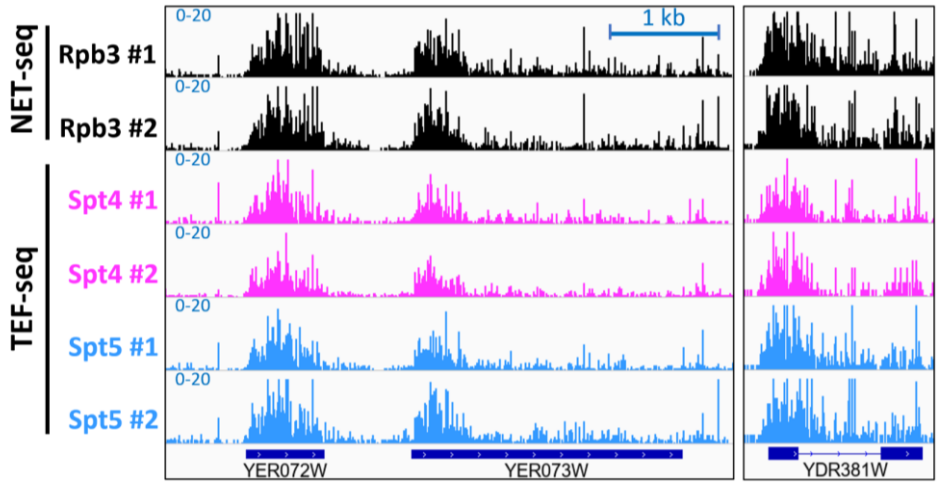


Figure 4.4) Rpb3, Spt4, and Spt5 occupancies at representative genes

IGV tracks of NET-seq (Rpb3) and TEF-seq (Spt4 and Spt5) reads of three representative genes transcribed from the positive strand: *YER072W, YER073W,* and *YDR381W* in two biological replicates. The dark blue boxes indicate the transcribed region of the genes (from TSS to PAS), while the blue line indicates the intronic region in *YDR381W.*

A metagene analysis for protein-coding genes (PCGs) was performed to demonstrate the genome-wide positions of Rpb3, Spt4, and Spt5. The two replicates of each experiment were combined, as they were highly reproducible. 4820 PCGs (>750 nt) were taken and genes with negative values (due to no tag normalisation, see the Methods) were discarded. To avoid

genes with incorrect TSS annotations, genes having 1.5x more reads upstream of the TSS (-150 to 0 nt) than downstream of the TSS (+1 to 150 nt) were also discarded. 4212 PGCs were plotted as metagenes relative to the TSS.

The metagene profiles for Spt4 and Spt5 are similar to that of Rpb3 around the TSS and over the gene bodies (**Figure 4.5**). Spt4 and Spt5 signals rise as early as the Rpb3 signal at the 5'-end, showing the involvement of the factors at the early stage of transcription. These results agree with the structural models predicting that Spt4 and Spt5 replace transcription initiation factors in the elongation complex (Bernecky et al. 2017; Grohmann et al. 2011). Over the gene bodies, the factors remain associated with RNAPII, and at the 3'-end of the genes Spt4 and Spt5 signals drop to the background levels before the PAS (**Figure 4.5**). This region coincides with the peak of Rpb3 where reduced speed and pausing of RNAPII is predicted to be important for the transition from elongation to termination complex (Mischo and Proudfoot 2013). In *S.pombe*, Spt5 has been shown to dissociate from RNAPII at the 3'-end to allow binding of termination factors (Kecman et al. 2018). The drop in the Spt5 signal, therefore, is likely to reflect dissociation of Spt5 from the RNAPII during the transition from the elongation to the termination complex and Spt4 follows the same pattern (**Figure 4.5**). Interestingly, just downstream of the PAS, after clear depletion of Spt4 and Spt5, the factors peak again. This will be further discussed in the section (4.2.4) .

Thus, to conclude, TEF-seq supports Spt4 and Spt5 joining RNAPII right after initiation, travelling with elongating RNAPII and dissociating from RNAPII upstream of the PAS.

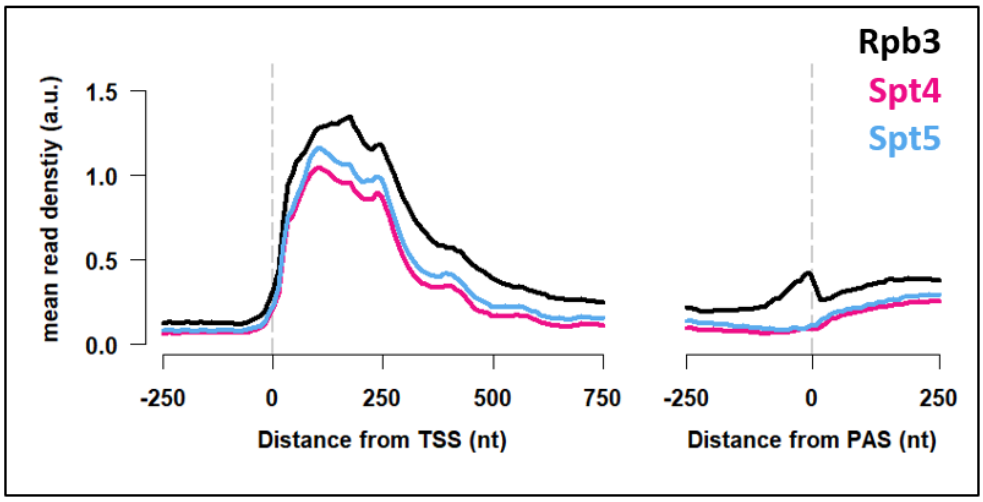


Figure 4.5) Spt4 and Spt5 travel with RNAPII until the PAS

Metagene plots of RNAPII (black), Spt4 (pink) and Spt5 (light blue) reads.

TEF-seq is performed in duplicate (See Appendix 3 for individual figures). The background is removed by subtracting the no tag control. RNAPII reads are normalised to *S.pombe* spike-ins.

4.2.4 Spt4 and Spt5 have a phased distribution on RNAPII

The TEF-seq signals presented above come from the native RNA attached to RNAPII that is in association with the elongation factors. Although overall Spt4 and Spt5 signals are similar to the Rpb3 signal, there are differences, particularly in the 0-150 nt window downstream from the TSS or around the PAS (**Figure 4.5**). To be able to demonstrate relative affinities of Spt4 and Spt5 on Rpb3, we plotted the TEF-seq signal over the NET-seq signal. As was mentioned before, NET-seq and TEF-seq experiments were spike-in normalised. This means that if every RNAPII was associated with Spt4 (or Spt5), Spt4 to RNAPII ratio would be 1. Additionally, it is worth mentioning that here, the factor profiles are independent of the speed of transcription. This is because the factors travelling with the elongating RNAPII would have the same speed as RNAPII and the normalisation would eliminate variations in the profile originated from the changes in the transcriptional speed.

Interestingly, NET-seq normalised TEF-seq plots of Spt4 and Spt5 show phased distribution of the factors on RNAPII across the gene bodies suggesting periodic changes in the affinities of the factors with RNAPII (**Figure 4.6A**). Moreover, phasing of the factors seems to oscillate with

~160 nt frequency, comparable to the nucleosome phasing in yeast previously described (Ocampo et al. 2019). This observation raised the question as to whether the Spt4 and Spt5 occupancies on RNAPII are related to nucleosome positions. This hypothesis will be tested in chapter 9 (**Figure 9.2** and **Figure 9.3**).

Once plotted as a ratio of factor levels on RNAPII level, the data suggest a dynamic interaction between Spt4, Spt5, and RNAPII, especially towards the 3'-end of genes (**Figure 4.6A**). Most notable is the loss of the factors as RNAPII approaches the PAS and re-establishment of Spt4 and Spt5 with RNAPII after the PAS. One possible explanation for the rapid re-association between the factors and RNAPII after the PAS would be initiation of transcription from a downstream gene. The yeast genome is densely organised with many genes being tandemly transcribed and some genes having overlapped transcription units. To test if the high Spt4 and Spt5 signals after the PAS were due to transcription of the spatially close downstream genes, genes were split into 2 groups: (i) as tandem genes (< 250 nt distance between the PAS and the TSS of the adjacent gene, n=1482) and (ii) non-tandem genes (> 250 nt distance between the PAS and the TSS of the adjacent gene, n=2730). Comparison between the tandem and non-tandem genes show that the association between the factors and RNAPII after the PAS are observed in both groups and cannot be explained by the transcription of downstream genes (**Figure 4.6B&C**).

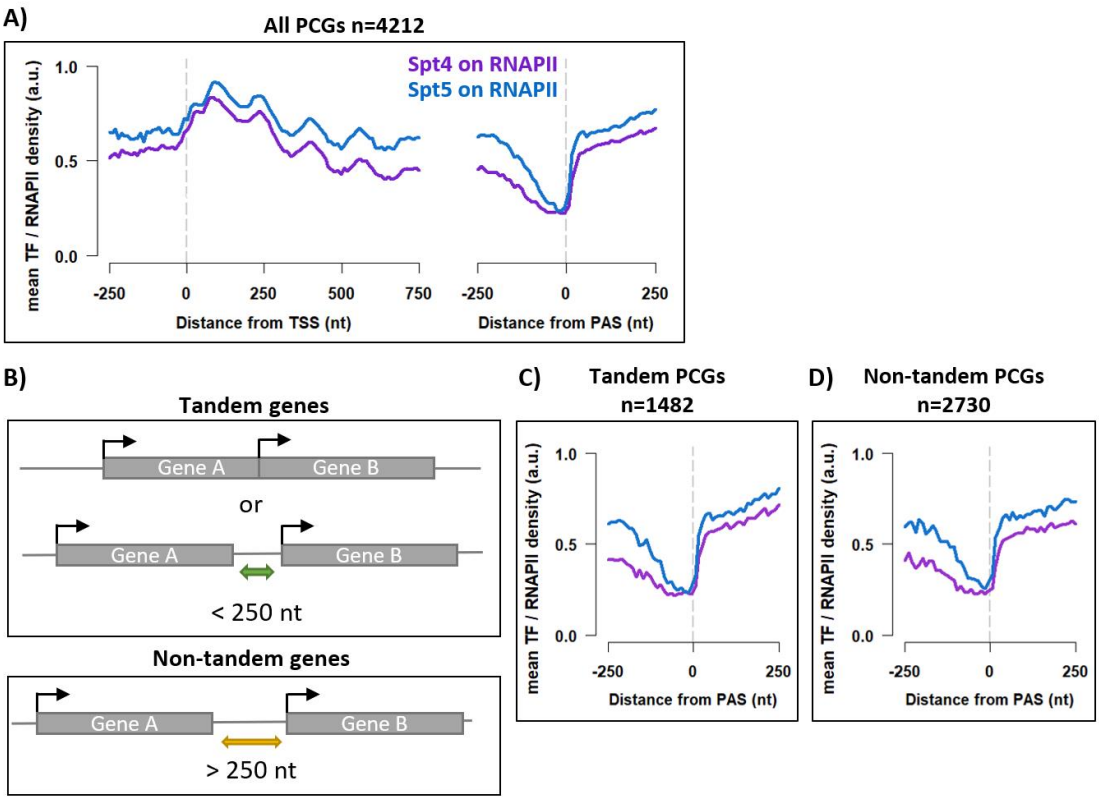


Figure 4.6) Spt4 and Spt5 oscillate on RNAPII over the bodies and re-engage with RNAPII after the PAS

A) Metagene profiles of Spt4 on RNAPII (purple) and Spt5 on RNAPII (dark blue) data. Spt4 on RNAPII was plotted by dividing Spt4-engaged RNAPII signal by RNAPII signal shown in **Figure 4.5**. The same is applied to Spt5 data. **B)** Schematic representation of tandem and nom-tandem genes. Tandem genes are defined as PCGs having <250 nt distance between two transcription units, including overlapping genes. Non-tandem genes are defined as PCGs having >250 nt distance between two transcription units. Metagene plots of Spt4 and Spt5 on RNAPII around the PAS plotted separately for tandem **(C)** and non-tandem genes **(D)**.

On the other hand, the high signals of Spt4 and Spt5 after the PAS could be due to direct factor-RNA interactions. There is no evidence showing a direct Spt4-RNA interaction. However, Spt5 is known to interact with RNA, especially downstream of the PAS (Baejen et al. 2017) and the Spt4 signal also might be high due to interactions through Spt5. As further explained in the discussion, the possibility of TEF-seq detecting some factor-RNA interaction cannot be ruled out and therefore, caution must be taken when interpreting the data. Overall, these data show that either by engaging with RNAPII or RNA, Spt4 and Spt5 re-localise at downstream of the PAS.

4.2.5 Spt4 and Spt5 occupancies on the genes are proportional to RNAPII occupancy

After defining the genome-wide position of Spt4 and Spt5, the levels of Spt4 and Spt5 at specific PCGs were examined to see if they were differentially enriched or depleted. To this end, reads for gene bodies (taken as TSS to PAS-250 nt) were counted from the two replicates of Spt4 and Spt5 TEF-seq and compared to the reads from the two replicates of RNAPII NET-seq. To compare the genome-wide correlations PCGs were ranked from high to low based on the average RNAPII counts and the read counts of Rpb3, Spt4, and Spt5 were plotted as heatmaps. Heatmap profiles show that the Spt4 and Spt5 levels match with RNAPII levels (**Figure 4.7A**).

A quantitative analysis of Spt4 and Spt5 occupancy on RNAPII was also performed by comparing Spt4 and Spt5 TEF-seq counts to NET-seq counts and applying differential expression algorithm in the DEseq2/R package (Love et al. 2014). 5579 PCGs were taken and genes with negative values (due to no tag normalisation, see material and methods) were discarded. To avoid genes with incorrect TSS annotation, genes having 1.5x more reads upstream of the TSS (-150 to 0 nt) than downstream of the TSS (+1 to 150 nt) were also discarded. Analysis of 4820 PGCs showed that the Spt4 levels are proportional to the RNAPII levels at most of the genes (> 99 %, adjusted p-value < 0.05, **Figure 4.7B**). Notably, the results did not change when non-coding genes were included (6037 genes in total, Appendix 4). The same holds true for Spt5 (**Figure 4.7C**). Therefore, the data suggest that both Spt4 and Spt5 associate with elongating RNAPII at most of the genes, and do not have specific enrichment for different gene groups.

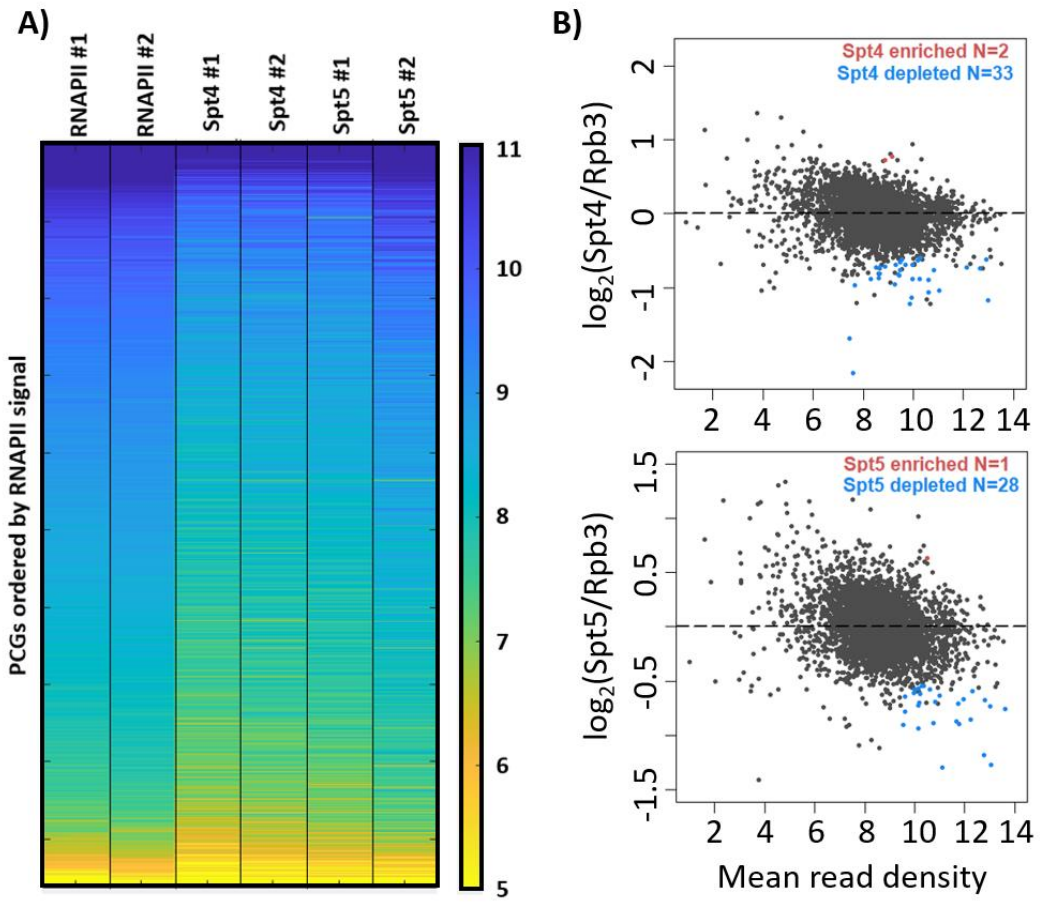


Figure 4.7) Spt4 and Spt5 occupancies on RNAPII

A) Heatmaps of two biological repeats of RNAPII NET-seq and Spt4 and Spt5 TEF-seq reads over the gene bodies (taken as TSS to PAS-250 nt) on log_2 scale. PCGs are ranked by RNAPII levels.

B) Differential enrichment of Spt4 (top) and Spt5 (bottom) on RNAPII. DEseq2 applied to the read counts from the gene body (TSS to TSS-250 nt) for two replicates of each data. Significantly enriched and depleted genes indicated in red and blue, respectively (p-adjusted <0.05).

4.2.6 The genome-wide localisation of Spt5 in the absence of Spt4

Previous studies suggest that Spt4 stabilises Spt5, and one study proposes that the Spt5 protein level is reduced by around 70% upon deletion of *SPT4* compared to controls (Ding, LeJeune, and Li 2010). To test if it was the case for the strains studied here, western blots on the FLAG-tagged Spt5 in WT and *spt4Δ* whole-cell extracts were performed in duplicate. Tubulin was used as a loading control and detected by anti-tubulin, and Spt5 was detected by anti-FLAG antibody (**Figure 4.8A**). Although western blotting does not provide an absolute quantification, to get an idea about the relative amounts of the proteins the Spt5 and tubulin

bands were quantified using ImageJ and calculated the ratios (Spt5/tubulin). The western blot analysis suggest that there was no significant change in the protein levels of Spt5 in *spt4Δ* cells (**Figure 4.8B**). Note that, in addition to protein levels, there was no change in the transcript levels of *SPT5* in *spt4Δ* cells as assessed by RNA-seq, which will be discussed in chapter 5. Overall, these results do not agree with the previous studies showing that Spt4 plays a role in the stabilisation of Spt5.

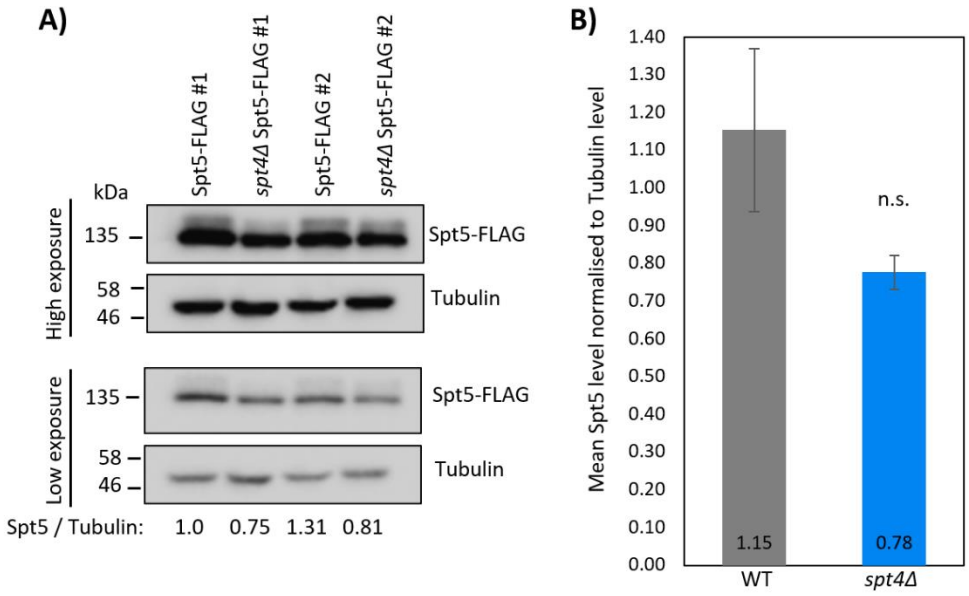


Figure 4.8) Spt5 level is reduced in the absence of Spt4

A) Western blot of WT and *spt4Δ* FLAG-tagged Spt5 with anti-FLAG (M2) and anti-Tubulin antibody run in 7.5 % SDS-PAGE.

B) Quantification of the western blot. The change in Spt5 levels are not significant a p-value = 0.131 (Student's t-test, unpaired, two-tailed).

In line with the results shown above, in *spt4Δ* cells, Spt5 should still be functional in transcription, as *spt5* knock out cells are inviable. Therefore, the localisation of Spt5 in *spt4Δ* cells was examined using TEF-seq. The IGV tracks show that the Spt5 profiles in WT and *spt4Δ* are similar to the Rpb3 profiles in their respective conditions, implying that Spt5 engages with RNAPII in the absence of Spt4 (**Figure 4.9**). Notably, here only the positions of Spt5 was examined. The reason for the *spt4Δ* Rpb3 signal being different from that of WT will be analysed and discussed in detail in chapter 5.

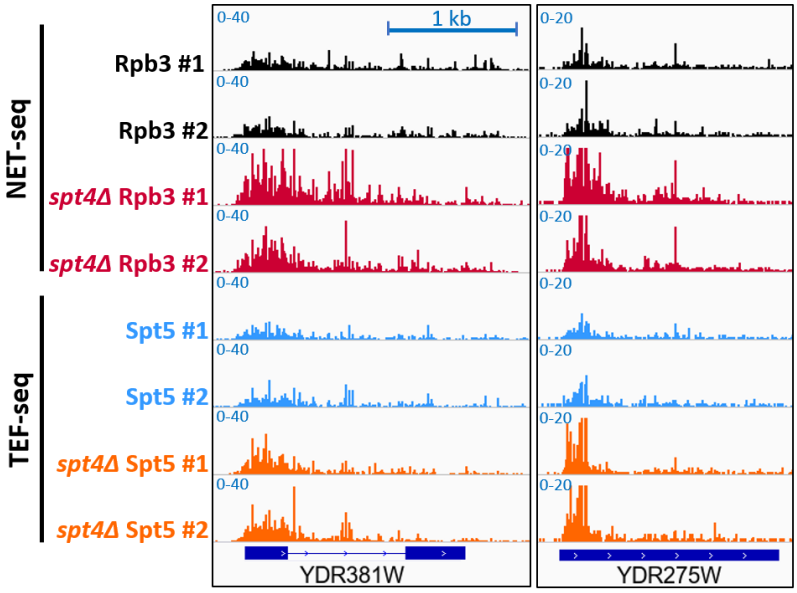


Figure 4.9) Rpb3, *spt4Δ Rpb3*, Spt5 and *spt4Δ* Spt5 occupancies at representative genes

IGV tracks of NET-seq (Rpb3 and *spt4Δ* Rpb3) and TEF-seq (Spt5 and *spt4Δ* Spt5) reads of two representative genes transcribed from the positive strand: *YDR381W* and *YDR275W* in two biological replicates. The dark blue boxes indicate the transcribed region of the genes (from TSS to PAS), while the blue line indicates the intronic region in *YDR381W.*

Metagene analysis of protein-coding genes (PCGs) was performed as described above. The metagene plots also show that the *spt4Δ* Spt5 profile is similar to that of *spt4Δ* Rpb3 around the TSS and over the gene bodies (**Figure 4.10**). After remaining associated with RNAPII over the gene bodies, the *spt4Δ* Spt5 signal drops to the background level at the 3'-end of the genes before the PAS. Overall, the data suggest that Spt5 joins RNAPII right after initiation, travels with elongating RNAPII, and dissociates from RNAPII upstream of the PAS even in the absence of Spt4.

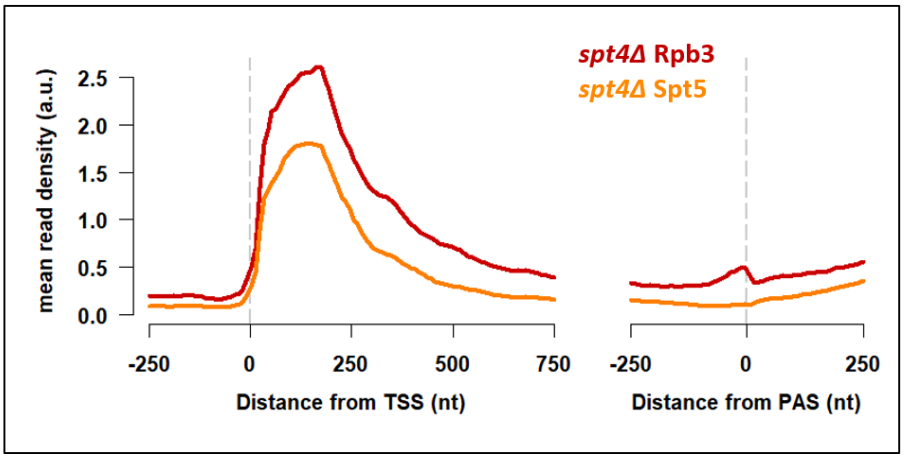


Figure 4.10) Spt5 travels with RNAPII in the absence of Spt4

Metagene plots of *spt4Δ* RNAPII (red) and *spt4Δ* Spt5 (orange) reads.

TEF-seq is performed in duplicate (See Appendix 3D&E for individual figures). The background is removed by subtracting the no tag control. RNAPII reads are normalised to *S.pombe* spike-ins.

4.2.7 In the absence of Spt4, Spt5 does not oscillate on RNAPII

Next, the relative affinity of Spt5 for RNAPII in the absence of Spt4 was addressed. To do this, the *spt4Δ* Spt5 signal over the *spt4Δ* Rpb3 signal was plotted together with that of the ratio in WT conditions for comparison. NET-seq normalised TEF-seq plot shows that in *spt4Δ,* Spt5 levels on RNAPII are reduced compared to that of in WT cells (**Figure 4.11A**). For a quantitative comparison, TEF-seq over NET-seq ratios were calculated for PCGs by counting the reads from the gene bodies (taken as TSS to PAS-250 nt) from the two replicates of Spt5 TEF-seq and dividing by the reads from the two replicates of RNAPII NET-seq both in WT and *spt4Δ* cells. The ratios were presented as boxplots (**Figure 4.11B**). The average Spt5 on RNAPII ratio drops from about 0.8 in WT to 0.6 in *spt4Δ*.

Interestingly, in *spt4Δ*, the phasing pattern of Spt5 on RNAPII observed in WT cells is substantially reduced, if not entirely lost (**Figure 4.11A**). This change in the Spt5 pattern on RNAPII indicates that even though Spt5 joins transcription, its interaction with RNAPII is altered in the absence of Spt4.

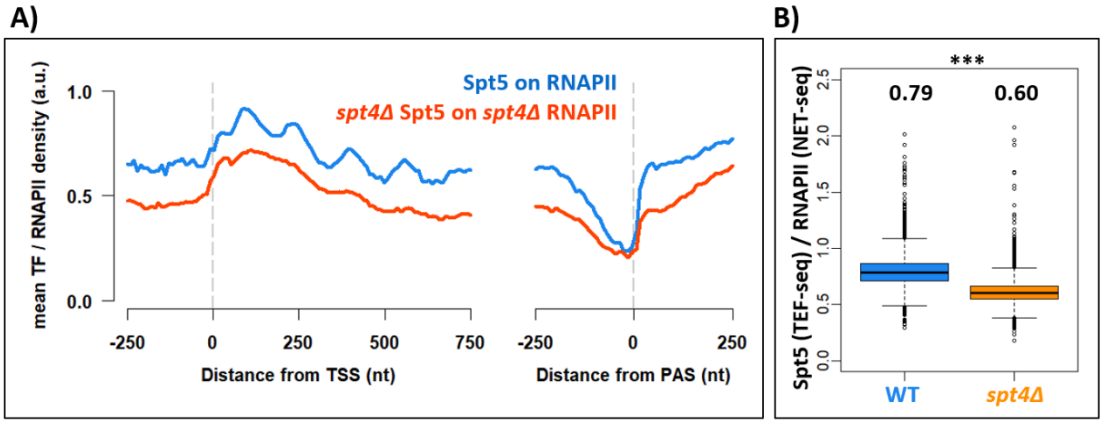


Figure 4.11) In *spt4Δ*, the association between the Spt5 and RNAPII is altered

A) Metagenes profiles of Spt5 on RNAPII (blue) and *spt4Δ* Spt5 over *spt4Δ* RNAPII (dark orange) data. Spt5 on RNAPII was plotted by dividing Spt5-engaged RNAPII signal by RNAPII signal as shown in **Figure 4.6**. The same is applied to *spt4Δ* Spt5 data.

B) Boxplots of the WT Spt5 over WT Rpb3 (blue) and *spt4Δ* Spt5 over *spt4Δ* Rpb3 (yellow) ratios for the protein-coding genes. The median of the ratios is indicated on the plot. The ratios were calculated by taking the reads from gene bodies (TSS to PAS-250 nt) of Spt5 TEF-seq and dividing by the reads from gene bodies of NET-seq both in WT and *spt4Δ*. (p-value < 0.001, Student's t-test, paired, two-tailed).

4.2.8 Spt5 occupancy on the genes is proportional to RNAPII occupancy in *spt4Δ*

The Spt5 levels on RNAPII seems to be globally reduced in *spt4Δ* compared to WT. This raises the question as to whether the lack of Spt4 caused differential association between Spt5 and RNAPII. To answer this question, a quantitative analysis was performed by comparing *spt4Δ* Spt5 TEF-seq counts with *spt4Δ* NET-seq counts and applying differential expression algorithm in the DEseq2/R package (Love et al. 2014) after filtering 5579 PCGs as in section 4.2.5. Analysis of 4820 PGCs showed that in *spt4Δ,* the Spt5 level is proportional to the RNAPII levels at most of the genes (> 99%, adjusted p-value < 0.05, **Figure 4.12**). In conclusion, the Spt5 association with elongating RNAPII was affected similarly at most genes, and there was no preferential enrichment of Spt5 for different gene groups in the absence of Spt4.

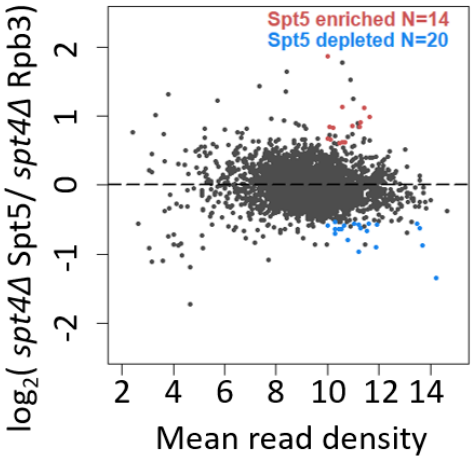


Figure 4.12) Spt5 is not differentially enriched on RNAPII in the absence of Spt4

No differential enrichment of Spt5 on RNAPII across all 5579 genes in the absence of Spt4. DEseq2 applied to the read counts from the gene body (TSS to TSS-250 nt) for two replicates of each data. Significantly enriched and depleted genes indicated in red and blue, respectively (p-adjusted <0.05).

4.3 Discussion

In this chapter, it was shown that Spt4 and its binding partner Spt5 join transcription as early as the RNAPII signal can be detected (20 nt from the TSS), travel together with RNAPII throughout the elongation step, and dissociate from RNAPII before reaching the 3'-ends of the genes. The distributions of Spt4 and Spt5 on RNAPII follow a phasic pattern and the levels of these factors are proportional with RNAPII levels. In the absence of Spt4, Spt5 still joins transcription, but stops oscillating on RNAPII suggesting that Spt4 is important for maintaining normal interaction between Spt5 and RNAPII.

ChIP-based experiments previously presented a broad view of the position of many TEFs suggesting that RNAPII dynamically interacts with different factors at different stages of transcription (Mayer et al. 2010). To get a more detailed insight into the involvement of Spt4 and Spt5 in transcription, RNAPII engaged Spt4 and Spt5 were mapped at single-nucleotide resolution.

One drawback of TEF-seq is that the IP of TEF could pull down RNA directly associated with the FLAG-tagged factor and it might be difficult to interpret data, as TEF-seq cannot

differentiate the direct factor-RNA interaction from RNAPII-engaged-factor interaction. This problem could be minimised by performing TEF-seq with sequential IP such that RNAPII could be IP-ed first and TEF could be IP-ed second. Sequential IP for TEF-seq was previously carried out in our lab for various TEFs, including Spt6, Spt16, Paf1 and Nrd1. The TEC was first immunoprecipitated with a FLAG-tagged Rpb3, and then re-precipitated with HA-tagged TEF to select only the TEF associating with RNAPII. Although most TEFs tested showed similar TEF-seq profiles carried out in single IP and sequential IPs (see Fischl *et al.*, 2017 for Paf1), we also observed an exception for Nrd1. Therefore, in TEF-seq protocol used here, the possibility of the IP of RNA that is directly in contact with Spt4 (or Spt5) cannot be ruled out. Performing sequential IP TEF-seq would improve the results.

With TEF-seq, the positions of Spt4 and Spt5 at single nucleotide resolution could be monitored. However, the questions of how and when Spt4 and Spt5 is recruited to RNAPII complex remain unanswered. Spt4 and Spt5 form a very stable heterodimeric complex through the NGN domain of Spt5 (Hartzog and Fu 2013). This is possibly why Spt4 and Spt5 are commonly studied as a complex and described to be recruited to RNAPII together. Indeed, experiments performed in human nuclear extracts concluded that there is no free SPT4 or SPT5 in the cell (Kim et al. 2003). A recent experiment in human cells challenged this idea by proposing that MYC also interacts with the NGN domain of SPT5 and recruits it to RNAPII (Baluapuri et al. 2019). The role of SPT4 in this process was not studied. Nevertheless, the possibility of Spt4 and Spt5 acting independently leads directly to the questions as to whether these factors are recruited to RNAPII together or sequentially.

An *in vitro* study suggests that the Spt4/5 complex binds to RNAPII once it has RNA > 20 nt (Missra and Gilmour 2010). Consistently, high signals of Spt4, Spt5 as well as RNAPII were detected 20 nt downstream of the TSS. However, NET and TEF-seq do not inform the position of Rbp3 and the factors in a short window at the TSS (TSS to TSS+20 nt) as the libraries are

prepared with the native RNAs > 20 nt. Moreover, genomic alignment of the short RNA fragments is limited to 15 nt even if library preparation allowed the processing of shorter RNAs. Consequently, it is difficult to conclude whether one of the factors joins transcription earlier than the other, and independent recruitment of Spt4 and Spt5 to RNAPII is still open to further studies.

Another question is whether Spt4 and Spt5 take a part in transcription termination. It is clear that these factors associate with RNAPII at early stages of transcription and maintain their interaction with RNAPII through the elongation. Their contribution to termination, on the other hand, is not completely understood. The results presented here indicate loss of both factors as RNAPII approaches the PAS. Around the same region, RNAPII appears to have relatively high signal, probably due to slowing down and/or pausing of RNAPII (Mayer, Landry, and Churchman 2017; Mischo and Proudfoot 2013). The RNAPII accumulation before the PAS is thought to regulate the elongation-to-termination transition and the loss of Spt4/5 around this region might be contributing to the RNAPII accumulation as well as allowing the recruitment of termination factors (Kecman et al. 2018).

After RNAPII transcribes the PAS, transcripts are cleaved and processed but RNAPII continues transcribing downstream of the PAS. One explanation for the increased engagement of Spt4 and Spt5 after the PAS would be that if the cleaved RNA downstream of the PAS keep associating with some of the termination factors, this could create an intermediate transcription complex remaining on the template DNA that could be re-bound by some of the elongation factors. This would also be consisted with the above mentioned observation that the Spt4/5 complex can bind to RNAPII having RNA > 20 nt *in vitro* (Missra and Gilmour 2010). Alternatively, or additionally, Spt4 and Spt5 might play a role in transcription termination. In fact, a recent study in *S.pombe* suggests that the dephosphorylation of Spt5 downstream of the PAS is important for the slowing down of RNAPII (Parua et al. 2018). Furthermore, the

torpedo model of termination states that RNAPII slows down after transcribing through the PAS. This model proposes that the reduced speed of RNAPII allows Rat1 (XRN2 in humans) to degrade downstream RNA, catch up with the transcribing RNAPII and dissociate RNAPII from the DNA template. Therefore, the re-association between Spt5 and RNAPII downstream of the PAS shown here might be required for slowing RNAPII down after the PAS. The role of Spt5 phosphorylation in this process could further be addressed by TEF-seq using antibodies against phosphorylated Spt5.

On the gene bodies, the relative affinity of the Spt4/Spt5 complex to RNAPII changes and oscillates with similar frequency of nucleosome phasing. It is tempting to speculate that the oscillation pattern of the factors is related to nucleosomes. This hypothesis will be tested and further discussed in chapter 9.

As Spt4 and Spt5 forms a complex , in the *spt4* knock out cells, which will be used in the next chapters, the stability and activity of Spt5 can be influenced by the deletion of *SPT4*. To address this, the occupancy of Spt5 on the genes was examined also in the absence of Spt4. Previously Spt4 was shown to stabilise Spt5 and deletion of *SPT4* caused a big drop in the Spt5 protein levels (Ding et al. 2010). Here the data suggest that the protein levels of Spt5 are not significantly lower in the absence of Spt4, and a large amount of Spt5 still exists in the cell. This discrepancy between the two studies could be explained by the difference in the tagging positions. The Spt5 used in the former study was N-terminally FLAG-tagged. Spt5 interacts with Spt4 through its NGN domain, which is near the N-terminus of the Spt5 (**Figure 1.3**) (Ehara et al. 2019) and in archaea Spt4 was shown to be required for the stability of NGN domain of Spt5 at high temperatures (Hirtreiter et al. 2010). Therefore, it is possible that the N-terminal FLAG-tag on Spt5 negatively impacts the stability of the NGN-domain once combined with the lack of Spt4. In contrast, in this study, a C-terminally FLAG-tagged Spt5 is used, and no impact on the stability of Spt5 is observed. Consistently, TEF-seq data presented

here supports Spt5 (with a FLAG-tag at the C-terminus) being able to be recruited to, and travelling with, RNAPII in *spt4Δ* cells. Moreover, in chapter 7 it will be demonstrated, using proteomics analysis that the levels of Spt5 (which does not have any tag) on the WT and *spt4Δ* transcription complex appears to be similar.

Interestingly, in the absence of Spt4, Spt5 does not oscillate on RNAPII and loses its ability to form strong interactions with RNAPII as the peaks in TEF-seq normalised NET-seq Spt5 profiles are not observed. This might be due to structural changes induced in the RNAPII complex in the absence of Spt4. The NGN domain of Spt5 and Spt4 together interact with the RNAPII near the DNA exit channel of the transcription complex (**Figure 1.3B**) (Ehara et al. 2019). Therefore, the absence of Spt4 might cause a reduced physical support at this location and cause a reduced interaction between Spt5 and RNAPII. Another explanation would be that if Spt4 interacts with other TEFs regulating the oscillating pattern of the Spt4/5 complex on RNAPII, the lack of Spt4 might also lead to the loss of interaction between Spt5 and RNAPII, hence the loss of the oscillating pattern of Spt5 on RNAPII. In either way, the data shows that Spt4 does not entirely inhibit the recruitment and binding of Spt5 to RNAPII, but it impacts the interaction between Spt5 and RNAPII.

5 RNAPII distribution in the absence of Spt4

5.1 Overview

In the previous chapter, it was shown that Spt4 engages with RNAPII at most genes. The association of Spt4 with RNAPII starts from early stages of transcription, continues up to a few hundred nt upstream of the PAS and seems to be re-established after the PAS. Therefore, Spt4 is likely to be involved in the transcription of all genes and take part in many stages of transcription.

In this chapter, the question of how Spt4 impacts the RNAPII distribution is addressed, using two different strain backgrounds: (i) complete deletion of Spt4 using *spt4* knock out (*spt4Δ*) cells to investigate absolute changes in RNAPII distribution and (ii) depletion of Spt4 from the nucleus using the anchor away (AA) system to observe immediate changes, as was described in chapter 3. NET-seq is used to map the position of RNAPII and a stochastic model of transcription on the NET-seq data used to better understand the changes in the distribution of RNAPII in the *spt4Δ* strain and compare with the WT conditions. Thus, the aims of this chapter are (1) to study the RNAPII distribution in *spt4Δ*, (2) to compare the changes in RNAPII distribution in *spt4Δ* using a stochastic transcription model and (3) to study RNAPII distribution in Spt4 depleted cells.

RNAPII distribution in the absence of Spt4 has been previously documented by precision run-on sequencing (PRO-seq) (Booth et al. 2016). As its name suggests, PRO-seq is a run-on based technique detecting elongation competent RNAPII. Briefly in this technique, yeast cells are chilled, permeabilised, internal NTPs are washed off and biotinylated-NTPs added. Then, transcription is run for 5 min (or longer) and biotinylated native RNA is pulled down. RNA is then sequenced and used as a reference to map RNAPII position. Although providing a great view of the engaged RNAPII, this technique does not map RNAPII that is still on the genes, but

not actively elongating during the 5 min run-on window. Consequently, stalling/backtracking events and long-term pauses of RNAPII are not reported by PRO-seq. Given that the possible function of Spt4 in transcription regulation is investigated in this thesis, all forms of RNAPII was desired to be detected as pausing/stalling/backtracking are known to be common features of transcription and thus, NET-seq was used.

A stochastic model of transcription was previously built as part of the DPhil work of Dr. Thomas Brown supervised by Drs. Andrew Angel and Jane Mellor (Brown 2019). Briefly, the model computationally simulated RNAPII dynamics by taking into account several variables including initiation rate, elongation rate, occlusion of RNAPII by a downstream RNAPII, stalling, backtracking, early termination and two distinct windows of transcription in which stalling and backtracking parameters can be different. The single nucleotide position of WT RNAPII provided by experimental NET-seq data in various strain backgrounds was then fitted to the shape of the transcription profile simulated by the model. This allowed key parameters and mechanisms that give rise to a particular experimental RNAPII distribution to be defined. Using the parameters that best explain the RNAPII distribution, WT and *spt4Δ* NET-seq data were compared. In this way, we obtain quantitative information on each parameter and test the impact of the deletion of *SPT4* on the RNAPII distribution.

5.2 Results

5.2.1 Steady-state RNA levels do not change in *spt4Δ* compared to WT

Previous studies showed that there was no change in the steady-state RNA levels in *spt4* deletion mutants compared to WT cells (Booth et al. 2016). First, whether this held true for the strains used in this study was tested, before examining the distribution of RNAPII. To this end, RNA-seq was carried out with the samples taken from the same cultures as the NET-seq experiments.

The reproducibility of the two experimental repeats was tested by calculating Spearman's correlation coefficient. Correlation plots show that WT and *spt4Δ* RNA-seq were reproducible (Spearman's ρ=0.993 and 0.985, respectively, **Figure 5.1A**). A quantitative analysis of steady-state transcript levels was then performed for PCGs (genes with > 0 reads; n=5538) by comparing WT and *spt4Δ* RNA-seq counts and applying the differential expression algorithm in the DEseq2/R package (Love et al. 2014). Thus, the steady-state levels of transcripts (> 99%) remain unchanged in *spt4Δ*. With RNA-seq we also verified that there is no defect in the transcription of any other transcription related gene other than *SPT4* (**Figure 5.1B**).

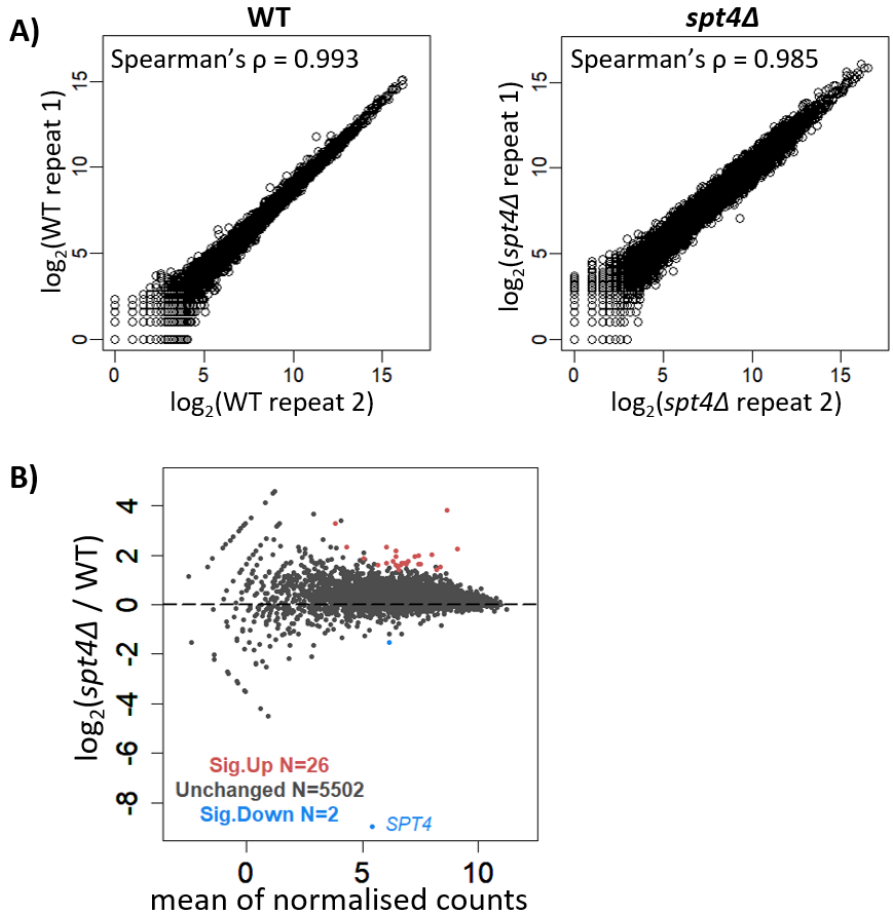


Figure 5.1) Steady-state transcript levels remain unchanged in WT and *spt4Δ*.

A) Correlations of the two biological repeats of RNA-seq (transcript) reads in WT (left) and *spt4Δ* (right).

B) DEseq2 analysis of RNA-seq reads in WT and *spt4Δ*. Significantly enriched and depleted genes indicated in red and blue, respectively (p-adjusted <0.05).

5.2.2 Immunoprecipitation of Rpb3 in *spt4Δ*

Next, changes in the distribution of RNAPII using the NET-seq data is explored in detail and compared between WT and *spt4Δ* conditions. As FLAG-tagged Rpb3 is used in NET-seq to detect the position of RNAPII, "Rpb3" and "RNAPII" will be used interchangeably from this point to describe the NET-seq results.

The efficiency of the NET-seq IPs was checked by western blotting using antibodies against the FLAG epitope as before. Here, we also used an antibody against Ser5-P of RNAPII-CTD to validate the successful IP of active RNAPII by western blotting. Strong Ser5-P and Rpb3 bands in the FLAG-tagged eluate samples are observed at expected sizes, verifying the effective pull-down of RNAPII (**Figure 5.2**). Ser5-P signals of the no tag eluate samples are very low compared to the FLAG-tagged eluate samples, indicating non-specific RNAPII pulldown was minimal. There is no FLAG signal in the no tag control samples as expected (**Figure 5.2**). The data analysis was performed, and the reproducibility of the replicates was tested as in sections 4.2.1 and 4.2.2, respectively.

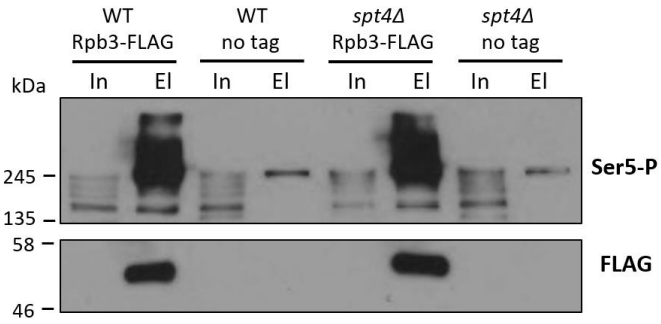


Figure 5.2) Immunoprecipitation of FLAG tagged cells

Western blot of Input (In) and eluate (El) samples with anti-Ser5-P (3E810) and anti-FLAG (M2) antibody run in 10 % SDS-PAGE. See Appendix 2A for the whole blots.

5.2.3 RNAPII distribution is altered in *spt4Δ*

At the single-gene level, IGV tracks show that *spt4Δ* RNAPII has a higher signal at the beginning of the gene compared to WT, indicating an accumulation of RNAPII early in transcription (**Figure 5.3**, also see **Figure 4.9**). Genome-wide analyses were then carried out by combining the two replicates of each experiment, as they were highly reproducible. Both heatmaps and metagene profiles were plotted to describe and compare RNAPII distribution in WT and *spt4Δ* as they provide different information. Heatmaps can visualise reads for many genes at the same time without averaging out the RNAPII signal, but it can be difficult to read the overall pattern from the heatmaps. Metagene plots can better summarise the general pattern, but it can be misleading as the averaged profiles might not be the actual pattern in most genes and can be biased by highly expressed genes. Therefore, here, both heatmap and metagene plots were presented to show the RNAPII distribution in WT and *spt4Δ* cells.

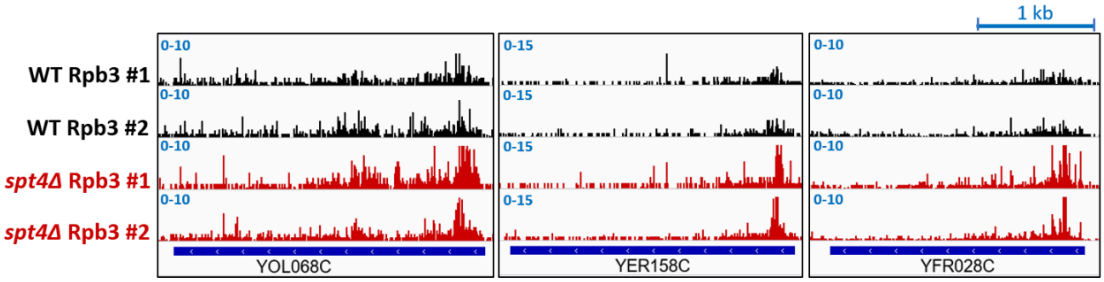


Figure 5.3) Rpb3 and *spt4Δ* Rpb3 occupancies at representative genes

IGV tracks of WT and *spt4Δ* NET-seq reads of three representative genes transcribed from the negative strand: *YOL068C, YER158C* and *YFR028C* in two biological replicates. The dark blue boxes indicate the transcribed region of the genes (from TSS to PAS).

5402 PCGs (>500 nt) were taken and genes with negative values (due to no tag normalisation) were discarded. To avoid genes with the wrong TSS annotation, genes having 1.5x more reads upstream of TSS (-150 to 0 nt) compared to downstream of the TSS (+1 to 150 nt) were also discarded. The remaining 4610 PCGs were ranked based on gene length and the Rpb3 signal was plotted as heatmaps from 250 nt downstream of the TSS to 4750 nt (TSS-250 to TSS+4750 nt) (**Figure 5.4**). Heatmaps show that for most of the genes, the RNAPII signal is higher around

the TSS in *spt4Δ* compared to WT. This can be more clearly seen in the heatmap plotted by subtracting the *spt4Δ* Rpb3 signal from the WT signal (**Figure 5.4B**). Moreover, heatmaps indicate that RNAPII accumulation is confined around the same region in *spt4Δ* regardless of the gene length.

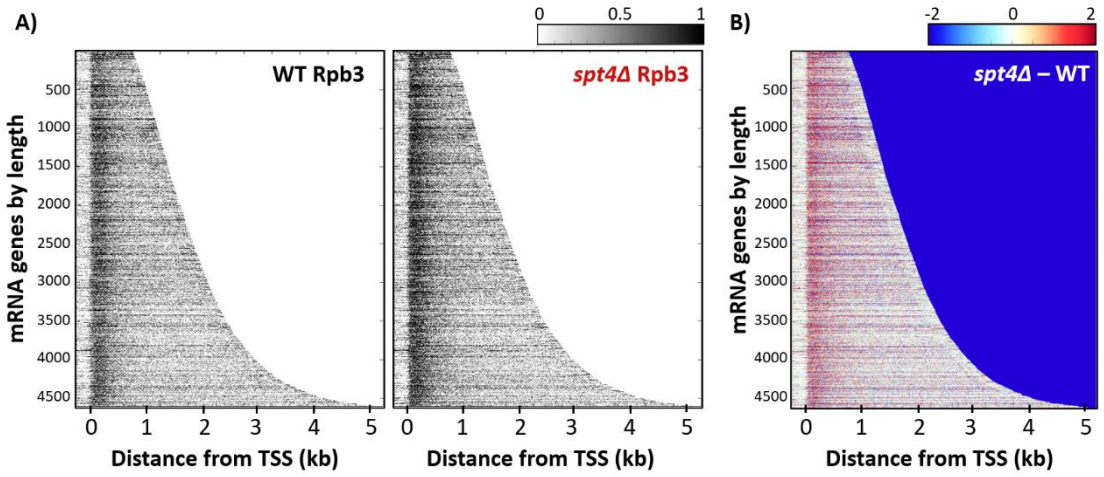


Figure 5.4) WT and *spt4Δ* RNAPII distributions are different in most of the PCGs

A) Heatmaps of NET-seq in WT (left) and *spt4Δ* (right). Each row indicates a PCG (n=4610), ranked by gene length. Different intensities of grey reflect RNAPII signal for each nucleotide position from TSS-250 nt to TSS+4750 nt (x-axis) as shown by the colour bar.

B) Heatmaps of the difference in *spt4Δ* and WT NET-seq signal (*spt4Δ* - WT). x and y axes are the same as in **A**. The colour code from red to blue reflects the changes in the RNAPII signal as shown in the colour bar.

Next, metagene profiles were plotted relative to the TSS for the PCGs that were filtered using the abovementioned criteria and were greater than 750 nt long (n=4212). Under normal conditions in WT cells, the Rpb3 signal is relatively high within the first 500 nt from the TSS, then it drops while transcribing over the gene body, and peaks again upstream of the PAS, as described previously (Churchman and Weissman 2011; Fischl et al. 2017) (**Figure 5.5A**). The nonuniform RNAPII distribution in WT cells reflects frequent pauses and changes in the speed of RNAPII due to transition from initiation to elongation and from elongation to termination as well as intrinsic transcriptional barriers such as nucleosomes, and co-transcriptional events such as splicing (Churchman and Weissman 2011; Jonkers and Lis 2015). Moreover, a

relatively high signal of RNAPII in the early stages of transcription might be an indication of some RNAPII failing to continue transcription in the later phases, perhaps due to early termination or pausing (Milligan et al. 2016).

In *spt4Δ* cells compared to WT, the metagene profiles give higher resolution and indicates that the most apparent change in the distribution and density of RNAPII is within the first 200 nt from the TSS, while the 3'-end of the genes is less affected (**Figure 5.5A**). For quantification, the NET-seq reads were counted over the gene bodies (taken from TSS to PAS-250 nt) in both conditions. The comparison between the NET-seq reads in WT and *spt4Δ* cells indicates that RNAPII occupancy was increased by 73% in *spt4Δ* cells ($p < 0.001$, Student's t-test, paired, two tailed, **Figure 5.5B**)

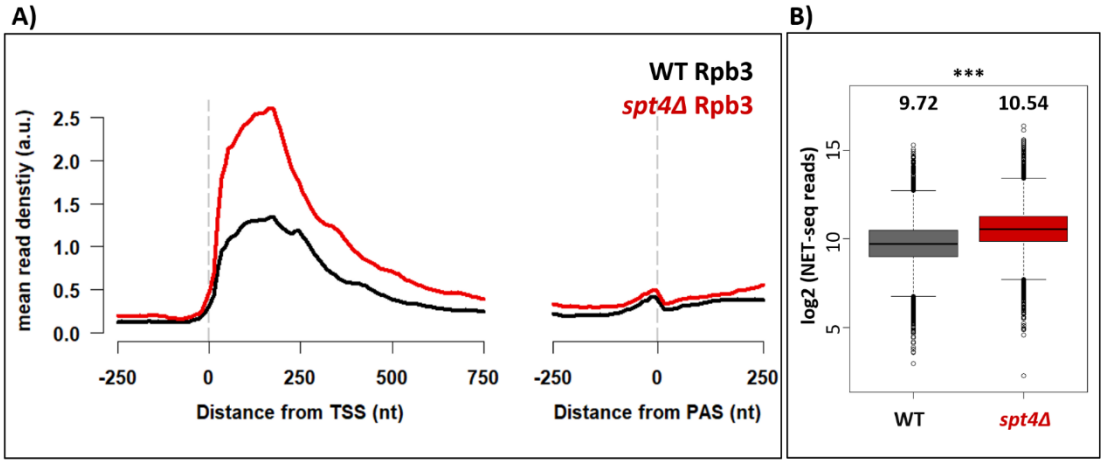


Figure 5.5) RNAPII density is higher at the 5'-end of the genes in *spt4Δ* compared to WT

A) Metagene plots of NET-seq reads in WT (black) and *spt4Δ* (red).

NET-seq is performed in duplicate (See Appendix 3A&D for individual plots). The background is removed by subtracting the no tag control. Reads are normalised to *S.pombe* spike-ins.

B) Boxplots of the NET-seq reads in WT (grey) and *spt4Δ* (red) cells on $\log_2$ scale. The median of the reads is indicated on the plot. The reads were counted for PCGs by taking the reads from gene bodies (TSS to PAS-250 nt) (p-value < 0.001, Student's t-test, paired, two-tailed).

What causes the high RNAPII signal in *spt4Δ* compared WT? NET-seq provides a snapshot of the position of RNAPII which can be influenced by changes in the level or speed of RNAPII, and can also reflect different forms of RNAPII. Because of the high complexity of transcription,

the data interpretation of NET-seq may not be straightforward, especially for the TEF mutants as they can impact more than one aspect of the transcription. To have a better insight into the transcriptional changes in *spt4Δ* cells, we employed mathematical modelling of the shape of the distribution of RNAPII and compared key parameters of transcription in WT and *spt4Δ*.

5.2.4 Mathematical modelling of transcription suggests that the absence of Spt4 changes the dynamics early in transcription

The NET-seq data used in the model was normalised to the mean and the standard deviation of each gene to account only for the shape of the RNAPII distribution, meaning that the model was not informed by the expression levels of the genes. It is worth mentioning that as part of the modelling study, using unsupervised k-means clustering, the RNAPII distribution across the genome was shown to have not one unified, but at least three different shapes originating from three different subpopulations of genes. These three subpopulations include; (1) a class of genes having a relatively more uniform (flat) distribution of RNAPII over the gene bodies, (2) a class with a 5'-end peak immediately downstream of the TSS which could be comparable to the promoter-proximal pause observed in metazoans, and (3) a class of genes with RNAPII accumulation within the first 400-500 nt from the TSS over the gene bodies (**Figure 5.6**). In *spt4Δ* cells, although the extent of the RNAPII accumulation seems to vary between the clusters, there is a similar shift in the shape of the distribution of RNAPII. In other words, the lack of Spt4 leads to an RNAPII accumulation within the first 200 nt from the TSS compared to that of WT in all groups (**Figure 5.6**) and thus, further analysis was performed on the combined metagene profiles.

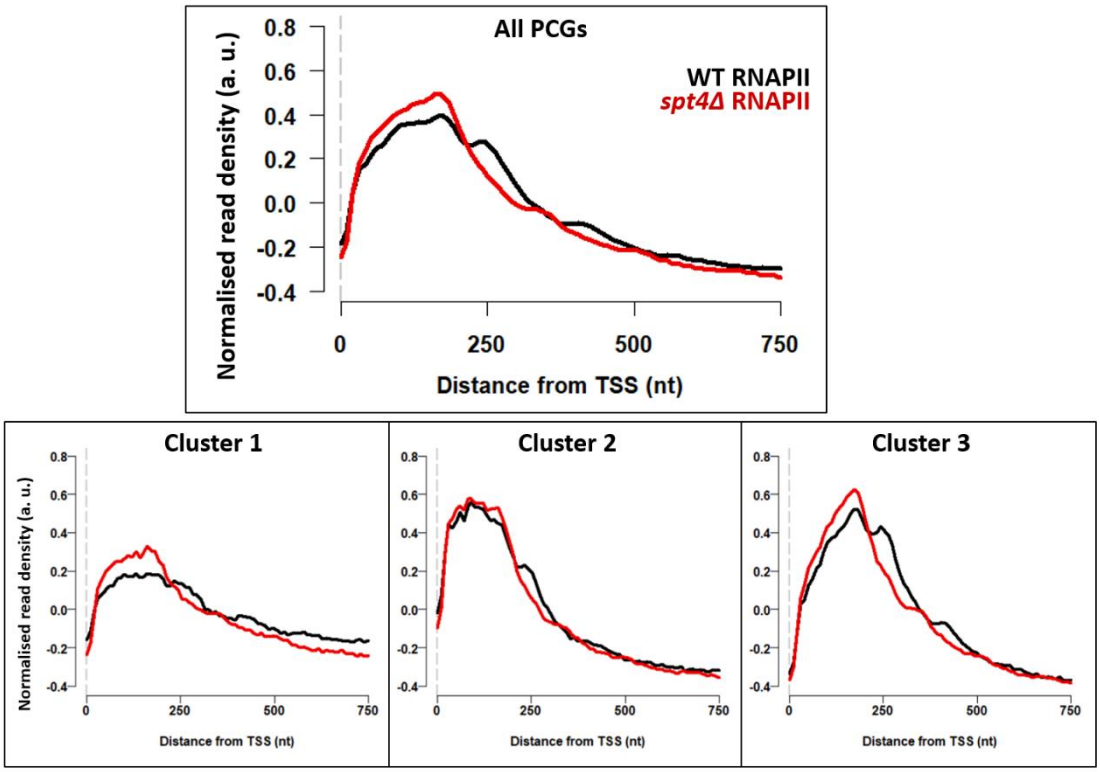


Figure 5.6) RNAPII distribution is influenced similarly by the lack of Spt4 in different gene clusters

NET-seq data were normalised to the mean and the standard deviation of each gene to account only for the shape of the RNAPII distribution. The top panel shows the RNAPII profile for all PCGs in the analysis (genes > 1000 nt; n=4137). The bottom panel shows three different clusters. The clusters of genes were defined by unsupervised k-means clustering analysis. n= 1132 in cluster 1, n= 1237 in cluster 2, and n=1768 in cluster 3.

The model assumes that there are sufficient RNAPIIs in the cells, the rate of RNAPII adding new nucleotide (elongation rate) is constant over the gene bodies, stalled/backtracked RNAPII can re-start transcription, and an early termination can take place with or without collision of an upstream RNAPII colliding with a downstream RNAPII. Under these assumptions, the best fitting model suggests that there are six key determinants of the shape of the WT RNAPII distribution; ratio of initiation to elongation rate, processivity, moving ratio in window 1, size of window 1, site of early termination, and moving ratio in window 2. Defining two windows of transcription was motivated by the notion that RNAPII is subject to distinct regulations in early and late stages of transcription (Peck et al. 2019). Therefore, RNAPII was allowed to have two different moving ratios (with different rates of backtracking and re-starting) in these two

distinct windows of transcription, and the sizes of these windows could change in each gene. To test the extent of the contribution of each parameter in the change of the distribution of RNAPII, the parameter values were obtained for each gene in WT and *spt4Δ,* and the two conditions were quantitatively compared. The significance of the changes was reported by calculating p-values and the magnitude of the changes was reported by calculating Cohen's d (Cohen, 1988). Cohen's d is computed by taking the mean difference between the WT and *spt4Δ* parameter value divided by the standard deviation of the differences:

$$d = \text{mean}\left(\frac{spt4\Delta - WT}{standard\ deviation\ (spt4\Delta - WT)}\right)$$

The value of Cohen's d is considered to be a measure of the effect size of the change such that the values between 0.2 to 0.5 indicate small changes, between 0.5 and 0.8 indicate medium changes, and > 0.8 large changes (Cohen, 1988). Moreover, positive and negative values indicate the direction of the change: a relative increase or decrease in the given parameter, respectively. The changes in the parameters between the WT and *spt4Δ* are given in **Table 5.1**.

Table 5.1) Differences of the difference in the distribution of RNAPII in *spt4Δ* compared WT.

	Magnitude of the difference (Cohen's d)	**Significance of the difference (p-value)**
Initiation / elongation rate	+ 0.82	<0.001
Processivity	- 0.54	<0.001
Site of early termination	- 0.64	<0.001
Moving ratio in window 1	- 1.46	<0.001
Size of window 1	- 0.20	<0.001
Moving ratio in window 2	+ 0.04	0.765

The stochastic model of transcription was developed and the comparison analysis was performed by Dr. Thomas Brown (Brown 2019). Large changes are highlighted in yellow, medium changes in green, small changes in cyan.

The model suggests that in *spt4Δ* cells, the ratio between initiation to elongation rate is higher. This can be explained either by a higher initiation or lower elongation rates in the

mutant. In the model, processivity was defined as a measure of successful completion of transcription, which is the proportion of initiating polymerase that successfully transcribe 1000 nt. The reduced processivity in *spt4Δ* cells, therefore, suggests a high proportion of RNAPII is falling off the DNA template before reaching the end of the genes. Consistently, the site of the early termination is found be more upstream in *spt4Δ* cells than that of WT.

The moving ratio is defined as the mean time of RNAPII spends in motion over static state, meaning that the moving ratio is negatively affected by stalling and backtracking events. As mentioned above, the model was allowed to have two different transcription windows with different moving ratios (with different rates of backtracking and re-starting) and sizes. Interestingly, in *spt4Δ* cells, the moving ratio in window 1 is found to be substantially lower suggesting more frequent and/or more slowly resolved stalling and backtracking events. The size of the window 1 tends to be smaller, which might indicate that transition from window 1 to window 2 happens more upstream in *spt4Δ* cells. Finally, after the transition to window 2, RNAPII movement appears to be similar in both WT and *spt4Δ* cells, as the moving ratios in window 2 are not different between WT and *spt4Δ*.

Overall, the model proposes that the changes in the shape of the distribution of RNAPII in *spt4Δ* cells are due to changes in the processivity and the dynamics of RNAPII in the early stages of transcription.

5.2.5 The PIC levels and position do not change in *spt4Δ* compared to WT

In *spt4Δ* cells, the RNAPII signal is higher at the beginning of the genes, and the model predicts an increased initiation to elongation ratio compared to WT cells. These observations raise the question as to whether transcription initiation rates were altered in *spt4Δ* cells. To address this question, the levels and positions of Sua7 (TFIIB) in WT and *spt4Δ* were examined using ChIP-seq.

Sua7 is a subunit of the PIC and it is recruited to promoters prior to RNAPII recruitment and dissociates from the transcription complex after transcription of 12-13 nt RNA (Sainsbury et

al. 2015). A recent *in vitro* study also shows that nearly all TFIIB dissociates from the RNAPII after synthesis of 7-9 nt RNA (Ly et al. 2020). Therefore, it was hypothesised that if transcription was more frequently initiated in *spt4Δ* compared to WT, the levels of Sua7 would be expected to change accordingly. Moreover, the positions of Sua7 could demonstrate if there was a notable change in the TSS in *spt4Δ*.

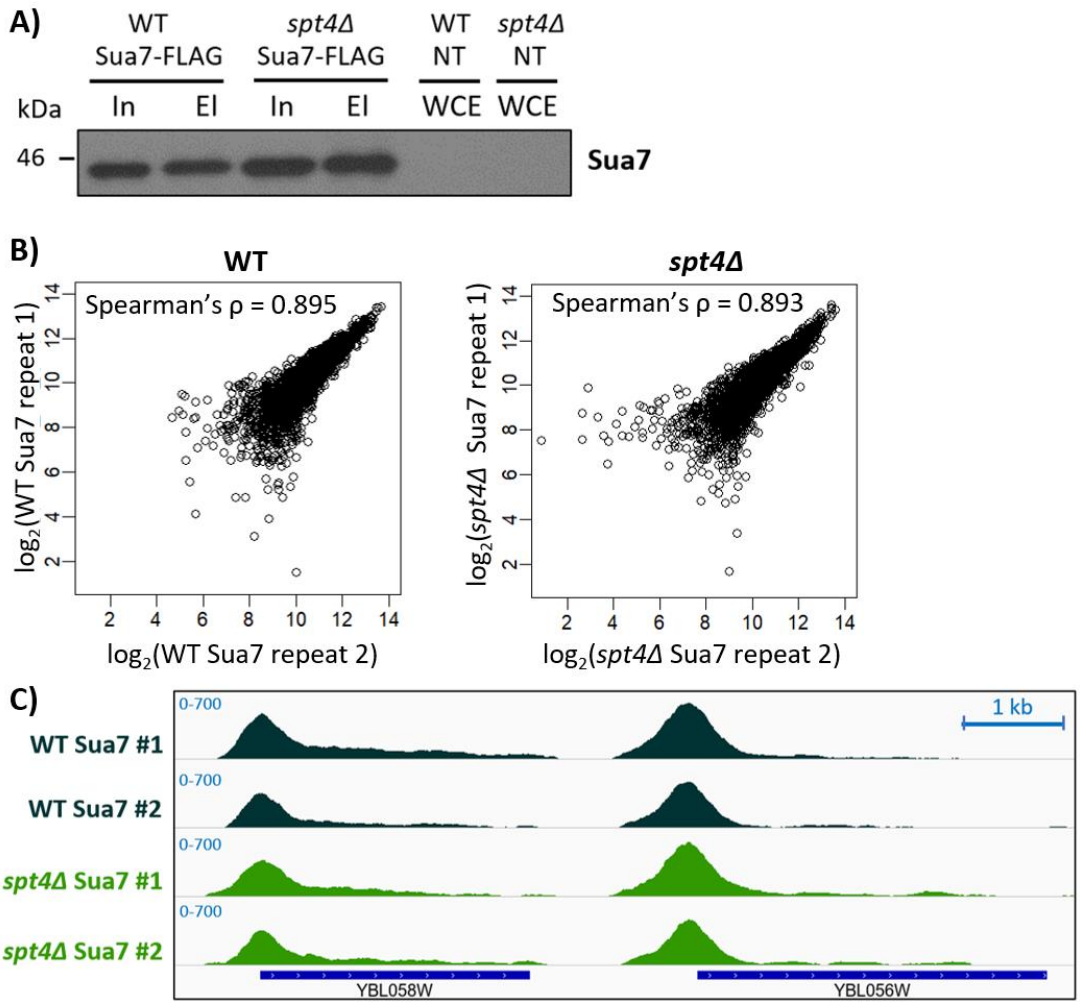


Figure 5.7) Sua7 and ChIP-seq in WT and *spt4Δ*

A) IP of FLAG-tagged Sua7. Western blot of Input (In) and eluate (El) samples with anti-FLAG (M2) antibody run in 10 % SDS-PAGE. See Appendix 5A for the whole blots.

B) Correlation plot of the two repeats of each experiment. Reads are counted around the TSS (TSS-100 to TSS+100) for each gene. log_2 transformed gene counts are correlated and Spearman's ρ calculated for each pair.

C) IGV tracks of WT and *spt4Δ* Sua7 ChIP-seq of two representative genes transcribed from the positive strand: *YBL058W* and *YBL056W* in two biological replicates. The dark blue boxes indicate the transcribed region of the genes (from TSS to PAS).

The efficiency of ChIP-seq IPs was checked by western blotting using antibodies against the FLAG epitope. Strong bands at expected sizes for Sua7 eluate samples verify the effective pull-down of the protein (**Figure 5.7A**). WT and *spt4Δ* Sua7 ChIP-seq were repeated twice. Data were corrected for input reads and analysed by a peak-calling algorithm MASC2 as described in the Methods (Feng et al. 2012). 3383 peaks were identified for PCGs. The reproducibility of the biological repeats was tested by counting reads around the peaks (taken from TSS-100 to TSS+100 nt) and calculating Spearman's correlation coefficient. Correlation plots and IGV tracks of representative genes show that WT and *spt4Δ* Sua7 repeats are reproducible (Spearman's ρ=0.895, 0.893, respectively, **Figure 5.7B&C**).

The position of Sua7 in WT and *spt4Δ* cells appears to be similar, and the level of Sua7 is slightly higher in WT cells (**Figure 5.8A**). To quantify the changes in the Sua7 levels, the ChIP-seq reads around the peaks (TSS-100 to TSS+100 nt) were counted in each gene. The DEseq2 algorithm in the R package was then applied to the data (Love et al. 2014). Out of 3269 PCGs, Sua7 levels in 3184 (97%) genes remain unchanged (p-adjusted < 0.05, **Figure 5.8B**). This data suggests that the levels of the PIC is not altered in *spt4Δ* and would not explain the increased RNAPII signal around the TSS in *spt4Δ.*

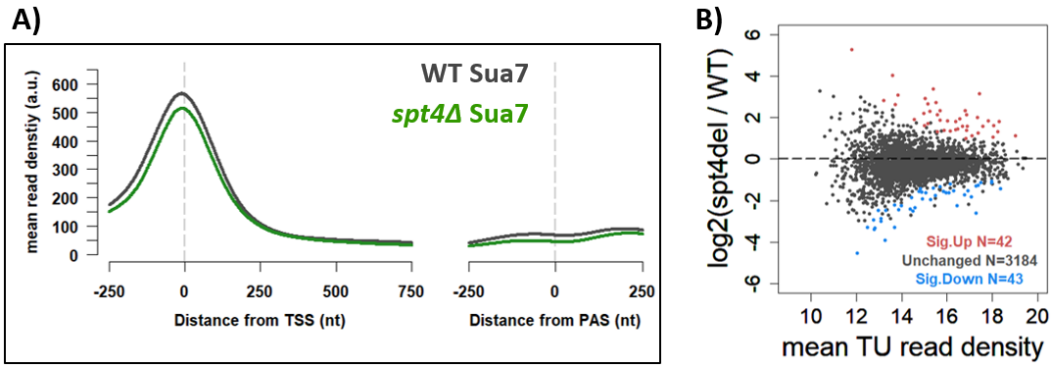


Figure 5.8) Sua7 positions and levels remain similar in WT and *spt4Δ*

A) Metagene plots of Sua7 ChIP-seq reads in WT (black) and *spt4Δ* (green).

B) Differential enrichment analysis of Sua7 in WT and *spt4Δ*. DEseq2 applied to the read counts around the TSS (TSS-100 to TSS+100 nt) for the two replicates of each data. Significantly enriched and depleted genes indicated in red and blue, respectively (p-adjusted <0.05).

ChIP-seq is performed in duplicate (see Appendix 6A for individual plots). The background is removed by subtracting input reads. Reads are normalised to *S.pombe* spike-ins.

5.2.6 RNAPII distribution is altered in Spt4 anchor away cells

As a TEF, it is not surprising that a complete deletion of *SPT4* causes alterations in the distribution of RNAPII. Consistently, the modelling analysis predicts changes in all tested parameters except the moving ratio in window 2. It is very likely that the changes in the RNAPII distribution do not happen all at once, but they reflect the long-term effects of the lack of Spt4. To detect short-term effects and capture the starting point of the changes in the distribution of RNAPII in the absence of Spt4, the AA system was used. Upon depletion of Spt4 from the nucleus, as described in chapter 3, the position of RNAPII was mapped using NET-seq.

The Spt4-FRB cells were grown in YPD supplemented with rapamycin or DMSO for 140 min. NET-seq was performed and analysed as described previously. The efficient IP of Rbp3 was verified by western blotting using antibodies against FLAG-tag (**Figure 5.9A**). NET-seq of rapamycin treated Spt4-FRB cells was performed in duplicate. The reproducibility of two repeats was tested by Spearman`s correlation coefficient and visualised by IGV tracks of representative genes (**Figure 5.9B&C**). NET-seq of DMSO treatment was performed once in Spt4-FRB and once in Spt5-FRB cells. DMSO treatment of the two strains produced almost identical results as quantified by Spearman`s correlation and shown by metagene plots (**Figure 5.9B**, see Appendix 7). Importantly, NET-seq was also performed in cells lacking FRB-tag (No-FRB) as a control to monitor the effect of rapamycin addition on RNAPII distribution. DMSO treated Spt4-FRB and rapamycin treated no-FRB cells give near identical results as indicated by Spearman`s correlation coefficient (**Figure 5.9B** also see **Figure 5.10C**). The NET-seq of the rapamycin and DMSO treated Spt4-FRB cells were performed in parallel, to allow comparison of these two conditions.

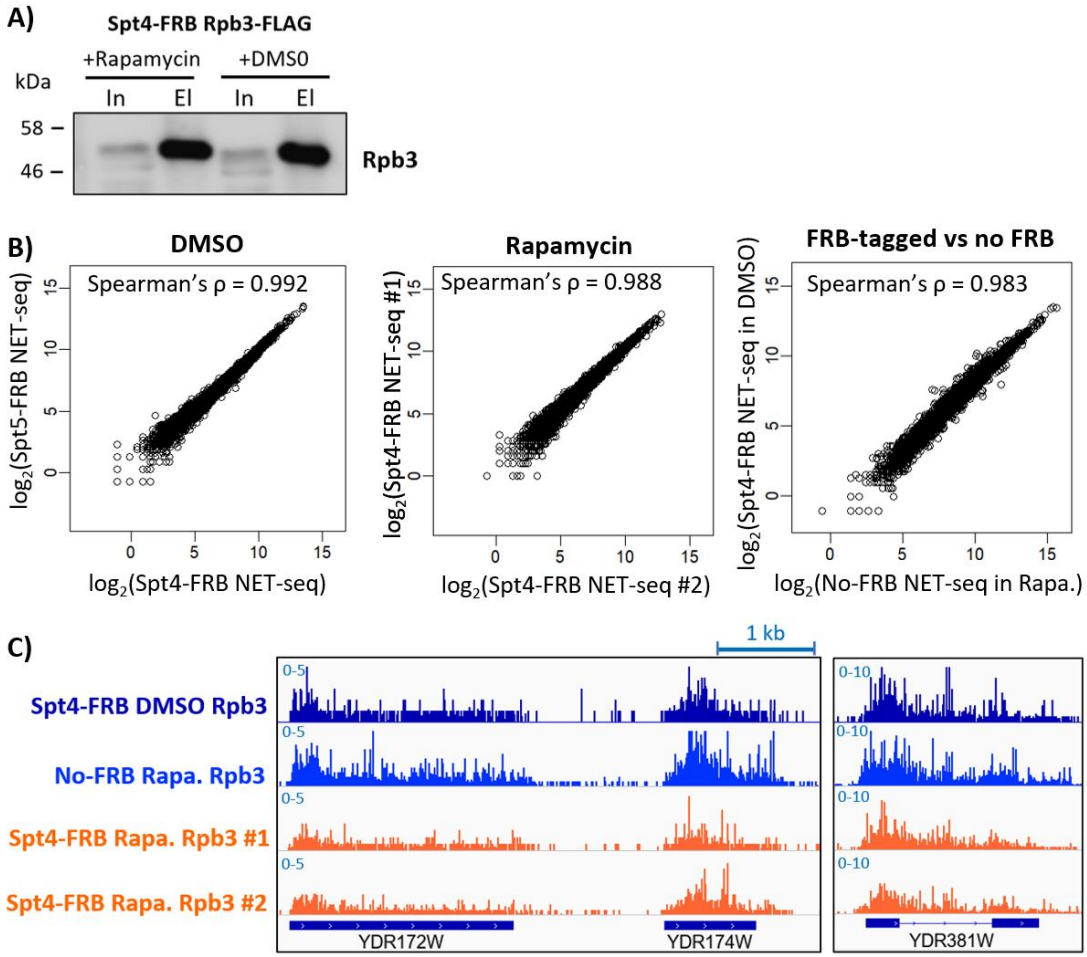


Figure 5.9) NET-seq in Spt4 Anchor Away cells upon rapamycin and DMSO treatment

A) NET-seq IP of FLAG-tagged Rpb3. Western blot of Input (In) and eluate (El) samples with anti-FLAG (M2) antibody run in 10 % SDS-PAGE. See Appendix 2C for the whole blots.

B) Correlation plots of the DMSO treated Spt4-FRB and Spt5-FRB, the two repeats of rapamycin-treated Spt4-FRB, or DMSO treated Spt4-FRB and rapamycin treated No FRB NET-seq data. Reads are counted from the TSS to the PAS for each gene. log_2 transformed gene counts are correlated and Spearman's ρ calculated for each pair.

C) IGV tracks of DMSO or rapamycin-treated Spt4-FRB or rapamycin treated No FRB NET-seq of three representative genes transcribed from the positive strand: *YDR172W, YDR174W* and *YDR381W*. Two biological replicates are shown for rapamycin treated cells. The dark blue boxes indicate the transcribed region of the genes (from TSS to PAS), while the blue line indicates the intronic region in *YDR381W.*

Comparison between the IGV tracks of the rapamycin-treated Sp4-FRB and control samples shows that RNAPII distributions look similar at all conditions at single-gene levels (**Figure 5.9C**). The genome-wide RNAPII distributions of the rapamycin and DMSO treated Spt4-AA cells were analysed by heatmaps and metagene plots. PCGs (n=4610) were ranked based on gene length and heatmaps were plotted for the RNAPII signal at single-nucleotide resolution as before. Heatmaps show a slightly higher read density close to the TSS in rapamycin-treated cells compared to DMSO control (**Figure 5.10A&B**). Metagene plots show that the RNAPII signals in both conditions look similar except around 200 nt downstream of the TSS (**Figure 5.10C**). The metagene profile of the No-FRB control is consistent with the DMSO control, verifying that the change in the RNAPII distribution in Spt4 depleted cells is not due to secondary effects of rapamycin treatment but due to the absence of Spt4 (**Figure 5.10C**). For quantification, the NET-seq reads were counted over the gene bodies (taken from TSS to PAS-250 nt) in both conditions. The comparison between the NET-seq reads in the DMSO and rapamycin treated Spt4-FRB cells indicates that RNAPII occupancy levels were not different in the two conditions (p-value = 0.716, Student's t-test, paired, two tailed **Figure 5.10D**).

The Spt4-AA NET-seq complements the *spt4Δ* NET-seq results and shows that the change in the distribution and dynamics of RNAPII in the absence of Spt4 first manifests itself around 200 nt downstream from the TSS, and becomes more pronounced in the continuous absence of Spt4 (in *spt4Δ* cells).

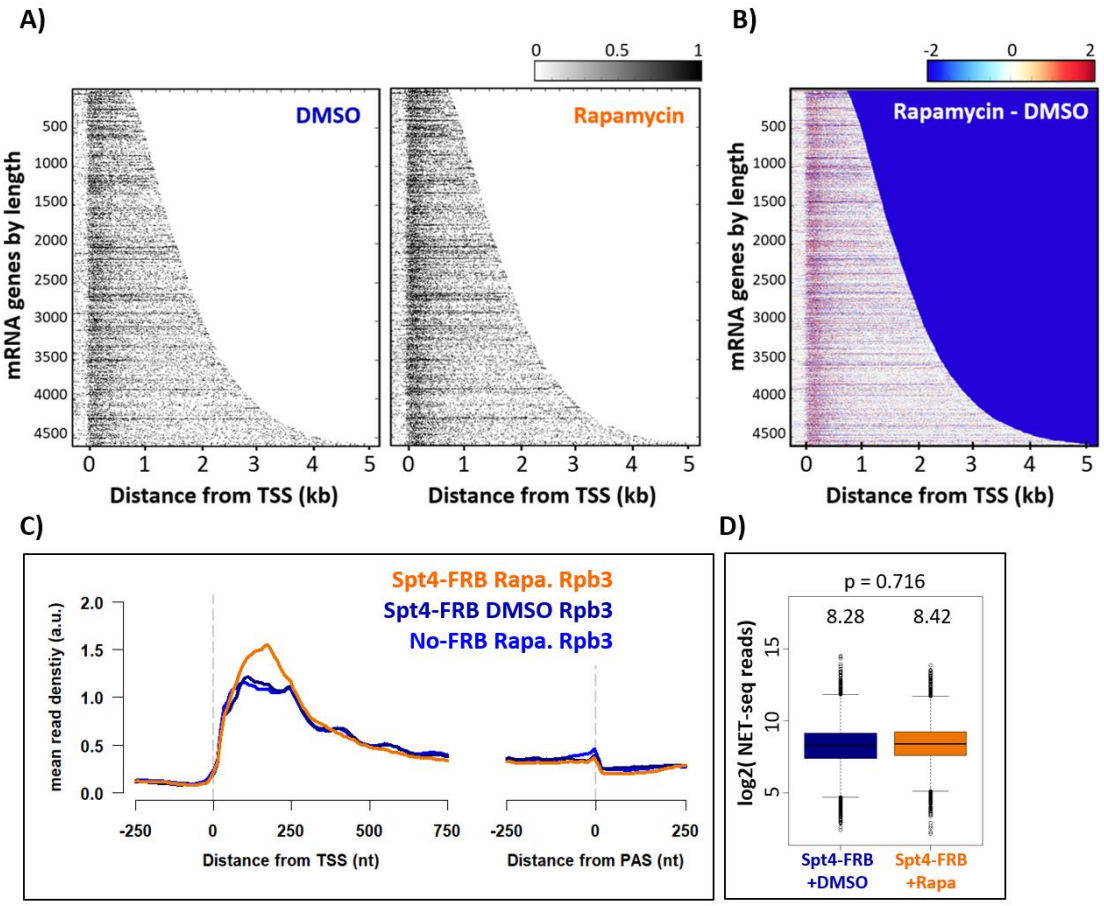


Figure 5.10) RNAPII accumulates at 200 nt downstream of the TSS in Spt4-FRB anchored away cells.

A) Heatmaps of NET-seq in DMSO (left) and rapamycintreated cells (right). Each row indicates an mRNA gene (n=5495), ranked by gene length. Amount of RNAPII is shown by different intensities of grey for each nucleotide position from TSS-250 nt to TSS+4750 nt (x-axis).

B) Heatmaps of the difference in rapamycin and DMSO treated Spt4-FRB signal (Rapamycin - DMSO). x and y axes are the same as in **A**. The colour code from red to blue reflects the changes in the RNAPII signal as shown in the colour bar.

C) Metagene plots of NET-seq reads in DMSO control (navy), rapamycin-treated Spt4-FRB (orange) and No-FRB cells (blue). Rapamycin treated NET-seq is performed in Spt4-FRB cells in duplicate and DMSO treated NET-seq performed once in Spt4-FRB and once in Spt5-FRB cells (See Appendix 8 for individual plots). Reads are normalised to *S.pombe* spike-ins.

D) Boxplots of the NET-seq reads in the DMSO (navy) and rapamycin (orange) treated Spt4-FRB cells on $\log_2$ scale. The median of the reads is indicated on the plot. The reads were counted for PCGs by taking the reads from gene bodies (TSS to PAS-250 nt) (p-value = 0.716, Student's t-test, paired, two-tailed).

Steady-state transcript levels were assessed in Spt4-AA cells taken from the same cultures as the NET-seq samples. The reproducibility of the two experimental repeats was tested by calculating Spearman's correlation coefficient. DMSO controls and Spt4-AA RNA-seq were reproducible (Spearman's ρ=0.989 and 0.980, respectively, **Figure 5.11A**). Next, differential expression analysis was carried out for PCGs (n=4272, transcripts with > 10 reads) using the DEseq2 package (Love et al. 2014). The transcript levels in most genes (3530 out of 4272, > 80%) remain unchanged (**Figure 5.11B**). For the 639 downregulated genes, a gene ontology (GO) analysis was done using the online tool yeastgenome.org/goTermFinder (with p-value cut off < 0.01). The downregulated genes were enriched for ribosomal genes, which are some of the most highly transcribed genes in the genome. Indeed, all the downregulated genes in Spt4-AA cells have generally high, or more than average, transcript levels in the control cells (**Figure 5.11C**). This might indicate that the absence of Spt4 first impacts the highly expressed genes, although there is no impact on transcript levels in *spt4Δ*, suggesting that the cells adapt to prolonged loss of Spt4. Additionally, given the growth rate is reduced in *spt4Δ* cells as shown in chapter 3, the depletion of Spt4 might trigger an initial negative feedback mechanism to reduce production of translation related genes. On the other hand, a small group of genes (103, < 3%) are upregulated in Spt4-AA cells and these genes are enriched for catabolic reaction and oxidative stress genes which might be triggered by depletion of Spt4. Note that here, the DMSO treated Spt4-FRB cells were used as a control to account for any change that might originate from the FRB-tagging of Spt4 and, the differences in the transcript levels due to rapamycin addition cannot be ruled out. A control experiment in No-FRB cells upon rapamycin addition would improve the results shown here.

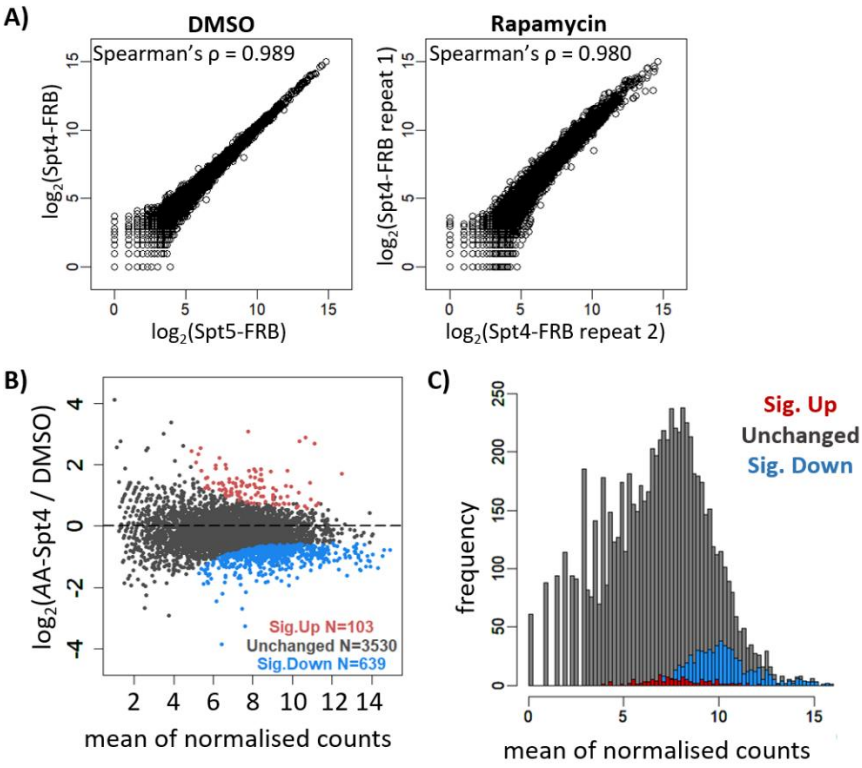


Figure 5.11) Steady-state transcript levels in Spt4-AA and DMSO control cells.

A) Correlations of the two biological repeats of RNA-seq (transcript) reads in DMSO (left) and rapamycin (right) treated cells.

B) DEseq2 analysis of RNA-seq reads in Spt4-AA and DMSO control. Significantly enriched and depleted genes indicated in red and blue, respectively (p-adjusted <0.05).

C) Histogram shows the distribution of transcripts levels in DMSO treated cells. Genes significantly downregulated upon Spt4 depletion are shown in blue, upregulated genes shown in red.

5.3 Discussion

This work demonstrates that Spt4 is required for normal distribution of RNAPII. Despite the high variability in the distribution of RNAPII across the genome, the deletion of *SPT4* similarly impacts the distribution of RNAPII at most genes. In *spt4Δ* cells, RNAPII accumulates early in transcription, regardless of gene length. The stochastic modelling of the shape of the distribution of RNAPII predicts lower processivity, higher frequency of stalling/backtracking events, and increased initiation/elongation ratio in *spt4Δ* compared to WT cells. Upon rapid depletion of Spt4, the RNAPII profile was less affected enabling us to show that the starting point of the changes in the distribution of RNAPII was around 200 nt downstream of the TSS.

Additionally, no changes in the levels and positions of the PIC was detected in *spt4Δ* compared to WT cells.

Interestingly, in both *spt4Δ* and Spt4 AA cells, RNAPII peaks around the same region: 200 nt downstream of the TSS. These observations suggest that Spt4 suppresses RNAPII accumulation around this region, perhaps by helping RNAPII pass a checkpoint or transcribe over a physical barrier. In this case, in Spt4 AA cells, RNAPII would not be efficiently released into gene bodies beyond a checkpoint/barrier and start accumulating. This accumulation could be due to long pauses, stalling or backtracking that could be normally resolved in WT cells. However, in *spt4Δ* cells, if RNAPII accumulates around the same region in each round of transcription, this would increase the probability of collision of RNAPII with an upstream RNAPII, higher frequencies of stalling and backtracking events progressively lead to a less permissive transcription into gene bodies. In turn, the RNAPII starts accumulating around 200 nt downstream of the TSS would pile up, spread more upstream toward the TSS and give rise to the high RNAPII signal in early transcription, in *spt4Δ* cells. The mathematical modelling of the NET-seq data would also support this interpretation as the moving ratio in early transcription (window 1) is highly and significantly decreased in *spt4Δ* compared to WT cells. If this is the case, the question is what is the checkpoint or barrier? One of the most likely candidates is a nucleosome barrier, and this hypothesis will be tested in chapters 8 and 9.

High stalling or backtracking early in transcription could also bring about a higher chance of the early termination in *spt4Δ*. There is emerging evidence showing that the early termination (also termed premature termination) is more wide-spread than initially thought and there is a high RNAPII turnover close to the TSS even under normal conditions (Kamieniarz-Gdula and Proudfoot 2019). A high rate of stalling, backtracking and probability of a collision between an upstream RNAPII in the absence of Spt4 could decrease the stability of RNAPII and end up in an early termination event. If more RNAPII is terminated before reaching the 3'-ends of the

genes in *spt4Δ*, this would also mean that the processivity is lower in *spt4Δ* cells as the model predicts. In fact, a previous study performed on a single gene also supports this model by suggesting a lower processivity of RNAPII in the *spt4* mutant cells (Mason and Struhl 2005).

An alternative explanation for the higher RNAPII signal at the beginning of genes would be because of more transcription initiation in *spt4Δ* cells compared to WT. At least two pieces of data argue against this possibility. First, as a result of the rapid depletion of Spt4, a high RNAPII signal is not detected immediately downstream of the TSS but instead, the RNAPII signal starts in the gene bodies. Therefore, the high RNAPII signal (from TSS to TSS+200 nt) in *spt4Δ* is more likely to be due to piling-up of the RNAPII which started accumulating in the gene bodies. Second, the levels of the PIC in WT and *spt4Δ* does not indicate changes in the transcription initiation rates. Note that here, the ChIP-seq on Sua7 is used to assess the levels of PIC and the conclusion is made based on the assumption that each initiation event would require binding of a new Sua7. In support of this, previous studies indicate that Sua7 dissociates from the transcription complex after synthesis of 8-13 nt RNA (Ly et al. 2020; Sainsbury et al. 2015). However, the possibility that some Sua7 stays bound at the promoters and recruits new RNAPII cannot be ruled out. If this is the case, the amount of transcription initiation would not directly correlate with the levels of Sua7. An alternative way of comparing the initiation events in WT and *spt4Δ* would be using one of the techniques relying on the detection of the 5'capped RNA, such as TSS-seq, PRO-cap, or CAGE-seq (Booth et al. 2016; Lu and Lin 2019; Qiu et al. 2020). Overall, having no changes in the RNAPII signals immediately downstream of the TSS upon depletion of Spt4, together with having no changes in the levels of Sua7 in WT and *spt4Δ* suggest that the RNAPII accumulation in the early stages of transcription in *spt4Δ* is not likely to be due to changes in the initiation rates.

Another explanation for the increased RNAPII signal in *spt4Δ* cells would be that the elongation rate (nt addition to RNA per min) of the RNAPII is altered in the mutant. Note that

the elongation rate defines the capability of RNAPII to add one nt to the RNA, and it is related to but different from the speed of RNAPII motion (which is often defined as a moving ratio). It is tempting to think that the reduced elongation rate is more likely to result in a high NET-seq signal, as the chances of immunoprecipitating a nascent transcript associated with a slow RNAPII are higher at any one position on a gene. Indeed, high NET-seq signals correlating with the reduced RNAPII speed have been reported for some genes and are generally observed at gene boundaries (Leng et al. 2020). However, a recent study performed with a fast RNAPII mutant in *Arabidopsis* suggested that the increased RNAPII speed also produces higher NET-seq signals in gene bodies (Leng et al. 2020). Therefore, caution must be taken when interpreting RNAPII speed from NET-seq data. Previously the Struhl Lab suggested that there is no difference in the RNAPII elongation rate in WT and *spt4* mutant cells (Mason and Struhl 2005). However, as this study was limited to a single gene and performed at one time point, it would be hard to draw a generalised conclusion. Given that the transcription rate in eukaryotes is predicted to be 2-6 kb/min (Gregersen, Mitter, and Svejstrup 2020) and the average yeast gene is 1.1 kb (almost all genes are < 10 kb), no attempt has been made to test the elongation rate in this study, since it would be technically challenging to get reliable results. More studies are needed to address the elongation rate of RNAPII in yeast and perhaps, a different organism with longer genes would be more appropriate to test the possible impact of Spt4 on the elongation rate.

It is worth noting that the mathematical model used here only considers the shape of the RNAPII distribution, meaning that it does not take into account the changes in the levels of RNAPII signal. Therefore, although the model predicts initiation to elongation ratio being increased in *spt4Δ* cells, implying either increase in the initiation or decrease in the elongation rate, it is still possible that the high RNAPII signal is mainly due to accumulated RNAPII without a substantial change in either initiation and elongation rates.

Structural studies showed that the Spt4/5 complex binds to the body of RNAPII and increases the stability of the transcription complex (Ehara et al. 2019). Thus, there is a possibility that the deletion of *SPT4* or the depletion of Spt4 induces conformational changes in the transcription complex which could increase the efficiency of the NET-seq IP and cause the high RNAPII signal. However, previously published PRO-seq data serves as an alternative approach here, and shows that the increased RNAPII density close to the TSS in *spt4Δ* cells could also be detected by immunoprecipitation of metabolically labelled transcripts (rather than RNAPII) (Booth et al. 2016). This supports a biological rather than technical reason for the RNAPII accumulation. To avoid such biases, experiments could be complemented with site-specific mutations in *SPT4* rather than the complete deletion of the gene.

The steady state RNA levels were less affected by the absence of Spt4. Upon rapid depletion of Spt4, 15 % of the genes were downregulated, but this was not observed in *spt4Δ* cells. This difference might be due altered the transcript stability in *spt4Δ* cells buffering the changes in the steady state transcript levels, as has been described by us before (Brown et al. 2018). The downregulation in the transcript levels could be monitored in Spt4 AA cells, perhaps before the transcript stability and the steady state transcript levels reach an equilibrium. Whether the amount of newly synthesis RNA is affected in *spt4Δ* cells could be investigated in more detail using metabolic labelling methods such as 4tU-seq (Shetty et al. 2017).

To conclude, the data supports a positive role for Spt4 in controlling the movement of RNAPII early in elongation by reducing backtracking or stalling events, and by helping RNAPII move over transcription obstacles or a checkpoint, before RNAPII moves into the processive elongation phase.

6 RNAPII distribution in Spt5 anchor away

6.1 Overview

In chapter 5, the changes in the distribution of RNAPII upon depletion of Spt4 could be detected using the AA system followed by NET-seq. In this chapter, the same approach was taken with Spt5, the binding partner of Spt4, to show the impact of loss of Spt5 on the distribution of RNAPII.

Like Spt4, Spt5 also associates with RNAPII in the early stages of transcription, travels with RNAPII, dissociates from the RNAPII upstream of the PAS as shown in chapter 3. However, unlike Spt4, Spt5 is an essential protein covering a large surface of the transcription complex. Spt5 has an unstructured C-terminal region (CTR) similar to the CTD of RNAPII, and the CTR of Spt5 is also subject to regulation by phosphorylation and mediates interactions with other TFs (Mayekar, Gardner, and Arndt 2013).

As Spt5 is an essential protein, the AA system enables studying the function of Spt5 in transcription. Therefore, the aim of this chapter is to study the distribution of RNAPII in Spt5 AA cells.

6.2 Results

6.2.1 Immunoprecipitation and data analysis

Spt5-FRB was depleted from the nucleus upon the addition of rapamycin for 180 min using the AA system as described in section 3.2.4. The DMSO treated Spt5-FRB cells were used as a control and NET-seq was performed as described previously. Note that here, no-tag IP control was not included in the AA coupled NET-seq experiments, so that the data analysis and the spike-in normalisation were performed as described in the Methods, except for the no-tag normalisation step.

The efficiency of the NET-seq IPs was checked by western blotting using antibodies against the FLAG epitope and against Ser5-P of RNAPII-CTD to confirm the successful IP of active RNAPII by western blotting. Strong Rpb3 and Ser5-P bands in the FLAG-tagged eluate samples are observed at expected sizes, verifying the effective pull-down of RNAPII (**Figure 6.1A**).

NET-seq of the rapamycin treated Spt5-FRB cells was performed in duplicate. The reproducibility of two repeats was tested by Spearman`s correlation coefficient and visualised by IGV tracks of representative genes (**Figure 6.1B&C**). Control experiments were performed in DMSO treated Spt4-FRB, Spt5-FRB cells as well as rapamycin treated No-FRB cells as previously presented in section 5.2.6. The DMSO treated Spt5-FRB and rapamycin treated no-FRB cells give near identical results as indicated by Spearman`s correlation (**Figure 6.1B,** also see **Figure 5.9B**). The NET-seq of rapamycin and DMSO treated Spt5-FRB cells were performed in parallel to compare these two conditions.

IGV tracks of the representative genes show that the DMSO treated Spt5-FRB cells and rapamycin treated no FRB cells produce very similar RNAPII patterns (**Figure 6.1C**). Importantly, the changes in the levels and distribution of RNAPII in the two replicates of rapamycin treated Spt5-FRB cells are reproducible. The RNAPII signal from the intronic region of *YGL103W* cells indicates that even though the native RNAs were detectable, the level of the RNAPII was substantially reduced in the Spt5 AA cells (**Figure 6.1C**).

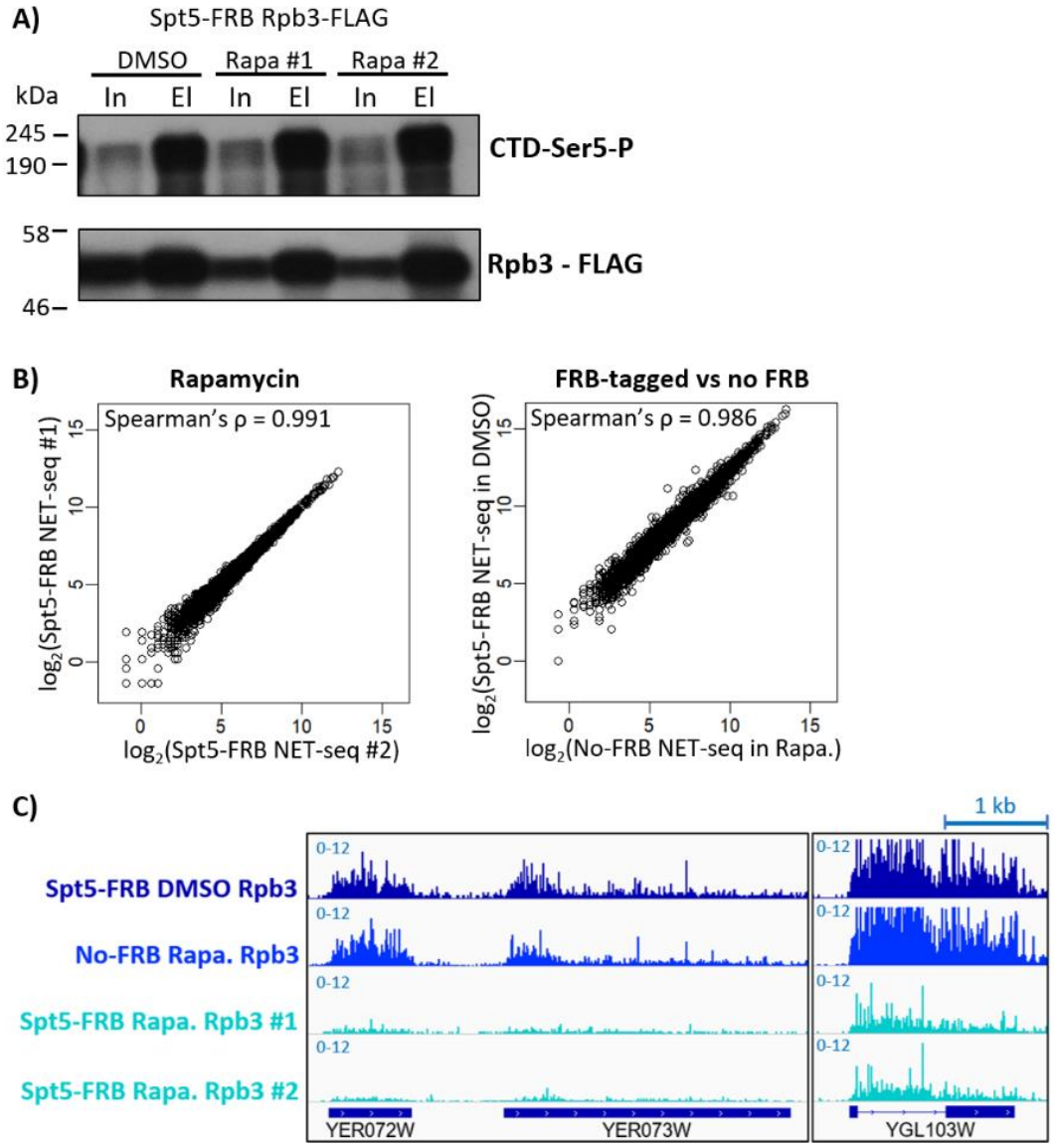


Figure 6.1) NET-seq in Spt5 Anchor Away cells upon rapamycin and DMSO treatment

A) NET-seq IP of FLAG-tagged Rpb3. Western blot of Input (In) and eluate (El) samples with anti CTD Ser5-P and anti-FLAG (M2) antibody run in 10 % SDS-PAGE. See Appendix 8 for the whole blots.

B) Correlation plots of the two repeats of rapamycin-treated Spt5-FRB, or DMSO treated Spt5-FRB and rapamycin treated No FRB NET-seq data. Reads are counted from the TSS to the PAS for each gene. log_2 transformed gene counts are correlated and Spearman's ρ calculated for each pair.

C) IGV tracks of DMSO and rapamycin-treated Spt5-FRB and rapamycin treated No FRB NET-seq of three representative genes transcribed from the positive strand: *YDR172W, YDR173W* and *YGL103W*. Two biological replicates are shown for rapamycin treated cells. The dark blue boxes indicate the transcribed region of the genes (from TSS to PAS), while the blue line indicates the intronic region in *YGL103W.*

6.2.2 The level and distribution of RNAPII are substantially affected in Spt5 AA cells

Next, genome-wide analysis were performed by plotting heatmaps for PCGs (n=4610) after depletion of Spt5. Importantly, as the data were normalised to spike-ins, global changes in the level of RNAPII in different conditions could be assessed. The heatmap of rapamycin treated Spt5-FRB cells compared to DMSO control shows that there is a reduction in the RNAPII read density at most genes (**Figure 6.2A**). Once the RNAPII signal of the rapamycin treated cells are subtracted from that of the DMSO control cells, it can be seen that the reduction in the levels of RNAPII is generally confined to the region around the TSS (**Figure 6.2B**). Although most genes indicate a reduction in the RNAPII read density represented by the blue signal, red signal is also observed in the subtracted heatmap for some genes (**Figure 6.2B**). This result suggests that the change in the RNAPII signal in the absence of Spt5, compared to the DMSO control, was not uniform across the genome.

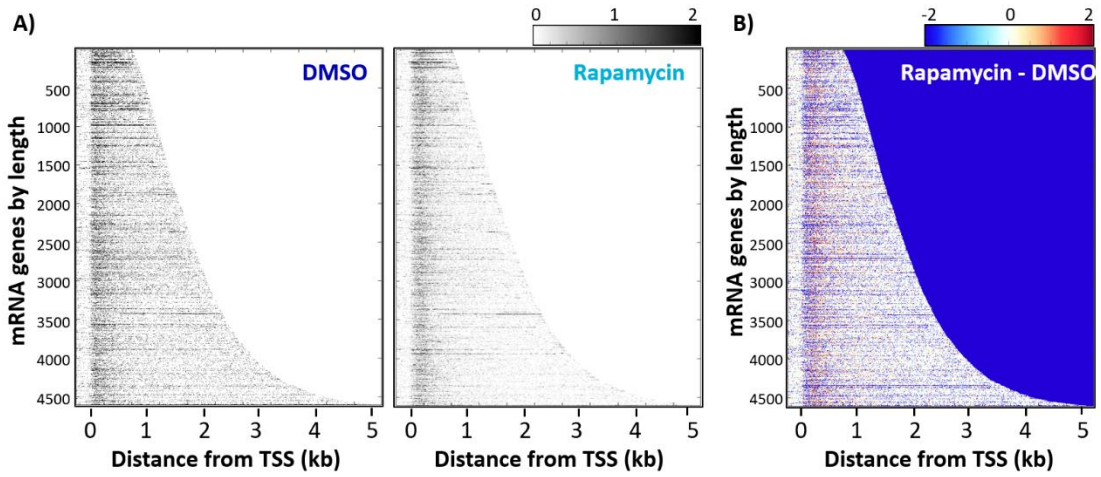


Figure 6.2) The distributions of RNAPII are different between rapamycin and DMSO treated Spt5-FRB cells in most of the PCGs

A) Heatmaps of NET-seq in DMSO treated (left) and rapamycin treated (right) Spt5-FRB cells. Each row indicates a PCG (n=4610), ranked by gene length. Different intensities of grey reflect RNAPII signal for each nucleotide position from TSS-250 nt to TSS+4750 nt (x-axis) as shown by the intensity bar.

B) Heatmaps of the difference in rapamycin and DMSO treated cells NET-seq signal (Rapamycin - DMSO). x and y axes are the same as in **A**. The colour code from red to blue reflects the changes in the RNAPII signal as shown in the colour bar.

To assess the extent of the heterogeneity in the change of the distribution of RNAPII, the genes were ranked based on the difference between the RNAPII signal in rapamycin and DMSO treated Spt5-FRB cells (rapamycin-DMSO) and heatmaps were re-plotted for a window from TSS-250 nt to TSS +750. In **Figure 6.3** the rapamycin-DMSO plot exhibits that the level of RNAPII is reduced for most genes. The reduction in the RNAPII levels seems to correlate with the levels of RNAPII in the DMSO treated Spt5-FRB (control) cells. A small portion of genes (top part of the rapamycin - DMSO plot) has a relatively higher RNAPII signal in rapamycin treated cells. The genes with the increased RNAPII signal upon Spt5 depletion are associated with osmotic stress and proteasome related genes as determined by GO enrichment analysis using online GOrilla tool by submitting the ranked gene list (Eden et al. 2009). This is possibly due to the transcriptional stress originated from the depletion of Spt5 in the cells.

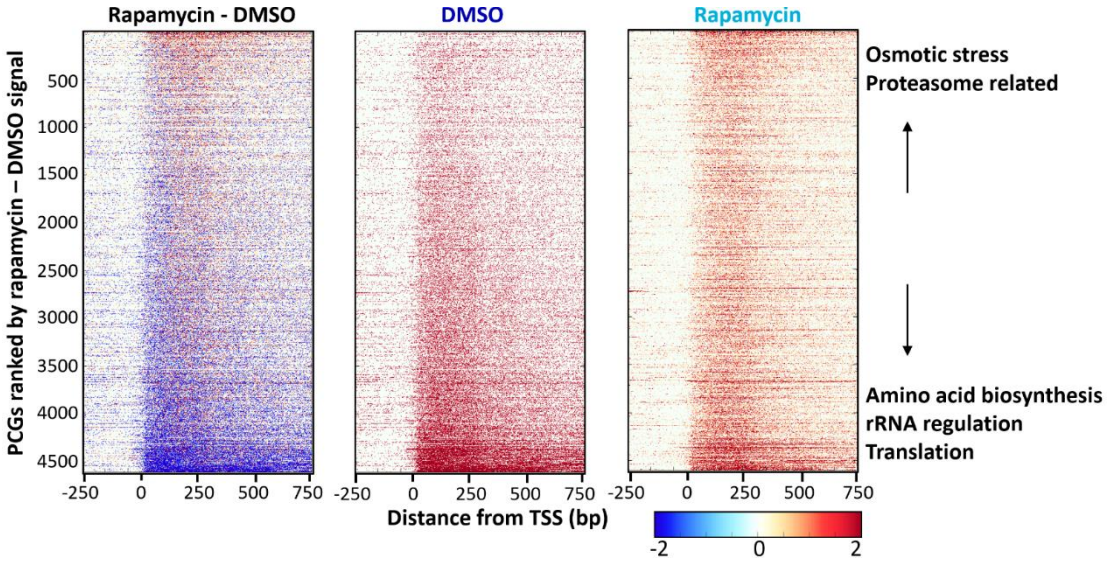


Figure 6.3) Depletion of Spt5 impacts the distribution of RNAPII

Heatmaps of the difference in rapamycin and DMSO treated cells NET-seq signal (Rapamycin – DMSO; left), in DMSO treated (middle) and rapamycin treated (right) Spt5-FRB cells. Each row indicates a PCG (n=4610), ranked by the signal intensities in the rapamycin-DMSO plot (i.e. from highest to lowest difference in the RNAPII signal between rapamycin and DMSO treated Spt5-FRB cells in the first 750 nt from the TSS).

Next, metagene profiles were plotted relative to the TSS for the PCGs > 750 nt (n=4212) to demonstrate the distribution of RNAPII across the genome. The metagene plots show that upon Spt5 depletion, the RNAPII signal is dramatically lowered compared to the DMSO control cells (**Figure 6.4A**). For quantification, the NET-seq reads were counted over the gene bodies (taken from TSS to PAS-250 nt) in both conditions. The comparison between the NET-seq reads in the rapamycin and DMSO treated Spt5-FRB cells indicates that RNAPII occupancy was significantly lower in Spt5 AA cells (by 49%, p <0.001, Student's t-test, paired, two tailed, **Figure 6.4B**). Interestingly, rather than a complete loss of the signal or a flattened distribution, the RNAPII density peaks around 250 nt downstream of the TSS and it drops again over the gene body including the PAS (**Figure 6.4A**). The metagene profile of the No-FRB control is consistent with the DMSO control verifying that the change in the RNAPII distribution in Spt5 depleted cells is not due to secondary effects of rapamycin treatment but due to the absence of Spt5 (**Figure 6.4A**).

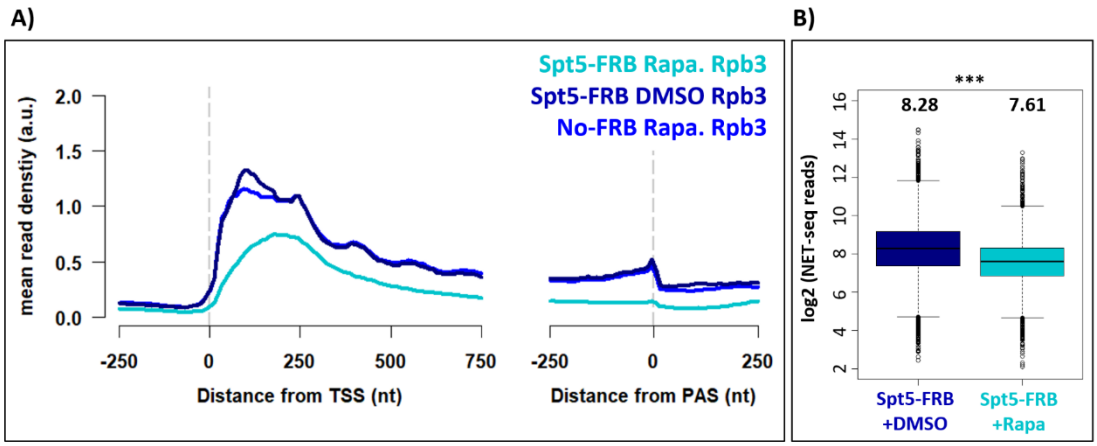


Figure 6.4) The level of RNAPII is decreased and RNAPII accumulates at 250 nt downstream of the TSS in Spt5-FRB anchored away cells.

A) Metagene plots of NET-seq reads in DMSO control (navy), rapamycin-treated Spt5-FRB (cyan) and No-FRB cells (blue). Rapamycin treated NET-seq is performed in Spt5-FRB cells in duplicate and DMSO treated NET-seq performed once in Spt4-FRB and once in Spt5-FRB cells (See Appendix 7 for individual plots). Reads are normalised to *S.pombe* spike-ins.

B) Boxplots of the NET-seq reads in DMSO (navy) and rapamycin (cyan) treated Spt5-FRB cells on $\log_2$ scale. The median of the reads is indicated on the plot. The reads were counted for PCGs by taking the reads from gene bodies (TSS to PAS-250 nt) (p-value < 0.001, Student's t-test, paired, two-tailed).

Overall, the data suggest that the depletion of Spt5 leads to a reduced RNAPII signal for most genes and impacts the distribution of RNAPII at every stage of transcription from initiation to termination. Furthermore, RNAPII has a distinct pattern 250 nt downstream of the TSS, suggesting a pausing or stalling point of RNAPII in the absence of Spt5.

6.2.3 Steady-state transcript levels are reduced in Spt5 AA cells

To test if the reduced RNAPII over the genes was reflected in the steady state transcript levels, RNA-seq was performed with the cells taken from the same cultures of NET-seq samples.

The reproducibility of the two biological repeats was tested by calculating Spearman's correlation coefficient. The DMSO controls and Spt5-AA RNA-seq were reproducible (Spearman's ρ=0.989 and 0.977, respectively, **Figure 6.5A**). Differential expression analysis was then carried out for PCGs (n=4835, transcripts with >10 reads) using the DEseq2 package (Love et al. 2014). 4065 out of 4835 PCGs (> 83%) are significantly downregulated upon depletion of Spt5 (**Figure 6.5B**) and a small group of transcripts (n=101) were significantly upregulated. The 101 upregulated genes are found to be enriched for catabolic reaction and oxidative stress genes as a result of gene ontology (GO) analysis performed by using online yeastgenome.org/goTermFinder (with p-value cut off <0.01). Catabolic reaction and oxidative stress genes might be triggered due to depletion of Spt5, indicating that the cells are stressed. Importantly, here the DMSO treated Spt5-FRB cells were used as a control to account for any change which might originate from the FRB-tagging of Spt5 and the differences in the transcript levels due to rapamycin addition cannot be ruled out. In conclusion, these data show that both transcription and transcript levels were severely affected upon 180 min depletion of Spt5.

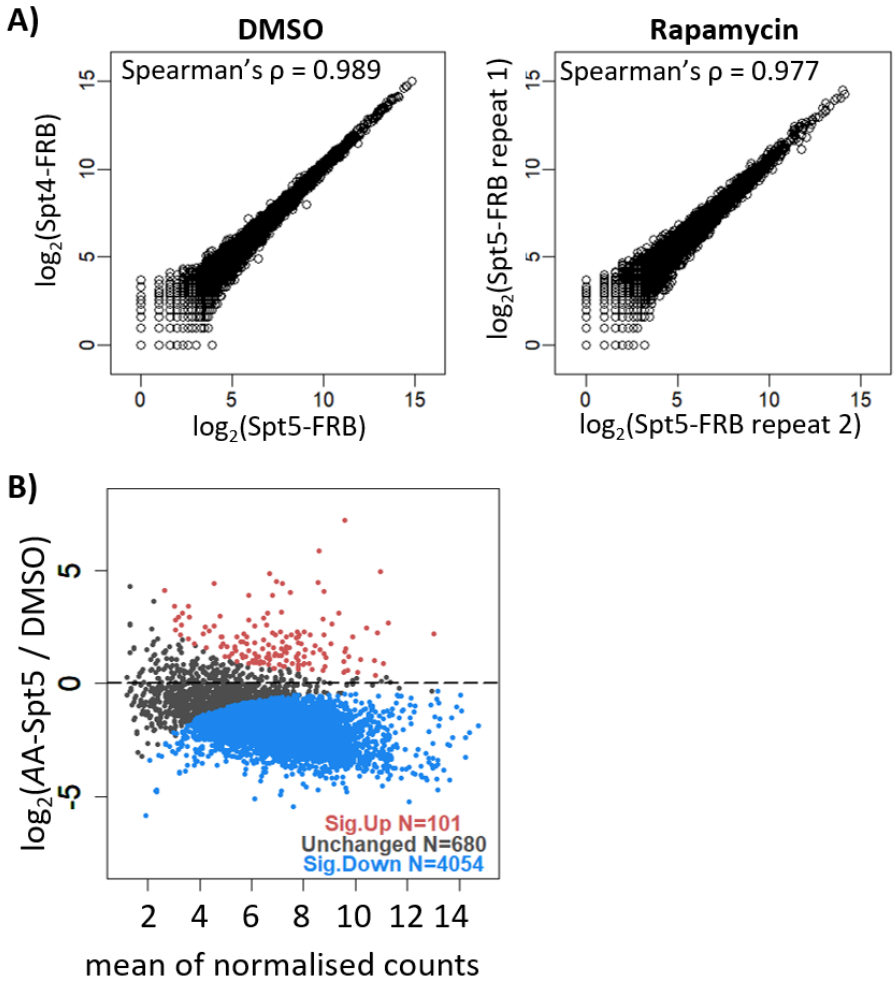


Figure 6.5) Steady-state transcript levels in Spt5-AA and DMSO control cells.

A) Correlations of the two biological repeats of RNA-seq (transcript) reads in DMSO (left) and rapamycin (right) treated cells.

B) DEseq2 analysis of RNA-seq reads in Spt5-AA and DMSO control. Significantly enriched and depleted genes indicated in red and blue, respectively (p-adjusted <0.05).

6.2.4 The depletion of Spt5 has a different impact on the distribution of RNAPII compared to that seen in Spt4 depleted cells

Next, the distribution of RNAPII upon Spt4 and Spt5 AA was compared by re-plotting the metagene profiles together.

The comparison between the depletion of Spt4 or Spt5 on the distribution of RNAPII demonstrates the differences and similarities of the two conditions. The most striking difference between Spt4 and Spt5 AA is that the levels of RNAPII are lower in Spt5 AA (**Figure**

6.6). Given that the steady state levels of transcripts are also greatly reduced and Spt5 is an essential TEF, the low NET-seq signal suggests a reduced transcriptional activity in Spt5 AA cells. Another difference between Spt4 and Spt5 AA is at the TSS. The earliest signals of RNAPII observed in the Spt4 AA and the DMSO control cells are similar. However, in the Spt5 AA cells, the first RNAPII signal arises further downstream of the TSS compared to Spt4 AA and DMSO cells, implying shifted and/or impaired transcription initiation upon depletion of Spt5 (**Figure 6.6**).

Additionally, both in the Spt4 and Spt5 AA cells, RNAPII peaks around 200-250 nt downstream of the TSS, suggesting an accumulation of RNAPII around this region, implying a common function for Spt4 and Spt5, or the function of Spt4/5 as a complex, probably in passing a transcriptional barrier.

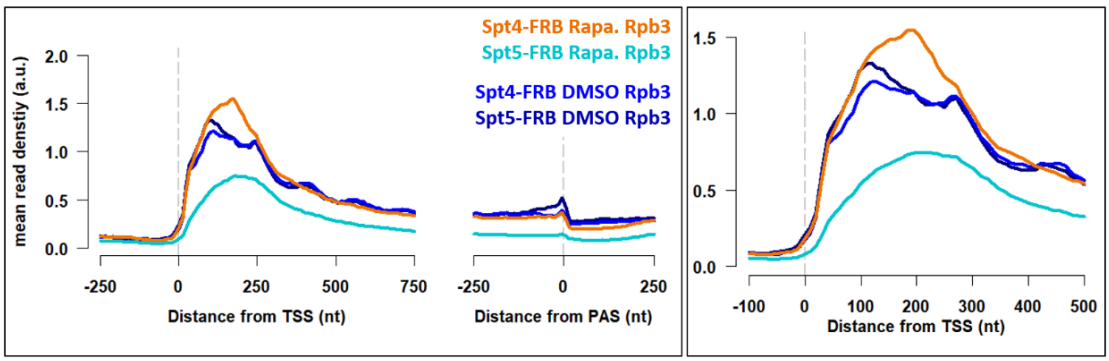


Figure 6.6) The depletion of Spt4 and Spt5 differentially impact the position and levels of RNAPII across the genome.

Metagene plots of NET-seq reads in rapamycin-treated Spt4-FRB (orange) or Spt5-FRB (cyan) or DMSO treated Spt4-FRB (blue) or Spt5-FRB (navy) cells. Close-up of the same plot (left).

Reads are normalised to *S.pombe* spike-ins.

6.3 Discussion

In this chapter, it was shown that the distribution of RNAPII is altered and the levels of RNAPII is dramatically reduced upon depletion of Spt5. The reduction in the levels of RNAPII was also reflected in the steady state transcript levels. The comparison between the Spt4 and Spt5 AA conditions shows that in both conditions RNAPII accumulates in the early stages of transcription.

The experiments performed here show the importance of the spike-in normalisation. The global reductions in the levels of the RNAPII signal and the steady state RNA in the Spt5 AA cells could only be assessed with the addition of an external and fixed amount of RNA before the IP step of the sequencing techniques.

Importantly, the RNAPII signal and the steady state transcript levels showing a parallel trend is not a general phenomenon. Studies show that transcript stability in cells is highly dynamic and can buffer the changes originating from altered transcription rates (Blasco-Moreno et al. 2019; Brown et al. 2018). However, here, given that the Spt5-FRB cells stop growing after 180 min of rapamycin treatment, it is clear that the transcription is severely impaired, and the cells cannot continue normal levels of transcription the absence of Spt5.

The results shown here agree with a recent study conducted in *S.pombe*. The study noted RNAPII accumulation in the early transcription using NET-seq upon conditional depletion of Spt5 by the AID system (Shetty et al. 2017). This study did not report a reduction in the RNAPII signal due to lack of quantitative normalisation. However, the researchers measured 4-thiouracil (4tU) incorporation rate into newly synthesised RNAs (4tU-seq) and concluded a significant decrease in the RNA synthesis rate upon degradation of Spt5. Notably, although possible technical problems about the use of the AID system was raised in chapter 3, the results presented in this study and the former study draw similar conclusions; RNAPII accumulates in early transcription and transcriptional activity is reduced in the absence of

Spt5. These results show that the essential function of Spt5 in transcription is conserved between *S.cerevisiae* and *S.pombe*.

Despite the dramatic changes in the levels of RNAPII in the Spt5 AA cells, the resolution of NET-seq afforded here make further observations on the distribution of RNAPII possible. These observations were also possible because the depletion time of Spt5 was kept minimal, enabling the early responses in the distribution of RNAPII to be observed. Extended time for the depletion of Spt5 would result in even lower RNAPII signals as downregulation in transcription would progressively decrease the protein levels and eventually shut down the transcription. Thus, it is important to find an effective time window for the AA system: long enough to allow sufficient depletion of the protein, yet short enough to capture the changes in the transcription.

A downstream shift at the TSS and a distinct peak around 250 nt from the TSS was detected in the Spt5 depleted cells compared to the DMSO control cells. Although NET-seq is not designed to report the transcription initiation sites, the downstream shift in the earliest RNAPII signal around the TSS might imply changes in the TSS. A recent study showed that catalytically slow mutants of RNAPII tend to have TSS more downstream compared to WT cells (Qiu et al. 2020). Consequently, the RNAPII shift at the TSS observed in the Spt5 AA cells might be a result of the lowered transcriptional activity in these cells.

The depletion of Spt4 and Spt5 have opposite effects on the densities of RNAPII. Spt4 depletion (and deletion) leads to an increased RNAPII density over the gene bodies, whereas Spt5 depletion leads to a decreased RNAPII density compared to the control cells. However, once only the shape of the distribution of the RNAPII is considered, in the Spt5 depleted cells, RNAPII shows local accumulation around 250 nt downstream of the TSS that is similar to the location of the RNAPII accumulation observed as results of Spt4 deletion and depletion, shown in chapter 5. This observation is not surprising, as the depletion of Spt5 is likely to

result in the depletion of Spt4 as well as we discussed in chapter 3. Although this has not been tested in this study, the structural studies in yeast, archaea, and mammalian cells point out that the interaction of Spt4 is mediated by Spt5 (Ehara et al. 2019; Martinez-Rucobo et al. 2011; Vos et al. 2018). Therefore, the reduction in transcriptional activities and the accumulation of RNAPII in the early stages of transcription could be interpreted as being as a result of depletion of Spt4-Spt5 as a complex.

Alternatively, as discussed in chapter 5, the reduced RNAPII signal can be explained by the changes in processivity and speed of RNAPII such that an increase in the speed of RNAPII due to less stalling and backtracking could lead to a lower RNAPII signal upon depletion of Spt5. However, as Spt5 is an essential TEF and the negative impacts of the depletion in both cell survival and transcript levels could be clearly detected, the increased speed of RNAPII is unlikely to be the case in the Spt5 AA cells.

7 The composition of transcription complex in *spt4Δ* cells

7.1 Overview

The composition of the transcription complex dynamically changes due to continuous association and dissociation of TFs with RNAPII as transcription progresses on chromatin. The dynamic association and dissociation of TFs is dependent on the stage of transcription and/or completion of the upstream events. For instance, TFIIF only associates with transcription initiation complex after the recruitment of Sua7 (TFIIB) and dissociates after promoter escape (Sainsbury et al. 2015). Another example is that termination factor Rai1 is recruited to transcription complex only after Ser2 phosphorylation of the RNAPII-CTD (Baejen et al. 2017). Given that Spt4 travels with the elongating RNAPII and is required for normal distribution of RNAPII, in this chapter, whether and how Spt4 impacts on the composition of the transcription complex were explored.

To answer this question, a proteomics approach was taken and affinity purification mass spectrometry (AP-MS) was performed (Burriss and Mosley 2019). The transcription complex is purified by immunoprecipitating FLAG-tagged Rpb3 from WT and *spt4Δ* cells, the proteomics of the eluted samples was performed by mass spectrometry (MS) and the protein abundances were assessed and compared by a label-free quantitative mass spectrometry analysis (Zhang et al. 2018) (**Figure 7.1**).

Although mass spectrometry is a sensitive technique, some proteins having transient interactions with RNAPII, such as kinases, are more challenging to IP without a crosslinking step (Joo et al. 2019). Therefore, the MS results were complemented by ChIP-seq performed on the Kin28 (CDK7) and Bur1 (CDK9) kinases that phosphorylate the CTD of RNAPII on Ser5 and Ser2, respectively, to assess the enrichment of these factors in the absence of Spt4.

Thus, this chapter aims to identify the changes in the composition of transcription complex and to examine the association of Kin28 and Bur1 with chromatin in *spt4Δ* compared to WT cells.

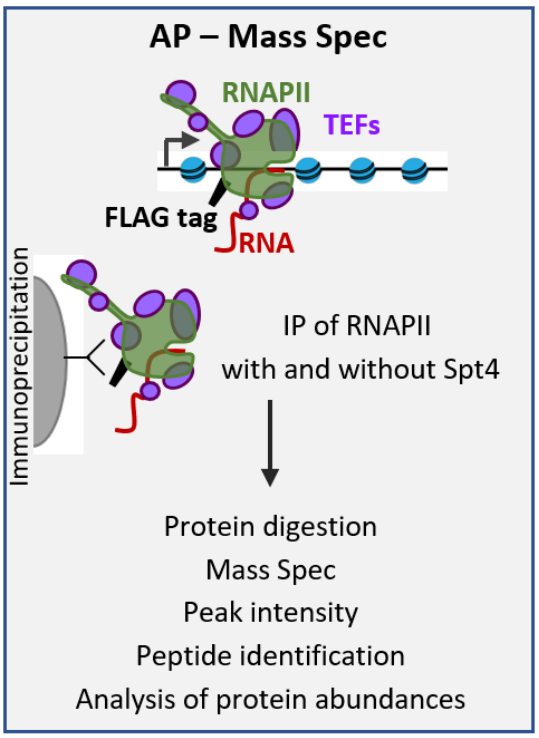


Figure 7.1) Affinity purified mass spectrometry

7.2 Results

7.2.1 Mass Spectrometry identifies TFs in the transcription complexes

The IP of RNAPII was carried out in WT and *spt4Δ* cells following the IP step of NET-seq protocol with some modifications. Like the NET-seq pulldown, the experiment was performed including an RNase inhibitor to avoid RNA degradation, meaning that the protein interactions supported by RNA were preserved in the transcription complex. The technique differed from the IP of NET-seq, as phosphatase inhibitors were included during the IP to avoid losing interactions maintained by phosphorylated proteins. Moreover, an increased number and duration of washing steps with higher salt concentrations were added to the protocol to reduce background as described in Methods.

The composition of the transcription complexes relies to a large extent on native interactions, and we wanted to get a holistic view of its composition. Therefore, similar to previous work by (Harlen *et al.*, 2016), non-stringent washes were performed during the IP of the transcription complexes. This risks the IP of non-specific proteins in the samples. To account for contamination and to detect non-specific proteins, purifications were also performed with no tag control samples. After purification of the complexes, a small portion of the purified samples was run on SDS-PAGE and the protein enrichment was checked using silver staining to test if the IP was successful. Silver staining exhibits the successful pulldown of many proteins, some presumably being the large subunits of RNAPII (Rpb1 and Rpb2) given the band sizes in WT and *spt4Δ* Rpb3-FLAG samples. Silver staining also verifies enrichment of many proteins over the no tag IPs (**Figure 7.2**). The purified samples were then sent for proteomics analysis.

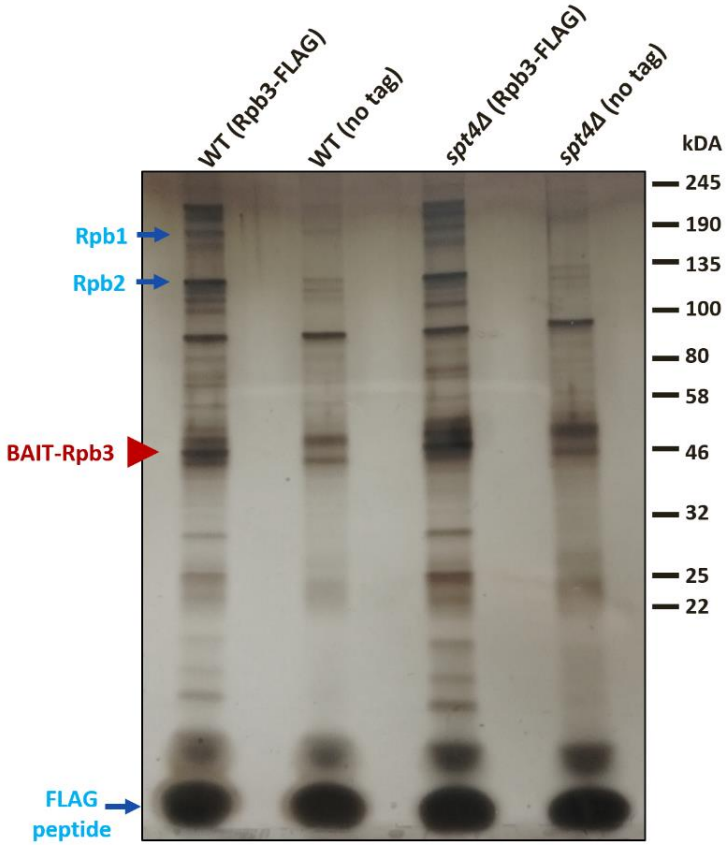


Figure 7.2) Silver staining of purified transcription complexes.

Purified samples run in 4-20% gradient SDS-PAGE. Proteins were detected using Silver staining (Pierce Silver Stain Kit).

The IP of the transcription complex was performed in biological duplicate. Proteomics of the purified samples was detected by MS and protein intensities were generated by MaxQuant software at the mass spectrometry facility. The proteins represented by fewer than two peptides in the IP specific (FLAG-tagged) samples were discarded. Protein enrichment analysis was then performed by using the DEP package / RStudio (Zhang et al. 2018). The analysis validates the successful enrichment of the bait protein, Rpb3 in WT and *spt4Δ* FLAG-tagged samples over their no tag controls in the both replicates (**Figure 7.3**).

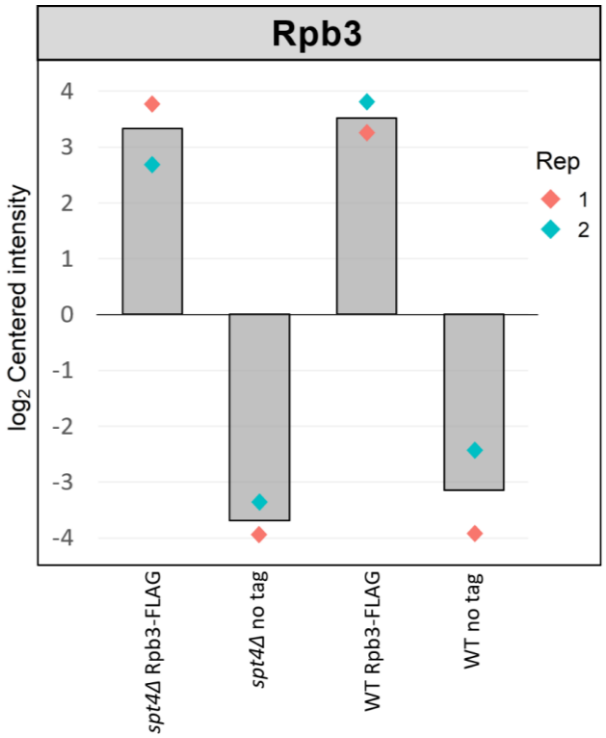


Figure 7.3) Rpb3 was enriched in FLAG-tagged IPs compared to their no tag controls

Log_2 centred intensities of Rpb3 (bait) in WT and *spt4Δ* samples are plotted. IP and MS were performed in two biological repeats.

First, the protein enrichments were analysed separately for WT and *spt4Δ* conditions to compare their FLAG-tagged and no tag IPs and remove non-specific proteins. In **Figure 7.4** volcano plots display the fold enrichment and the significance of all identified proteins in each condition (336 proteins for WT and 352 for *spt4Δ*). TEFs, splicing proteins, histones, and all subunits of RNAPII were significantly enriched in both WT and *spt4Δ* FLAG-tagged samples.

No tag controls were enriched for cytoskeleton, ribosomal, and highly expressed metabolic proteins. Proteins with $\log_2$(fold change FLAG/no tag) < 2 and p-value > 0.05 were considered as contaminant proteins and removed from the data sets for further analysis. The background removal left 68 IP specific proteins in WT and 78 proteins in *spt4Δ* conditions (**Figure 7.4** and **Figure 7.5**). Some proteins were only enriched in one condition, indicating that the composition of the transcription complexes was altered in the absence of Spt4 (**Figure 7.5**). However, caution must be taken while drawing conclusions from these lists of proteins as not all of them were similarly abundant and equally consistent between the replicates. Next, the protein enrichment analysis was applied to compare the abundances of the proteins on WT and *spt4Δ* transcription complexes.

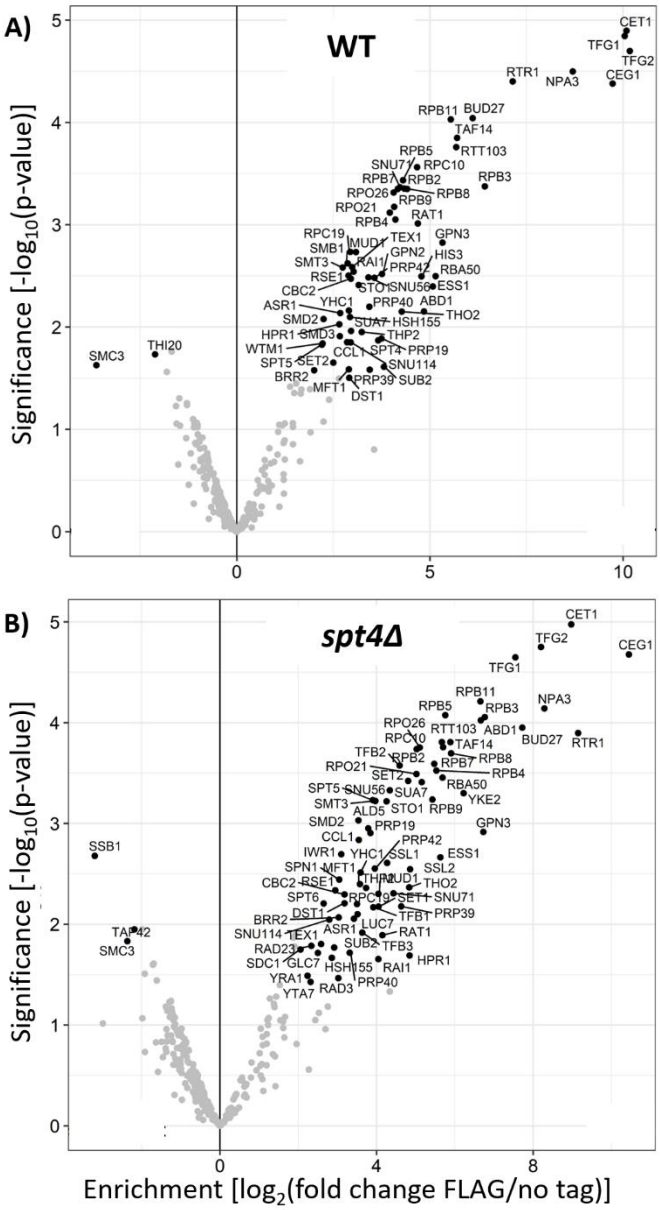


Figure 7.4) Protein enrichments in WT and *spt4Δ*

A) Volcano plots of 336 identified proteins in WT FLAG-tagged and no tag samples.

B) Volcano plots of 352 identified proteins in *spt4Δ* FLAG-tagged and no tag samples.

Proteins with p-value < 0.05 are tagged.

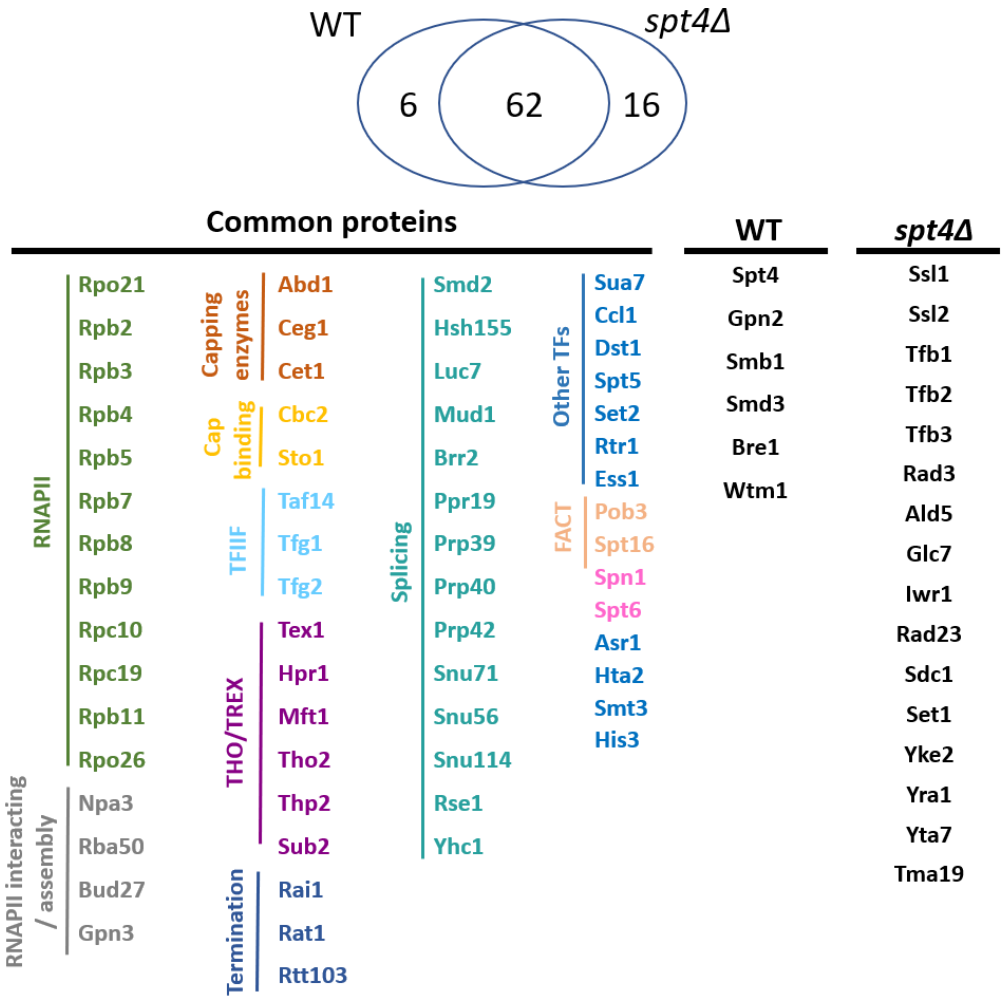
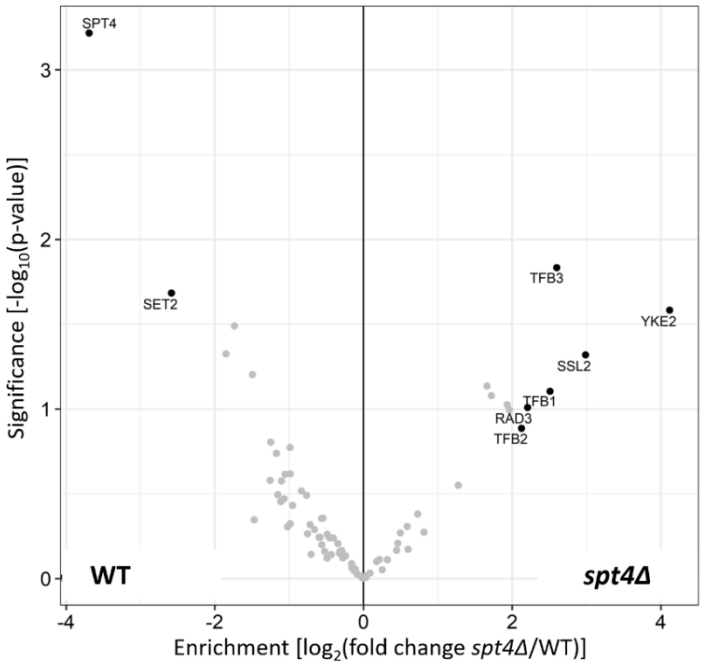


Figure 7.5) IP specific proteins detected by mass spectrometry

68 IP specific proteins are found in WT IP and 78 proteins in *spt4Δ*. IP specific proteins are selected based on their fold enrichments and significance: $\log_2$(fold change FLAG/no tag) > 2 and p-value < 0.05.

7.2.2 The transcription complex composition changes in *spt4Δ*

The protein enrichment analysis was applied to the IP specific proteins left after the background filtering (n = 84). The analysis was visualised by a volcano plot showing the fold enrichments and the significance of the differentially abundant proteins (**Figure 7.6**). The full list of proteins is given in the Appendix (See Appendix table 2).



Protein	p-value	p-adjusted
Spt4	< 0.001	< 0.001
Tfb3	0.015	< 0.001
Set2	0.021	< 0.001
Yke2	0.026	< 0.001
Ssl2	0.048	0.003
Tfb1	0.079	0.012
Rad3	0.099	0.019
Tfb2	0.130	0.039

Figure 7.6) Transcription complex composition changes in *spt4Δ*

Volcano plot shows proteins with significantly changed RNAPII association upon *SPT4* deletion. log_2(fold change *spt4Δ*/WT) > 2 and p-adj< 0.05. P-values and p-adjusted vales are also given on the right for the significant proteins.

The data suggest that the transcription complex composition changes in the absence of Spt4. Unsurprisingly, Spt4 is the most significantly depleted protein from the WT complex verifying the deletion of the protein and validating that the analysis can successfully pick up differently enriched proteins (**Figure 7.6**). Interestingly, in the absence of Spt4, a lower amount of Set2 was found to interact with the transcription complex compared to that of in WT cells (**Figure 7.6**). Set2 is a methyltransferase which deposits methylation marks on lysine36 on histone H3 (Bilokapic and Halic 2019) and the interaction between Set2 and RNAPII was also previously detected by MS studies (Burriss and Mosley 2019). This result suggests that recruitment of Set2 to the transcription complex is directly or indirectly dependent on Spt4.

The abundances of six TEFs is increased in the absence of Spt4. One of them is a prefoldin protein, Yke2 (also named as Pfd1) (**Figure 7.6**). Yke2 takes part in the co-translational assembly of proteins in the cytoplasm, but one study claims that Yke2 is transported to the nucleus and can involve in the chromatin dynamics by helping histone eviction during

transcription (Millán-Zambrano et al. 2013). Importantly, in WT, the abundance of Yke2 was below the background levels which is in agreement with previous work analysing the profile of RNAPII associated factors using proteomics (Harlen *et al.*, 2016; Joo *et al.*, 2019). It is worth pointing out that the transcript levels of the gene encoding Yke2 (*YLR200W*) were similar between WT and *spt4Δ* cells, as well as Spt4 AA and DMSO control cells as assessed by RNA-seq in chapter 5, so that it does not seem as if the higher association of Yke2 with transcription complex is due to overexpression of *YLR200W* in *spt4Δ* cells. One explanation for the involvement of YKe2 in the *spt4Δ* transcription complex might be that Yke2 becomes important in the absence of Spt4, perhaps to improve histone eviction.

Five other abundant proteins (Tfb1, Tfb2, Tfb3, Ssl2, and Rad3) on the *spt4Δ* transcription complex belong to 10 subunit TFIIH complex. TFIIH has two modules: the core (Ssl1, Ssl2, Rad3, Tfb1, Tfb2, Tfb4, and Tfb5) and the kinase module (Kin28, Tfb3, Ccl1) (**Figure 7.7**). The kinase module is mobile and takes a part in the transcription initiation by phosphorylating the RNAPII-CTD. The core module associates with the RNAPII and functions both in promoter melting in during transcription initiation and the formation of the DNA repair bubble during transcription coupled DNA repair (Kokic et al. 2019). Given its dual role in transcription, the high abundance of TFIIH on *spt4Δ* RNAPII brings about two non-mutually exclusive predictions. First, a relatively larger portion of the *spt4Δ* transcription complex is at the initiation stage compared to WT cells. Second, the DNA damage response pathway is activated in the absence of Spt4. The latter possibility will be further discussed in the discussion section.

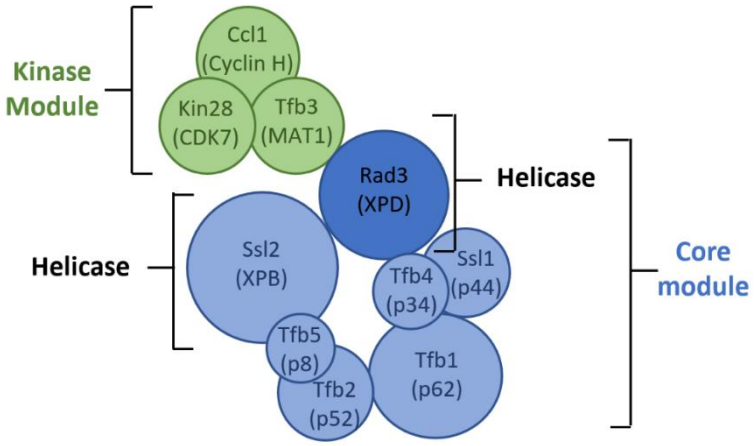


Figure 7.7) Structural organisation of the 10 subunits of TFIIH

Human counterparts of the proteins are given in brackets (Gibbons et al. 2012).

The MS analysis showed that the *spt4Δ* transcription complex was enriched for subunits of TFIIH that are related to initiation stage of transcription and depleted from Set2 that joins transcription at later stages of elongation (**Figure 7.6**). This observation might indicate that a bigger portion of the transcription complex in *spt4Δ* cells are at initiation stage compared to that of in WT cells, however, it did not hold true for all TEFs detected. For example, TFIIF (Taf14, Tfg1, Tfg2), capping enzymes (Abd1, Ceg1, Cet1), and cap-binding proteins (Sto1 and Cbc2), and some splicing factors (such as Luc2, Mud1) play roles in the early stages of transcription (Baejen et al. 2014; Mayer et al. 2010), but their abundances on both transcription complexes were found to be similar (**Figure 7.8**). Likewise, transcription factors travelling with RNAPII such as Spt5 and Dst1, or factors joining into transcription at later stages such as THO/TREX complex subunits (such as Hpr1, Tho2), and termination related proteins (Rtt103, Rai1, Rat1) were similarly enriched in both complexes (Baejen et al. 2014, 2017) (**Figure 7.8**). In conclusion, these results suggest that in *spt4Δ* cells, the composition of transcription complex is changed, and this is not merely due to a general bias towards a certain stage of transcription.

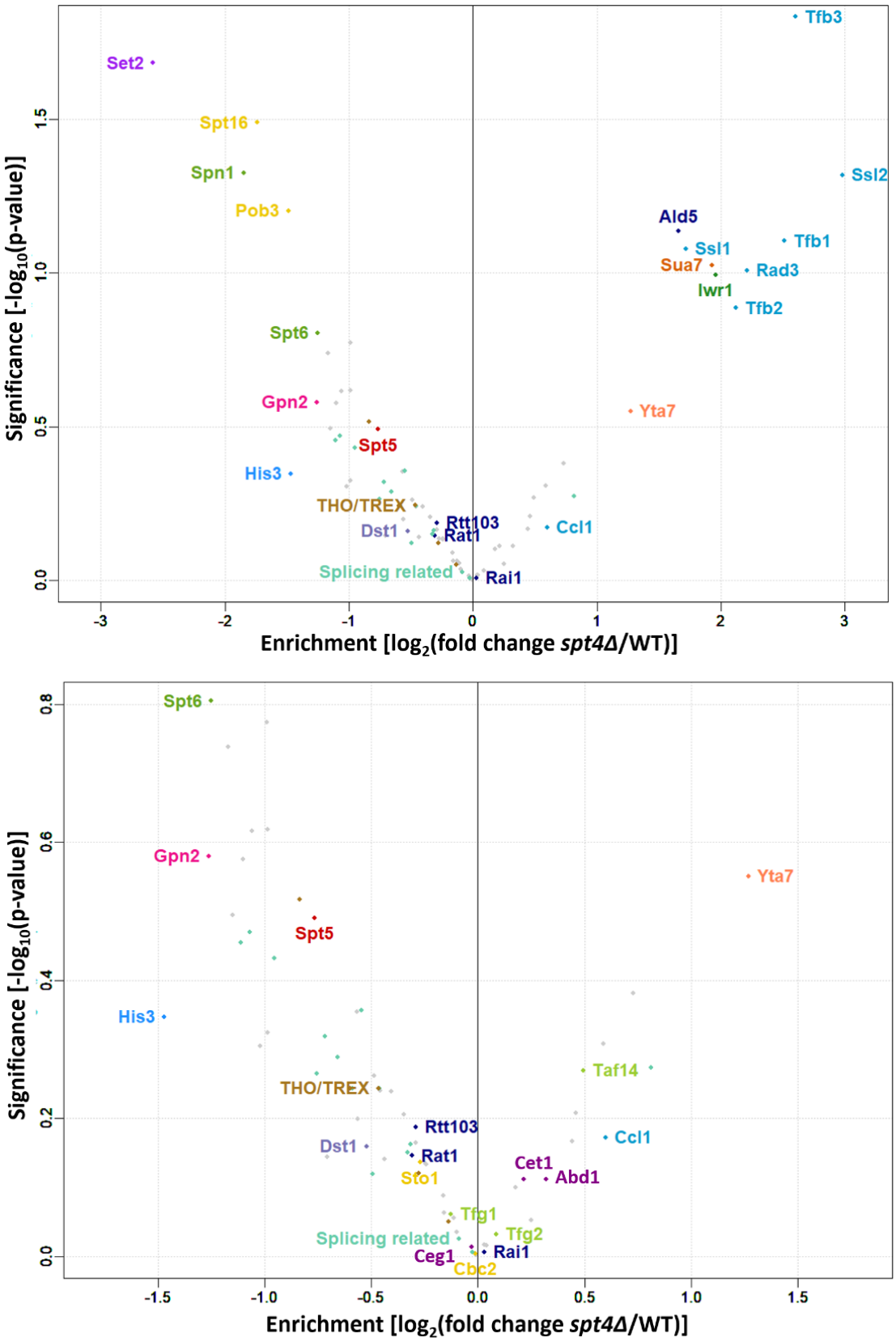


Figure 7.8) Close-up volcano plots

Close up volcano plots at two different scales to show the abundances and significance of TFs on the WT and *spt4Δ* transcription complexes.

7.2.3 Genome-wide localisation of Kin28 and Bur1 in WT and *spt4Δ* cells

In addition to the composition of transcription complexes in WT and *spt4Δ*, the interaction of Kin28 (CDK7) and Bur1 (CDK9) with RNAPII in WT and *spt4Δ* cells were of interest for several reasons. Kin28 phosphorylates Ser5 which is important for the transition of the RNAPII from initiation to elongation (Zaborowska et al. 2016). Moreover, as mentioned above, some subunits of TFIIH was found to be enriched in the *spt4Δ* transcription complex, but Kin28 was not detected in the MS analysis despite being a subunit of TFIIH. Bur1 is another kinase phosphorylating Ser2 of the RNAPII-CTD that associates with the elongating RNAPII over the gene bodies. Although Ser2 can also be phosphorylated by Ctk1 (CDK12) (Bowman and Kelly 2014), here, Bur1 was chosen for further investigation as it is also responsible for the phosphorylation of the Spt5-CTR.

Kin28 and Bur1 (and Ctk1) were not detected in the MS analysis, possibly due to their transient association with RNAPII making them difficult to co-IP with the transcription complexes. Therefore, instead of testing their interaction with RNAPII, ChIP-seq was performed for Kin28 and Bur1 in WT and *spt4Δ* cells, as the crosslinked position and amount of Kin28 and Bur1 could reflect their recruitment to chromatin. Additionally, ChIP-seq was performed for Rpb3 to test if the changes in the distribution of RNAPII could be detected via this technique.

The IPs of Kin28, Bur1, and Rpb3 were checked by western blotting using antibodies against the FLAG epitope. Strong bands at expected sizes for the eluate samples verify the successful pull-down of the protein (**Figure 7.9**).

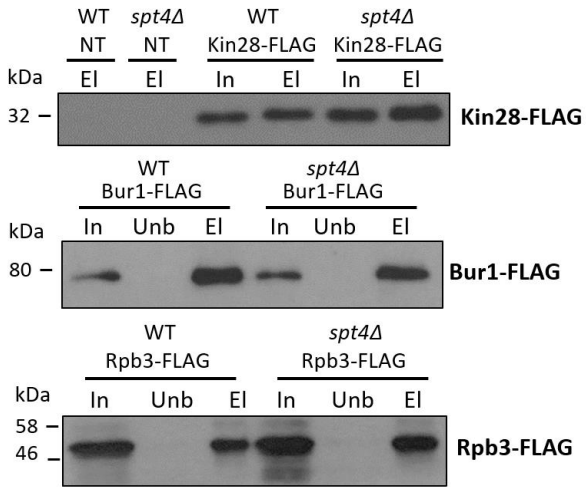


Figure 7.9) Immunoprecipitation of FLAG-tagged Kin28, Bur1, and Rpb3 in WT and *spt4Δ*

IP of FLAG-tagged Kin28, Bur1 and Rpb3. Western blot of Input (In), unbound (Unb), and eluate (El) samples with anti-FLAG (M2) antibody run in 10 % SDS-PAGE. See Appendix 5 for the whole blots.

ChIP-seq was performed in biological duplicate and the data were corrected for input reads as described in the Methods. The reproducibility of the experimental repeats was tested by counting reads in the first 500 nt from the TSS (the region where most of the ChIP-seq signal is detected) and calculating Spearman's correlation coefficient. Correlation plots and IGV tracks of representative genes show that the repeats are reproducible (**Figure 7.10** and **Figure 7.11**).

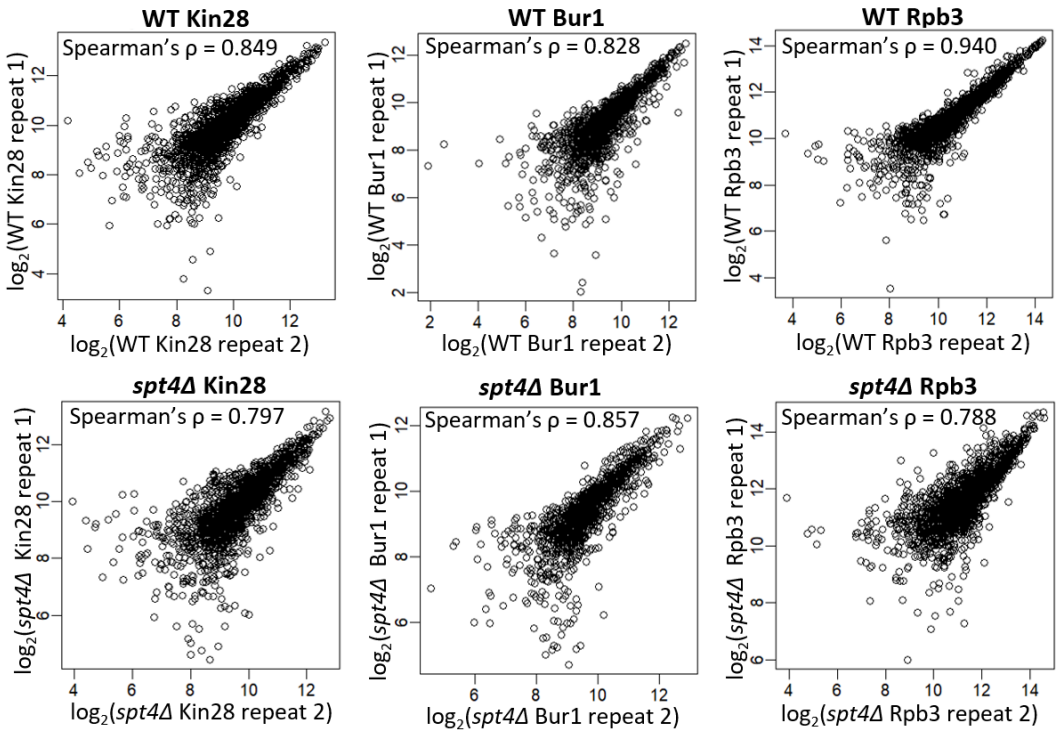


Figure 7.10) Correlations between ChIP-seq repeats

Correlation plots of the two biological repeats of each experiment. Reads are counted over the first 500 nt from the TSS for each gene. Log_2 transformed gene counts are correlated and Spearman's ρ calculated for each pair.

IGV tracks of three representative genes were given as examples of the ChIP-seq on Bur1, Kin28 and Rpb3 in WT or *spt4Δ* cells. In general, IGV tracks show that there are no major differences in the levels of Bur1 and Kin28 in the early stages of transcription between WT and *spt4Δ* cells (**Figure 7.11**). The Rpb3 signal does not seem to change much at *ALD5* locus but shows higher crosslinking at *CCT6* and *BAP2* in *spt4Δ* compared to WT cells. Additionally, the shape of the distribution of the ChIP-seq signals show moderate changes (**Figure 7.11**).

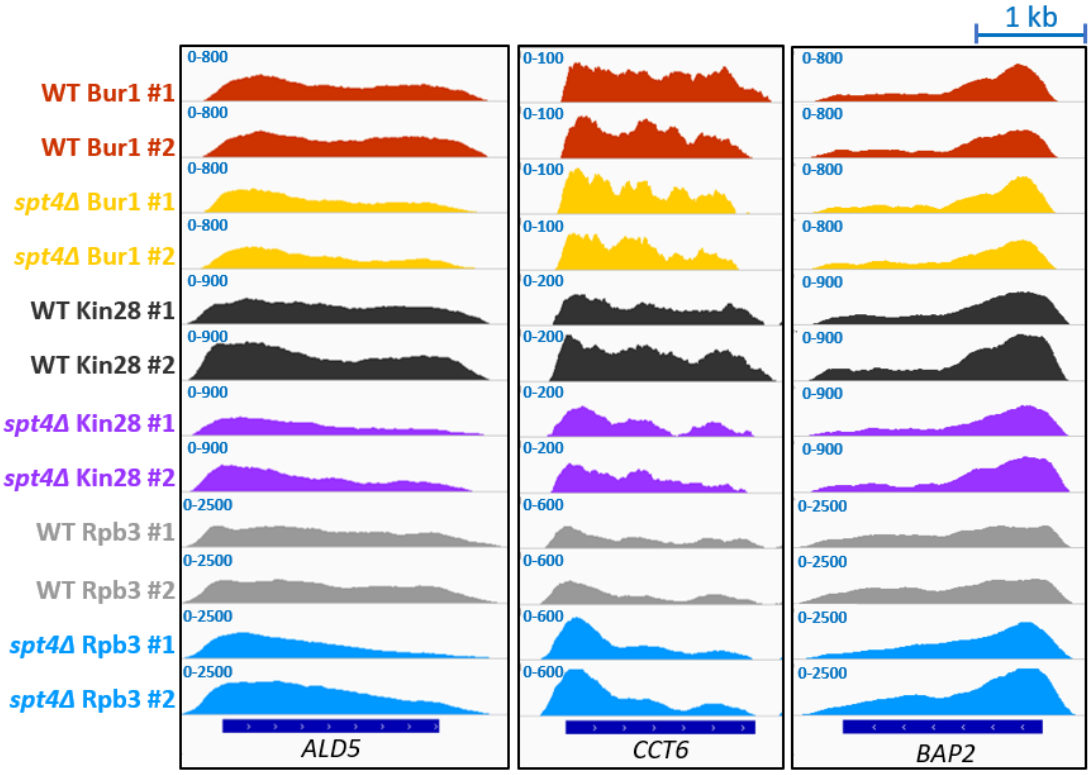


Figure 7.11) Bur1, Kin28, and Rpb3 densities at representative genes

IGV tracks of ChIP-seq reads of three representative genes *ALD5*, *CTC6* and *BAP2* in two biological replicates. The dark blue boxes indicate the transcribed region of the genes (from TSS to PAS).

Next, to examine the genome-wide distributions of Kin28, Bur1, and Rpb3, metagene profiles were plotted for the genes having signal above the background in all conditions (n=1124). In *spt4Δ* compared to WT cells, the metagene plots show that there is a decrease in Kin28 levels although Bur1 levels seems mostly unchanged, with a slight drop in the signal in *spt4Δ* cells over the gene bodies (**Figure 7.12**). Quantification of the ChIP-seq signals were performed by counting the reads in the first 500 nt from the TSS and applying the DEseq2 algorithm in R package (Love et al. 2014). Out of 1124 PCGs, Kin28 and Bur1 levels seem to remain unchanged in most of the genes (94% and 92%, respectively, p-adjusted <0.05, **Figure 7.12**). A small subset of genes is detected to have lower Kin28 and Bur1 levels in *spt4Δ* cells, 59 and 57 genes, respectively. Interestingly, 43 of those genes were common. A gene ontology (GO) enrichment analysis was performed using online GOrilla tool by submitting the 43 common genes (target list) against the initial list of 1124 proteins (background list) (Eden et al. 2009).

Among the 43 genes, 10 of them are found to be related to glycolytic cycle, ADP and pyruvate metabolic processes (**Table 7.1**). Interestingly, the Rpb3 ChIP-seq levels were unchanged in these 10 genes. Moreover, the steady state transcript levels of these genes were not changed in *spt4Δ* cells as assessed by RNA-seq in chapter 5. These observations might indicate that in the absence of Spt4, a negative feedback mechanism is triggered in some metabolic genes, perhaps due to the reduced growth rate, but the levels of Rpb3 on those genes do not seem to be affected and thus, their transcription might not be completely switched off.

As the sensitivity of NET-seq and ChIP-seq are quite different, Rpb3 ChIP-seq was also included to test if the change in the distribution of RNAPII in *spt4Δ* cells could be captured by this method, which is important for drawing conclusions from the Kin28 and Bur1 ChIP-seq data. The ChIP-seq results suggest an increase in the Rpb3 signal in *spt4Δ* cells for a quarter of the genes, as opposed to the most genes having a higher Rpb3 levels by the NET-seq (73%) (**Figure 7.12**). Note that, the technical variability in the ChIP-seq was higher than that of the NET-seq (Spearman's correlation being around 0.8 for ChIP-seq and 0.95 for NET-seq), which could also lower the precision and power of the quantification of the ChIP-seq results. Remarkably, despite the lower resolution in ChIP-seq, the highest Rpb3 (ChIP-seq) signal was observed around 250 nt downstream of the TSS in *spt4Δ* cells, supporting the idea of RNAPII accumulation early in transcription (**Figure 7.12**).

Overall, for the most genes, the cross-linked levels of Rpb3, Kin28 and Bur1 were similar in WT and *spt4Δ* cells. ChIP-seq shows an increased level of Rpb3 in *spt4Δ* cells for a quarter of the genes analysed here, but the levels of Kin28 and Bur1 remain unchanged in these group. This could mean that in this one quarter of genes, there is less Kin28 and Bur1 per RNAPII. Alternatively, if one kinase molecule can phosphorylate more than one RNAPII, the unchanged levels of Kin28 and Bur1 could still accommodate normal association between Kin28-RNAPII as well as Bur1-RNAPII in WT and *spt4Δ* cells.

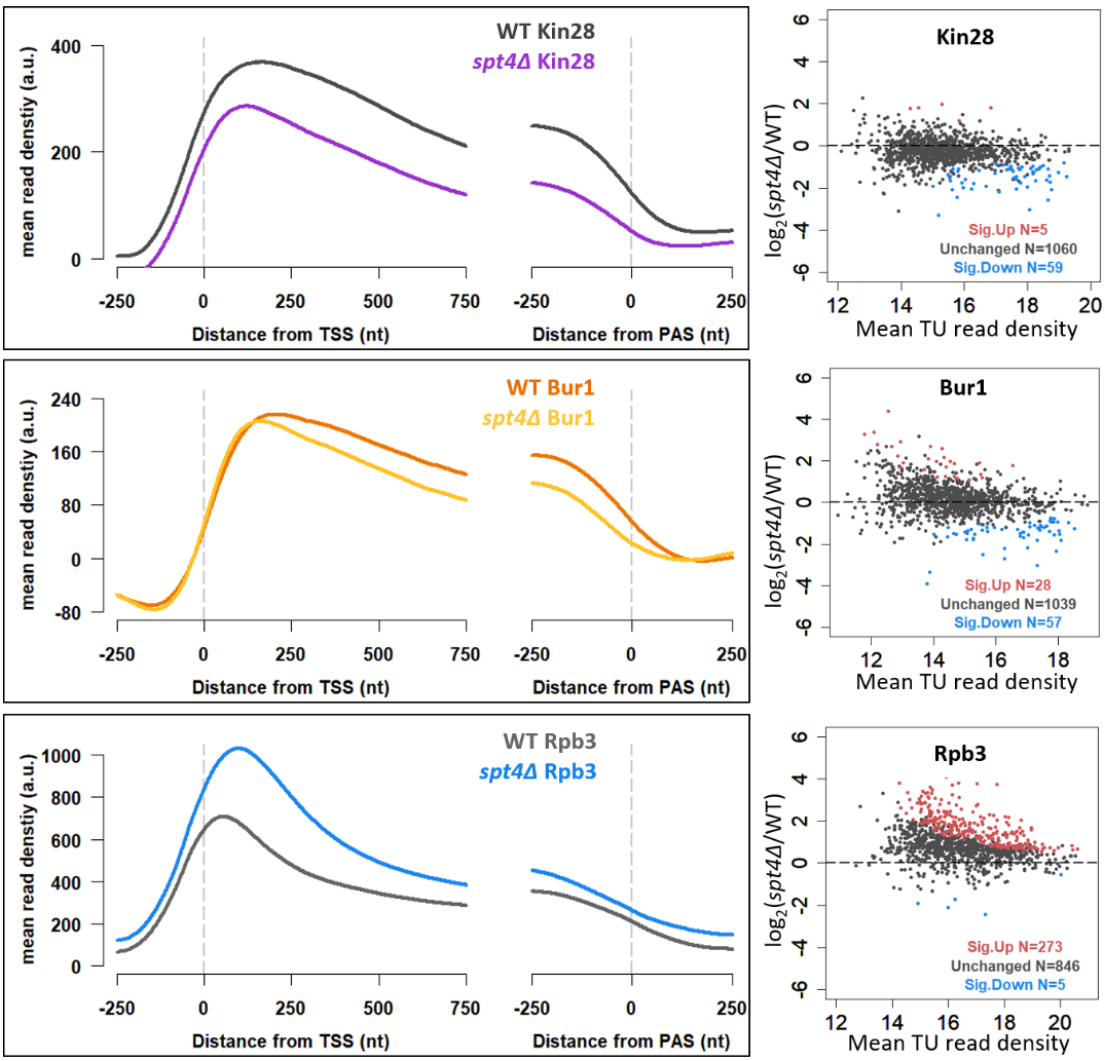


Figure 7.12) Genome-wide positions of Kin28, Bur1, and Rpb3 in WT and *spt4Δ*

Metagene plots of ChIP-seq reads in WT and *spt4Δ* cells are given on the left for the 1142 common genes with signal above the background in Kin28, Bur1, and Rpb3 ChIP-seq. Differential enrichment analyses of Kin28, Bur1, and Rpb3 in WT and *spt4Δ* are given on the right. DEseq2 applied to the read counts over the gene bodies (TSS to TSS+500 nt) for the two replicates of each data. Significantly enriched and depleted genes indicated in red and blue, respectively (p-adjusted <0.05). ChIP-seq are performed in experimental duplicate (see Appendix 6B,C&D for individual plots). The background is removed by subtracting input reads. Reads are normalised to *S.pombe* spike-ins.

Table 7.1) Gene ontology term enrichment analysis for the 43 genes having significantly low Kin28 and Bur1 levels in *spt4Δ* cells.

GO term	description	p-value	FDR	Enrichment	Genes
GO:0046031 GO:0006090	ADP and pyruvate metabolic process	$2x10^{-10}$	<0.001	13.53 (1055,20,39,10)	CDC19 - pyruvate kinase
					ENO1 - phosphopyruvate hydratase
					ENO2 - phosphopyruvate hydratase
					FBA1 - fructose-bisphosphate aldolase
					GPM1 - phosphoglycerate mutase
					PDC1 - indolepyruvate decarboxylase
					PGK1 - phosphoglycerate kinase
					TDH2 - glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
					TDH3 - glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
					TPI1 - triose-phosphate isomerase

Enrichment means that 1055 and 39 genes in the background and target lists, respectively, were considered for the analysis; 20 genes from the background list was associated with the indicated GO terms and 10 of those were enriched in the target list.

7.3 Discussion

In this chapter, it is found that Spt4 is important for the effective association and dissociation of TEFs on RNAPII. In *spt4Δ* cells, the levels of Rpb3 crosslinked to chromatin is observed to be higher in 25% of the genes, while the levels of Kin28 and Bur1 do not seem to change.

The proteomics analysis performed here provides a useful insight into the composition of the transcription complexes in WT and *spt4Δ* cells. The analysis could be improved or complemented in several ways. To begin with, MS is a sensitive technique with high stochasticity and the quantification power of the experiment increases as the number of replicates are increased. To avoid false positive hits, background removal was applied and gave rise to the removal of some TEFs known to interact with RNAPII such as Paf1C. Therefore, the quantitative power and the depth of the analysis could be enhanced either by increasing the number of replicates or using isotope labelling strategies (SILAC) such that the cells could be grown in light and heavy amino acid containing medium in WT and mutant conditions,

respectively, and the protein abundances could be compared based on the labelled peptides. Another point is that the IP of the MS analysis was performed on the Rpb3 subunit, meaning that a mixed population of RNAPII was pull down regardless of its phosphorylation status. To specifically focus on certain stages of transcription, RNAPII could be immunoprecipitated using antibodies against differentially phosphorylated RNAPII-CTD, such as Ser5-P to target RNAPII in early transcription or Ser2-P to pull down more downstream RNAPII. This approach was used previously to define the composition of the WT transcription complex (Harlen *et al.*, 2016). Finally, here, the transcription complexes were IP-ed in the presence of RNase inhibitors, meaning that protein interactions mediated through native RNA were included in the IP. This is a common strategy used in the MS studies performed in yeast (Harlen *et al.*, 2016; Joo *et al.*, 2019). In mammalian cells, similar experiments are usually carried out with the addition of benzonase at the cell lysis steps, which eventually degrades both RNA and DNA (Tufegdžić Vidaković et al. 2020). This is probably because the native RNA is too long to be pulled down in mammalian cells. It would be a good complementary experiment to perform MS experiments with benzonase addition to degrade RNA in the yeast system too. It could possibly lower the background levels and increase the depth of the analysis.

In chapter 4, the genome-wide positions of Spt4 and Spt5, and their associations with RNAPII were assessed at single nucleotide resolution using TEF-seq. As Kin28 and Bur1 do not co-IP with Rpb3 under native conditions, TEF-seq was not an option to map these factors, and the genome-wide positions of Kin28, Bur1, and Rpb3 were assessed by ChIP-seq. Although general observations were possible, the conclusions drawn from the ChIP-seq data were limited due to the high background signal and low resolution. A lower background signal and higher resolution for the crosslinked factors could be obtained using improved ChIP based techniques, such as ChIP-exo or CUT&Tag (Kaya-Okur et al. 2019; Rossi, Lai, and Pugh 2018).

No change in the levels of Kin28 and Bur1 in *spt4Δ* cells might lead to the assumption of no major changes in Ser5-P and Ser2-P of the RNAPII-CTD. However, the positions of Kin28 and Bur1 are not direct indications of Ser5-P and Ser2-P on the RNAPII-CTD, respectively, as these modifications are not limited to Kin28 and Bur1 activities, and the activities of these kinases are not strictly restricted to the phosphorylation of specific residues (Jeronimo et al. 2016). Therefore, in addition to Kin28 and Bur1, genome-wide Ser5-P and Ser2-P profiles could be studied to investigate the impact of Spt4 on the transition of the transcription complexes between different stages of transcription. Indeed, NET-seq could be a powerful technique to carry out for this purpose using antibodies against RNAPII-CTD modifications, as was done previously in mammalian cells (Nojima et al. 2016). Additionally, no change in the levels of Bur1 in *spt4Δ* cells would lead to the conclusion that phosphorylation of Spt5-CTR is not changed. However, this could be further studied using TEF-seq with an antibody specific to phosphorylated (and/or unphosphorylated) version of the Spt5-CTR.

Interestingly, the lack of Spt4 negatively impacts the association of chromatin related TEFs with RNAPII. These TEFs include Set2 and to a lesser extent, Spt6-Spn1 and FACT. The reduction in the abundance of Spt6-Spn1 and FACT (from here described generally as histone chaperones) were not significant. However, the MS analysis reports the protein abundances in a mixed population of RNAPII, so that if there is a specific region on the TUs or a specific subset of genes in which the association of these histone chaperones was differentially influenced, this could only be partially detected in the analysis carried out here. The genome-wide association of these histone chaperones in *spt4Δ* cells should be tested in the future to further address if the RNAPII accumulation in *spt4Δ* could be explained by the differential association of these histone chaperones.

The low levels of Set2 on the *spt4Δ* transcription complex would not explain the accumulation of RNAPII observed *spt4Δ* cells, as previous work showed using NET-seq that even the

complete deletion of *SET2* does not change the distribution of RNAPII (Churchman and Weissman 2011). However, the significantly reduced abundance of Set2 in the *spt4Δ* transcription complex might indicate a problem early in transcription, as Set2 joins transcription after completion of several upstream events such as the association of Spt6, Bur1, and Paf1C with RNAPII and Ser2-P of the RNAPII-CTD (Mcdaniel and Strahl 2017). This brings about the question as to whether these upstream events are impaired in *spt4Δ* cells. The ChIP-seq analysis suggests no major difference in the recruitment of Bur1. Moreover, indirect evidence from the MS data suggest no defect in the Ser2-P in *spt4Δ* cells, as there was no difference in the abundances of the termination factors, which are also recruited by Ser2-P. On the other hand, the reduced abundances of the histone chaperones might impact the recruitment of Set2 to the transcription complex. Future experiments should be performed first to validate the reduced abundance of Set2 on the *spt4Δ* transcription complex, and second to further address the relationship between the Spt4 and Set2 recruitment.

Besides the TFs related to chromatin regulation, Sua7 was more abundant in *spt4Δ* transcription complex compared to that of WT, although this was not a significant increase. It is important to recall that in chapter 5, it was shown that the levels and positions of Sua7 crosslinked to chromatin were not different in WT and *spt4Δ.* Note, however, ChIP-seq reflects the position of Sua7 on DNA, even in the absence of RNAPII, while the MS analysis only reports the interaction of Sua7 with the transcription complex. Therefore, the increased abundance of Sua7 in the *spt4Δ* transcription complex, while its amount on the chromatin is similar to WT levels, might indicate that the association between the Sua7 and RNAPII is somehow increased in the absence of Spt4. Alternatively, this result might be due to a technical variability in the MS, as the increase in the Sua7 amount in the *spt4Δ* transcription complex was not statistically significant.

Interestingly, the abundances of some TFs related to RNA processing were not differentially enriched in either WT or in *spt4Δ* transcription complex. These TFs include capping enzymes, cap binding proteins, splicing factors, components of the THO/TREX complex, and termination factors. This observation suggests no major defects in Ser5-P and Ser2-P on the RNAPII-CTD, as the recruitment of the capping enzymes and splicing factors are dependent on the Ser5-P (Harlen *et al.*, 2016; Zaborowska, Egloff and Murphy, 2016) and the recruitment of the termination factors are dependent on Ser2-P of the RNAPII-CTD (Nemec et al. 2017). Consequently, the changes in the composition of the transcription complex in the *spt4Δ* strain would not be explained by problems in the phosphorylation status of the RNAPII-CTD.

Given that the abundance of TFIIH increased in the *spt4Δ* transcription complex, the MS results presented here raises a possibility of an activated DNA repair mechanism in the mutant cells. Indeed, the notion of a possible role of Spt4 in the regulation of transcription coupled repair is not new (Jansen et al. 2000). Previous studies showed the UV-induced DNA lesions in *RAD26* deletion mutants are more rapidly and efficiently repaired in *rad26Δ spt4Δ* double mutants (Duan et al. 2020; Jansen et al. 2000). Rad26 (Cockayne syndrome group B, CSB in mammals) is an ATPase-dependent translocase that is required for transcription coupled nucleotide excision repair (TC-NER). How Rad26 is recruited to RNAPII is not clear, however it is known that Rad26 interacts with RNAPII upon DNA damage and somehow induces the recruitment of TFIIH (Lans et al. 2019). Intriguingly, a recent structural study shows that Rad26 and Spt4 bind to the same region of RNAPII, hence their interaction with RNAPII is mutually exclusive (Xu et al. 2017). It is, therefore, tempting to speculate that one function of Rad26 is replacing Spt4, or Spt4/5 as a complex, which in turn, could activate TC-NER by recruiting TFIIH (Gregersen and Svejstrup 2018). The increased abundance of TFIIH in the *spt4Δ* transcription complex shown here, therefore, could be an explanation for how *rad26Δ spt4Δ* cells could induce TC-NER even in the absence of Rad26. Alternatively, if Rad26 takes a direct role in the recruitment of TFIIH, the increased TFIIH-RNAPII association might be a result of the deletion

of Spt4, allowing Rad26-RNAPII interaction (van der Weegen et al. 2020). However, this would not explain the restoration of TC-NER in the *rad26Δ spt4Δ* double mutant cells.

It should be further verified whether TC-NER is activated in *spt4Δ* cells. This could be done by testing the phosphorylation status of some proteins known to be involved in the DNA repair pathways and phosphorylated upon DNA damage, such as Mec1 and Rad26 (Taschner et al. 2010). Indeed, the phosphorylation of Rad26 was aimed to be tested using western blotting, but the results were not conclusive (Appendix 9). The investigation into the DNA damage response and repair in *spt4Δ* cells, therefore, could be addressed with more sensitive and alternative techniques, which is beyond the scope of this thesis.

Although not verified, the possibility of activated TC-NER in *spt4Δ* cells, brings about two alternative explanations. First, there might be actual DNA damage in *spt4Δ* cells. As the NGN domain of Spt5 interacts with non-template DNA (Crickard et al. 2016), the Spt4/5 complex might be suppressing DNA damage, perhaps by reducing the formation of R-loops during transcription. Whether this could be a reason for the RNAPII accumulation observed early in transcription in the absence of Spt4 remains unknown. Second, activated TC-NER in *spt4Δ* cells might not be a result of DNA damage, but might be due to RNAPII accumulation as TC-NER is triggered by paused RNAPII even though the mechanism is not clear (Gregersen and Svejstrup 2018).

8 Nucleosome positions in *spt4Δ* cells

8.1 Overview

Interestingly, once Spt4 and Spt5 were discovered, *spt4* and *spt5* mutants were shown to have similar phenotypes to *spt6* and *spt16* mutants, which lack TEFs acting as histone chaperones (Hartzog et al. 1998; Sims, Belotserkovskaya, and Reinberg 2004). Despite the early observations categorising Spt4 and Spt5 in the histone group genes, to our knowledge, the role of Spt4 and Spt5 in chromatin architecture has not been addressed by a genome-wide study *in vivo.* Therefore, the aim of this chapter is to investigate the role of Spt4 in nucleosome positioning.

The nucleosome positions in WT and *spt4Δ* was examined by Micrococcal nuclease (MNase) digestion followed by DNA sequencing (MNase-seq). MNase preferentially digests internucleosomal regions and leaves nucleosome bound DNA intact. The isolated and sequenced DNA then shows the nucleosome positions across the genome (Chereji and Clark 2018) (**Figure 8.1**).

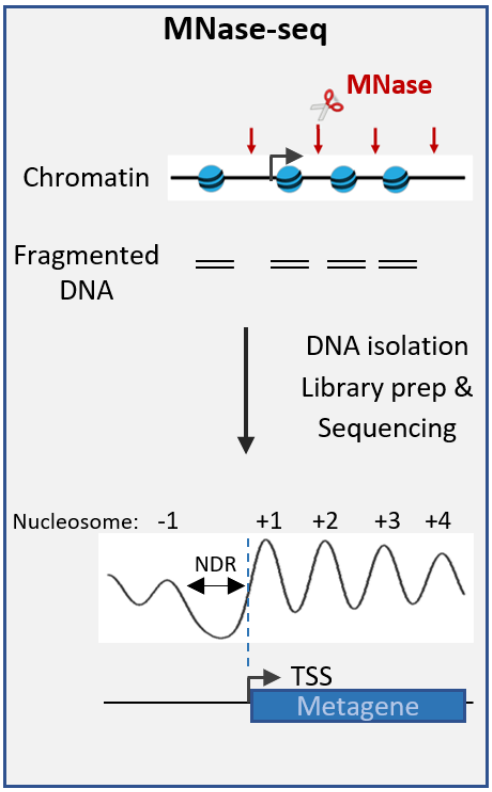


Figure 8.1) MNase-seq to map nucleosome protected DNA

Micrococcal nuclease digestion followed by sequencing (MNase-seq) detects genome-wide nucleosome protected DNA.

8.2 Results

8.2.1 Nucleosome positions in WT and *spt4Δ* are detected by MNase-seq

At appropriate titrations of MNase digestion, DNA can be detected in an agarose gel reflecting mono-nucleosome (~150 bp), di-nucleosome (~300 bp) and multi-nucleosome protected bands. The generally accepted criteria for MNase digestion is to select a concentration with a bright mono-nucleosome band to ensure sufficient digestion for high resolution but with faint di- and tri-nucleosome bands to indicate and avoid over-digestion (Gutiérrez et al. 2017). Based on these criteria, different dilutions of MNase from 320 to 10 U in WT and *spt4Δ* strains were tested. 80U produced the optimal digestion pattern (**Figure 8.2**). The mono-nucleosomal bands are then gel extracted, sequenced and aligned to the yeast genome.

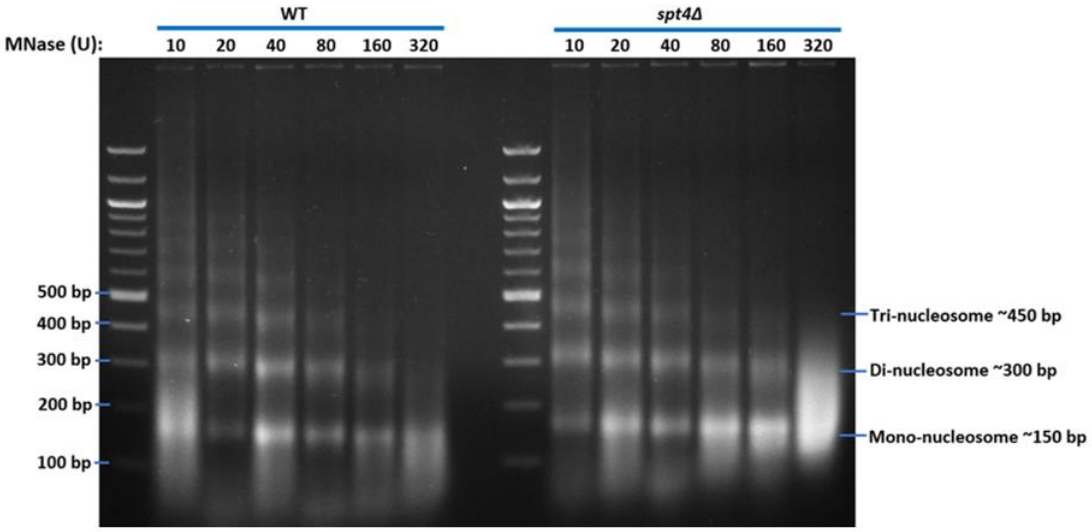


Figure 8.2) Nucleosome protected DNA is detectable after MNase digestion.

Chromatin was digested with different amounts of MNase ranging between 10-320 U. Digested samples were separated by 1.5% agarose-TBE gel electrophoresis. The lowest concentration of MNase (10 U) presents under-digested chromatin with detectable bands longer than 450 bp. The highest concentration of MNase (320 U) presents over-digestion with undetectable bands longer than 300 bp.

An analysis of MNase-seq reads was performed using a peak-calling software DANPOS2 (Chen et al. 2013). Estimates for the nucleosome centres obtained from this software were then used to visualise and compare nucleosomal peaks across the genome. In a typical MNase-seq profile, NDRs are detected at promoters, and nucleosomes are regularly arrayed in gene bodies relative to the TSS. IGV tracks of single-genes show expected MNase-seq profiles and allow a visualise assessment of the reproducibility of the experiments (**Figure 8.3**).

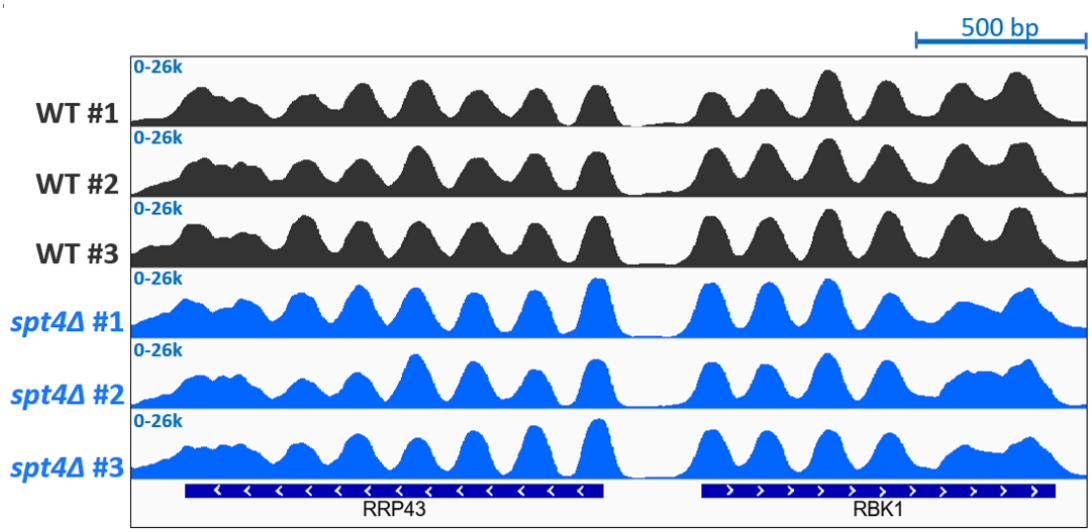


Figure 8.3) MNase-seq is reproducible

IGV tracks of MNase-seq reads of two representative genes transcribed from the negative (*RRP43*) and positive strand (*RBK1*) in WT and *spt4Δ* in 3 biological replicates. The dark blue boxes indicate the transcribed region of the genes (from TSS to PAS), while the white arrows indicate transcription direction.

Next, WT and *spt4Δ* nucleosome positions of PCGs were plotted as heatmaps. The three replicates of each experiment were combined as they were reproducible. 5191 PCGs (> 600 nt) were taken and ranked based on their +1 nucleosome position in WT. Both in WT and *spt4Δ* cells, nucleosomes are arrayed over the gene bodies and NDRs are distinguishable immediately upstream of the TSS, with a small group of genes having shorter NDRs than the rest of the genome as previously shown (Chereji and Clark 2018) (**Figure 8.4**). Overall, these data show that the MNase-seq created the expected digestion pattern across the genome.

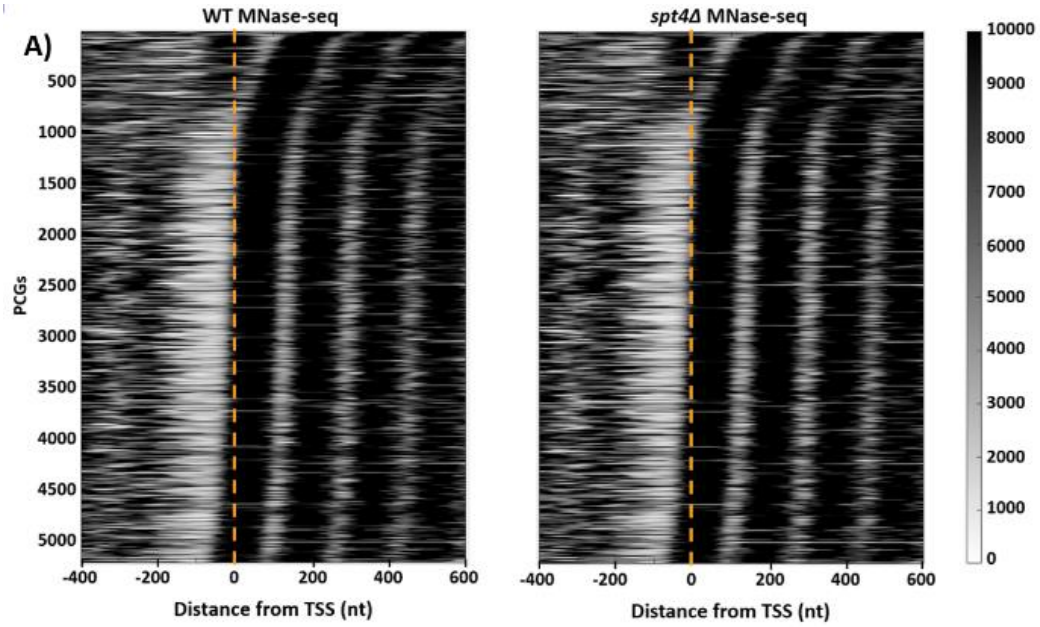


Figure 8.4) MNase-seq detects nucleosome protected DNA

Heatmaps of MNase-seq reads in WT (left) and *spt4Δ* (right). PCGs ordered based on the position of +1 nucleosome in WT. Dashed line in orange indicates the TSS.

8.2.2 Nucleosome positions are altered in *spt4Δ* compared to WT

As it is difficult to judge overall changes in the nucleosome positions from the heatmaps, next, metagene profiles were plotted relative to the TSS. Interestingly, nucleosome positions are shifted to the right (towards the 3'-end) in *spt4Δ* (**Figure 8.5**). This shift was clearer at downstream nucleosomes, especially starting from the +3 nucleosome (**Figure 8.5**).

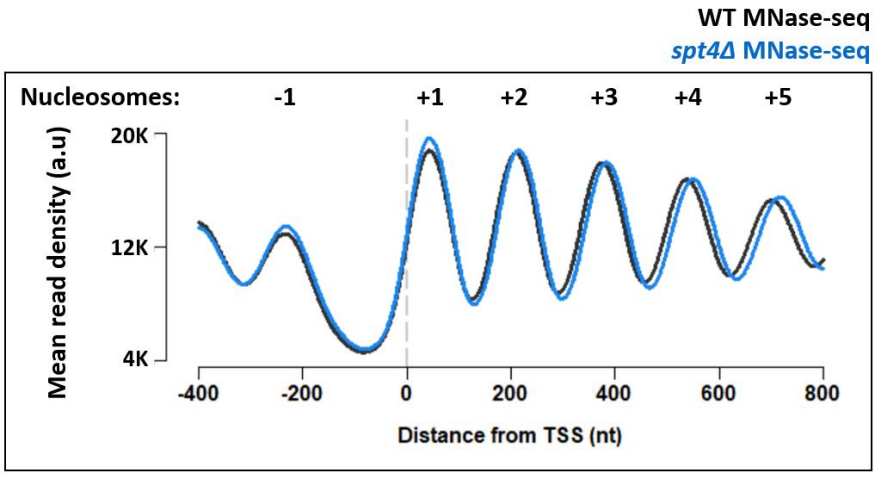


Figure 8.5) Nucleosome positions change in *spt4Δ* compared to WT.

Metagene plots of MNase-seq reads in WT (black) and *spt4Δ* (blue). Nucleosome numbers are indicated above each peak. The dashed line in grey indicates the TSS.

MNase-seq is performed in triplicate (see Appendix 10 for individual metagene plots). Reads are normalised to total reads.

8.2.3 The position of the +1 nucleosome and NDR length do not change in *spt4Δ*

More detailed analysis was performed to explain the downstream shift in nucleosome positions in *spt4Δ* cells. To this end, the position of the +1 nucleosomes, NDR length, and nucleosome spacing were assessed in WT and *spt4Δ* cells.

5579 PCGs were taken to analyse the position of the +1 nucleosome. Genes with too few or too many MNase peaks (< 2 and > 5 peaks) within the first 600 nt from the TSS (+1 to 600 nt) were discarded to avoid genes with poorly phased nucleosomes. Genes shorter than 600 nt were also discarded. 5053 PCGs were left for the analysis. The position of each +1 nucleosome was defined as the distance between the nucleosome centre and TSS. Analysis shows that the median positions of the +1 nucleosome in WT (49 nt) and *spt4Δ* (50 nt) cells are not significantly different (p-value=0.189, Student's t-test, paired, two-tailed) (**Figure 8.6B**).

Next, NDR lengths in WT and *spt4Δ* cells were compared. NDR length was defined as the distance between the centres of the +1 nucleosome and -1 nucleosome. Analysis shows that NDR lengths in WT (209 nt) and *spt4Δ* (208 nt) cells are not significantly different (p-value=0.604, Wilcoxon rank-sum test, paired, two-tailed). Taken together, the position of the +1 nucleosomes and the NDR lengths failed to explain the downstream shift in nucleosome positions in *spt4Δ.*

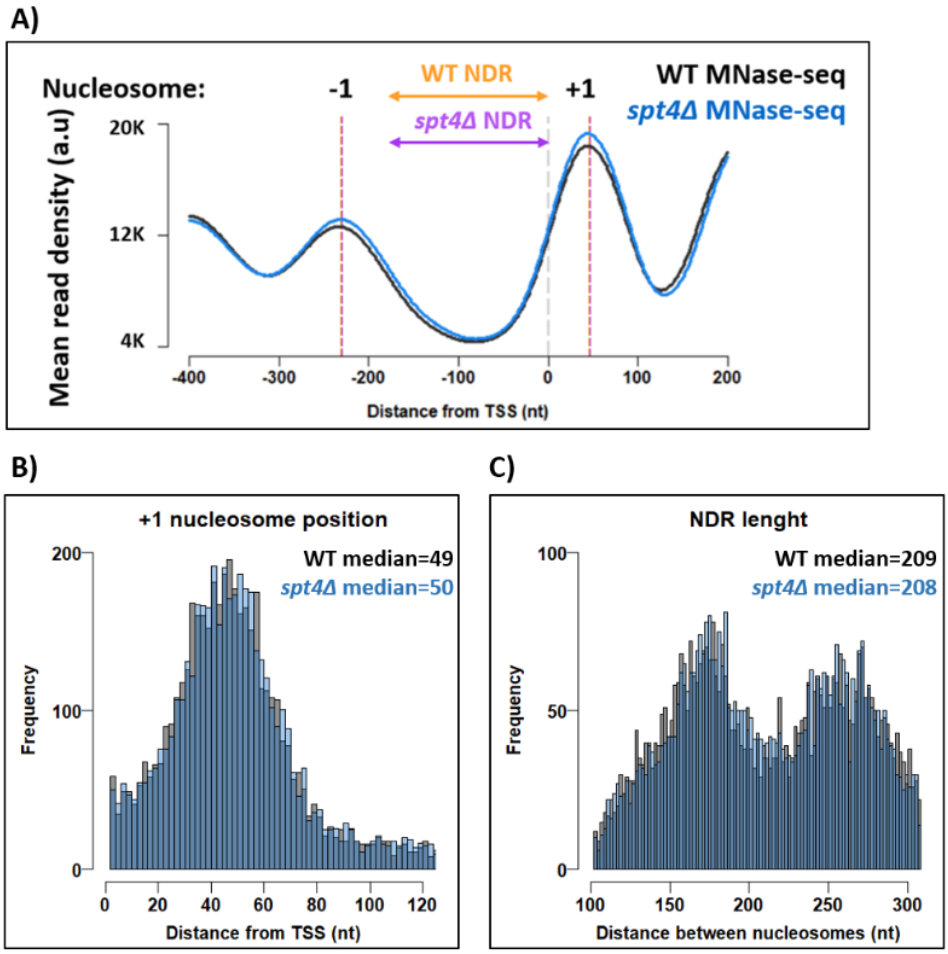


Figure 8.6) Position of +1 nucleosomes and NDR length are unaltered in WT and *spt4Δ*

A) (Close-up of figure 4.8) Metagene plots of MNase-seq reads in WT (black) and *spt4Δ* (blue). Nucleosome numbers are indicated above each peak. Dashed lines indicate centre of respective nucleosomes in WT (orange) and *spt4Δ* (purple).

B) Histograms of the distance of +1 nucleosomes relative to the TSS in WT and *spt4Δ*.

The difference of the +1 nucleosome positions are not significant with p-value = 0.189 (Student's t-test, paired, two-tailed).

C) Histograms of NDR lengths in WT and *spt4Δ*. NDR length is calculated by subtracting the peak position of +1 nucleosome from -1 nucleosome for each gene in WT and *spt4Δ*.

The difference of the NDR lengths are not significant with p-value = 0.604 (Wilcoxon sum-rank test, paired, two-tailed).

8.2.4 Nucleosome spacing is increased in *spt4Δ* compared to WT

Next, the spacing of nucleosomes were investigated in WT and *spt4Δ*. Nucleosome spacing was defined as the distance between the centres of adjacent nucleosomes. This analysis was limited up to the +4 nucleosome, as nucleosome phasing is less well defined for the nucleosomes further away from the TSS, as also described by others (Clapier et al. 2017).

The analysis revealed that in WT cells, the median and mean nucleosome spacing between the +1 and +2 nucleosomes are 165 and 166.1 nt, respectively, agreeing with the previously published MNase-seq results (Ocampo et al. 2019) (**Figure 8.7B**). The mean nucleosome spacing between the +2 and +3 is similar to, and between the +3 and +4 nucleosomes are slightly shorter than, the mean spacing between the +1 and +2 nucleosomes. When compared to WT, nucleosome spacing between the +1 and +2 nucleosomes in *spt4Δ* cells is significantly increased by 4 nt with the median and mean being 169 and 170.1 nt, respectively (p-value $<2.10^{-16}$, Student's t-test, paired, two-tailed). The increased spacing is also observed for the following nucleosome pairs, but to a lesser extent, although the difference was still significant (p-value $<2.10^{-16}$, Student's t-test, paired, two-tailed) (**Figure 8.7C&D**). Taken together, the data support that the downstream shift of nucleosomes in *spt4Δ* is explained by the increased nucleosome spacing and, importantly, the shift is independent from the position of the +1 nucleosome and NDR length.

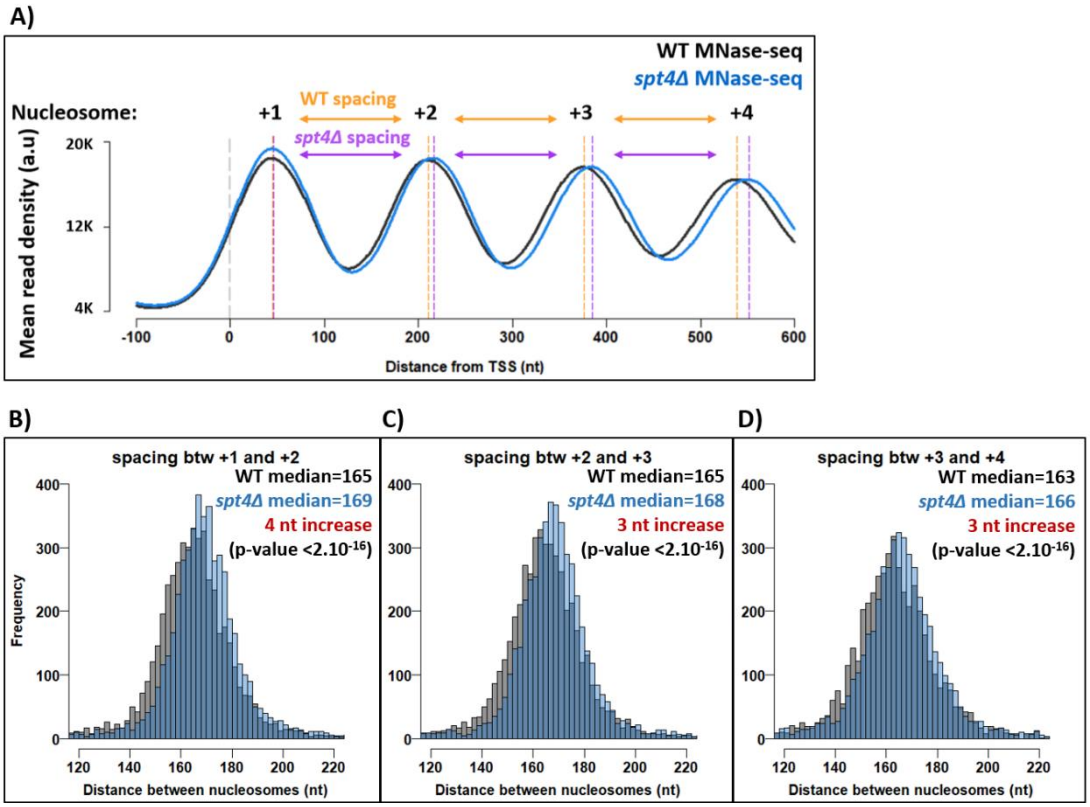


Figure 8.7) Distance between nucleosomes is increased in *spt4Δ* compared to WT.

A) (Close-up of figure 4.8) Metagene plots of MNase-seq reads in WT (black) and *spt4Δ* (blue). Nucleosome numbers are indicated above each peak. Dashed lines indicate centre of respective nucleosomes in WT (orange) and *spt4Δ* (purple).

Histograms of spacing between nucleosomes **B)** +1 and +2 **C)** +2 and +3 **D)** +3 and +4 in WT and *spt4Δ*. Spacing is calculated by subtracting the peak position of n+1 nucleosome from n nucleosome for each gene in WT and *spt4Δ*.

The difference in spacing between the nucleosome pairs presented here are significant p-value < 2.10^{-16} for each pair (Wilcoxon signed-rank test, paired, two-tailed).

8.3 Discussion

In this chapter, it was shown that Spt4 is required for normal nucleosome positioning. In *spt4Δ* cells, the distance between nucleosomes is increased compared to that of WT, while the NDR lengths and the +1 nucleosome positions remain the same.

MNase-seq was used to detect nucleosome positions, and there was no immunoprecipitation step in the protocol. Consequently, the conclusion is drawn based on the assumption that the DNA digestion pattern reflects the positions of nucleosomes. To make sure the DNA

fragments directly correspond to nucleosome positions, experiments could be performed by immunoprecipitating a certain histone, such as H3, following MNase-seq (MNase-ChIP-seq) (Lai and Pugh 2017). Additionally, using MNase-seq, the levels of nucleosome occupancy could also be addressed (Chereji, Bryson, and Henikoff 2019). The MNase-seq data shown here was normalised to sequencing depth, and no reliable quantification of global nucleosome levels between conditions was possible. One other improvement to this MNase-seq dataset, therefore, would have been including spike-ins to quantify the levels of the nucleosomes and collect additional information about the impact of Spt4 on the nucleosome occupancy. Notably, although no spike-ins were used here, if there was a differential nucleosome occupancy at specific positions rather than a global change, it could be detected relative to the rest of the genome. No such differences were observed.

Nucleosome positioning is tightly regulated in yeast mainly, but not only, by chromatin remodellers and TFs. After replication, three consecutive events define the nucleosome positions: NDR formation, positioning of the +1 nucleosome, and alignment of downstream nucleosomes. The first two steps of the nucleosome positioning are coordinated by DNA sequence features around promoters as well as RSC and Ino80 chromatin remodellers, respectively (Baldi, Korber, and Becker 2020). Given that NDR length and the position of the +1 nucleosome is not changed in *spt4Δ* cells, Spt4 is unlikely to play a part in the initial positioning of the -1 and +1 nucleosomes.

Alignment of the downstream nucleosomes after the +1 nucleosome is regulated by Isw1, Chd1, Isw2 remodellers (known as spacing enzymes). The main function of the spacing enzymes seems to push nucleosomes apart since deletion mutations (and combination of deletion mutations) in these enzymes tend to reduce nucleosome spacing both *in vitro* and *in vivo* (Ocampo et al. 2019). Consequently, the increased nucleosome spacing in *spt4Δ* cells cannot be explained by a problem in the recruitment or functioning of the spacing enzymes.

Another factor influencing the nucleosome positioning is transcription. The movement of RNAPII is destructive for the chromatin structure, as nucleosomes are needed to be partially or fully disassembled to allow transcription, and then re-assembled after the RNAPII passage to maintain the chromatin compaction. The increase in nucleosome spacing in *spt4Δ* cells, therefore, might be an indication of Spt4 directly or indirectly regulating histone/nucleosome dynamics during transcription. This possibility will be further discussed in chapter 9.

9 The interplay between the position and distribution of nucleosomes, Spt4 and RNAPII in WT or *spt4Δ* cells

9.1 Overview

Previous chapters provided several observations about Spt4 and Spt5. First, these factors oscillate on and off RNAPII (chapter 4). Second, in the absence of Spt4, RNAPII accumulates over the gene bodies in early transcription (chapter 5), and third, nucleosome spacing is increased in *spt4Δ* compared to WT cells (chapter 8). In this chapter, the data are evaluated together and the cross correlation between the nucleosomal architecture and the position of Spt4, Spt5, and RNAPII are investigated. The aims of this chapter are 1) to examine the correlation between the nucleosome positions and the position of Spt4/5, 2) to examine the correlation between the nucleosome positions and the changes in the RNAPII distribution in the absence of Spt4, and 3) to examine the correlation between the change in nucleosome spacing in *spt4Δ* and transcriptional activity or the amount of RNAPII accumulation in *spt4Δ* cells.

These examinations were made by analysing the TEF/NET/MNase-seq metagene profiles with respect to nucleosome positions in WT and *spt4Δ* conditions and testing the correlations between the certain nucleosome positions and amount of the NET-TEF-seq reads.

9.2 Results

9.2.1 Spt4 and Spt5 oscillation profiles are related to nucleosome positions

The relative affinities of Spt4/5 on RNAPII periodically change and their periodicity resembles the frequency of nucleosome phasing (**Figure 4.6**). To test whether nucleosome positions correlate with the phasing patterns of Spt4 and Spt5, the NET-seq normalised TEF-seq profiles and nucleosome positions were re-plotted as metagene plots, relative to the centre of the +1

nucleosome. For the metagene analysis, PCGs (> 600 nt) with positive TEF and NET-seq values (as described in section 4.2.3) and with well-defined nucleosome positions (as described in section 8.2.3) were taken (n=4417 PCGs in total). In **Figure 9.1** nucleosome positions (WT MNase-seq) relative to the +1 dyad are demonstrated. The dashed lines (dyad centres), highlighted background (nucleosomal DNA +/- 70 nt from the dyads), and graphical representation indicate nucleosome positions. This representation will be used in the next sections to mark the nucleosome positions and display the position of Spt4, Spt5 and Rpb3 relative to the +1 nucleosome dyad.

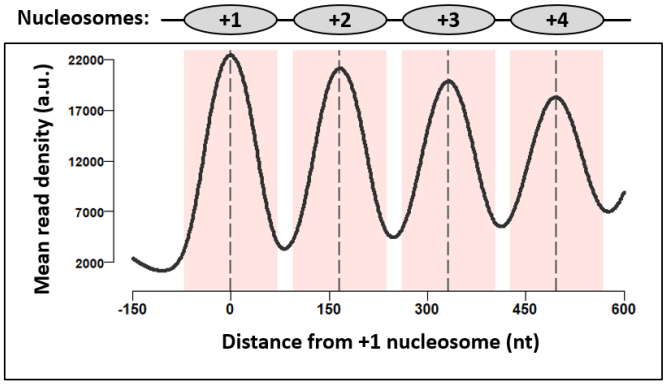


Figure 9.1) Nucleosome positions in WT

MNase-seq data in WT were plotted relative to the +1 nucleosome dyad. Dashed lines (black) through the peaks indicate the centres of the nucleosomes. The nucleosomal DNA (+/- 70 nt around the nucleosome centres) is highlighted in light pink. The position of nucleosomes graphically shown above the plot.

Once plotted relative to the +1 nucleosome, the NET-seq normalised TEF-seq metagene plots show that the oscillation pattern of the Spt4 and Spt5 on RNAPII is offset with respect to the nucleosome phasing (**Figure 9.2**). To get a more detailed view, the metagene profiles were plotted around the +1, +2, +3 and +4 nucleosome dyads at a higher resolution, separately. Overall, the Spt4 and Spt5 occupancies on RNAPII appear to be relatively lower around 30-40 nt upstream of the nucleosome dyads and higher around 30-40 nt downstream of the nucleosome dyads (**Figure 9.3**). The significance of this pattern was tested by counting the factor/RNAPII values from a window upstream of the nucleosome centres (taken as -60 to -

10 nt from each nucleosome dyad) and from a window downstream (taken as +10 to +60 nt from each nucleosome dyad) for each gene. The upstream and downstream factor occupancies (on RNAPII) were then compared using Student's t-test (paired, one-tailed). The analysis verifies that the association of Spt4 and Spt5 with RNAPII is significantly different upstream and downstream of the +1, +2 and +3 nucleosome centres across the genome (**Figure 9.3**). Although the factors also seem to oscillate around the +4 nucleosome, the differential occupancy of Spt4 or Spt5 on RNAPII before and after the transcription through the +4 nucleosome dyad was not statistically significant (**Figure 9.3**). This is probably due to less well-defined nucleosome occupancy and/or lower TEF-seq reads towards the 3'end of the genes.

In short, the data suggest that the interaction of Spt4/5 with RNAPII periodically changes depending on the position of the transcription complex relative to nucleosome positions.

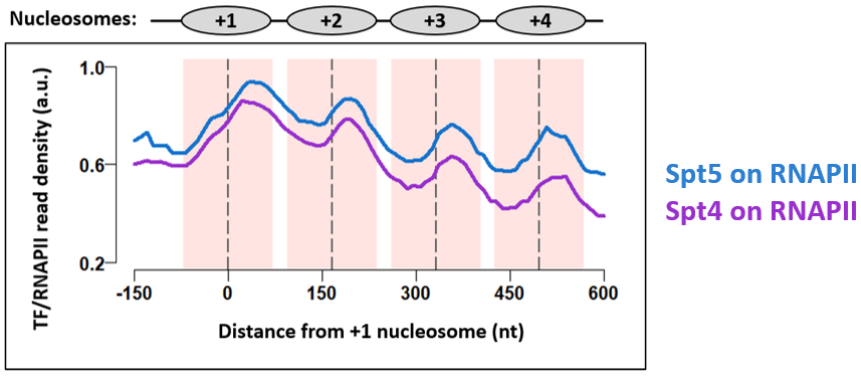


Figure 9.2) Spt4 and Spt5 occupancy on RNAPII relative to the nucleosome positions

Metagene plots of Spt4 on RNAPII (purple) and Spt5 on RNAPII (dark blue) reads. Data from **Figure 4.6** were re-plotted relative to the +1 nucleosome at the top panel. Dashed lines (black) through the peaks indicate the centres of the nucleosomes and the nucleosomal DNA (+/- 70 nt around the centre) is highlighted in light pink. Position of nucleosomes graphically shown above the metagene plot.

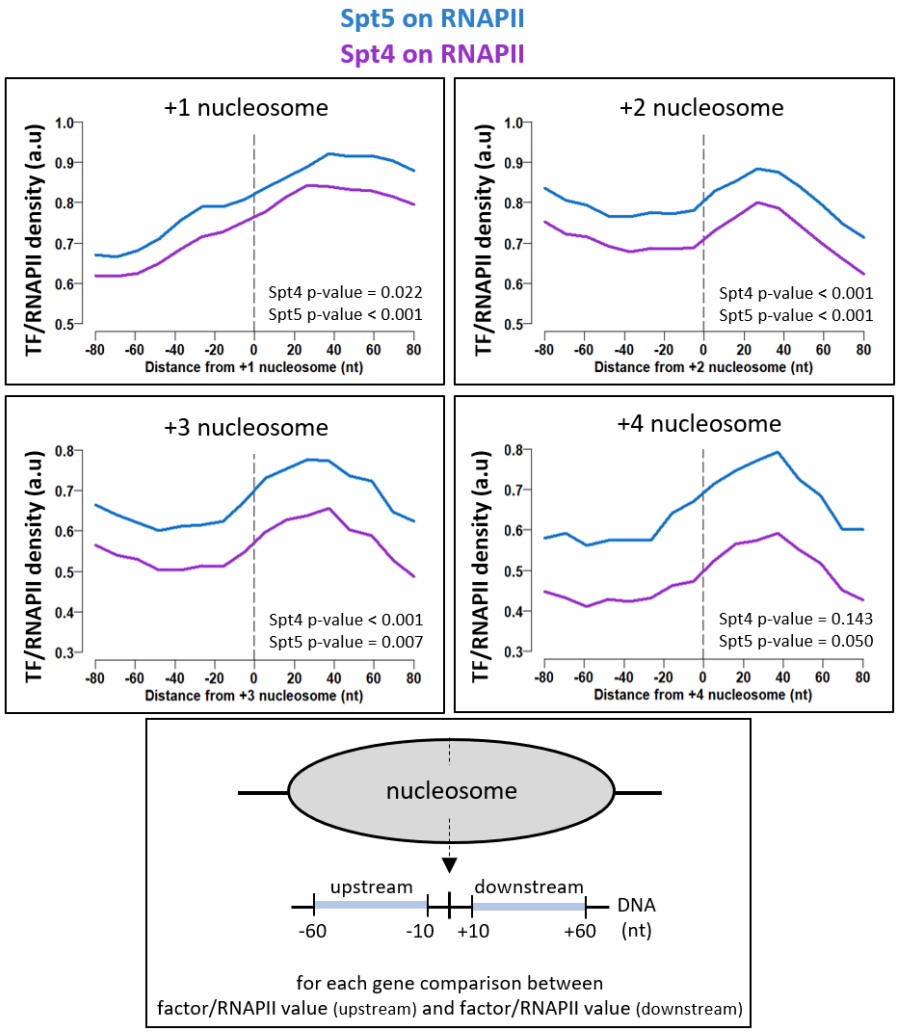


Figure 9.3) Spt4 and Spt5 oscillate on RNAPII according to nucleosome positions

Spt4 and Spt5 occupancies on RNAPII are shown around individual nucleosome dyads +1,+2,+3, and +4. Significance of the change in the factor occupancies around the nucleosomes were tested by Student's t-test (one tailed, paired) by comparing the factor/RNAPII values from upstream of the dyad (-60 to -10 nt from the dyad) and downstream of the dyad (+10 to +60 nt from the dyad) for each gene.

9.2.2 Altered RNAPII distributions in *spt4Δ* and Spt4 anchor away cells are related to nucleosome positions

Next, the NET-seq data were revisited and why RNAPII particularly accumulates 200 nt from the TSS in *spt4Δ* and Spt4-AA cells was addressed considering nucleosome positions. One possible explanation for the RNAPII accumulation would be physical barriers created by nucleosomes due to histone-DNA interactions that are required to be overcome during

transcription. To test this possibility, the distribution of RNAPII was compared with nucleosome positions.

4417 PGCs were taken for the analysis as explained above. *spt4Δ* and Spt4 AA NET-seq metagene profiles and their controls were re-plotted relative to the +1 nucleosome position. The NET-seq profiles were plotted in two ways using different normalisation methods to demonstrate different aspects of the data. First, the spike-in normalised NET-seq data were plotted to point out the extent of the RNAPII accumulation over the genes. Second, the NET-seq reads were normalised to the mean and standard deviation (sd) of each gene, so that each gene contributes similarly to the metagene plots. In this way, the shape of the distribution of RNAPII could be seen more clearly regardless of the expression level differences between the genes.

In WT and the DMSO treated Spt4-FRB cells, the spike-in normalised metagene plots show that the distribution of RNAPII seems to correlate with nucleosome positions even under normal conditions. Especially around the +2 nucleosome, the RNAPII signals in WT, and DMSO treated Spt4-FRB cells have peaks upstream and downstream of the nucleosome dyads (**Figure 9.4**). Similarly, the RNAPII signals downstream of the +3 and +4 nucleosome dyads were also observed to be locally high. The mean and sd normalised plots also show that the shape of the distribution of RNAPII appears to have a distinct pattern around nucleosome dyads (**Figure 9.4**).

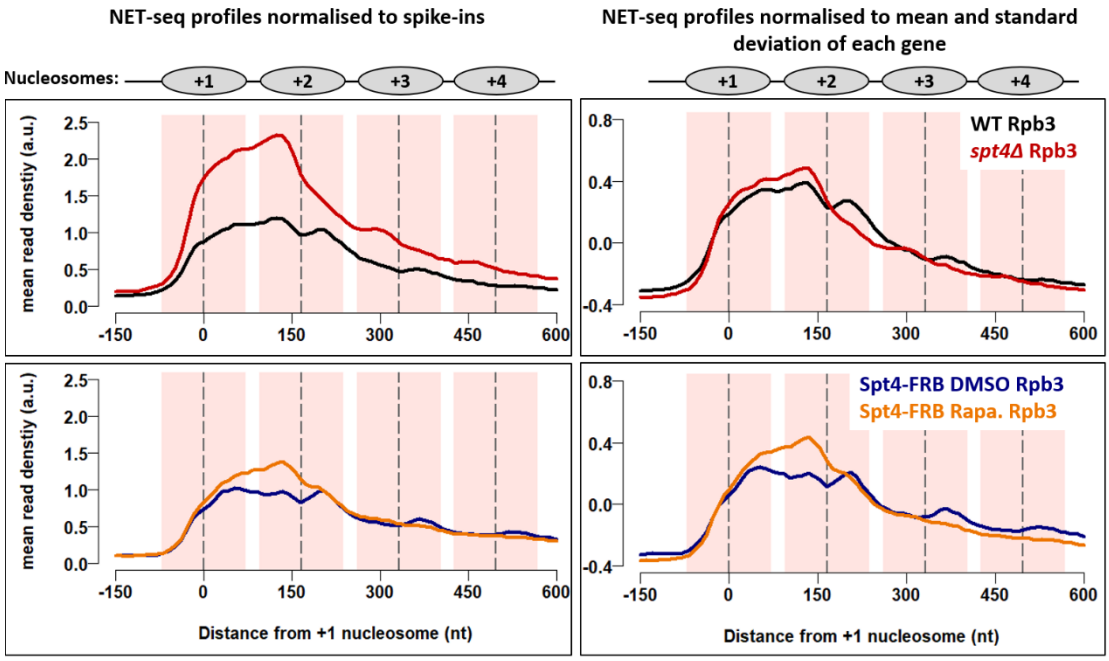


Figure 9.4) The distribution of RNAPII relative to the +1 nucleosome centre in the absence of Spt4 and the control cells

Metagene plots of NET-seq reads in WT (black), *spt4Δ* (red), DMSO treated Spt4-FRB (navy) and rapamycin-treated Spt4-FRB cells (orange). The NET-seq profiles given on the left are normalised to spike-ins to show the change in the signal densities. The profiles on the right are normalised to the mean and standard deviation of each gene to show change in the shape of the distribution of RNAPII. Dashed lines (black) through the peaks indicate the centres of the nucleosomes and the nucleosomal DNA (+/- 70 nt around the centre) is highlighted in light pink. Position of nucleosomes graphically shown above the plots.

In the absence of Spt4 (both in *spt4Δ* and Spt4 AA cells), the most notable difference in the RNAPII distribution is seen around the +2 nucleosome, such that the RNAPII accumulation is more concentrated upstream of the +2 dyad compared to their respective controls (**Figure 9.4**). In *spt4Δ*, the spike-in normalised metagene plot demonstrates that RNAPII accumulates in a region from the TSS to upstream of the +2 nucleosome dyad, while in Spt4 AA cells, the level of accumulation is not as high. Once only the shape of the distribution of RNAPII is considered, it becomes more apparent that in the absence of Spt4 (both in *spt4Δ* and Spt4 AA cells), RNAPII has a distinct peak upstream of the +2 nucleosome dyad, and the peaks around downstream of the nucleosome dyads (observed in the control cells) are diminished. A similar trend is also observed around +3 and +4 nucleosomes, but a lesser extent (**Figure 9.4**).

Interestingly, the highest RNAPII peaks in *spt4Δ* and Spt4-AA are around similar positions, at 135 nt and 138 nt from the +1 dyad, respectively. Notably, the highest peak of the averaged metagene profiles only reflects the average maximum, and the position of the RNAPII accumulation is likely to vary at single gene levels. To gain a greater insight into the accumulation of RNAPII in the absence of Spt4, the NET-seq signals were plotted as heatmaps around the +1, +2, +3 and +4 nucleosomes. Importantly, the PCGs were not ranked (presented as in their genomic order) to avoid bias in the visualisation. Additionally, the metagene plots of the RNAPII signals in *spt4Δ*, Spt4 AA and their control cells were plotted around the +1, +2, +3, and +4 nucleosomes, separately, at a higher resolution. The mean and sd normalised NET-seq profiles were used for these plots to highlight the changes in the shape of the distribution of RNAPII.

First, the NET-seq profiles in WT and *spt4Δ* cells were analysed. Around the +1 nucleosome, the distribution of RNAPII was similar across the genome in both conditions (**Figure 9.5** and **Figure 9.6**). Around the +2 nucleosome, in WT cells, the heatmap plots demonstrate the RNAPII signal varying across the genome with frequent pauses both up- and downstream of the dyad. In *spt4Δ* cells, the RNAPII signal is higher upstream of the +2 dyad for most genes (**Figure 9.5**). Once plotted as metagene profiles, it is seen that in WT cells, the RNAPII peaks are at - 40 nt and + 30 nt relative to the nucleosome dyad. In *spt4Δ* cells, the RNAPII peak upstream of the dyad is around similar position and more notable, while the downstream peak is lost (**Figure 9.5**).

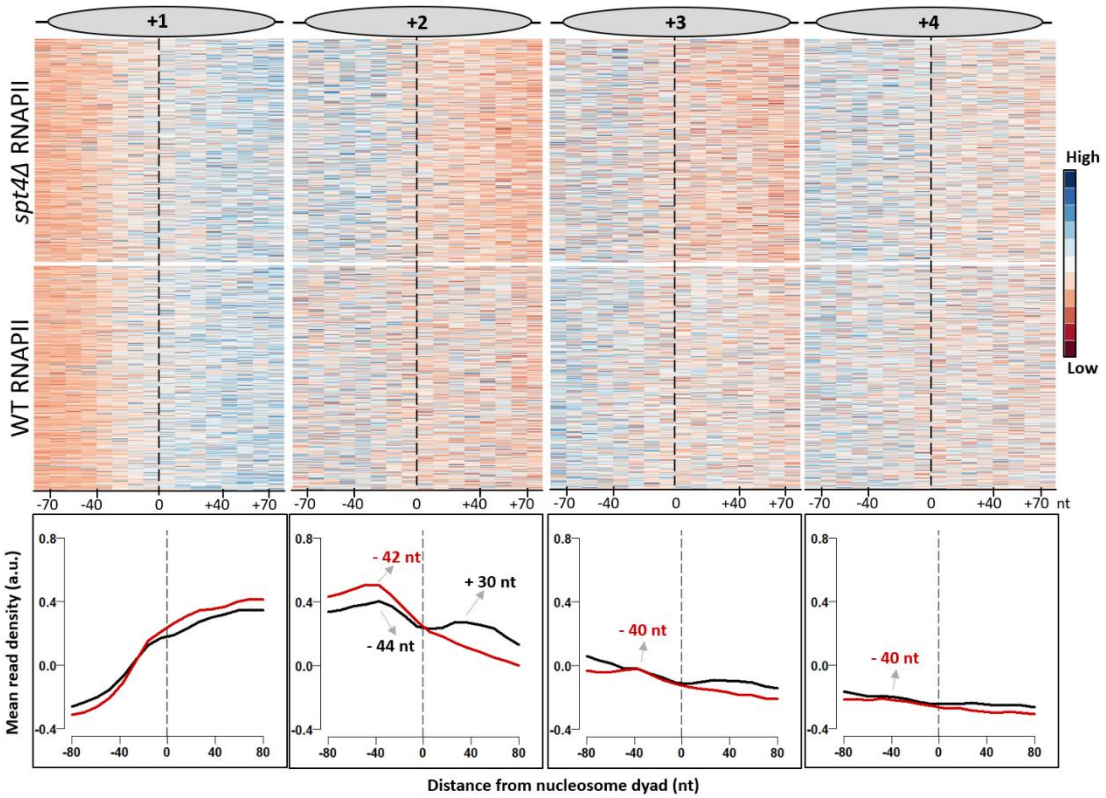


Figure 9.5) WT and *spt4Δ* NET-seq profiles around the +1, +2, +3 and, +4 nucleosomes.

Heatmaps of NET-seq in *spt4Δ* (top) and WT (middle). Each row indicates a PCG (n=4417). RNAPII signal is shown in 10 nt bins around the indicated nucleosome dyads (-/+ 80 nt from the dyad; x-axis). Close-up metagene profiles around the nucleosome dyads are given on at the bottom.

The significance of the WT and *spt4Δ* RNAPII densities at different positions was tested using Student's t-test. The NET-seq reads from -60 to -10 nt (upstream) and from +10 to +60 nt (downstream) relative to the +2 dyad were counted for each gene (**Figure 9.6**). The upstream and downstream read counts were then compared to test whether the RNAPII density was significantly higher upstream of the +2 dyad (Student's t-test, paired, one-tailed) in each condition. The analysis implies that there is significantly more RNAPII detected upstream of the +2 dyad in both WT and *spt4Δ* conditions suggesting that in both conditions RNAPII accumulates while transcribing into the +2 nucleosome (**Figure 9.6**).

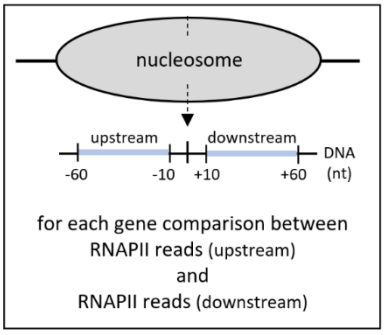
The Student's t-test reports if there is a significant change in the RNAPII density, however, it does not indicate the magnitude of the change. To test the magnitude of the change in the

RNAPII distribution, next, Cohen's d was used as introduced in chapter 5, section 5.2.4 (Cohen, 1988). The value of Cohen's d is a measure of the effect size of the change such that the values between 0.2 to 0.5 indicate small changes, between 0.5 and 0.8 indicate medium, and > 0.8 large changes (Cohen, 1988). Cohen's d is calculated by taking the mean difference between the upstream and downstream NET-seq read counts divided by the standard deviation of the differences in WT or *spt4Δ* cells:

$$d = \text{mean}\left(\frac{\text{upstream} - \text{downstream}}{\text{standard deviation (upstream} - \text{downstream)}}\right)$$

The analysis shows that the RNAPII accumulation upstream of the +2 dyad in WT cells was a medium (d=0.406) change, while it was a large (d=1.045) change in *spt4Δ* cells across the genome (**Figure 9.6**).

The Student's t-test and Cohen's d test were also applied for the WT and *spt4Δ* NET-seq data around the +3 and +4 nucleosome positions. The analysis is summarised in **Figure 9.6** and overall, it shows that RNAPII accumulates while transcribing into the +2, +3, and +4 nucleosomes in both WT and *spt4Δ* cells to different degrees. In *spt4Δ* cells, the magnitude of the RNAPII accumulation upstream of the dyads is greater compared to WT cells, with the largest change being around the +2 nucleosome.



	+ 1 nucleosome			
	WT	*spt4Δ*	DMSO Spt4-FRB	Rapa. Spt4-RFB
p-value	1	1	1	1
Cohen's d	- 0.945	- 0.972	- 0.949	- 1.040
	+ 2 nucleosome			
	WT	*spt4Δ*	DMSO Spt4-FRB	Rapa. Spt4-RFB
p-value	< 0.001	< 0.001	0.398	< 0.001
Cohen's d	0.406	1.045	0.004	0.732
	+ 3 nucleosome			
	WT	*spt4Δ*	DMSO Spt4-FRB	Rapa. Spt4-RFB
p-value	< 0.001	< 0.001	1	< 0.001
Cohen's d	0.337	0.667	- 0.119	0.298
	+ 4 nucleosome			
	WT	*spt4Δ*	DMSO Spt4-FRB	Rapa. Spt4-RFB
p-value	< 0.001	< 0.001	1	< 0.001
Cohen's d	0.249	0.450	- 0.125	0.165

Figure 9.6) Testing the significance and magnitude of the RNAPII accumulation upstream of the nucleosome dyads compared to downstream of the dyads.

P-values (Student's t-test, paired, one tailed) indicates the significance of the accumulation of RNAPII in *spt4Δ* relative to WT in the upstream of the +2 nucleosome dyad in compared to that of downstream. Upstream of the +2 dyad was taken as -60 to -10 nt from the dyad and downstream was taken as+10 to +60 nt from the dyad.

Next, the same analyses were repeated with the DMSO and rapamycin treated Spt4-FRB NET-seq data (**Figure 9.7**). Around the +1 nucleosomes, the distribution of RNAPII was similar across the genome in both conditions. Around the +2, in DMSO control cells, the RNAPII signal was not higher in the upstream of the dyad (p=0.398, d=0.004) (**Figure 9.6**). This is not consistent with the RNAPII distribution in WT cells, showing that different cell backgrounds can result in different RNAPII profiles. Nevertheless, the impact of the loss of Spt4 was similar in both backgrounds. In the Spt4 AA cells, the RNAPII reads were significantly higher upstream of the dyad compared to downstream, and the change was moderate (p< 0.001, d=0.732) (**Figure 9.6**). Around the +3 and +4 nucleosomes, the analysis indicates that there is a similar but a weaker trend in the RNAPII signal being higher upstream of the nucleosome dyads in Spt4 AA cells (**Figure 9.6** and **Figure 9.7**).

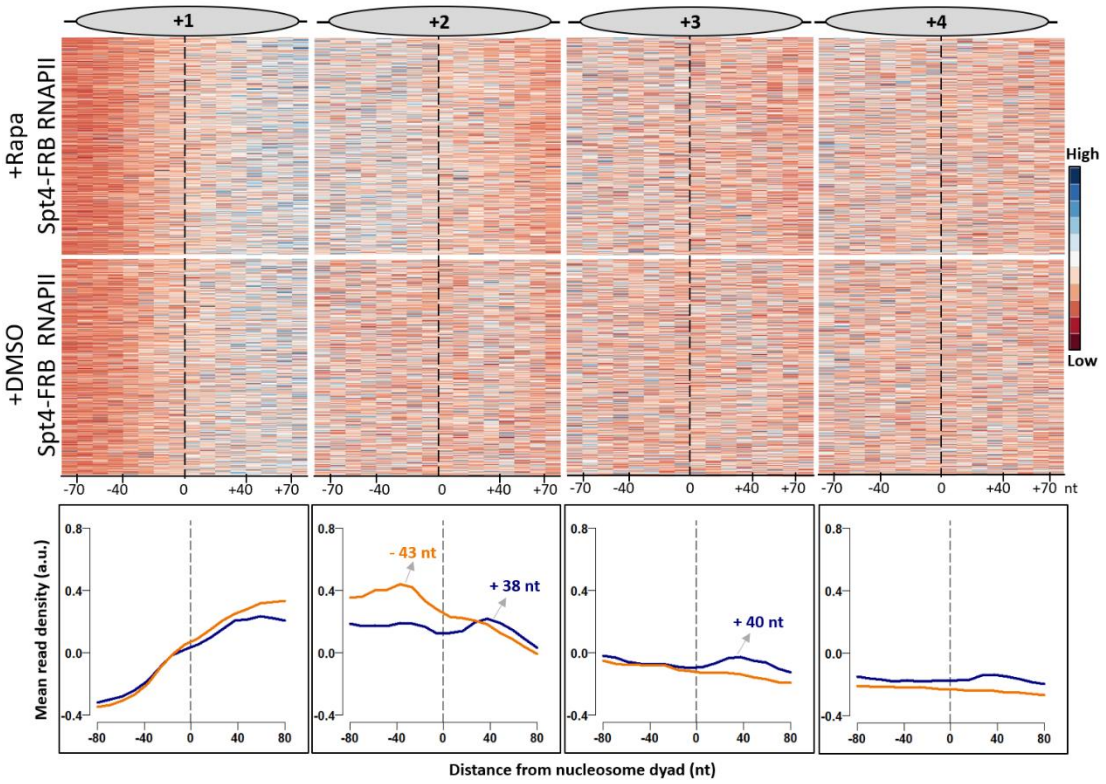


Figure 9.7) DMSO and rapamycin treated Spt4-FRB NET-seq profiles around the +1, +2, +3 and, +4 nucleosomes.

Heatmaps of NET-seq in Spt4 AA (top) and DMSO control (middle). Each row indicates a PCG (n=4417). RNAPII signal was shown in 10 nt bins around the indicated nucleosome dyads (-/+ 80 nt from the dyad; x-axis). Close-up metagene profiles around the nucleosome dyads are given on at the bottom.

Overall, the data show that the changes in the RNAPII profile over the gene bodies coincide with the position of nucleosomes. In two different cell backgrounds, the absence of Spt4 impacts the distribution of RNAPII around the +2 to +4 nucleosomes in a similar way. Additionally, a second RNAPII accumulation point was observed in control cells (WT and DMSO treated Spt4-FRB), around 30-40 nt downstream of the nucleosome centres. This suggests that in normal conditions, RNAPII encounters a second barrier around these regions, perhaps while nucleosomes are being re-assembled. Interestingly, this second peak was removed in the absence of Spt4 meaning that Spt4 might be playing a two-step role in the movement of RNAPII. Alternatively, in the absence of Spt4, the second peak could be masked because of the upstream high peak.

In conclusion, the results support that Spt4 helps RNAPII overcome nucleosome barriers during transcription, especially around the +2 nucleosome.

9.2.3 Comparing RNAPII distribution changes in *dst1Δ* and *spt4Δ:* Spt4 specifically impacts nucleosomes downstream of the +1 nucleosomes

Interestingly, the distribution of RNAPII in the absence of Spt4 was altered mostly around the +2 nucleosome (and the downstream nucleosomes, to a lesser degree). Around the +1 nucleosome, on the other hand, the shape of the distribution of RNAPII does not change. The high RNAPII signal immediately downstream of the +1 nucleosome seems to be due to a spread over of the RNAPII accumulation upstream of the +2 dyad once evaluated together with the impact of the rapid depletion of Spt4 on RNAPII. This observation raised a question as to whether Spt4 specifically regulates the distribution of RNAPII around nucleosomes except for the +1 nucleosome, or whether the RNAPII signal around the +1 nucleosome was not detectable by NET-seq. To further verify the impact of Spt4 on RNAPII distribution around the nucleosomes, a comparison was made between the RNAPII distribution in *spt4Δ* cells and cells lacking Dst1. Dst1 (TFIIS) is a TF that helps rescue backtracked RNAPII by triggering the cleavage activity of RNAPII (Zatreanu et al. 2019). In the absence of Dst1, backtracked RNAPII is not efficiently resolved and it accumulates upstream of the nucleosome dyads, including the +1 dyad (Churchman and Weissman 2011).

NET-seq on Dst1 knock-out cells (BY4741 *dst1::Kan Rpb3-FLAG*, *dst1Δ*) was performed, the data was analysed, and the reproducibility was verified as described before. *dst1Δ* NET-seq presented here was also consistent with the previous data (**Figure 9.8B**).

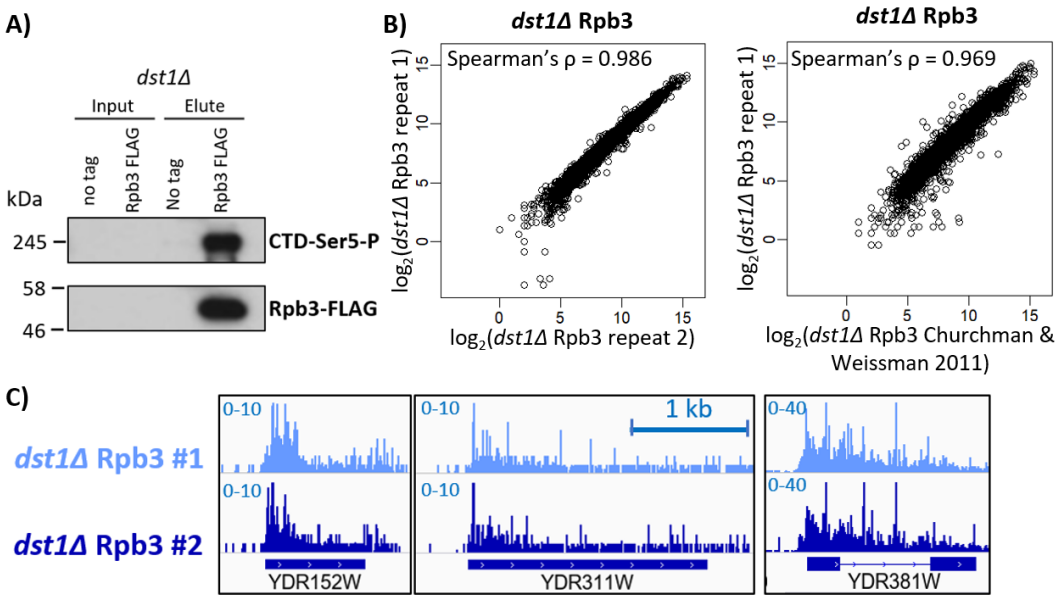


Figure 9.8) NET-seq in *dst1Δ*

A) NET-seq IP of no-tag control and FLAG-tagged Rpb3 in *dst1Δ*. Western blot of Input (In) and eluate (El) samples with anti-FLAG (M2) antibody run in 10 % SDS-PAGE. See Appendix 11A for the whole blots.

B) Reads are counted from the TSS to the PAS for each gene. log_2 transformed gene counts are correlated and Spearman's ρ calculated for the two repeats of *dst1Δ* Rpb3 and the published data.

C) IGV tracks of *dst1Δ* NET-seq reads of the three representative genes transcribed from the positive strand: *YDR152W, YDR311W* and *YDR381W* in two biological replicates. The dark blue boxes indicate the transcribed region of the genes (from TSS to PAS), while the blue line indicates the intronic region in *YDR381W.*

The *spt4Δ* and *dst1Δ* NET-seq metagene profiles demonstrates that around the +1 nucleosome, RNAPII has a distinct peak in *dst1Δ* condition. This means that the accumulation of RNAPII in very early stages of transcription could be detected using NET-seq (**Figure 9.9**). Such a difference in the distribution of RNAPII was not observed in *spt4Δ* cells. Therefore, although the accumulation of RNAPII in early transcription could eventually affect the transcription dynamics around the +1 nucleosome, Spt4 does not seem to take a part in overcoming the +1 nucleosome barrier. Consistently, Spt4 does not impact the position of the +1 nucleosome as shown in chapter 8.

Downstream of the +1 nucleosome, RNAPII accumulation patterns are similar in both mutants (**Figure 9.9**). Given that the altered NET-seq profile in *dst1Δ* is because of the unresolved backtracking events, the resemblance in the distribution of RNAPII around the +2 and downstream nucleosomes in the two mutant cells suggests RNAPII accumulation in *spt4Δ* is, maybe partially, due to backtracked RNAPII (**Figure 9.9**).

Therefore, it is concluded that the impact of Spt4 on the distribution of RNAPII is specific to nucleosomes downstream of the +1 nucleosome.

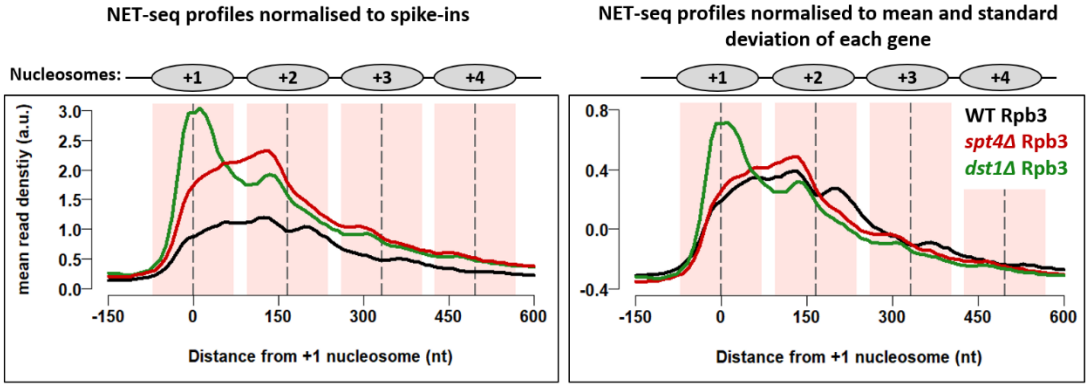


Figure 9.9) RNAPII distribution is differently affected in different TF deletion mutants

Metagene plots of RNAPII in WT (black), *spt4Δ* (red), and *dst1Δ* (green). The NET-seq profiles given on the left are normalised to spike-ins to show the change in the signal densities. The profiles on the right are normalised to the mean and standard deviation of each gene to show change in the shape of the distribution of RNAPII. Dashed lines (black) through the peaks indicate the centres of the nucleosomes and the nucleosomal DNA (+/- 70 nt around the centre) is highlighted in light pink. Position of nucleosomes graphically shown above the plots. *dst1Δ* NET-seq was perfomed in duplicate, see Appendix 11B for individual plots.

9.2.4 Comparison between the PRO-seq and NET-seq profiles

As mentioned earlier, the distribution of RNAPII in the absence of Spt4 was previously studied by PRO-seq. Given that PRO-seq detects only actively elongating RNAPII, while NET-seq detects all forms of RNAPII, the results from the two techniques can complement each other and provide information about the state of the RNAPII detected.

To compare the results, the PRO-seq data in WT and the absence of Spt4 were plotted as metagene profiles relative to the +1 nucleosome dyad (**Figure 9.10**). In *spt4Δ* cells, both techniques indicate a high RNAPII occupancy early in transcription. The PRO-seq has a sharp peak ~30 nt downstream of the +1 nucleosome dyad, whereas the NET-seq has a broader high signal spanning from around the +1 nucleosome to 40 nt upstream of the +2 nucleosome dyad (**Figure 9.10**). PRO-seq reports short term pauses which can resume within the 5 min transcription run-on window. Consequently, the high PRO-seq signal in the absence of Spt4 is likely to show that RNAPII progresses slower than WT RNAPII do, due to either reduced speed or an increased frequency of short term pauses just downstream of the +1 nucleosome. Further downstream of the +1 nucleosome, the PRO-seq signal progressively decreases (**Figure 9.10**). The progressive decrease in the levels of PRO-seq could be explained in two ways. First, the slow progress of RNAPII is alleviated over the gene bodies, meaning that either the frequency of the pauses is diminished, or the speed of transcription is less affected. Second, a portion of the RNAPII is not moving due to stalling or backtracking events, hence PRO-seq signal is not produced. The latter option is more likely as the *spt4Δ* NET-seq data indicates a high occupancy of RNAPII over the gene bodies, especially around the +2 nucleosome. Therefore, the comparison between the PRO and NET-seq data supports that in the absence of Spt4, RNAPII is likely to stall or backtrack upstream of the +2 nucleosome giving rise (only) to NET-seq reads. This accumulation in RNAPII (due to stalling and/or backtracking), in turn, might lead to slowing down of the upstream RNAPII, perhaps by collision events between upcoming and stalled RNAPII that can be detected by both NET and PRO-seq. Notably, PRO-seq was performed in WT303 background as opposed to BY4741 background used in this thesis, and differences originating from the background strains cannot be ruled out (Booth et al. 2016).

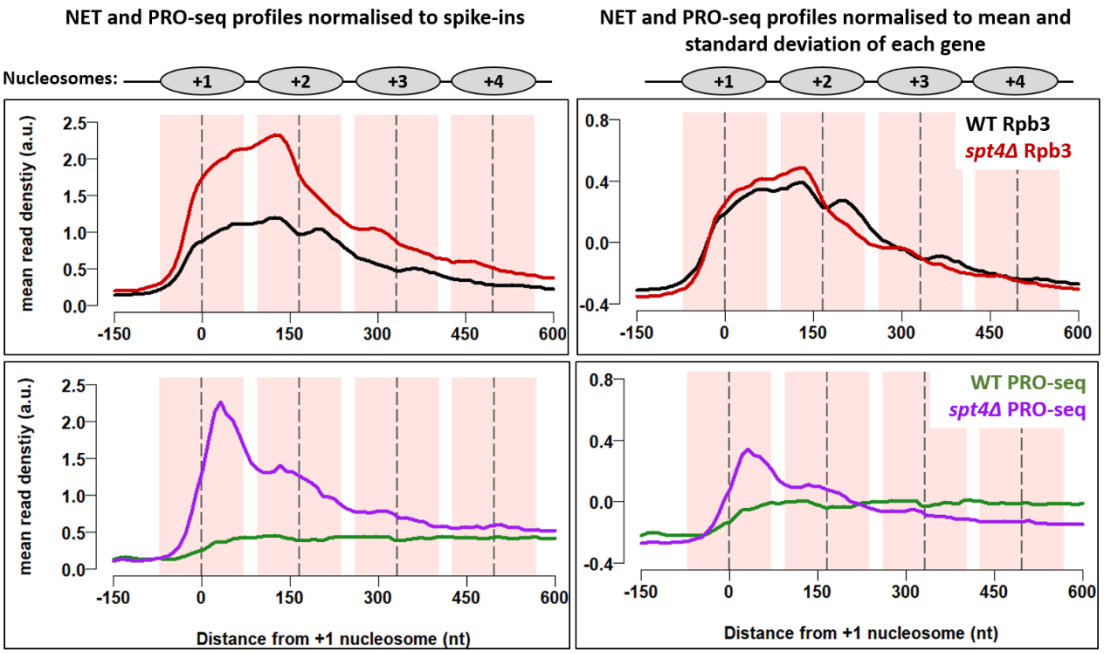


Figure 9.10) NET-seq and PRO-seq RNAPII distributions in the absence of Spt4 relative to the +1 nucleosome centre

Metagene plots of NET-seq reads in WT (black) and *spt4Δ* (red) in the top panel and previously published PRO-seq reads in WT (green) and *spt4Δ* (purple) in the bottom panel. PRO-seq data was taken from NCBI database GSE76142. The NET and PRO-seq profiles given on the left are normalised to spike-ins to show the change in the signal densities. The profiles on the right are normalised to the mean and standard deviation of each gene to show change in the shape of the distribution of RNAPII. Dashed lines (black) through the peaks indicate the centres of the nucleosomes and the nucleosomal DNA (+/- 70 nt around the centre) is highlighted in light pink. Position of nucleosomes graphically shown above the plots.

9.2.5 The interplay between nucleosome spacing and RNAPII in WT and *spt4Δ*

In addition to the changes in the distribution of RNAPII, nucleosome positions were also altered in *spt4Δ*. To test these two observations were somehow related, the cross-talk between the RNAPII accumulation and the change in nucleosome spacing in *spt4Δ* cells was investigated.

If Spt4 directly or indirectly, and co-transcriptionally, regulates the nucleosome positioning the deletion of Spt4 should affect nucleosome spacing to a greater extent in highly active genes compared to less active genes. Moreover, if the RNAPII accumulation and the increased

nucleosome spacing in *spt4Δ* cells are connected, then the amount of RNAPII accumulation should reflect the degree of increase in nucleosome spacing. To test these hypotheses, the correlation between nucleosome spacing and RNAPII densities in WT and *spt4Δ* cells were analysed. RNAPII density was assessed by counting the WT NET-seq reads within the first 500 nt from the TSS. PCGs were then ranked by the WT NET-seq densities and split into 10 groups. For each group, the median nucleosome spacing between the +1 and +2, as well as between the +2 and +3 nucleosomes in WT and *spt4Δ* were calculated and plotted together with the median NET-seq read density for comparison.

In WT conditions, nucleosome spacing is shorter in highly expressed genes and it progressively increases for the genes having lower expression levels (**Figure 9.11**). In *spt4Δ*, on the other hand, nucleosome spacing is less variable (**Figure 9.11**). The increased distance between the nucleosomes is observed at all groups, but the amount of the increase is larger for the highly expressed genes. Importantly, once the analysis was repeated using *spt4Δ* NET-seq read density (instead of the WT data) or transcript levels from WT or *spt4Δ* RNA-seq data, similar results were obtained (**Figure 9.12**).

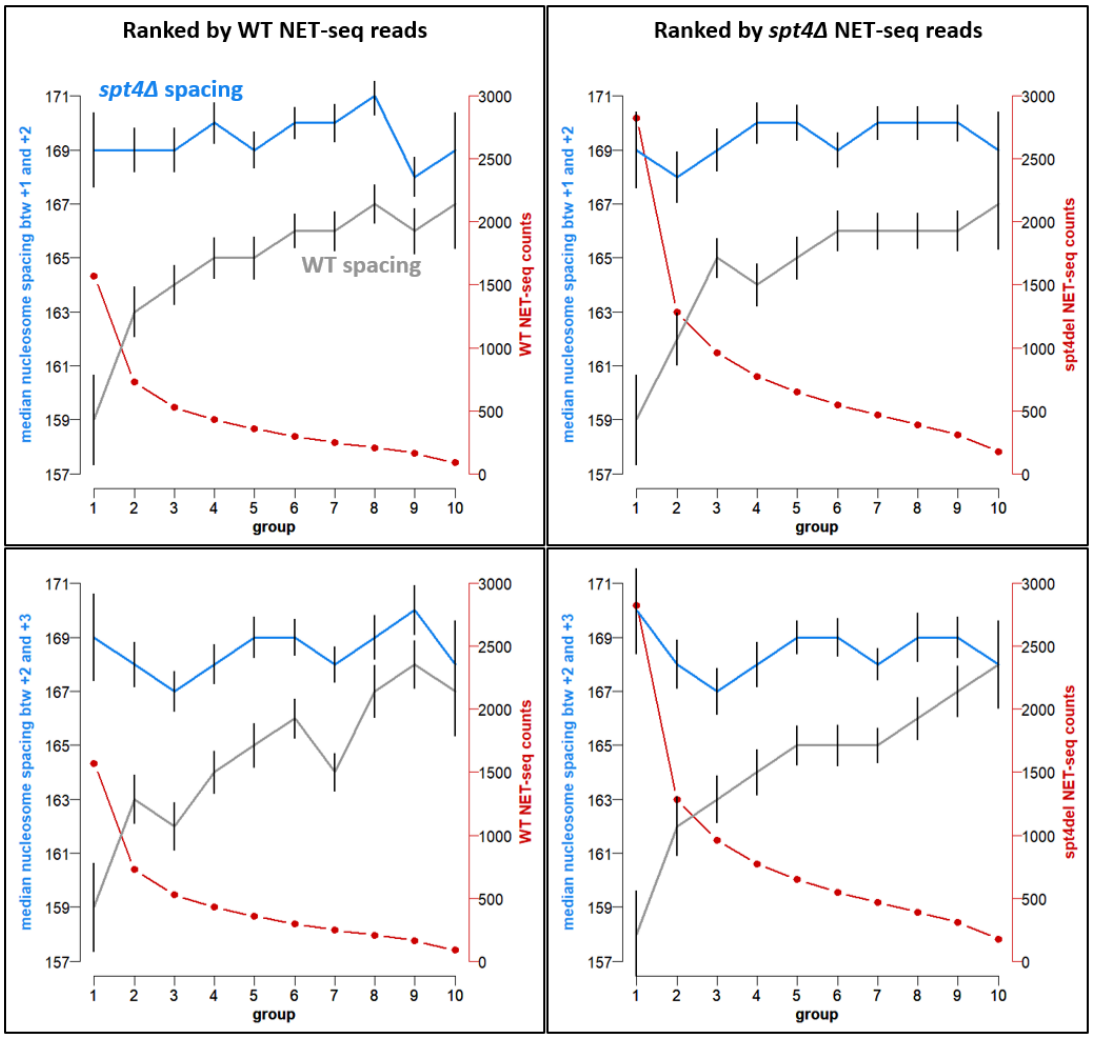


Figure 9.11) Comparison between the WT and *spt4Δ* nucleosome spacing in groups of genes with different levels of NET-seq signals.

PCGs were split into 10 groups based on WT (left) or *spt4Δ* (right) NET-seq reads in the first 500 nt from the TSS (red, y-axis). The median NET-seq reads and the median nucleosome spacing in *spt4Δ* (blue) and WT (grey) were plotted for each group. The black bars around nucleosome spacing data points indicate one standard deviation.

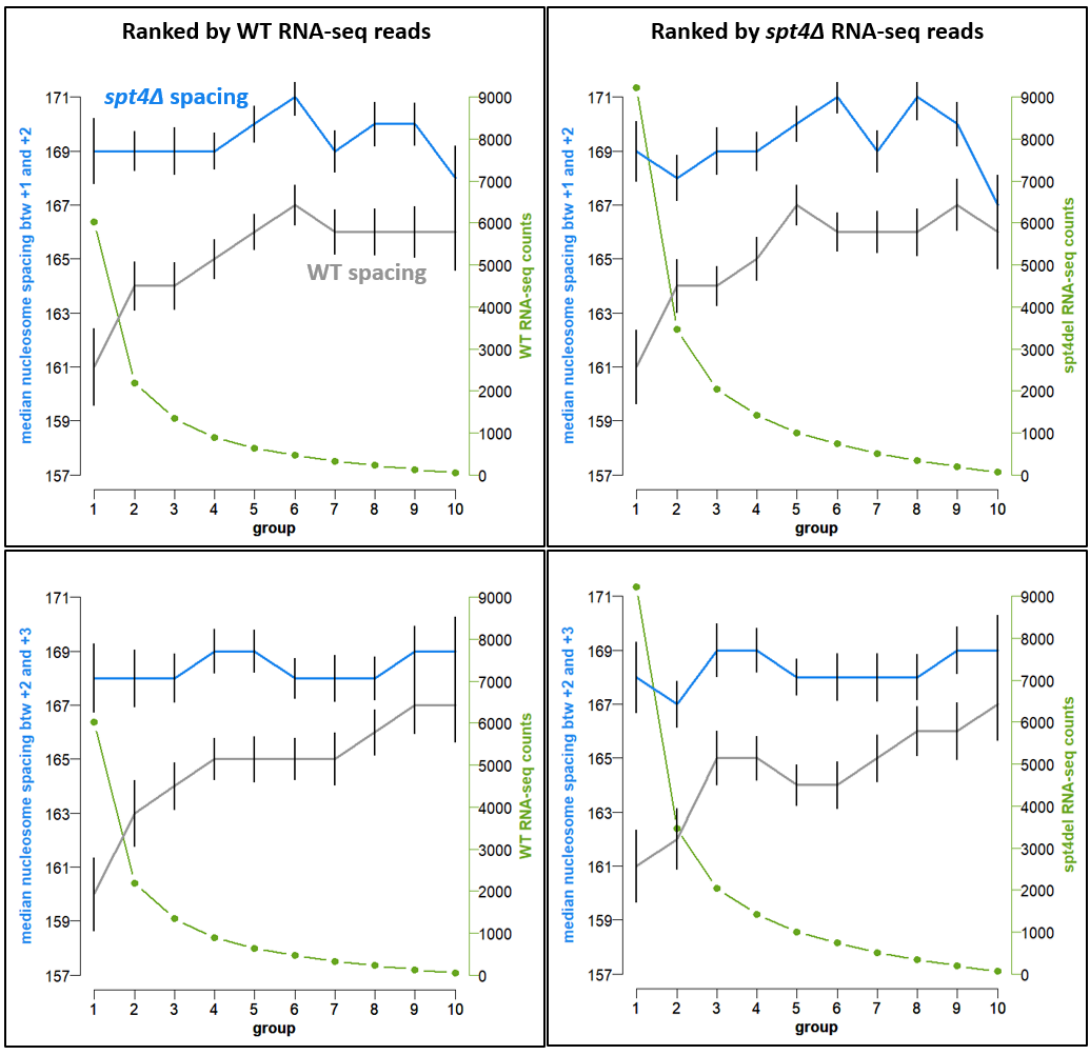


Figure 9.12) Comparision between the WT and *spt4Δ* nucleosome spacing in groups of genes with different amount of RNA-seq signals.

PCGs were split into 10 groups based on WT (left) or *spt4Δ* (right) RNA-seq reads (green, y-axis). The median RNA-seq reads and the median nucleosome spacing in *spt4Δ* (blue) and WT (grey) were plotted for each group. The black bars around nucleosome spacing data points indicate one standard deviation.

To test the association between transcriptional activity (assessed by NET-seq or RNA-seq) and the distance between the nucleosomes, the Pearson's correlation coefficient was calculated by comparing the median NET-seq (or RNA-seq) reads with the median nucleosome spacing in ten groups for each pair described above (**Table 9.1**). The analysis shows that, in WT cells, there is a negative correlation between nucleosome spacing and transcriptional activity, and this correlation is lost in *spt4Δ* cells. In conclusion, the data indicate that the lack of Spt4 leads to an increase in nucleosome spacing across the genome. The increase is greater in highly

expressed genes, and this could be explained by a loss of close nucleosome spacing in WT cells.

Table 9.1) Transcription activity and nucleosome spacing are negatively correlated in WT, and poorly correlated in *spt4Δ* cells.

	Spacing btw + 1 and +2		Spacing btw + 2 and +3	
	WT	*spt4Δ*	WT	*spt4Δ*
Ranked by				
WT NET-seq	-0.99	-0.19	-0.89	-0.02
spt4Δ NET-seq	-0.97	-0.42	-0.96	+0.35
WT RNA-seq	-0.94	-0.22	-0.94	-0.38
spt4Δ RNA-seq	-0.93	-0.20	-0.84	-0.32

Pearson's correlation coefficient is calculated to compare the median NET or RNA-seq reads and the median nucleosome spacing for each pair in ten groups of genes described in figure 9.11 and 9.12.

9.3 Discussion

In this chapter, the data show that the oscillation of Spt4 and Spt5 on and off RNAPII correlates with the position nucleosomes. Spt4 seems to assist transcription through nucleosomes by suppressing RNAPII accumulations upstream of the nucleosome dyads, especially at the +2 nucleosome. In WT cells, nucleosome spacing is inversely proportional to the amount of transcription. Such correlation is absent in *spt4Δ* cells, and the increased nucleosome spacing originates from the loss of closely spaced nucleosomes in well expressed genes.

9.3.1 Spt4 and Spt5 oscillate on and off RNAPII related to the position of RNAPII around the nucleosomes

The association between Spt4/5 and RNAPII periodically decreases during transcription of the nucleosomal DNA upstream of the dyads and increases downstream, after passing the dyads. Interestingly, similar to Spt4 and Spt5, the oscillating pattern of TEFs on RNAPII was also described for histone chaperones Spt6 and Spt16 (Fischl et al. 2017). This pattern was thought to originate from the changes in the preferential affinities of the TEFs to histones as transcription progresses over the chromatin template. Spt6 assists the reassembly of nucleosomes after RNAPII passage and physically interacts with histone H3 (Bortvin and

Winston 1996). During the transcription over the nucleosome dyads, Spt6 would have a relatively stronger interaction with histone H3 than that of with RNAPII. After passing the dyad, Spt6-H3 interaction would be lost and Spt6-RNAPII interaction would increase. Consequently, the differential interaction between Spt6-RNAPII is predicted to create the periodic TEF-seq pattern of Spt6 on RNAPII (Appendix 12). Similarly, Spt16 oscillates on RNAPII, but with a shifted pattern compared to Spt6, consistent with its preferential interaction with H2A/H2B (Fischl et al. 2017; Hondele et al. 2013) (Appendix 12).

The occupancies of Spt4/5 on RNAPII resemble that of Spt6 (Appendix 12). Consequently, one interpretation of the oscillation pattern of Spt4/5 on RNAPII would be that these factors directly or indirectly take roles in the nucleosome arrangements, perhaps by interacting with histone H3. Indeed, Spt4/5 was shown to stabilise partially unwrapped nucleosome-RNAPII intermediate during *in vitro* transcription, suggesting that these factors interact differently with the TEC at different confirmations and/or stages (Crickard et al. 2017). Additionally, cryo-EM images reveal that the Spt4/5 complex on RNAPII occupies a space between the DNA-exit site of the TEC and the nucleosome downstream of the transcription site (Ehara et al. 2019). Therefore, spatially, the interaction between Spt4/5 and a nucleosome could be possible. Similar to histone chaperones, Spt5 bears an acidic domain that could interact with basic histone proteins (Ehara et al. 2019). Spt4 does not have an acidic domain and this makes Spt4 a less likely candidate for having a direct interaction with the histones. However, the affinity of Spt4 for RNAPII could change indirectly through Spt5. Alternatively, if the Spt4/5 complex interacts with unwrapped nucleosomal DNA or other TFs more strongly than it does with RNAPII upstream of the nucleosome dyads, this in turn, could also result in a phased occupancy of the factors on RNAPII.

9.3.2 Spt4 suppresses RNAPII accumulation upstream of the nucleosome dyads, except the +1 nucleosome

As mentioned before, the movement of RNAPII on a nucleosomal DNA was recently documented in yeast system using *in vitro* transcription assays (Crickard et al. 2017; Ehara et al. 2019; Kujirai et al. 2018). Studies show that the Spt4/5 complex reduces RNAPII pausing upstream of a purified nucleosomes (Crickard et al. 2017; Ehara et al. 2019). Importantly, the position of the reduced RNAPII pausing was different in these studies. Both studies reveal multiple RNAPII pausing sites on a purified nucleosome. However, while Crickard *et al.* show a reduced RNAPII pausing around 50 nt into the nucleosome upon addition of purified Spt4/5, Ehara *et al.* demonstrate a similar impact, but around 20 nt into nucleosome. Although the positions of the reduced RNAPII pausing are inconsistent in these studies, perhaps because of different DNA templates being used, both studies demonstrated that even in the absence of other TFs (except TFIIS and Elf1), the Spt4/5 complex promotes transcription by reducing pauses upstream of the nucleosome centres.

The NET-seq data presented here would support these observations. The accumulation of RNAPII upstream of the nucleosome dyads in the absence of Spt4 provides *in vivo* and genome-wide evidence for a role of the Spt4/5 complex promoting transcription through nucleosomes and complements the *in vitro* results. *In vivo*, the position of RNAPII accumulation on nucleosomes is likely to be influenced by multiple factors, such as occupancy of the other TEFs, histone PTMs impacting the nucleosomal DNA-histone interactions, or whether there is transcription on the opposite strand. That is probably why, in the absence of Spt4, the position of RNAPII accumulation varies between the genes. However, interestingly and consistently, both in *spt4Δ* and Spt4 AA cells, the highest peak of the RNAPII signal on average was around SHL (-4), ~40 nt upstream of the +2 nucleosome dyad. Around SHL (-4), nucleosomal DNA makes contact with H2A/H2B and around SHL (-3), H4/H3 interaction starts (Kujirai and Kurumizaka 2020) (**Figure 1.4**). The accumulation around SHL (-4), therefore,

might indicate that RNAPII without Spt4 is less competent to move through the DNA-H3/H4 interactions. As speculated above, this could be due to a direct or indirect interaction of Spt4 with nucleosomes or other factors influencing DNA-histone interactions. Alternatively, Spt4 could promote RNAPII transcription by structurally supporting the elongation complex as proposed before (Crickard et al. 2016).

Why is the region most affected by the absence of Spt4 immediately upstream of the +2 nucleosome? This might be because the composition of the transcription complexes is different while transcribing over the +1, +2, compared to downstream nucleosomes. Although Spt4 is recruited to RNAPII early in transcription (at least after the synthesis of ~20 nt RNA as shown in Chapter 4), in most genes TSS is already ~20 nt into the +1 nucleosome (as shown in figures 9.6 by heatmaps and 9.10). Consequently, the +1 nucleosome barrier could be passed by the transcription initiation complex. It was also speculated that the +1 nucleosome might be disassembled prior to initiation (Tramantano et al. 2016). In either case, the +2 nucleosome would be the first nucleosome barrier that the TEC encounters.

Further downstream of the +2 nucleosome, as RNAPII transcribes over the gene bodies, the composition of the TEC changes. The changes in the TEC are likely to provide additional mechanisms to keep the RNAPII transcription over nucleosomes (more) effective and productive. For instance, Paf1C, which is a complex generally associated with productive elongation, joins transcription around the +3 nucleosome and its level on RNAPII progressively increases towards the 3'-end of the genes (Fischl et al. 2017). The involvement of Paf1C starts a cascade of histone PTMs. Although the role of histone PTMs on passing nucleosomal barriers are not clear, histone PTMs are likely to affect nucleosome compaction and the recruitment of factors further helping transcription on chromatin. For example, H2Bub1 induced by Paf1C was proposed to promote histone disassembly by Spt16 (Pavri et al. 2006). Therefore, around the +2 nucleosome, before transitioning to productive transcription, RNAPII progress over

nucleosomes might rely more on the function of Spt4. Additionally, in the absence of Spt4, there is a milder change in the distribution of RNAPII around the +3 and +4 nucleosomes. This result implies that Spt4 still contributes to the removal of the nucleosome barriers downstream of the +2 nucleosome, perhaps acting redundantly. This could be also an explanation of the synthetic lethality observed in double mutants of *spt4* and genes encoding one of the five Paf1C components (Squazzo et al. 2002).

9.3.3 Spt4 seems to contribute to close nucleosome spacing associated with highly expressed genes

Interestingly, in WT cells the distance between the nucleosomes was not the same across the genome. When split into groups, it could be seen that nucleosome spacing tends to be shorter as gene expression increases. Indeed, this correlation appears to be conserved between different species as it was also described in *Drosophila* as well as yeast (Baldi et al. 2018; Ocampo et al. 2016), and it is thought to be a result of a reduced regularity in the nucleosome positions due to heavy transcription. The impact of the transcription levels on nucleosome dynamics could be explained by the different degrees of nucleosome decompaction. More specifically, more than one RNAPII passing over the genes might require complete removal of histones/nucleosomes. This could create more disruption in chromatin structure, and it would take a longer time to restore nucleosome positions after RNAPII passage. On the contrary, when there is less RNAPII trafficking over the genes, one RNAPII could progress around a partially unwrapped nucleosome. In this case, nucleosome reassembly and repositioning would be relatively easier and faster behind transcribing RNAPII. Consequently, the full and partial histone disassembly could give rise to different MNase digestion patterns in genes with different transcription levels.

In *spt4Δ*, the increase in nucleosome spacing appears to be due to a lack of the reduced nucleosome spacing at highly expressed genes, which can be explained in two ways. First, if

the RNAPII accumulation in *spt4Δ* does not allow high transcription frequency, nucleosomes would not be completely removed, and normal nucleosome spacing could be maintained as in lowly expressed genes. Arguably, RNAPII accumulation does not always change the nucleosome architecture. For example, *dst1Δ* causes similar RNAPII accumulation profile around the +2 dyad as *spt4Δ* does, but previously published MNase-seq data suggests no change in nucleosome spacing in *dst1Δ* cells compared to WT conditions (Gutiérrez et al. 2017). This means that the impact of Spt4 on the nucleosome positioning might not be a result of RNAPII accumulation.

The second possibility is that, Spt4 co-transcriptionally regulates the positioning of the +2 and downstream nucleosomes, perhaps by directly or indirectly helping histone removal. If this is the case, highly expressed genes would be more affected in *spt4Δ* cells, as complete histone disassembly would be more compromised than the partial disassembly, and transcription would continue at sub-optimal conditions, possibly with help of redundant mechanisms overcoming nucleosomal barriers. In that case, RNAPII accumulation in *spt4Δ* could be a result of the changed nucleosome dynamics.

Taken together, the data presented in this thesis support a case for Spt4 contributing to co-transcriptional nucleosome disassembly downstream of the +1 nucleosomes, especially around the +2 nucleosome (**Figure 9.13**). In WT conditions, higher transcriptional activity correlates with shorter nucleosome spacing potentially due to more disruptive impact of RNAPII on chromatin. The impaired nucleosome disassembly in *spt4Δ* cells, therefore, results in increased nucleosome spacing, especially in highly expressed genes. Additionally, in the absence of Spt4, RNAPII accumulates upstream of nucleosomes, as nucleosomes are not efficiently disassembled. This contribution of Spt4 to transcription is likely to involve Spt5, as Spt4 also regulates the interaction between Spt5 and RNAPII.

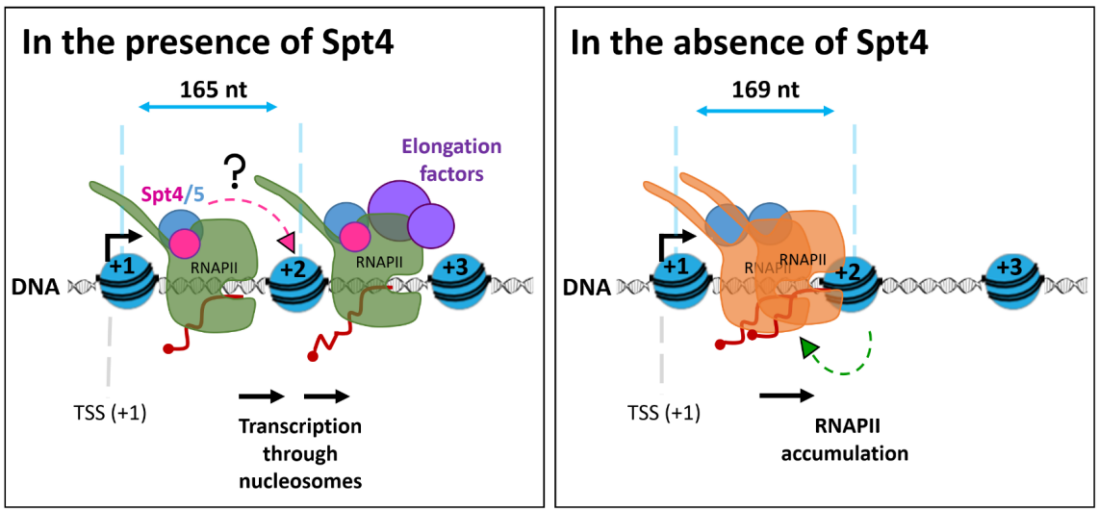


Figure 9.13) Model

Spt4 helps RNAPII pass nucleosome barriers *in vivo*, especially around the +2 nucleosome.

10 Conclusions and future approaches

10.1 Summary

This thesis provides genome-wide investigations into the contribution of Spt4 to RNAPII transcription in *S.cerevisiae*.

First, Spt4 and Spt5 were mapped on the genes across the genome. The high resolution afforded here demonstrates Spt4 and Spt5 joining transcription when just 20 nt of RNA have been synthesised and traveling with elongating RNAPII until the PAS. Once the position of Spt4 and Spt5 are examined relative to the RNAPII (i.e divided by RNAPII signal), the interaction of Spt4/5 with RNAPII was found to be dynamic over the genes. The occupancy of Spt4/5 on RNAPII is lower during transcription of the upstream of the nucleosome dyads, and higher downstream of the dyads. This dynamic behaviour of Spt4/5 on RNAPII around the nucleosome dyads implies that these factors might be periodically interacting with other surrounding transcriptional components such as TEFs, histones or chromatin remodellers, and the function of the Spt4/5 complex might be related to regulation of the nucleosome dynamics.

The data also show that, in *spt4Δ* cells, Spt5 is still recruited to RNAPII, but the oscillating pattern of Spt5 on RNAPII is lost. This observation implies that at least one function of Spt4 is regulating the periodic interaction of Spt5 with RNAPII.

Second, RNAPII was mapped in *spt4Δ* and Spt4 depleted (Spt4 AA) cells to differentiate between the long and short term effects of the lack of Spt4 on the distribution of RNAPII. In *spt4Δ* cells, RNAPII has an increased occupancy early in transcription, mostly upstream of the +2 nucleosome. In Spt4 AA, RNAPII exhibits a peak around 40 nt upstream of the +2 nucleosome dyad (i.e. 30-35 nt into the +2 nucleosomal DNA). Together these data suggest that in the absence of Spt4, the change in the distribution of RNAPII starts around 40 nt

upstream of the +2 dyad (in Spt4 AA), RNAPII progressively accumulates around this region and ends up having a broader accumulation profile in *spt4Δ* cells. Similar changes in the distribution of RNAPII were also observed upstream of the +3 and +4 nucleosomes in both *spt4Δ* and Spt4 AA cells, to a lesser extent. Therefore, the RNAPII accumulation in the absence of Spt4 is very likely to be due to transcriptional barriers created by nucleosome-DNA interactions upstream of the dyads.

The shape of the distribution of *spt4Δ* RNAPII was further analysed by using mathematical modelling to gain a greater insight. The model predicts a lower moving ratio in early transcription in *spt4Δ* cells due to stalling and backtracking of RNAPII and suggests reduced processivity and increased early termination in *spt4Δ* compared to WT cells, that might be a result of higher backtracking and stalling.

Third, RNAPII was mapped in Spt5 depleted (Spt5 AA) cells. The comparison between the distribution of RNAPII in Spt4 and Spt5 AA cells showed similar accumulation pattern early in transcription, suggesting that these factors act together to overcome a transcriptional barrier. However, the overall impact of the depletion Spt5 was severe and the level of RNAPII was significantly reduced, consistent with the essential function of Spt5.

Fourth, the composition of the transcription complex was investigated in WT and *spt4Δ* cells. The most significant result was the increased abundance of TFIIH on RNAPII in *spt4Δ* cells, implying TC-NER might be activated in the absence of Spt4. Additionally, the abundances of histone chaperones (Spt6, Spn1, Spt16, and Pob3) on the *spt4Δ* transcription complex were slightly reduced, suggesting that Spt4 might be important for the recruitment of histone chaperones in some cases, perhaps in a subpopulation of genes.

Finally, mapping of nucleosomes showed that in *spt4Δ* cells, the distances between the nucleosomes were larger compared to that of in WT cells. The increase in nucleosome spacing was observed for most genes across the genome with the largest increase being between the

+1 and +2 nucleosomes by 4 nt on average. More detailed analysis showed that in WT cells, highly active genes associates with shorter nucleosome spacing. In *spt4Δ* cells, this correlation was lost, meaning that the amount of increase in nucleosome spacing was larger for highly expressed genes.

In conclusion, the findings presented in this thesis propose that the function of Spt4 involves helping RNAPII pass nucleosome barriers by contributing to disassembly of nucleosomes downstream of the +1 nucleosome, especially at the +2 nucleosome. Given that Spt4 also regulates the interaction between Spt5 and RNAPII, the contribution of Spt4 is likely involve Spt5.

10.2 Future directions

The full mechanism of how Spt4/5 contributes to removal of nucleosome barriers is unclear. One possibility is that Spt4/5 directly interacts with histones. The model proposed here would not necessarily require such interaction. Nevertheless, structurally, the position of the Spt4/5 complex being in between the RNAPII DNA exit site and a downstream nucleosome suggests that this speculative interaction is possible. Therefore, one future experiment would be testing the histone-Spt4/5 interactions. This could be addressed *in vitro* by examining the protein affinities with purified Spt4, Spt5 and histones. Alternatively, *in vivo* protein-protein interactions could be studied using site specific crosslinking techniques that recently proved useful to detect direct interactions between H2A and TEFs (Cucinotta et al. 2019). In this technique, a photo-activatable unnatural amino acid p-benzoyl-L-phenylalanine (BPA), is integrated into the protein of interest to allow crosslinking. The BPA containing protein is then UV-crosslinked to interactors, and the proteomics of the interactome is analysed. Applying this technique to Spt4 to identify its interactome would provide unbiased approach to test whether Spt4 directly or indirectly interacts with histones or proteins involved in histone regulation.

Additionally, one outcome of the work performed here was the finding that TFIIH is more abundant on the transcription complex lacking Spt4. This finding is in line with the idea of Spt4 suppressing TC-NER (Duan et al. 2020). Questions arising from this function of Spt4 are whether Spt4 suppresses DNA damage or DNA damage-sensing mechanisms. Therefore, another line of study would be investigating the impact of Spt4 in TC-NER.

10.3 Conclusion remarks

The work presented here adds to the understanding of the function of the Spt4/5 complex and proposes that Spt4/5 helps RNAPII pass nucleosome barriers by contributing to disassembly of nucleosomes, especially the first nucleosome that the TEC encounters (that is the +2 nucleosome in budding yeast).

Does the mammalian counterpart of Spt4/5 (DSIF) contribute to transcription over nucleosomes? It is likely as Spt4/5 is conserved in all organisms having nucleosomes, its binding position to RNAPII is also conserved, and human SPT4 complements the *spt4* mutant phenotype in yeast (Hartzog et al. 1996; Osman and Cramer 2020). However, DSIF might have an additional function in different contexts.

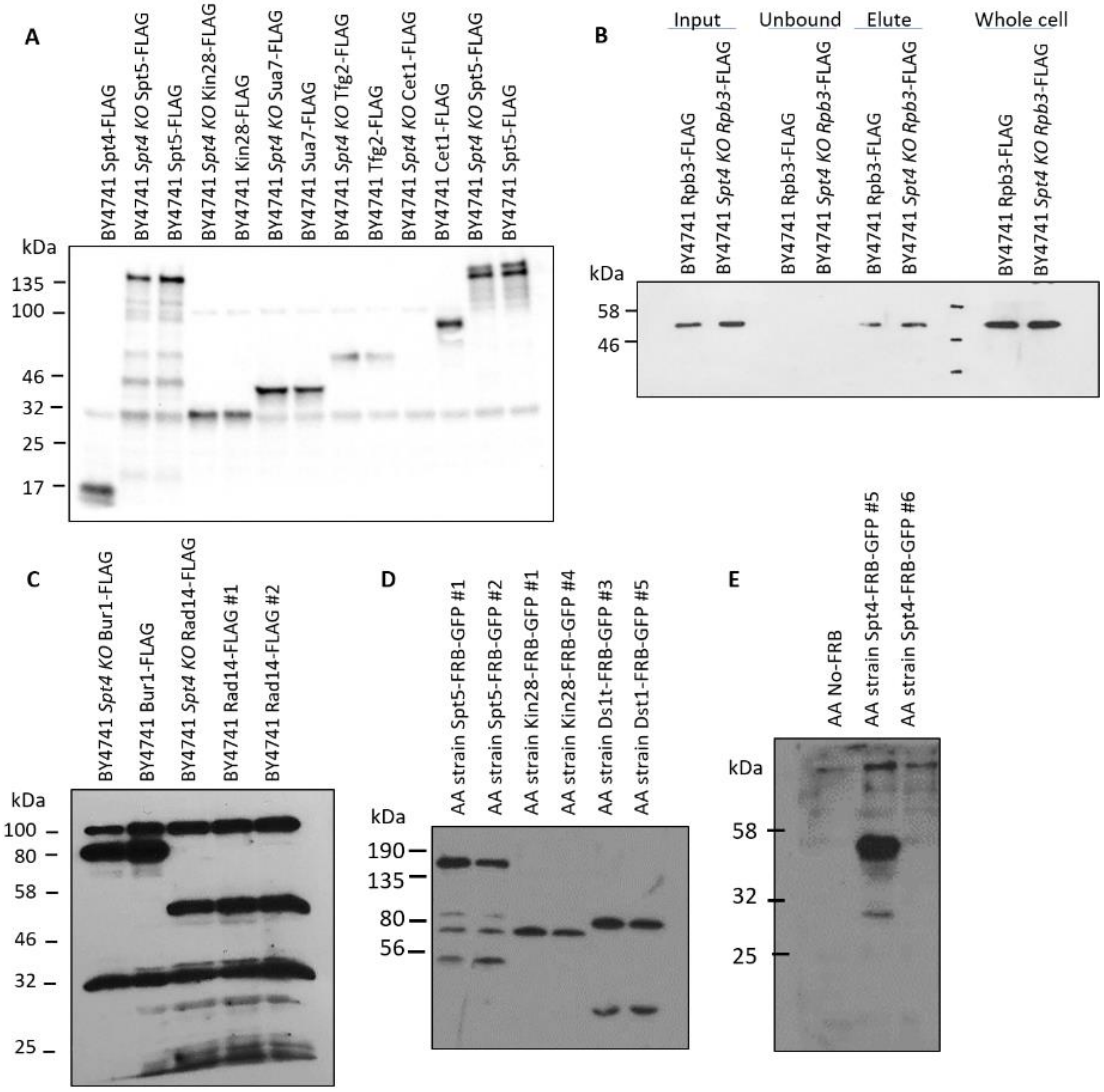
In budding yeast, Spt4/5 only acts as a positive factor. In mammalian cells, however, DSIF was discovered as a factor that can induce DRB-sensitivity, and it also has a negative role. Shortly after transcription initiation in mammalian cells, RNAPII pauses around 20-80 nt downstream of the TSS that is defined as promoter proximal pause. The current model suggests promoter-proximal pause being induced by two complexes acting as negative elongation factors; unphosphorylated DSIF and NELF. Following the pause, both DSIF and NELF are phosphorylated by P-TEFb which then triggers dissociation of NELF, turns DSIF into a positive elongation factor, and releases the promoter proximally paused RNAPII (Core and Adelman 2019).

Could the negative role of DSIF be related to (lack of) its positive function? In other words, could promoter proximal pause be induced or stabilised by nucleosomes because DSIF does not exert its positive function?

In mammalian cells, transcription does not start within a nucleosome, and the +1 nucleosome is ~140 nt downstream of the TSS meaning that the promoter proximal pause occurs between the TSS and the +1 nucleosome. Whether the +1 nucleosome plays a part in the establishment of promoter-proximal pausing is controversial. Experiments using PRO-seq and GRO-seq show that the pausing site is well upstream of the +1 nucleosome, and the pause is released before RNAPII starts transcribing into the +1 nucleosome (Aoi et al. 2020; Kwak et al. 2013; Williams et al. 2015). Therefore, one possibility is that DSIF is a negative factor before encountering any nucleosome, and thus its negative function is not related to nucleosome dynamics. However, as we showed and discussed above, the high RNAPII signal obtained by run-on based methods, which is accepted to be due to short term pauses, could originate from a downstream backtracked/stalled RNAPII (detected by NET-seq). This means that the starting position of the pause might be hindered by the tools used to study it, and it remains possible for the +1 nucleosome being, at least partially or in some cases, related to the establishment of the promoter proximal pause in mammals.

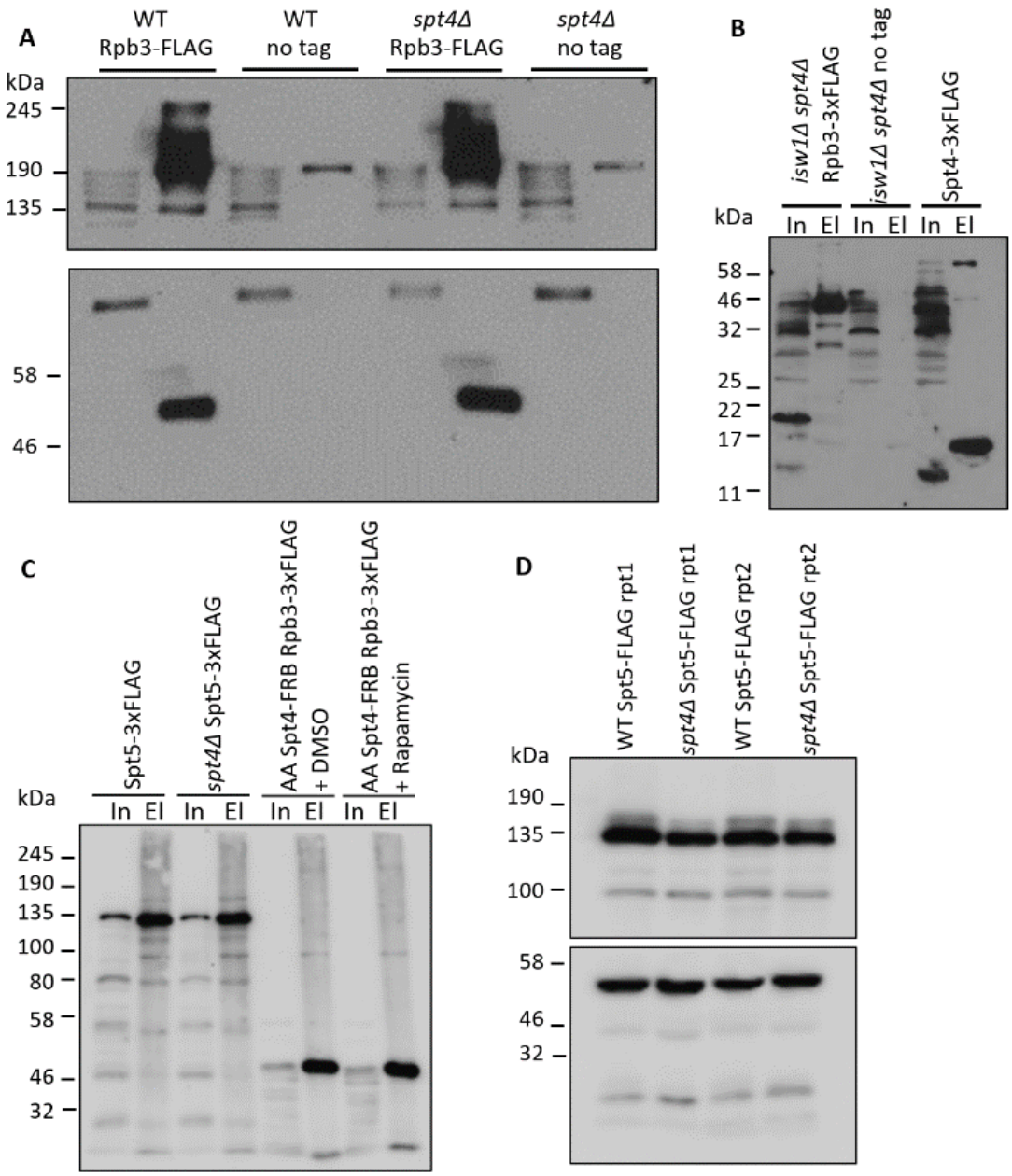
The whole mechanism of promoter proximal pausing or the function of the DSIF complex in transcription are far from being completely understood. However, the findings about the function of Spt4 presented here provides insight into the RNAPII transcription over the nucleosome barriers that is important for understanding the regulation of transcription in other organisms, in different contexts as well.

11 Appendix



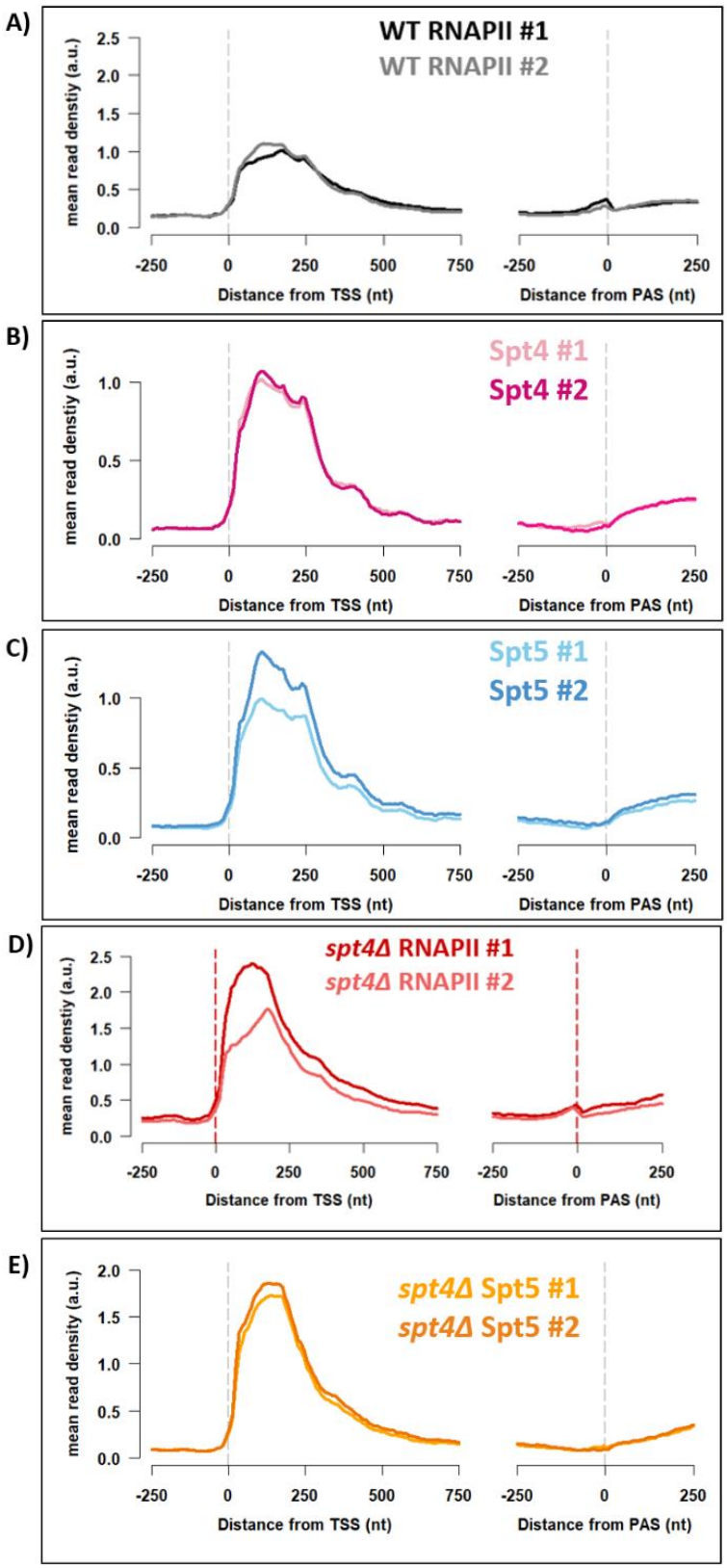
Appendix 1) Western blots

A, B &C) Against FLAG D&E) Against GFP

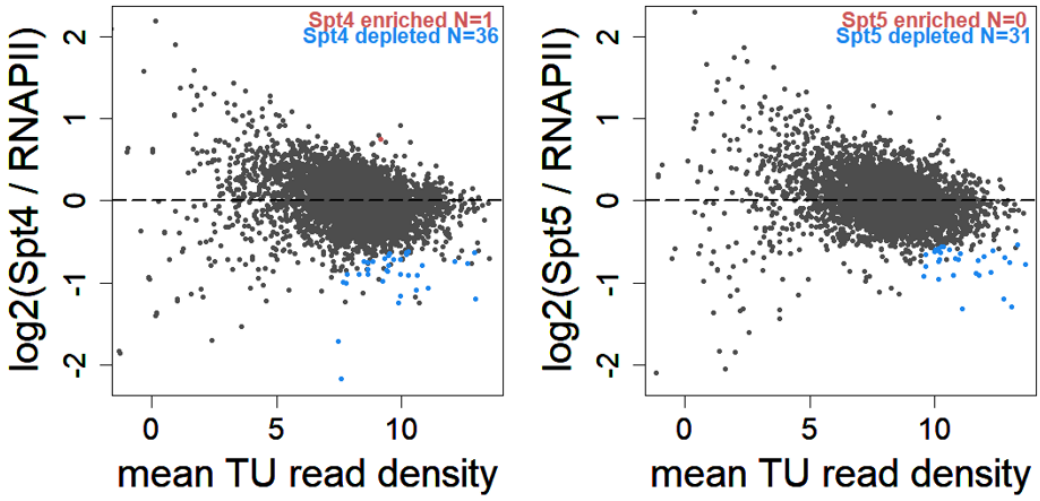


Appendix 2) Western blots

A-bottom, B, C & D-top) against FLAG. A-top) against Ser5P. D-bottom) against Tubulin.

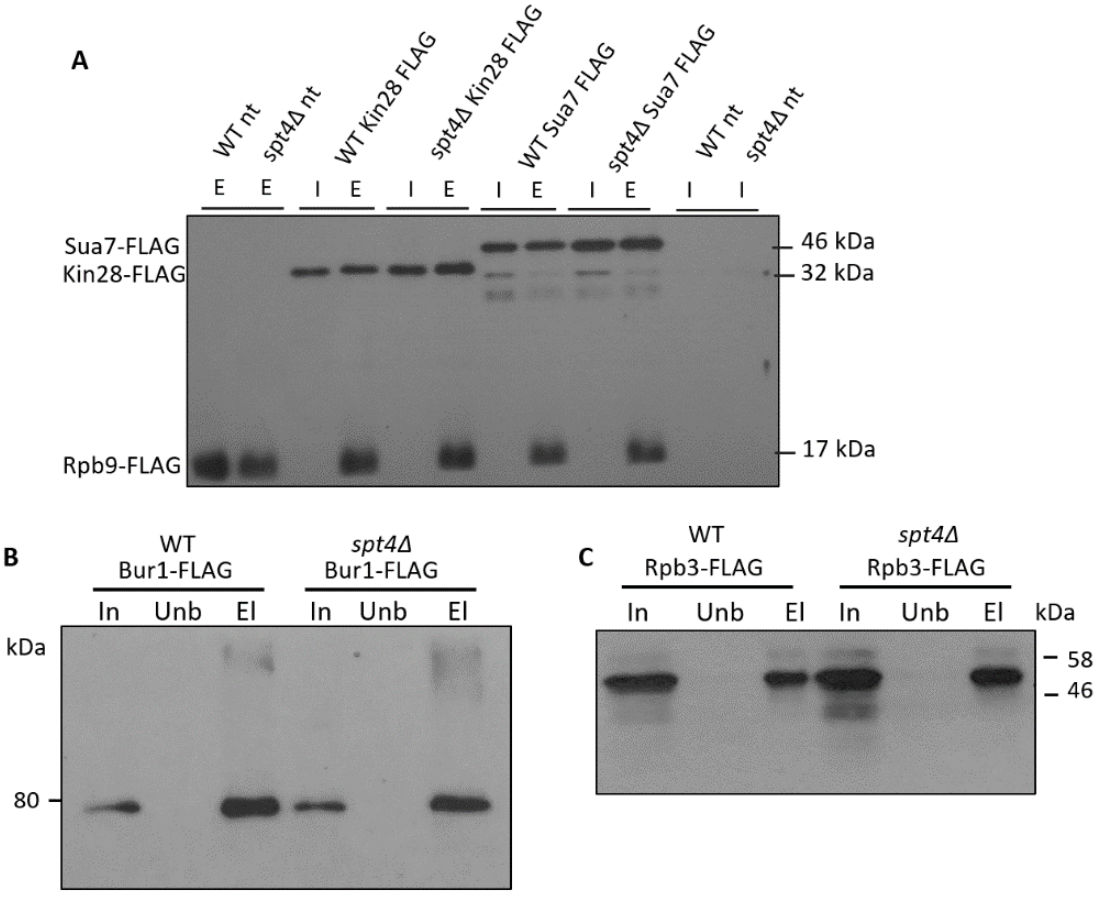


Appendix 3) NET&TEF-seq metagene profiles for each repeat



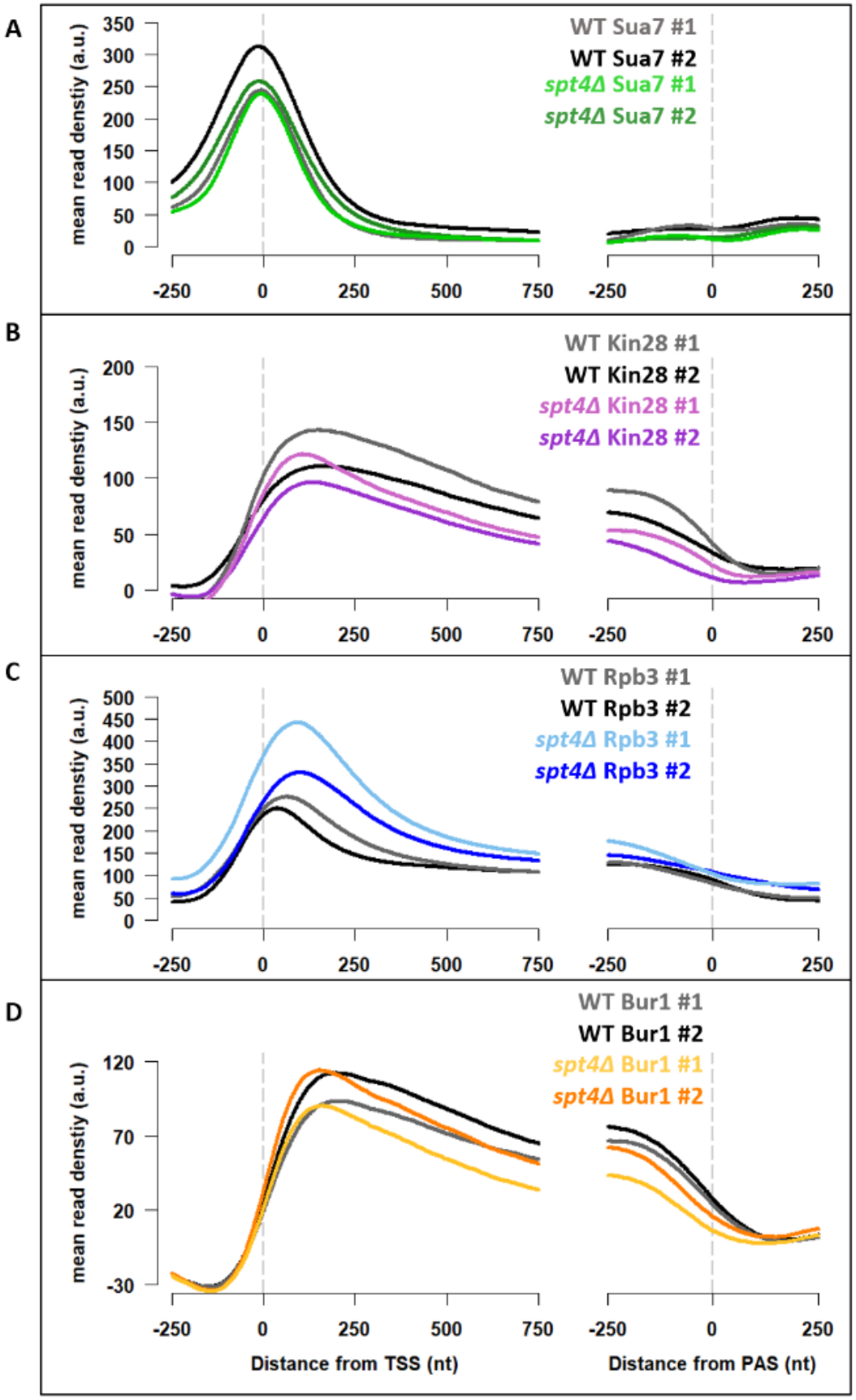
Appendix 4) Spt4 and Spt5 occupancies on RNAPII for all genes n=6037

Differential enrichment of Spt4 (top) and Spt5 (bottom) on RNAPII. DEseq2 applied to the read counts from the gene body (TSS to TSS-250 nt) for two replicates of each data. Significantly enriched and depleted genes indicated in red and blue, respectively (p-adjusted <0.05).

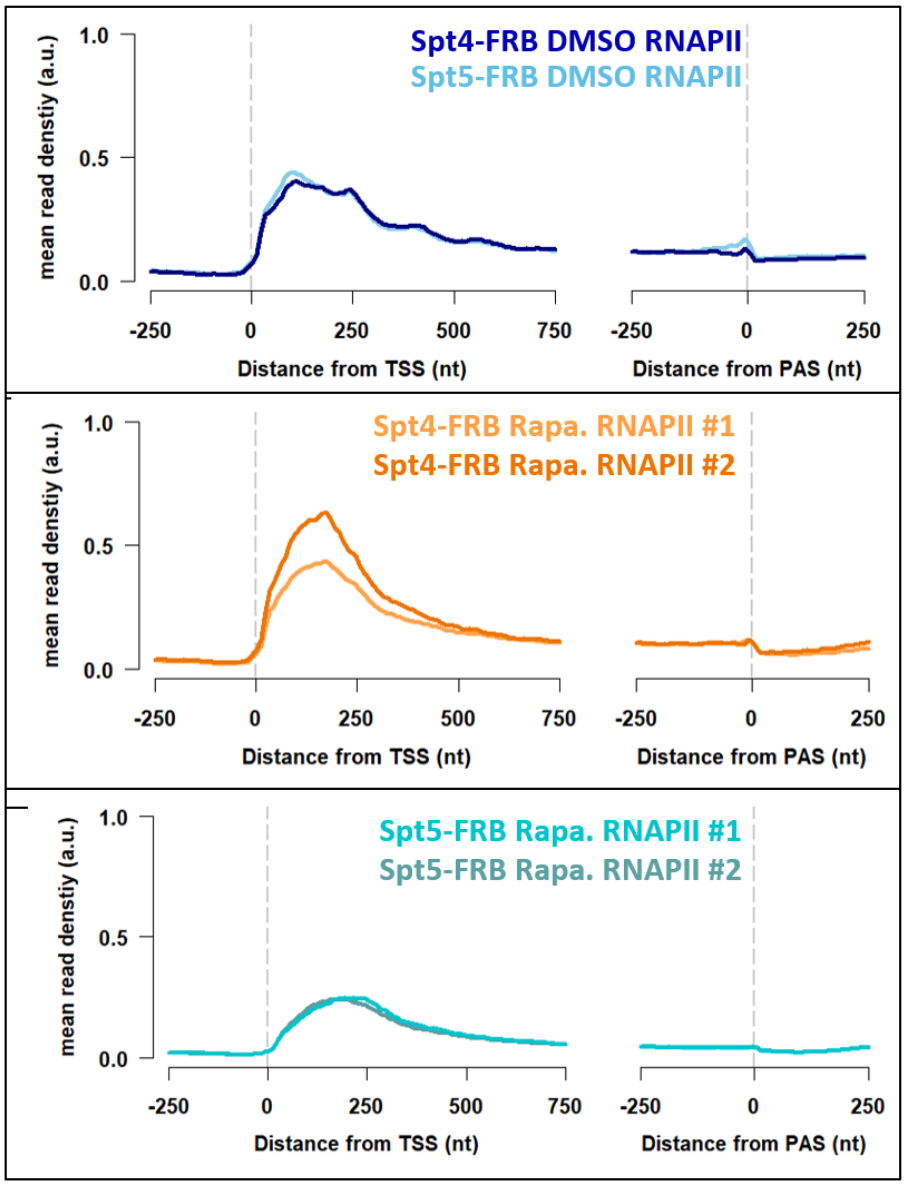


Appendix 5) Western blots of ChIP-seq IPs.

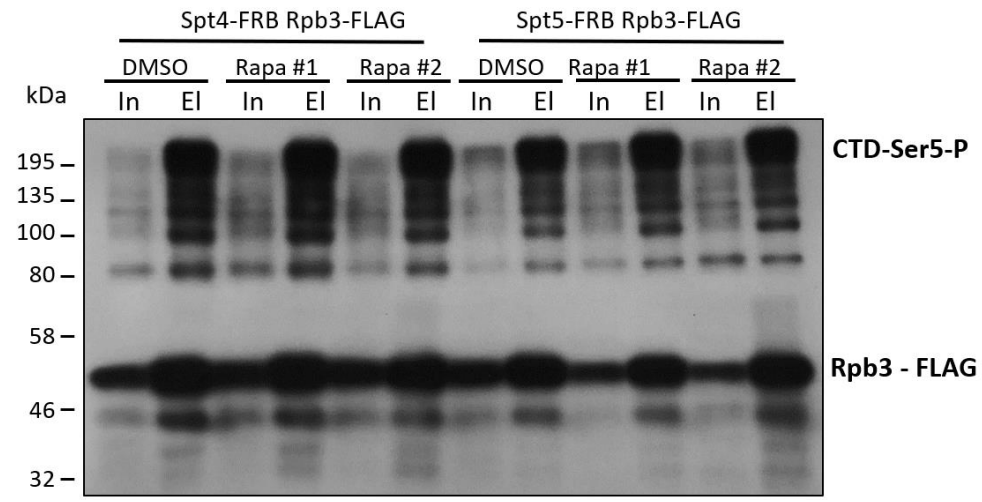
Against FLAG antibody



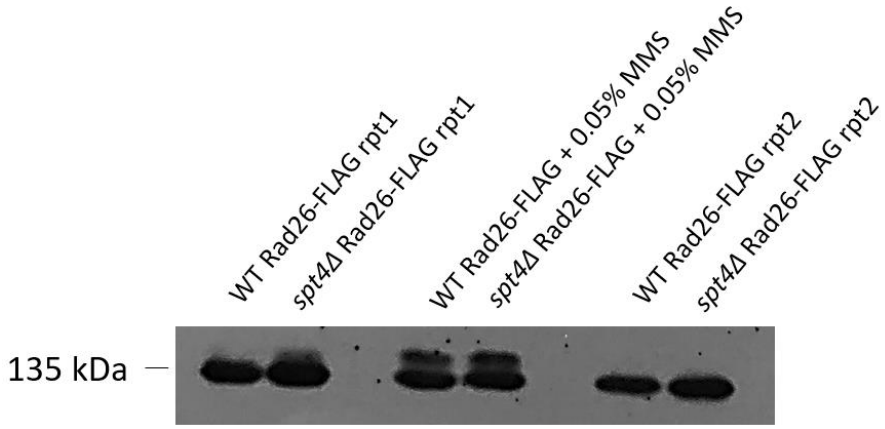
Appendix 6) Metagene plots of ChIP-seq repeats



Appendix 7) NET-seq metagene profiles of each repeats of DMSO or rapamycin treated samples

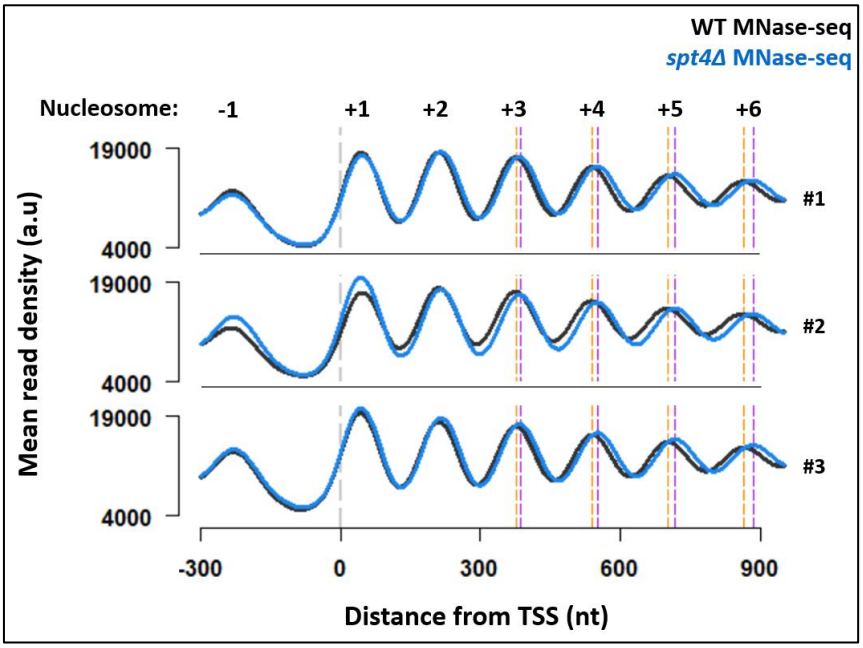


Appendix 8) Western blots of DMSO and rapamycin treated NET-seq IP samples

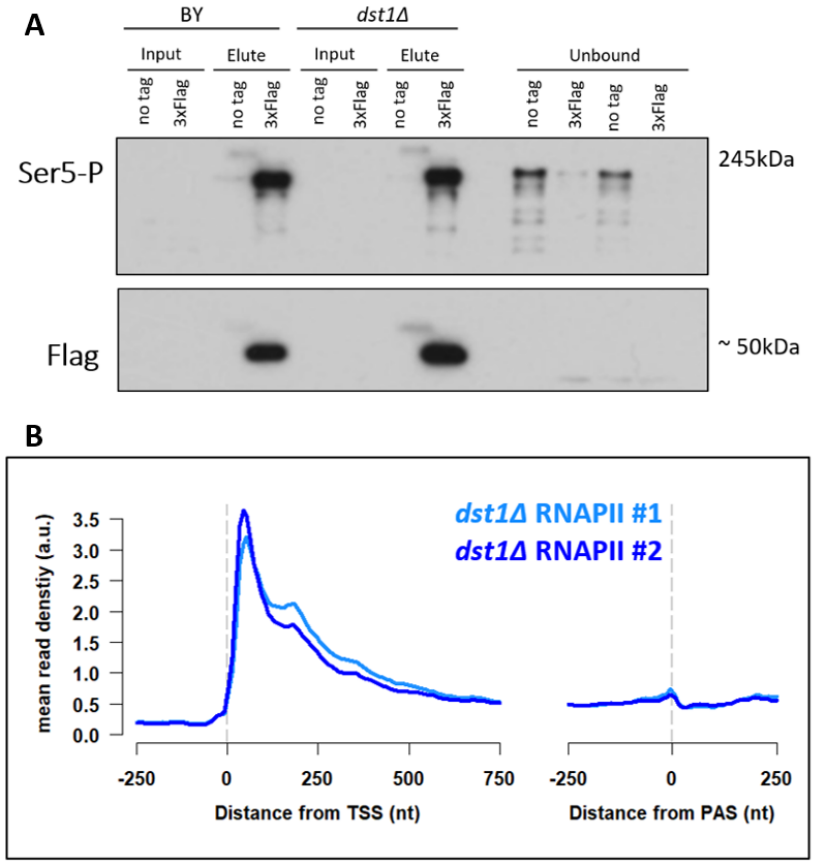


Appendix 9) Rad26 western blot to check phosphorylation.

0.05% MMS treated cells are expected to have phosphorylated Rad26 and serve as positive controls. No phosphorylation detected on Rad26 in WT cells. Rad26 seems to be phosphorylated in the absence of Spt4 in repeat 1, but not in repeat 2.



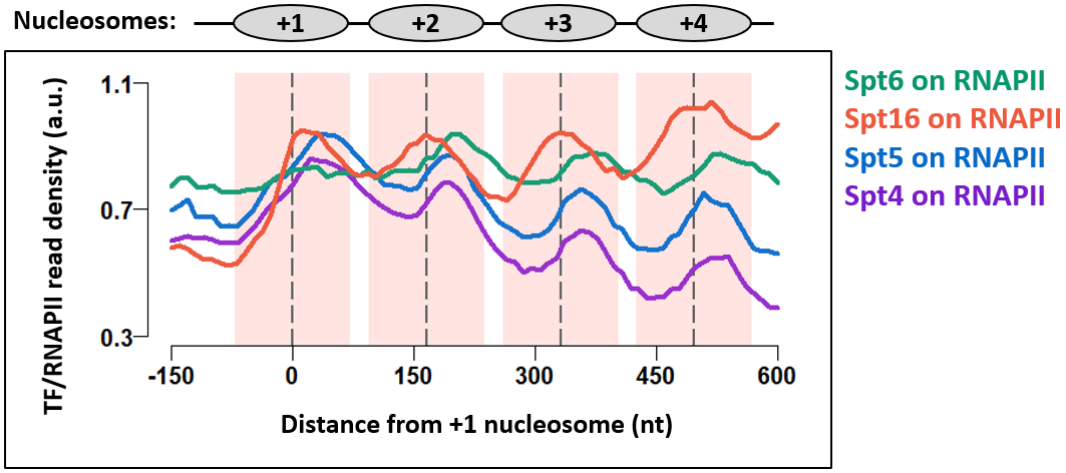
Appendix 10) MNase-seq metagene profiles for each repeat



Appendix 11) *dst1Δ* NET-seq is performed in duplicate.

A) Western blots of NET-seq IP samples against Ser5-P and FLAG.

B) Metagene profiles for each repeat



Appendix 12) TEF-seq metagene profiles

Metagene plots of Spt4 on RNAPII (purple), Spt5 on RNAPII (dark blue), Spt6 on RNAPII (green), and Spt16 on RNAPII (orange). Spt6, Spt16 TEF-seq data and their control RNAPII NET-seq data were taken from ArrayExpress: E-MTAB-4568.

Appendix table 1) NET-seq and TEF-seq read counts, alignment and spike-in ratios

Sample	Total reads	Uniquely aligned reads	% of uniquely aligned reads	% of *S.pombe* reads
Rpb3 #1	65825472	15542590	23.6	18.3
Rpb3 #2	63731533	15459921	24.3	18.0
Spt4 #1	53144178	9281278	17.5	27.6
Spt4 #2	73834296	15982773	21.7	24.5
Spt5 #1	64786577	13268751	20.8	23.9
Spt5 #2	47288902	8663001	18.3	23.1
WT no tag	65690548	12368537	18.8	69.5
***spt4Δ* Rpb3 #1**	66269665	14833491	22.4	12.1
***spt4Δ* Rpb3 #2**	87285765	20887292	23.9	15.1
***spt4Δ* no tag**	60251158	10589267	17.6	75.0
***dst1Δ* Rpb3 #1**	63588801	15854661	24.9	16.7
***dst1Δ* Rpb3 #2**	59471460	14179620	23.8	15.8
***dst1Δ* no tag**	52196209	9062729	17.4	68.9
***spt4Δ* Spt5 #1**	57895501	12712909	22.0	24.8
***spt4Δ* Spt5 #2**	43787211	8269081	18.9	22.3
Spt4-FRB Rapa #1	23665978	4738145	20.0	19.1
Spt4-FRB Rapa #2	15221859	3410956	22.4	16.6
Spt4-FRB DMSO	33480207	7026593	21.0	18.0
Spt5-FRB Rapa #1	22416495	3962731	17.7	28.2
Spt5-FRB Rapa #2	29302558	5623042	19.2	28.0
Spt5-FRB DMSO	25011250	5606801	22.4	17.3
No-FRB Rapa	86428986	18519429	21.4	19.1

Appendix table 2) Mass spec log fold enrichment analysis results

name	*spt4Δ* vs WT p.val	*spt4Δ* vs WT p.adj	*spt4Δ* vs WT ratio
SPT4	0.001	0.000	-3.690
SET2	0.021	0.000	-2.580
SPN1	0.047	0.003	-1.850
SPT16	0.032	0.001	-1.740
POB3	0.063	0.007	-1.490
HIS3	0.449	0.414	-1.470
GPN2	0.263	0.185	-1.260
SPT6	0.157	0.062	-1.250
TMA19	0.183	0.087	-1.170
SUB2	0.320	0.252	-1.150
PRP39	0.351	0.290	-1.110
YRA1	0.266	0.188	-1.100
PRP42	0.339	0.275	-1.070
RPB3	0.242	0.158	-1.060
BRE1	0.496	0.466	-1.020
RPC10	0.169	0.073	-0.989
WTM1	0.241	0.157	-0.986
ASR1	0.474	0.442	-0.985
HSH155	0.370	0.315	-0.953
TEX1	0.304	0.234	-0.835
SPT5	0.323	0.256	-0.767
RSE1	0.543	0.513	-0.754
SNU71	0.480	0.449	-0.718
SET1	0.718	0.645	-0.706
PRP40	0.514	0.485	-0.659
THO2	0.571	0.539	-0.599
SMB1	0.571	0.539	-0.587
RPB11	0.442	0.405	-0.566
RBA50	0.632	0.588	-0.563
SMD2	0.440	0.403	-0.547
DST1	0.692	0.630	-0.522
LUC7	0.759	0.668	-0.492
RPO21	0.548	0.518	-0.488
HPR1	0.571	0.538	-0.466
PRP19	0.571	0.539	-0.461
RPB2	0.575	0.542	-0.459
SNU114	0.722	0.648	-0.436
HTA2	0.576	0.543	-0.407
RPB5	0.622	0.580	-0.346
YHC1	0.707	0.638	-0.328
SMD3	0.688	0.627	-0.316
RAT1	0.714	0.643	-0.309
RTT103	0.685	0.625	-0.291
MFT1	0.757	0.667	-0.278
CBC2	0.731	0.653	-0.271
RPB4	0.733	0.654	-0.245
RPB8	0.735	0.655	-0.241
RPO26	0.816	0.697	-0.162
RPC19	0.864	0.717	-0.157
THP2	0.891	0.728	-0.136
TFG1	0.868	0.719	-0.127
RPB9	0.880	0.724	-0.113
SDC1	0.921	0.738	-0.100
SNU56	0.941	0.746	-0.087
CEG1	0.968	0.754	-0.031
MUD1	0.986	0.760	-0.025
STO1	0.991	0.761	-0.011

ESS1	0.994	0.762	-0.008
RAI1	0.985	0.759	0.030
RPB7	0.963	0.752	0.032
NPA3	0.964	0.753	0.040
TFG2	0.929	0.741	0.086
SMT3	0.794	0.686	0.178
CET1	0.772	0.675	0.215
RTR1	0.887	0.726	0.250
ABD1	0.774	0.676	0.320
GPN3	0.680	0.622	0.443
RAD23	0.620	0.579	0.460
TAF14	0.538	0.508	0.494
GLC7	0.492	0.462	0.588
CCL1	0.672	0.616	0.598
BUD27	0.416	0.374	0.729
BRR2	0.532	0.503	0.814
YTA7	0.281	0.207	1.270
ALD5	0.073	0.010	1.660
SSL1	0.083	0.013	1.720
SUA7	0.094	0.017	1.930
IWR1	0.102	0.020	1.960
TFB2	0.130	0.039	2.120
RAD3	0.098	0.019	2.210
TFB1	0.079	0.012	2.510
TFB3	0.015	0.000	2.600
SSL2	0.048	0.003	2.980
YKE2	0.026	0.000	4.120

12 References

Aoi, Yuki, Edwin R. Smith, Avani P. Shah, Emily J. Rendleman, Stacy A. Marshall, Ashley R. Woodfin, Fei X. Chen, Ramin Shiekhattar, and Ali Shilatifard. 2020. "NELF Regulates a Promoter-Proximal Step Distinct from RNA Pol II Pause-Release." *Molecular Cell* 78(2):261-274.e5.

Artsimovitch, Irina. 2018. "Rebuilding the Bridge between Transcription and Translation." *Molecular Microbiology* 108(5):467–72.

Baejen, Carlo, Jessica Andreani, Phillipp Torkler, Sofia Battaglia, Bjoern Schwalb, Michael Lidschreiber, Kerstin C. Maier, Andrea Boltendahl, Petra Rus, Stephanie Esslinger, Johannes Söding, and Patrick Cramer. 2017. "Genome-Wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes." *Molecular Cell* 66(1):38-49.e6.

Baejen, Carlo, Phillipp Torkler, Saskia Gressel, Katharina Essig, Johannes Söding, and Patrick Cramer. 2014. "Transcriptome Maps of MRNP Biogenesis Factors Define Pre-MRNA Recognition." *Molecular Cell* 55(5):745–57.

Baldi, Sandro, Philipp Korber, and Peter B. Becker. 2020. "Beads on a String—Nucleosome Array Arrangements and Folding of the Chromatin Fiber." *Nature Structural and Molecular Biology* 27(2):109–18.

Baldi, Sandro, Stefan Krebs, Helmut Blum, and Peter B. Becker. 2018. "Genome-Wide Measurement of Local Nucleosome Array Regularity and Spacing by Nanopore Sequencing." *Nature Structural and Molecular Biology* 25(9):894–901.

Baluapuri, Apoorva, Julia Hofstetter, Nevenka Dudvarski Stankovic, Theresa Endres, Pranjali Bhandare, Seychelle Monique Vos, Bikash Adhikari, Jessica Denise Schwarz, Ashwin Narain, Markus Vogt, Shuang-Yan Wang, Robert Düster, Lisa Anna Jung, Jens Thorsten Vanselow, Armin Wiegering, Matthias Geyer, Hans Michael Maric, Peter Gallant, Susanne Walz, Andreas Schlosser, Patrick Cramer, Martin Eilers, and Elmar Wolf. 2019. "MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation." *Molecular Cell* 74(4):674-687.e11.

Baptista, Tiago, Sebastian Grünberg, Nadège Minoungou, Maria J. E. Koster, H. T. Mar. Timmers, Steve Hahn, Didier Devys, and László Tora. 2017. "SAGA Is a General Cofactor for RNA Polymerase II Transcription." *Molecular Cell* 68(1):130-143.e5.

Bengtsson, Henrik. 2018. “Read and Write MAT Files and Call MATLAB from Within R, R Package ‘R.Matlab.’” 3.

Bernecky, Carrie, Jürgen M. Plitzko, and Patrick Cramer. 2017. “Structure of a Transcribing RNA Polymerase II–DSIF Complex Reveals a Multidentate DNA–RNA Clamp.” *Nature Structural & Molecular Biology* 24(10).

Bharati, Akhilendra Pratap, Neha Singh, Vikash Kumar, Md Kashif, Amit Kumar Singh, Priyanka Singh, Sudhir Kumar Singh, Mohammad Imran Siddiqi, Timir Tripathi, and Md Sohail Akhtar. 2016. “The MRNA Capping Enzyme of Saccharomyces Cerevisiae Has Dual Specificity to Interact with CTD of RNA Polymerase II.” *Scientific Reports* 6(August):1–12.

Bilokapic, Silvija and Mario Halic. 2019. “Nucleosome and Ubiquitin Position Set2 to Methylate H3K36.” *Nature Communications* 10(1):1–9.

Blankenberg, Daniel, Assaf Gordon, Gregory Von Kuster, Nathan Coraor, James Taylor, Anton Nekrutenko, and Galaxy Team. 2010. “Manipulation of FASTQ Data with Galaxy.” *Bioinformatics* 26(14):1783–85.

Blasco-Moreno, Bernat, Leire de Campos-Mata, René Böttcher, José García-Martínez, Jennifer Jungfleisch, Danny D. Nedialkova, Shiladitya Chattopadhyay, María Eugenia Gas, Baldomero Oliva, José E. Pérez-Ortín, Sebastian A. Leidel, Mordechai Choder, and Juana Díez. 2019. “The Exonuclease Xrn1 Activates Transcription and Translation of MRNAs Encoding Membrane Proteins.” *Nature Communications* 10(1):1–15.

Booth, Gregory T., Pabitra K. Parua, Miriam Sansó, Robert P. Fisher, and John T. Lis. 2018. “Cdk9 Regulates a Promoter-Proximal Checkpoint to Modulate RNA Polymerase II Elongation Rate in Fission Yeast.” *Nature Communications* 9(1):543.

Booth, Gregory T., Isabel X. Wang, Vivian G. Cheung, and John T. Lis. 2016. “Divergence of a Conserved Elongation Factor and Transcription Regulation in Budding and Fission Yeast.” *Genome Research* 26(6):799–811.

Bortvin, Alex and Fred Winston. 1996. “Evidence That Spt6p Controls Chromatin Structure by a Direct Interaction with Histones.” *Science* 272(5267):1473–76.

Bowman, Elizabeth A. and William G. Kelly. 2014. “RNA Polymerase II Transcription Elongation and Pol II CTD Ser2 Phosphorylation.” *Nucleus* 5(3):224–36.

Bresson, Stefan, Alex Tuck, Desislava Staneva, and David Tollervey. 2017. “Nuclear RNA Decay Pathways Aid Rapid Remodeling of Gene Expression in Yeast.” *Molecular Cell* 65(5):787-

800.e5.

Brown, Thomas. 2019. “Study of the Dynamics of Gene Expression by Mathematical Modelling and Systems Approaches Submitted for the Degree of Doctor of Philosophy.”

Brown, Thomas, Françoise S. Howe, Struan C. Murray, Meredith Wouters, Philipp Lorenz, Emily Seward, Scott Rata, Andrew Angel, and Jane Mellor. 2018. “Antisense Transcription-dependent Chromatin Signature Modulates Sense Transcript Dynamics.” *Molecular Systems Biology* 14(2):e8007.

Burriss, Katlyn Hughes and Amber L. Mosley. 2019. “Methods Review: Mass Spectrometry Analysis of RNAPII Complexes.” *Methods* 159–160(January):105–14.

Chen, Kaifu, Yuanxin Xi, Xuewen Pan, Z. Li, Klaus Kaestner, Jessica Tyler, Sharon Dent, Xiangwei He, and Wei Li. 2013. “DANPOS: Dynamic Analysis of Nucleosome Position and Occupancy by Sequencing.” *Genome Research* 23(2):341–51.

Chereji, Răzvan V., Terri D. Bryson, and Steven Henikoff. 2019. “Quantitative MNase-Seq Accurately Maps Nucleosome Occupancy Levels.” *Genome Biology* 20(1):198.

Chereji, Răzvan V. and David J. Clark. 2018. “Major Determinants of Nucleosome Positioning.” *Biophysical Journal* 114(10):2279–89.

Churchman, L. S. and J. S. Weissman. 2011. “Nascent Transcript Sequencing Visualizes Transcription at Nucleotide Resolution.” *Nature* 469(7330):368–73.

Clapier, Cedric R., Janet Iwasa, Bradley R. Cairns, and Craig L. Peterson. 2017. “Mechanisms of Action and Regulation of ATP-Dependent Chromatin-Remodelling Complexes.” *Nature Reviews Molecular Cell Biology* 18(7):407–22.

Cole, Hope A., Josefina Ocampo, James R. Iben, Razvan V. Chereji, and David J. Clark. 2014. “Heavy Transcription of Yeast Genes Correlates with Differential Loss of Histone H2B Relative to H4 and Queued RNA Polymerases.” *Nucleic Acids Research* 42(20):12512–22.

Core, Leighton and Karen Adelman. 2019. “Promoter-Proximal Pausing of RNA Polymerase II: A Nexus of Gene Regulation.” *Genes & Development* 33(15–16):960–82.

Cramer, P., D. A. Bushnell, and R. D. Kornberg. 2001. “Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution.” *Science* 292(5523):1863–76.

Crickard, J. Brooks, Jianhua Fu, and Joseph C. Reese. 2016. “Biochemical Analysis of Yeast Suppressor of Ty 4/5 (Spt4/5) Reveals the Importance of Nucleic Acid Interactions in the

Prevention of RNA Polymerase II Arrest." *Journal of Biological Chemistry* 291(19):9853–70.

Crickard, John B., Jaehyoun Lee, Tae-hee Lee, and Joseph C. Reese. 2017. "The Elongation Factor Spt4 / 5 Regulates RNA Polymerase II Transcription through the Nucleosome." 45(11):6362–74.

Cucinotta, Christine E., A. Elizabeth Hildreth, Brendan M. McShane, Margaret K. Shirra, and Karen M. Arndt. 2019. "The Nucleosome Acidic Patch Directly Interacts with Subunits of the Paf1 and FACT Complexes and Controls Chromatin Architecture in Vivo." *Nucleic Acids Research* 47(16):8410–23.

Ding, Baojin, Danielle LeJeune, and Shisheng 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(8):5317–26.

Doamekpor, Selom K., Ana M. Sanchez, Beate Schwer, Stewart Shuman, and Christopher D. Lima. 2014. "How an MRNA Capping Enzyme Reads Distinct RNA Polymerase II and Spt5 CTD Phosphorylation Codes." *Genes and Development* 28(12):1323–36.

Doris, Stephen M., James Chuang, Olga Viktorovskaya, Magdalena Murawska, Dan Spatt, L. Stirling Churchman, and Fred Winston. 2018. "Spt6 Is Required for the Fidelity of Promoter Selection." *Molecular Cell* 72(4):687-699.e6.

Dronamraju, Raghuvar and Brian D. Strahl. 2014. "A Feed Forward Circuit Comprising Spt6, Ctk1 and PAF Regulates Pol II CTD Phosphorylation and Transcription Elongation." *Nucleic Acids Research* 42(2):870–81.

Duan, Mingrui, Kathiresan Selvam, John J. Wyrick, and Peng Mao. 2020. "Genome-Wide Role of Rad26 in Promoting Transcription-Coupled Nucleotide Excision Repair in Yeast Chromatin." *Proceedings of the National Academy of Sciences of the United States of America* 1–9.

Eaton, Joshua D., Laura Francis, Lee Davidson, and Steven West. 2020. "A Unified Allosteric/Torpedo Mechanism for Transcriptional Termination on Human Protein-Coding Genes." *Genes & Development* 34(1–2):132–45.

Eden, Eran, Roy Navon, Israel Steinfeld, Doron Lipson, and Zohar Yakhini. 2009. "GOrilla: A Tool for Discovery and Visualization of Enriched GO Terms in Ranked Gene Lists." *BMC Bioinformatics* 10(1):48.

Ehara, Haruhiko, Tomoya Kujirai, Yuka Fujino, Mikako Shirouzu, Hitoshi Kurumizaka, and Shun ichi Sekine. 2019. “Structural Insight into Nucleosome Transcription by RNA Polymerase II with Elongation Factors.” *Science* 363(6428):744–47.

Ehara, Haruhiko, Takeshi Yokoyama, Hideki Shigematsu, Shigeyuki Yokoyama, Mikako Shirouzu, and Shun-ichi Sekine. 2017. “Structure of the Complete Elongation Complex of RNA Polymerase II with Basal Factors.” *Science* 8552(August):1–8.

Faitar, Silviu L., Seth A. Brodie, and Alfred S. Ponticelli. 2001. “Promoter-Specific Shifts in Transcription Initiation Conferred by Yeast TFIIB Mutations Are Determined by the Sequence in the Immediate Vicinity of the Start Sites.” *Molecular and Cellular Biology* 21(14):4427–40.

Fassler, Jan S. and Fred Winston. 1988. “Isolation and Analysis of a Novel Class of Suppressor of Ty Insertion Mutations In.” 212:203–12.

Feng, Jianxing, Tao Liu, Bo Qin, Yong Zhang, and Xiaole Shirley Liu. 2012. “Identifying ChIP-Seq Enrichment Using MACS.” *Nature Protocols* 7(9):1728–40.

Feng, Jianxun, Haiyun Gan, Matthew L. Eaton, Hui Zhou, Shuqi Li, Jason A. Belsky, David M. MacAlpine, Zhiguo Zhang, and Qing Li. 2016. “Noncoding Transcription Is a Driving Force for Nucleosome Instability in Spt16 Mutant Cells.” *Molecular and Cellular Biology* 36(13):1856–67.

Fischl, Harry, Françoise S. Howe, Andre Furger, and Jane Mellor. 2017. “Paf1 Has Distinct Roles in Transcription Elongation and Differential Transcript Fate.” *Molecular Cell* 685–98.

Fujiwara, Rina, Nivedita Damodaren, Jeremy E. Wilusz, and Kenji Murakami. 2019. “The Capping Enzyme Facilitates Promoter Escape and Assembly of a Follow-on Preinitiation Complex for Reinitiation.” *Proceedings of the National Academy of Sciences of the United States of America* 116(45):22573–82.

Gibbons, Brian J., Edward J. Brignole, Maia Azubel, Kenji Murakami, Neil R. Voss, David A. Bushnell, Francisco J. Asturias, and Roger D. Kornberg. 2012. “Subunit Architecture of General Transcription Factor TFIIH.” *Proceedings of the National Academy of Sciences of the United States of America* 109(6):1949–54.

Gregersen, Lea H., Richard Mitter, and Jesper Q. Svejstrup. 2020. “Using TTchem-Seq for Profiling Nascent Transcription and Measuring Transcript Elongation.” *Nature Protocols* 15(2):604–27.

Gregersen, Lea H. and Jesper Q. Svejstrup. 2018. “The Cellular Response to Transcription-Blocking DNA Damage.” *Trends in Biochemical Sciences* 43(5):327–41.

Grohmann, Dina, Julia Nagy, Anirban Chakraborty, Daniel Klose, Daniel Fielden, Richard H. Ebright, Jens Michaelis, and Finn Werner. 2011. “The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation.” *Molecular Cell* 43(2):263–74.

Grünberg, Sebastian and Steven Hahn. 2013. “Structural Insights into Transcription Initiation by RNA Polymerase II.” *Trends in Biochemical Sciences* 38(12):603–11.

Gurova, Katerina, Han Wen Chang, Maria E. Valieva, Poorva Sandlesh, and Vasily M. Studitsky. 2018. “Structure and Function of the Histone Chaperone FACT – Resolving FACTual Issues.” *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1861(9):892–904.

Gutiérrez, Gabriel, Gonzalo Millán-Zambrano, Daniel A. Medina, Antonio Jordán-Pla, José E. Pérez-Ortín, Xenia Peñate, and Sebastián Chávez. 2017. “Subtracting the Sequence Bias from Partially Digested MNase-Seq Data Reveals a General Contribution of TFIIS to Nucleosome Positioning.” *Epigenetics & Chromatin* 10(1):58.

Han, Yan and Yuan He. 2016. “Eukaryotic Transcription Initiation Machinery Visualized at Molecular Level.” *Transcription* 7(5):203–8.

Han, Zhong, Olga Jasnovidova, Nouhou Haidara, Agnieszka Tudek, Karel Kubicek, Domenico Libri, Richard Stefl, and Odil Porrua. 2020. “Termination of Non-coding Transcription in Yeast Relies on Both an RNA Pol II CTD Interaction Domain and a CTD-mimicking Region in Sen1.” *The EMBO Journal* e101548.

Harlen, Kevin M., Kristine L. Trotta, Erin E. Smith, Mohammad M. Mosaheb, Stephen M. Fuchs, and L. Stirling Churchman. 2016. “Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue.” *Cell Reports* 15(10):2147–58.

Hartzog, G. A., M. A. Basrai, S. L. Ricupero-Hovasse, P. Hieter, and F. Winston. 1996. “Identification and Analysis of a Functional Human Homolog of the SPT4 Gene of Saccharomyces Cerevisiae.” *Molecular and Cellular Biology* 16(6):2848–56.

Hartzog, G. A., T. Wada, H. Handa, and F. Winston. 1998. “Evidence That Spt4, Spt5, and Spt6 Control Transcription Elongation by RNA Polymerase II in Saccharomyces Cerevisiae.” *Genes & Development* 12(3):357–69.

Hartzog, Grant A. and Jianhua Fu. 2013. “The Spt4-Spt5 Complex: A Multi-Faceted Regulator of Transcription Elongation.” *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1829(1):105–15.

Haruki, Hirohito, Junichi Nishikawa, and Ulrich K. Laemmli. 2008. “Technique The Anchor-Away Technique : Rapid , Conditional Establishment of Yeast Mutant Phenotypes.” 925–32.

Heath, Catherine G., Nicolas Viphakone, and Stuart A. Wilson. 2016. “The Role of TREX in Gene Expression and Disease.” *Biochemical Journal* 473(19):2911–35.

Hirtreiter, Angela, Gerke E. Damsma, Alan C. M. Cheung, Daniel Klose, Dina Grohmann, Erika Vojnic, Andrew C. R. Martin, Patrick Cramer, and Finn Werner. 2010. “Spt4/5 Stimulates Transcription Elongation through the RNA Polymerase Clamp Coiled-Coil Motif.” *Nucleic Acids Research* 38(12):4040–51.

Hondele, Maria, Tobias Stuwe, Markus Hassler, Felix Halbach, Andrew Bowman, Elisa T. Zhang, Bianca Nijmeijer, Christiane Kotthoff, Vladimir Rybin, Stefan Amlacher, Ed Hurt, and Andreas G. Ladurner. 2013. “Structural Basis of Histone H2A-H2B Recognition by the Essential Chaperone FACT.” *Nature* 499(7456):111–14.

Hossain, M. A., C. Chung, S. K. Pradhan, and T. L. Johnson. 2013. “The Yeast Cap Binding Complex Modulates Transcription Factor Recruitment and Establishes Proper Histone H3K36 Trimethylation during Active Transcription.” *Molecular and Cellular Biology* 33(4):785–99.

Irizarry, Maintainer Rafael A. 2016. “Convenience Functions for Routine Data Exploration Description, R Package ‘ Rafalib .’”

Ivanovska, I., P. E. 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(3):531–41.

Jaenicke, Laura A., Björn von Eyss, Anne Carstensen, Elmar Wolf, Wenshan Xu, Ann Katrin Greifenberg, Matthias Geyer, Martin Eilers, and Nikita Popov. 2016. “Ubiquitin-Dependent Turnover of MYC Antagonizes MYC/PAF1C Complex Accumulation to Drive Transcriptional Elongation.” *Molecular Cell* 61(1):54–67.

Jansen, L. E., H. den Dulk, R. M. Brouns, M. de Ruijter, J. A. Brandsma, and J. Brouwer. 2000. “Spt4 Modulates Rad26 Requirement in Transcription-Coupled Nucleotide Excision

Repair." *The EMBO Journal* 19(23):6498–6507.

Jeronimo, Célia, Pierre Collin, and François Robert. 2016. "The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain." *Journal of Molecular Biology* 428(12):2607–22.

Jeronimo, Célia, Christian Poitras, and François Robert. 2019. "Histone Recycling by FACT and Spt6 during Transcription Prevents the Scrambling of Histone Modifications." *Cell Reports* 28(5):1206-1218.e8.

Jeronimo, Célia and François Robert. 2014. "Kin28 Regulates the Transient Association of Mediator with Core Promoters." *Nature Structural and Molecular Biology* 21(5):449–55.

Jonkers, I. and John T. Lis. 2015. "Getting up to Speed with Transcription Elongation by RNA Polymerase II." *Nat Rev Mol Cell Biol* 16(3):167–77.

Joo, Yoo Jin, Scott B. Ficarro, Yujin Chun, Jarrod A. Marto, and Stephen Buratowski. 2019. "In Vitro Analysis of RNA Polymerase II Elongation Complex Dynamics." *Genes & Development* 578–89.

Joo, Yoo Jin, Scott B. Ficarro, Luis M. Soares, Yujin Chun, Jarrod A. Marto, and Stephen Buratowski. 2017. "Downstream Promoter Interactions of TFIID TAFs Facilitate Transcription Reinitiation." *Genes and Development* 31(21):2162–74.

Kachaev, Zaur M., Lyubov A. Lebedeva, Eugene N. Kozlov, and Yulii V. Shidlovskii. 2020. "Interplay of MRNA Capping and Transcription Machineries." *Bioscience Reports* 0(January):1–10.

Kamieniarz-Gdula, Kinga and Nick J. Proudfoot. 2019. "Transcriptional Control by Premature Termination: A Forgotten Mechanism." *Trends in Genetics* 35(8):553–64.

Kang, Jin Young, Tatiana V. Mishanina, Robert Landick, and Seth A. Darst. 2019. "Mechanisms of Transcriptional Pausing in Bacteria." *Journal of Molecular Biology* 431(20):4007–29.

Kaya-Okur, Hatice S., Steven J. Wu, Christine A. Codomo, Erica S. Pledger, Terri D. Bryson, Jorja G. Henikoff, Kami Ahmad, and Steven Henikoff. 2019. "CUT&Tag for Efficient Epigenomic Profiling of Small Samples and Single Cells." *Nature Communications* 10(1):1–10.

Kecman, Tea, Krzysztof Kuś, Dong Hyuk Heo, Katie Duckett, Adrien Birot, Sabrina Liberatori, Shabaz Mohammed, Lucia Geis-Asteggiante, Carol V. Robinson, and Lidia Vasiljeva. 2018. "Elongation/Termination Factor Exchange Mediated by PP1 Phosphatase

Orchestrates Transcription Termination." *Cell Reports* 25(1):259-269.e5.

Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L. Salzberg. 2013. "TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions." *Genome Biology* 14(4):R36.

Kim, Dong Ki, Naoto Inukai, Tomoko Yamada, Akiko Furuya, Hiroe Sato, Yuki Yamaguchi, Tadashi Wada, and Hiroshi Handa. 2003. "Structure-Function Analysis of Human Spt4: Evidence That HSpt4 and HSpt5 Exert Their Roles in Transcriptional Elongation as Parts of the DSIF Complex." *Genes to Cells* 8(4):371–78.

Kokic, Goran, Aleksandar Chernev, Dimitry Tegunov, Christian Dienemann, Henning Urlaub, and Patrick Cramer. 2019. "Structural Basis of TFIIH Activation for Nucleotide Excision Repair." *Nature Communications* 10(1):1–9.

Kornberg, R. D. 1974. "Chromatin Structure: A Repeating Unit of Histones and DNA." *Science* 184(4139):868–71.

Kujirai, Tomoya, Haruhiko Ehara, Yuka Fujino, Mikako Shirouzu, Shun-ichi Sekine, and Hitoshi Kurumizaka. 2018. "Structural Basis of the Nucleosome Transition during RNA Polymerase II Passage." *Science* 362(6414):595–98.

Kujirai, Tomoya and Hitoshi Kurumizaka. 2020. "Transcription through the Nucleosome." *Current Opinion in Structural Biology* 61:42–49.

Kumar, Ananthanarayanan, Marcello Clerici, Lena M. Muckenfuss, Lori A. Passmore, and Martin Jinek. 2019. "Mechanistic Insights into MRNA 3'-End Processing." *Current Opinion in Structural Biology* 59:143–50.

Kwak, Hojoong, Nicholas J. Fuda, Leighton J. Core, and John T. Lis. 2013. "Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing." *Science* 339(6122):950–53.

Lai, William K. M. and B. Franklin Pugh. 2017. "Understanding Nucleosome Dynamics and Their Links to Gene Expression and DNA Replication." *Nature Reviews Molecular Cell Biology* 18(9):548–62.

Langmead, Ben. 2011. "Alignment with Bowtie." 1–24.

Lans, Hannes, Jan H. J. Hoeijmakers, Wim Vermeulen, and Jurgen A. Marteijn. 2019. "The DNA Damage Response to Transcription Stress." *Nature Reviews Molecular Cell Biology*

20(12):766–84.

Leng, Xueyuan, Maxim Ivanov, Peter Kindgren, Indranil Malik, Axel Thieffry, Peter Brodersen, Albin Sandelin, Craig D. Kaplan, and Sebastian Marquardt. 2020. “Organismal Benefits of Transcription Speed Control at Gene Boundaries.” *EMBO Reports* 21(4):1–16.

Li, Bing, Michael Carey, and Jerry L. Workman. 2007. “The Role of Chromatin during Transcription.” *Cell* 128(4):707–19.

Li, Heng. 2011. “Improving SNP Discovery by Base Alignment Quality.” *Bioinformatics* 27(8):1157–58.

Longtine, Mark S., Amos McKenzie, Douglas J. Demarini, Nirav G. Shah, Achim Wach, Arndt Brachat, Peter Philippsen, and John R. Pringle. 1998. “Additional Modules for Versatile and Economical PCR-Based Gene Deletion and Modification in Saccharomyces Cerevisiae.” *Yeast* 14(10):953–61.

Lorch, Yahli, Barbara Maier-Davis, and Roger D. Kornberg. 2014. “Role of DNA Sequence in Chromatin Remodeling and the Formation of Nucleosome-Free Regions.” *Genes and Development* 28(22):2492–97.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15(12):1–21.

Lu, Huasong, Yuahua Xue, Guoying K. Yu, Carolina Arias, Julie Lin, Susan Fong, Michel Faure, Ben Weisburd, Xiaodan Ji, Alexandre Mercier, James Sutton, Kunxin Luo, Zhenhai Gao, and Qiang Zhou. 2015. “Compensatory Induction of MYC Expression by Sustained CDK9 Inhibition via a BRD4-Dependent Mechanism.” *ELife* 4(JUNE2015):1–16.

Lu, Zhaolian and Zhenguo Lin. 2019. “Pervasive and Dynamic Transcription Initiation in Saccharomyces Cerevisiae.” *Genome Research* 29(7):1198–1210.

Ly, Elina, Abigail E. Powell, James A. Goodrich, and Jennifer F. Kugel. 2020. “Release of Human TFIIB from Actively Transcribing Complexes Is Triggered upon Synthesis of 7- and 9-Nt RNAs.” *Journal of Molecular Biology* 432(14):4049–60.

Malone, E. A., C. D. Clark, A. Chiang, and F. Winston. 1991. “Mutations in SPT16/CDC68 Suppress Cis- and Trans-Acting Mutations That Affect Promoter Function in Saccharomyces Cerevisiae.” *Molecular and Cellular Biology* 11(11):5710–17.

Malone, Elizabeth A., Jan S. Fassler, and Fred Winston. 1993. “Molecular and Genetic

Characterization of SPT4, a Gene Important for Transcription Initiation in Saccharomyces Cerevisiae." *Molecular & General Genetics : MGG* 237(3):449–59.

Martinez-Rucobo, F. W. and P. Cramer. 2013. "Structural Basis of Transcription Elongation." *Biochim Biophys Acta* 1829(1):9–19.

Martinez-Rucobo, Fuensanta W., Sarah Sainsbury, Alan C. M. Cheung, and Patrick Cramer. 2011. "Architecture of the RNA Polymerase-Spt4/5 Complex and Basis of Universal Transcription Processivity." *EMBO Journal* 30(7):1302–10.

Mason, Paul B. and Kevin Struhl. 2005. "Distinction and Relationship between Elongation Rate and Processivity of RNA Polymerase II in Vivo." *Molecular Cell* 17(6):831–40.

Mayekar, Manasi K., Richard G. Gardner, and Karen M. Arndt. 2013. " The Recruitment of the Saccharomyces Cerevisiae Paf1 Complex to Active Genes Requires a Domain of Rtf1 That Directly Interacts with the Spt4-Spt5 Complex ." *Molecular and Cellular Biology* 33(16):3259–73.

Mayer, Andreas, Heather M. Landry, and L. Stirling Churchman. 2017. "Pause & Go: From the Discovery of RNA Polymerase Pausing to Its Functional Implications." *Current Opinion in Cell Biology* 46:72–80.

Mayer, Andreas, M. Lidschreiber, M. Siebert, K. Leike, J. Söding, and P. Cramer. 2010. "Uniform Transitions of the General RNA Polymerase II Transcription Complex." *Nature Structural & Molecular Biology* 17(10):1272–78.

McCullough, Laura, Zaily Connell, Charisse Petersen, and Tim Formosa. 2015. "The Abundant Histone Chaperones Spt6 and FACT Collaborate to Assemble, Inspect, and Maintain Chromatin Structure in Saccharomyces Cerevisiae." *Genetics* 201(3):1030–45.

Mcdaniel, Stephen L. and Brian D. Strahl. 2017. "Shaping the Cellular Landscape with Set2 / SETD2 Methylation." *Cellular and Molecular Life Sciences* 74(18):3317–34.

Mcdaniel, Stephen L., Brian D. Strahl, Chapel Hill, Chapel Hill, and Chapel Hill. 2018. "Shaping the Cellular Landscape with Set2/SETD2 Methylation." 74(18):3317–34.

Millán-Zambrano, Gonzalo, Alfonso Rodríguez-Gil, Xenia Peñate, Lola de Miguel-Jiménez, Macarena Morillo-Huesca, Nevan Krogan, and Sebastián Chávez. 2013. "The Prefoldin Complex Regulates Chromatin Dynamics during Transcription Elongation." *PLoS Genetics* 9(9).

Milligan, Laura, Vân A. Huynh-Thu, Clémentine Delan-Forino, Alex Tuck, Elisabeth Petfalski, Rodrigo Lombraña, Guido Sanguinetti, Grzegorz Kudla, and David Tollervey. 2016. "Strand-specific, High-resolution Mapping of Modified RNA Polymerase II." *Molecular Systems Biology* 12(6):874.

Mischo, Hannah E. and Nick J. Proudfoot. 2013. "Disengaging Polymerase: Terminating RNA Polymerase II Transcription in Budding Yeast." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1829(1):174–85.

Missra, Anamika and David S. Gilmour. 2010. "Interactions between DSIF (DRB Sensitivity Inducing Factor), NELF (Negative Elongation Factor), and the Drosophila RNA Polymerase II Transcription Elongation Complex." *Proceedings of the National Academy of Sciences of the United States of America* 107(25):11301–6.

Morillon, Antonin, Nickoletta Karabetsou, Justin O'Sullivan, Nicholas Kent, Nicholas Proudfoot, and Jane Mellor. 2003. "Isw1 Chromatin Remodeling ATPase Coordinates Transcription Elongation and Termination by RNA Polymerase II." *Cell* 115(4):425–35.

Moteki, Shin and David Price. 2002. "Functional Coupling of Capping and Transcription of MRNA." 10:599–609.

Nemec, Corey M., Fan Yang, Joshua M. Gilmore, Corinna Hintermair, Yi Hsuan Ho, Sandra C. Tseng, Martin Heidemann, Ying Zhang, Laurence Florens, Audrey P. Gasch, Dirk Eick, Michael P. Washburn, Gabriele Varani, and Aseem Z. Ansari. 2017. "Different Phosphoisoforms of RNA Polymerase II Engage the Rtt103 Termination Factor in a Structurally Analogous Manner." *Proceedings of the National Academy of Sciences of the United States of America* 114(20):E3944–53.

Nishimura, Kohei and Masato T. Kanemaki. 2014. "Rapid Depletion of Budding Yeast Proteins via the Fusion of an Auxin-Inducible Degron (AID)." *Current Protocols in Cell Biology* 2014:20.9.1-20.9.16.

Nojima, Takayuki, Tomás Gomes, Maria Carmo-Fonseca, and Nicholas J. Proudfoot. 2016. "Mammalian NET-Seq Analysis Defines Nascent RNA Profiles and Associated RNA Processing Genome-Wide." *Nature Protocols* 11(3):413–28.

Ocampo, Josefina, Razvan V. Chereji, Peter R. Eriksson, and David J. Clark. 2016. "The ISW1 and CHD1 ATP-Dependent Chromatin Remodelers Compete to Set Nucleosome Spacing in Vivo." *Nucleic Acids Research* 44(10):4625–35.

Ocampo, Josefina, Razvan V. Chereji, Peter R. Eriksson, and David J. Clark. 2019. “Contrasting Roles of the RSC and ISW1/CHD1 Chromatin Remodelers in RNA Polymerase II Elongation and Termination.” *Genome Research* 29(3):407–17.

Osman, Sara and Patrick Cramer. 2020. “Structural Biology of RNA Polymerase II Transcription: 20 Years On.” 1–34.

Van Oss, S. Branden, Christine E. Cucinotta, and Karen M. Arndt. 2017. “Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation.” *Trends in Biochemical Sciences* 42(10):788–98.

Parenteau, Julie, Laurine Maignon, Mélodie Berthoumieux, Mathieu Catala, Vanessa Gagnon, and Sherif Abou Elela. 2019. “Introns Are Mediators of Cell Response to Starvation.” *Nature* 565(7741):612–17.

Parua, Pabitra K., Gregory T. Booth, Miriam Sansó, Bradley Benjamin, Jason C. Tanny, John T. Lis, and Robert P. Fisher. 2018. “A Cdk9-PP1 Switch Regulates the Elongation-Termination Transition of RNA Polymerase II.” *Nature* 558(7710):460–64.

Pavri, Rushad, Bing Zhu, Guohong Li, Patrick Trojer, Subhrangsu Mandal, Ali Shilatifard, and Danny Reinberg. 2006. “Histone H2B Monoubiquitination Functions Cooperatively with FACT to Regulate Elongation by RNA Polymerase II.” *Cell* 125(4):703–17.

Peck, Sarah A., Katlyn D. Hughes, Jose F. Victorino, and Amber L. Mosley. 2019. “Writing a Wrong: Coupled RNA Polymerase II Transcription and RNA Quality Control.” *Wiley Interdisciplinary Reviews: RNA* 10(4):1–22.

Pelechano, Vicent, Wu Wei, and Lars M. Steinmetz. 2013. “Extensive Transcriptional Heterogeneity Revealed by Isoform Profiling.” *Nature* 497(7447):127–31.

Ponting, C. P. 2002. “Novel Domains and Orthologues of Eukaryotic Transcription Elongation Factors.” *Nucleic Acids Research* 30(17):3643–52.

Porrua, Odil and Domenico Libri. 2015. “Transcription Termination and the Control of the Transcriptome: Why, Where and How to Stop.” *Nature Reviews Molecular Cell Biology* 16(3):190–202.

Prather, Donald, Nevan J. Krogan, Andrew Emili, Jack F. Greenblatt, and Fred Winston. 2005. “Identification and Characterization of Elf1, a Conserved Transcription Elongation Factor in Saccharomyces Cerevisiae.” *Molecular and Cellular Biology* 25(22):10122–35.

Proudfoot, Nick J. 2016. “Transcriptional Translation in Mammals: Stopping de RNA Polymerase II Juggernaut.” *Science* 352(6291):1–22.

Pugh, B. F. and R. Tjian. 1991. “Transcription from a TATA-Less Promoter Requires a Multisubunit TFIID Complex.” *Genes and Development* 5(11):1935–45.

Qiu, Chenxi, Chenxi Qiu, Huiyan Jin, Irina Vvedenskaya, Irina Vvedenskaya, Jordi Abante Llenas, Jordi Abante Llenas, Tingting Zhao, Indranil Malik, Indranil Malik, Alex M. Visbisky, Scott L. Schwartz, Ping Cui, Pavel Čabart, Pavel Čabart, Kang Hoo Han, William K. M. Lai, William K. M. Lai, Richard P. Metz, Charles D. Johnson, Sing Hoi Sze, Sing Hoi Sze, B. Franklin Pugh, B. Franklin Pugh, Bryce E. Nickels, Bryce E. Nickels, and Craig D. Kaplan. 2020. “Universal Promoter Scanning by Pol II during Transcription Initiation in Saccharomyces Cerevisiae.” *Genome Biology* 21(1):1–31.

Quan, Tiffani Kiyoko and Grant Ashley Hartzog. 2010. “Histone H3K4 and K36 Methylation, Chd1 and Rpd3S Oppose the Functions of Saccharomyces Cerevisiae Spt4-Spt5 in Transcription.” *Genetics* 184(2):321–34.

Ramanathan, Anand, G. Brett Robb, and Siu Hong Chan. 2016. “MRNA Capping: Biological Functions and Applications.” *Nucleic Acids Research* 44(16):7511–26.

Renner, Dan B., Yuki Yamaguchi, Tadashi Wada, Hiroshi Handa, and David H. Price. 2001. “A Highly Purified RNA Polymerase II Elongation Control System.” *Journal of Biological Chemistry* 276(45):42601–9.

Rhee, Ho Sung and B. Franklin Pugh. 2012. “Genome-Wide Structure and Organization of Eukaryotic Pre-Initiation Complexes.” *Nature* 483(7389):295–301.

Rondón, Ana G., María García-Rubio, Sergio González-Barrera, and Andrés Aguilera. 2003. “Molecular Evidence for a Positive Role of Spt4 in Transcription Elongation.” *Embo J* 22(3):612–20.

Rossi, Matthew J., William K. M. Lai, and B. Franklin Pugh. 2018. “Simplified ChIP-Exo Assays.” *Nature Communications* 9(1):1–13.

Sainsbury, Sarah, Carrie Bernecky, and Patrick Cramer. 2015. “Structural Basis of Transcription Initiation by RNA Polymerase II.” *Nature Reviews Molecular Cell Biology* 16(3):129–43.

Sathyan, Kizhakke Mattada, Brian D. McKenna, Warren D. Anderson, Fabiana M. Duarte, Leighton Core, and Michael J. Guertin. 2019. “An Improved Auxin-Inducible Degron System Preserves Native Protein Levels and Enables Rapid and Specific Protein

Depletion.” *Genes and Development* 33(19–20):1441–55.

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 of the United States of America* 103(34):12707–12.

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 of the United States of America* 103(34):12707–12.

Schüller, Roland, Ignasi Forné, Tobias Straub, Amelie Schreieck, Yves Texier, Nilay Shah, Tim Michael Decker, Patrick Cramer, Axel Imhof, and Dirk Eick. 2016. “Heptad-Specific Phosphorylation of RNA Polymerase II CTD.” *Molecular Cell* 61(2):305–14.

Shetty, Ameet, Scott P. Kallgren, Carina Demel, Kerstin C. Maier, Dan Spatt, Burak H. Alver, Patrick Cramer, Peter J. Park, and Fred Winston. 2017. “Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation.” *Molecular Cell* 66(1):77-88.e5.

Sims, Robert J., Rimma Belotserkovskaya, and Danny Reinberg. 2004. “Elongation by RNA Polymerase II: The Short and Long of It.” *Genes and Development* 18(20):2437–68.

Squazzo, Sharon L., Patrick J. Costa, Derek L. Lindstrom, Kathryn E. Kumer, Rajna Simic, Jennifer L. Jennings, Andrew J. Link, Karen M. Arndt, and Grant A. Hartzog. 2002. “The Paf1 Complex Physically and Functionally Associates with Transcription Elongation Factors in Vivo.” *EMBO Journal* 21(7):1764–74.

Stadelmayer, Bernd, Gaël Micas, Adrien Gamot, Pascal Martin, Nathalie Malirat, Slavik Koval, Raoul Raffel, Bijan Sobhian, Dany Severac, Stéphanie Rialle, Hugues Parrinello, Olivier Cuvier, and Monsef Benkirane. 2014. “Integrator Complex Regulates NELF-Mediated RNA Polymerase II Pause/Release and Processivity at Coding Genes.” *Nature Communications* 5.

Sun, Pei Feng, Wei Ta Fang, Li Ying Shin, Jyuan Yu Wei, Shih Feng Fu, and Jui Yu Chou. 2014. “Indole-3-Acetic Acid-Producing Yeasts in the Phyllosphere of the Carnivorous Plant Drosera Indica L.” *PLoS ONE* 9(12):1–22.

Swanson, M. S. and F. Winston. 1992. “SPT4, SPT5 and SPT6 Interactions: Effects on Transcription and Viability in Saccharomyces Cerevisiae.” *Genetics* 132(2):325–36.

Tachiwana, Hiroaki, Wataru Kagawa, Akihisa Osakabe, Koichiro Kawaguchi, Tatsuya Shiga, Yoko Hayashi-Takanaka, Hiroshi Kimura, and Hitoshi Kurumizaka. 2010. “Structural Basis of Instability of the Nucleosome Containing a Testis-Specific Histone Variant, Human H3T.” *Proceedings of the National Academy of Sciences of the United States of America* 107(23):10454–59.

Taschner, Michael, Michelle Harreman, Yumin Teng, Hefin Gill, Roy Anindya, Sarah L. Maslen, J. Mark Skehel, Raymond Waters, and Jesper Q. Svejstrup. 2010. “A Role for Checkpoint Kinase-Dependent Rad26 Phosphorylation in Transcription-Coupled DNA Repair in Saccharomyces Cerevisiae.” *Molecular and Cellular Biology* 30(2):436–46.

Thorvaldsdóttir, Helga, James T. Robinson, and Jill P. Mesirov. 2013. “Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration.” *Briefings in Bioinformatics* 14(2):178–92.

Tramantano, Michael, Lu Sun, Christy Au, Daniel Labuz, Zhimin Liu, Mindy Chou, Chen Shen, and Ed Luk. 2016. “Constitutive Turnover of Histone H2A.Z at Yeast Promoters Requires the Preinitiation Complex.” *ELife* 5(JULY):1–30.

Tufegdžić Vidaković, Ana, Richard Mitter, Gavin P. Kelly, Michelle Neumann, Michelle Harreman, Marta Rodríguez-Martínez, Anna Herlihy, Juston C. Weems, Stefan Boeing, Vesela Encheva, Liam Gaul, Laura Milligan, David Tollervey, Ronald C. Conaway, Joan W. Conaway, Ambrosius P. Snijders, Aengus Stewart, and Jesper Q. Svejstrup. 2020. “Regulation of the RNAPII Pool Is Integral to the DNA Damage Response.” *Cell* 180(6):1245-1261.e21.

Turtola, Matti and Georgiy A. Belogurov. 2016. “NusG Inhibits RNA Polymerase Backtracking by Stabilizing the Minimal Transcription Bubble.” *ELife* 5(OCTOBER2016):1–27.

Venkatesh, Swaminathan and Jerry L. Workman. 2015. “Histone Exchange, Chromatin Structure and the Regulation of Transcription.” *Nature Reviews Molecular Cell Biology* 16(3):178–89.

Vos, Seychelle M., Lucas Farnung, Henning Urlaub, and Patrick Cramer. 2018. “Structure of Paused Transcription Complex Pol II–DSIF–NELF.” *Nature* 560(7720):601–6.

Wada, Tadashi, Toshiyuki Takagi, Yuki Yamaguchi, Anwarul Ferdous, Takeshi Imai, Susumu

Hirose, Seiji Sugimoto, Keiichi Yano, Grant A. Hartzog, Fred Winston, Stephen Buratowski, and Hiroshi 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(3):343–56.

van der Weegen, Yana, Hadar Golan-Berman, Tycho E. T. Mevissen, Katja Apelt, Román González-Prieto, Joachim Goedhart, Elisheva E. Heilbrun, Alfred C. O. Vertegaal, Diana van den Heuvel, Johannes C. Walter, Sheera Adar, and Martijn S. Luijsterburg. 2020. “The Cooperative Action of CSB, CSA, and UVSSA Target TFIIH to DNA Damage-Stalled RNA Polymerase II.” *Nature Communications* 11(1):1–16.

Williams, Lucy H., George Fromm, Nolan G. Gokey, Telmo Henriques, Ginger W. Muse, Adam Burkholder, David C. Fargo, Guang Hu, and Karen Adelman. 2015. “Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks.” *Molecular Cell* 58(2):311–22.

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(2):179–97.

Winston, F., C. Dollard, E. A. Malone, J. Clare, J. G. Kapakos, P. Farabaugh, and P. L. Minehart. 1987. “Three Genes Are Required for Trans-Activation of Ty Transcription in Yeast.” *Genetics* 115(4):649–56.

Wong, Koon Ho, Yi Jin, and Kevin Struhl. 2014. “TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape.” *Molecular Cell* 54(4):601–12.

Xu, Jun, Indrajit Lahiri, Wei Wang, Adam Wier, Michael A. Cianfrocco, Jenny Chong, Alissa A. Hare, Peter B. Dervan, Frank DiMaio, Andres E. Leschziner, and Dong Wang. 2017. “Structural Basis for the Initiation of Eukaryotic Transcription-Coupled DNA Repair.” *Nature* 551(7682):653–57.

Yamaguchi, Yuki, Toshiyuki Takagi, Tadashi Wada, Keiichi Yano, Akiko Furuya, Seiji Sugimoto, Jun Hasegawa, and Hiroshi Handa. 1999. “NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation.” *Cell* 97(1):41–51.

Yesbolatova, A., Toyoaki Natsume, Ken ichiro Hayashi, and Masato T. Kanemaki. 2019. “Generation of Conditional Auxin-Inducible Degron (AID) Cells and Tight Control of Degron-Fused Proteins Using the Degradation Inhibitor Auxinole.” *Methods* 164–

165(January):73–80.

Zaborowska, Justyna, Sylvain Egloff, and Shona Murphy. 2016. “The Pol II CTD: New Twists in the Tail.” *Nature Structural and Molecular Biology* 23(9):771–77.

Zatreanu, Diana, Zhong Han, Richard Mitter, Emanuela Tumini, Hannah Williams, Lea Gregersen, A. Barbara Dirac-Svejstrup, Stefania Roma, Aengus Stewart, Andres Aguilera, and Jesper Q. Svejstrup. 2019. “Elongation Factor TFIIS Prevents Transcription Stress and R-Loop Accumulation to Maintain Genome Stability.” *Molecular Cell* 76(1):57-69.e9.

Zhang, Xiaofei, Arne H. Smits, Gabrielle B. A. Van Tilburg, Huib Ovaa, Wolfgang Huber, and Michiel Vermeulen. 2018. “Proteome-Wide Identification of Ubiquitin Interactions Using UbIA-MS.” *Nature Protocols* 13(3):530–50.

Zhou, Keda, Guillaume Gaullier, and Karolin Luger. 2019. “Nucleosome Structure and Dynamics Are Coming of Age.” *Nature Structural and Molecular Biology* 26(1):3–13.

Zorio, Diego A. R. and David L. Bentley. 2001. “Transcription Elongation: The ‘Foggy’ Is Lifting...” *Current Biology* 11(4):144–46.