Investigations into transcription and transcripts in *Saccharomyces cerevisiae*



Harry Fischl

Lincoln College

Trinity Term 2015

Doctor of Philosophy, University of Oxford

Investigations into transcription and transcripts in *Saccharomyces cerevisiae*

A thesis submitted to the University of Oxford for the degree of Doctor of Philosophy

Harry Fischl, Lincoln College, Trinity Term 2015

Word count: 45,750

Abstract

Transcription of the *Saccharomyces cerevisiae* genome by RNA polymerase II (Pol II) is pervasive, occurring on the sense and antisense strands of protein-coding genes and at intergenic regions. Factors associate with Pol II at particular positions along transcribed regions (transcription units (TUs)) with functions including facilitating the transition of Pol II through chromatin, controlling termination and processing the nascent transcript. By developing a new technique, these associations were mapped strand-specifically with nucleotide resolution. This revealed increased transitions in the binding of factors along TUs, periodic binding as Pol II transcribes through nucleosomes and differential binding of particular factors to particular types of TU. These associations are regulated to influence the stability and subcellular location of the transcripts produced. Analysis of transcription levels upon a change in conditions showed widespread and rapid changes resulting from changes in the proportion of cells in particular phases of the yeast metabolic cycle. At individual loci, clustered TUs showed reciprocal changes in transcription. Overlap between coding and non-coding TUs on both strands was shown to regulate these changes, repressing out of phase transcription. While the process of non-coding transcription is generally thought to be the regulatory agent, a role for non-coding transcripts should not be overlooked. Non-coding RNAs were shown to be able to bind double-stranded DNA sequence-specifically at several locations providing a mechanism by which these RNAs could interact with chromatin and mediate a function.

Acknowledgements

There are many people I would like to thank for making this thesis possible. Special thanks go to Jane Mellor for giving me the opportunity to join her lab and for her encouragement, guidance and support throughout the course of my studies. Next, I would like to thank past and present members of the lab for creating a very enjoyable place to work and for their assistance through any problems. Particular thanks go to Françoise Howe and Anitha Nair for their advice on many experiments and the smooth running of the lab. Thanks also to Ronja Woloszczuk and Struan Murray for their bioinformatics advice. I would also like to acknowledge Tania Nguyen and Françoise Howe for carrying out the Northern blot and RNA FISH experiments in Chapter 4.

As NET-seq has been central to much of the work in this thesis, thanks go to Stirling Churchman and to Shannon Hogan from her group for giving me support when setting this technique up in the lab. In addition, I am thankful for the financial support of the Wellcome Trust.

Finally, thanks to all my friends who have made for a great past four years and thanks to my parents who will almost certainly enjoy reading this thesis.

Table of Contents

1 Introduction 2
1.1 Gene Expression and Transcription 2
1.2 Pervasive transcription 3
1.3 Pol II structure 4
1.4 Transcription in the context of chromatin 6
1.5 Transcription initiation 8
1.6 Transcription elongation 10
1.6.1 Post-translational modification of Rpb1 10
1.6.2 The roles and recruitment profiles of elongation factors 13
1.7 Transcription termination 19
1.7.1 The PAS dependent pathway 19
1.7.2 The Nrd1-Nab3-Sen1 dependent (NNS) pathway 21
1.8 Splicing 22
1.9 RNA fate 26
1.9.1 Transcript release and RNA export 26
1.9.2 RNA degradation 29
1.10 The role of non-coding transcription 31
1.11 Aims 33

2 Methods 36
2.1 Yeast culture 36
2.2 Yeast strains 36
2.3 DNA extraction 38
2.4 RNA extraction 38
2.5 Reverse transcription 38
2.6 Western blotting 39
2.7 Nuclear Extraction 40
2.8 *In vitro* transcription 40
2.9 Triple-stranded structure formation assay 40
2.10 Restriction enzyme digestion protection assay 41
2.11 Complementary single stranded DNA competition assay 41
2.12 Northern blotting 41
2.13 RNA fluorescence in situ hybridisation (FISH) 42
2.14 NET-seq 43
2.14.1 Yeast culture 43
2.14.2 Immunoprecipitation 43
2.14.3 Library generation 44
2.14.4 Sequencing 46
2.15 Bioinformatics 47
2.15.1 NET-seq/TEF-seq data processing 47
2.15.2 mRNA and ncRNA annotations 48
2.15.3 Metagene analysis 48
2.15.4 Differential occupancy analysis 50
2.15.5 Differential transcription analysis 51

3 Mapping the location of transcription elongation complex components genome-wide 53
3.1 Introduction 53
3.1.1 Chromatin immunoprecipitation 53
3.1.2 Problems with ChIP 53
3.1.3 Improving on ChIP 55
3.1.4 Aim 56

3.2 Results 59
3.2.1 Establishing the standard NET-seq technique 59
3.2.2 Improving the NET-seq technique 62
3.2.3 Co-immunoprecipitation of Pol II with elongation factors 65
3.2.4 Applying the adapted NET-seq technique to elongation factors 69
3.2.5 Replicates show adapted NET-seq is highly reproducible 70
3.2.6 Analysis of factor occupancy profiles shows the advantages and disadvantages compared to ChIP 72
3.2.7 Pol II-normalised elongation factor average profiles reveal periodic factor binding and an increased number of transitions in factor binding 86
3.2.8 TU length does not affect the position of factor binding transitions but affects the level of factors at 3' ends 97
3.2.9 Splicing intermediates co-immunoprecipiate with Pol II 100
3.2.10 Cet1 co-immunoprecipitates a TEC without splicing intermediates 105
3.2.11 Differential factor binding to non-coding and protein- coding gene TUs 106
3.2.12 Differential binding of factors relative to Pol II to particular protein coding TUs 113
3.2.13 FLAG-tagged strains show no growth or expression differences 116
3.2.14 mRNAs synthesised from enriched Paf1 TUs are enriched for Yra1 and Mex67 binding 117
3.2.15 mRNAs synthesised from Paf1-enriched TUs accumulate in the nucleus in a *Δpaf1* strain 118
3.3 Discussion 123

4 The dynamics and regulation of transcriptional changes upon a change in environmental conditions 130
4.1 Introduction 130
4.1.1 Co-ordinated regulation of the expression of neighbouring genes 130
4.1.2 Regulation by overlapping transcription units 130
4.1.3 Problems with population studies in steady state conditions 131
4.1.4 The yeast metabolic cycle 131
4.1.5 The environmental stress response 134
4.1.6 Regulation of the YMC 135
4.1.7 Aims 135
4.2 Results 136
4.2.1 Transcriptional changes upon a change in carbon source have links to the YMC 136
4.2.2 Gene clusters with transcriptional changes upon a shift to galactose share common features 138
4.2.3 Transcription patterns at the *HMS2-BAT2* locus 142
4.2.4 *HMS2* and *SUT650* show temporally segregated expression 143
4.2.5 *HMS2* transcription into the promoter of *BAT2* represses *BAT2* transcription 145
4.2.6 A model for transcriptional regulation at the *HMS2-BAT2* locus 147
4.2.7 The dynamics of transcriptional changes 148
4.3 Discussion 162

5 Assessing the ability of RNA to act by binding double-stranded DNA. 168
5.1 Introduction 168
5.1.1 RNAs and chromatin 168
5.1.2 Non-coding transcript function in *S. cerevisiae* 168
5.1.3 Non-coding transcript function in other organisms 170
5.1.4 Mechanisms through which ncRNAs may interact with DNA 171

5.1.5 Transcription of a ncRNA regulates recombination at the rDNA locus *in S. cerevisiae* 172
5.1.6 Aims 174
5.2 Results 175
5.2.1 RNA interacts with double-stranded DNA, sequence-specifically, at the rDNA locus 175
5.2.2 The RNA:DNA interaction is strongest over a short approximately 30 bp region 178
5.2.3 There is an optimal DNA length for triple-stranded structure formation 181
5.2.4 The interaction of RNA with dsDNA increases logarithmically with increasing RNA concentration 182
5.2.5 Using RNase H to determine the type of triple-stranded structure formed 185
5.2.6 Triple-stranded structure formation at the *GAL1-10* locus 186
5.3 Discussion 187
6 Discussion 191
6.1 Summary 191
6.2 Future directions 192
6.3 Closing remarks 194
7 Appendix 195
8 Publications 196
9 References 197

Abbreviations

Abbreviation	Term in full
ATP	adenosine tri-phosphate
bp	base pair
BSA	bovine serum albumin
CBC	cap binding complex
CF	cleavage factor
CHART	capture hybridisation of RNA targets
ChIP	chromatin immunoprecipitation
ChIRP-seq	chromatin isolation by RNA purification
CID	CTD interacting domain
CPF	cleavage and polyadenylation factor
CRE	*cis*-regulatory element
CTD	c-terminal domain (of Rpb1)
CTR	c-terminal repeat domain (of Spt5)
CUT	cryptic unstable transcript
Cy3	cyanine 3 fluorescent dye
Cy5	cyanine 5 fluorescent dye
DSB	double-stranded break
dsDNA	double-stranded DNA
ERC	extra-chromasomal rDNA circle
ESCE	equal sister chromatid exchange
ESR	environmental stress response
FISH	fluorescence *in situ* hybridisation
GTF	general transcription factor
H2Bub1	mono ubiquitination of lysine 123 of Histone H2B
H3K4/36/79me1/2/3	mono-, di- or tri- methylation at lysine 4, 36 or 79 of histone H3
IGS1/2	intergenic spacer region 1 or 2
iSp18	18 carbon spacer
ITC	initially transcribing complex
KOAc	potassium acetate
NaOAc	sodium acetate
NDR	nucleosome depleted region
NET-seq	native elongating transcript sequencing
NMD	nonsense-mediated decay
NNS	Nrd1-Nab2-Sen1
Non-coding RNA	ncRNA
NPC	nuclear pore complex
NRD	Nrd1 complex
nt	nucleotide
OD_{600}	optical density measured at wavelength 600 nm
ORF	open reading frame
OX	oxidative phase of the YMC
Paf1C	Paf1 complex
PAR-CLIP	photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation

PAS	polyadenylation site
PCG	protein-coding gene
PIC	preinitiation complex
Pol	RNA polymerase
polyA	polyadenylation/polyadenylated
Pro3/6	proline 3 or 6
PTM	post-translational modification
RB	reductive building phase of the YMC
RC	reductive charging phase of the YMC
RFB	replication fork block
rRNA	ribosomal RNA
RT	reverse transcription
SAP	sense-antisense pair
Ser2/5/7P	serine 2, 5 or 7 phosphorylation/phosphorylated
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
ssDNA	single-stranded DNA
SUT	stable unannotated transcript
T1-10	elongation factor binding transition 1-10
TBP	TATA box binding protein
TEC	transcription elongation complex
TEF-seq	transcription elongation factor sequencing
Thr4P	threonine 4 phosphorylation/phosphorylated
TIF-seq	transcript isoform sequencing
TSS	transcription start site
TTS	transcription termination site
TU	transcription unit
Tyr1P	tyrosine 1 phosphorylation/phosphorylated
UESCE	unequal sister chromatid exchange
wt	wild-type
XUT	Xrn1-sensitive unstable transcript
YMC	yeast metabolic cycle
YPD	yeast extract, bactopeptone, glucose media
YPG	yeast extract, bactopeptone, galactose media

Chapter 1.

Introduction

1 Introduction

1.1 Gene Expression and Transcription

Gene expression is the process of converting genetic information stored within the sequence of DNA into a functional product. Transcription, the first step of this process, uses the DNA template to synthesise a corresponding RNA molecule. This molecule can then either act as the functional product in the case of non-coding RNA (ncRNA) or act as an intermediary, known as messenger RNA (mRNA), which acts as the template for the synthesis of a functional protein in translation. Complex networks regulate each step of gene expression. This ensures the cell can integrate the network of environmental signals to which it is exposed and express particular genes at levels and with timing necessary to mediate an appropriate response. Highlighting the importance of study in this area, misregulation of gene expression in higher eukaryotes results in many diseases from cancers to developmental disorders, while a greater understanding of gene expression will advance the ability to engineer organisms for technological processes in the growing field of synthetic biology.

As the initial step, transcription plays a highly regulated central role in the control of gene expression. The key enzyme at the centre of this process is a DNA-directed RNA polymerase of which, in eukaryotic cells, there are three: RNA polymerases I, II and III (Pol I, Pol II and Pol III). In *Saccharomyces cerevisiae* (*S. cerevisiae*), Pol I synthesises the polycistronic 35S ribosomal RNA (rRNA) transcript which makes up the majority of RNA in the cell. Pol III synthesises 5S rRNA as well as transfer RNAs (tRNAs), some small nuclear/nucleolar RNAs (sn/snoRNAs) and a few additional ncRNAs. Pol II synthesises all mRNAs and the majority of snoRNAs and other ncRNAs.

1.2 Pervasive transcription

The yeast transcriptome has been analysed genome-wide using tiling arrays and RNA-seq in numerous studies. These have revealed that the yeast genome is extensively transcribed with transcripts being synthesised from 85 % of the genome in rich media (David et al., 2006). As well as transcripts mapping to the sense strand of known protein coding genes (PCG) and well characterised ncRNAs such as rRNAs, tRNAs, snoRNAs and snRNAs, there are also many other, apparently non-coding transcripts mapping to intergenic regions and the antisense strand of PCGs, which have been termed stable unannotated transcripts (SUTs). Transcriptome analysis in wild type cells is unable to detect very unstable transcripts. By using strains deficient for particular RNA degradation proteins such transcripts are stabilised, leading to their detection. These have been classified according to the main pathway by which they are degraded, including cryptic unstable transcripts (CUTs) and Xrn1 unstable transcripts (XUTs), stabilised by deletion of the nuclear exosome subunit Rrp6 and the cytoplasmic 5'-3' exonuclease Xrn1, respectively (van Dijk et al., 2011; Neil et al., 2009; Xu et al., 2011).

Chromatin immunoprecipitation (ChIP) coupled to tiling arrays (ChIP-chip) or sequencing (ChIP-seq) has confirmed the presence of Pol II bound to much of the yeast genome (van Dijk et al., 2011; Steinmetz et al., 2006). The development of native elongating transcript sequencing (NET-seq), a technique which gives a strand-specific nucleotide resolution map of the position of engaged Pol II, has improved this (Churchman and Weissman, 2011; Nguyen et al., 2014), revealing extensive non-coding transcription at antisense and intergenic regions (Murray et al., 2012).

This thesis will focus on this pervasive transcription by Pol II in *S. cerevisiae*, including analysis of the mechanism of transcription, as well as exploring possible roles for non-coding transcription and the non-coding transcripts produced.

1.3 Pol II structure

Pol II consists of 12 subunits (Rpb1-12). The crystal structure has been elucidated in various conformations, including in the act of transcription elongation, in complex with RNA and DNA (Gnatt et al., 2001; Kettenberger et al., 2004). The largest subunits, Rpb1 and Rpb2, are positioned either side of a cleft which accommodates the incoming DNA. Charged residues positioned before the active site unwind the DNA duplex and feed the template strand into the active site, which is positioned below a pore for incoming nucleotides. Conserved structures, including the bridge helix and trigger loop, coordinate the nucleotide addition cycle consisting of selection of the appropriate free nucleotide, addition of this to the 3' end of the RNA and translocation of RNA and DNA to allow the binding of the next nucleotide (Martinez-Rucobo and Cramer, 2013). Upstream of the active site, an RNA:DNA hybrid forms for 8-9 bp, which is then separated, enabling the DNA to reanneal with the coding strand. The RNA is fed through the RNA exit channel which is in proximity to the C-terminal domain (CTD) of Rpb1 (Douziech et al., 1999). This largely unstructured domain, consisting of 26 repeats in *S. cerevisiae* (52 repeats in mammals) of a $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$ consensus sequence, is post-translationally modified and acts as a recruitment platform for various factors involved in transcriptional processes (more below) (Figure 1.1) (Eick and Geyer, 2013).

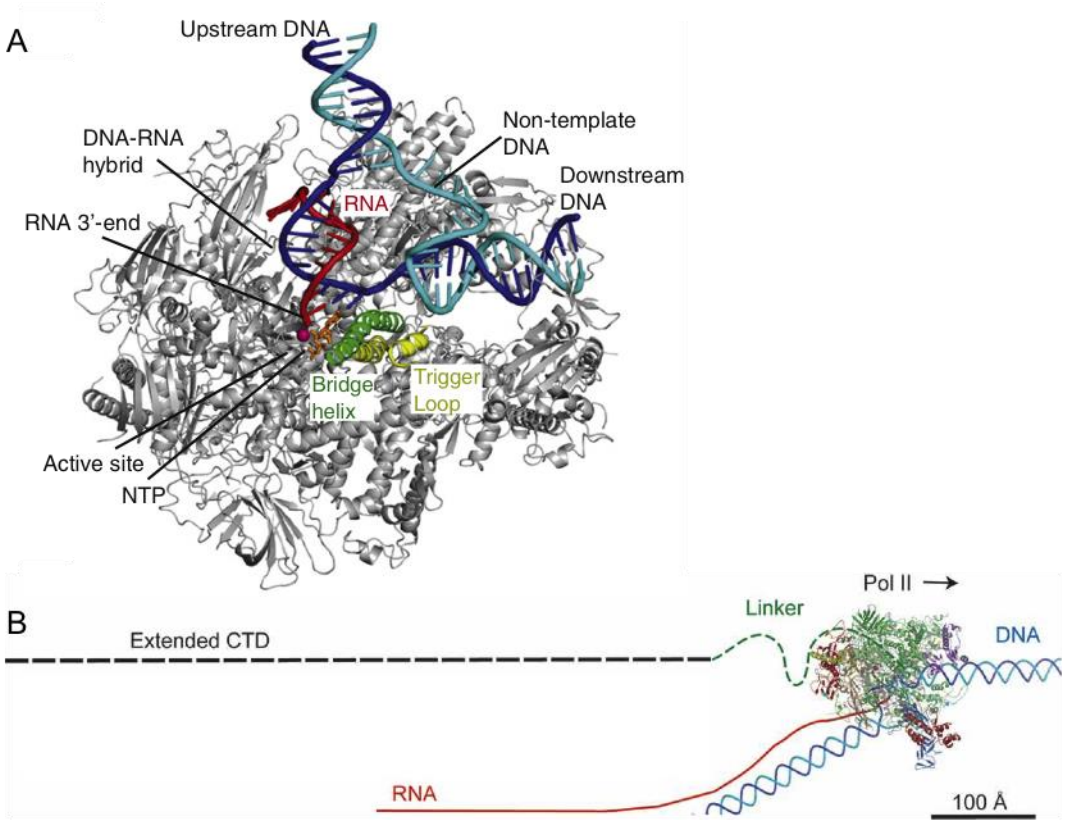


Figure 1.1. Structure of Pol II. A) Crystal structure of Pol II during transcription elongation. Figure adapted from Martinez-Rucobo et al., (2013). B) Schematic showing the Rpb1 CTD in an extended form. Figure adapted from Meinhart et al., (2005).

To simplify the analysis of Pol II transcription, it has been split into three stages: initiation, elongation and termination. Initiation consists of recruitment of Pol II to the DNA sequence at the start (the promoter) of the region to be transcribed (the transcription unit (TU)), selection of the transcription start site (TSS), and the transcription of its first few nucleotides. Elongation consists of conversion of this polymerase to a processive form and its progression to the end of the TU. Termination consists of the release of the transcript and disengagement of the polymerase from the TU.

1.4 Transcription in the context of chromatin

In order to understand the stages of transcription it is important to consider the DNA template upon which they occur. To fit the approximate 4 mm of *S. cerevisiae* DNA into the nuclear volume of ~3 μm^3 (Jorgensen et al., 2007) the DNA, as in all eukaryotic cells, is extensively packaged in the nucleus into a complex nucleoprotein structure known as chromatin. The core component of this is the histone consisting of an octamer of one H3 homodimer, one H4 homodimer and two H2A-H2B heterodimers. 146 bp of DNA is wrapped 1.65 times around the octamer to form a nucleosome with a total of 14 contact points between the positively charged histone surface and the negatively charged DNA (Luger et al., 1997). Nucleosomes occur in arrays resembling beads on a string with one located every 160-200 bp (Clark, 2010). These arrays are then further arranged into higher order structures ranging from more compacted (heterochromatin) to less compacted (euchromatin) states (Bi, 2014). In general, this packaging of DNA means its accessibility is reduced and the ability of Pol II to find sites to transcribe and progress along DNA is reduced (Izban and Luse, 1992).

Mapping of nucleosome locations has shown that their positions are partly determined by DNA sequence. As DNA is bent around the histone octamer, the minor groove is compressed when it faces into the histone octamer and stretched when it faces outwards (Richmond et al., 2003). Average nucleosomal DNA sequences show a ~10 bp periodic pattern, matching the helical repeat length of DNA, of AA/AT/TA/TT dinucleotides and CC/CG/GC/GG dinucleotides offset from each other by ~5 bp (Brogaard et al., 2012; Kaplan et al., 2009). The physical properties of these dinucleotides favour their placement in the compressed or stretched minor grooves, respectively (Travers and Klug, 1987). Additionally, other sequences are favoured at the centre of the nucleosome and others towards the edges. Furthermore, particular sequences such as polydA:polydT tracts are refractory to nucleosome binding, leading to

nucleosome depleted regions (NDR) (Kunkel and Martinson, 1981; Suter et al., 2000). As well as *cis*-acting signals, there are many *trans*-acting factors, which play a role in determining nucleosome locations, including DNA-binding factors, which can block nucleosome binding and chromatin remodelling enzymes, which use ATP to actively position nucleosomes (Hartley and Madhani, 2009; Yen et al., 2012). DNA sequence and *trans*-acting factors act in combination to determine the stability of nucleosome binding and preciseness with which each nucleosome is located (Hughes and Rando, 2014).

Despite the repressive nature of chromatin, it is ordered in such a way that transcription can still occur at specific locations. Fundamental to this is the typical pattern of nucleosomes along active TUs. This consists of a NDR of between ~30 bp to ~150 bp within the promoter flanked upstream and downstream by well-positioned nucleosomes, termed -1 and +1 nucleosomes respectively (Cairns, 2009; Zaugg and Luscombe, 2012). Downstream of the +1 nucleosome, nucleosomes adopt progressively less well-defined positions until near the site of termination, which is often in another NDR (Mavrich et al., 2008).

An additional layer of regulation is provided by post-translational modification of histones. A range of different modifications, including acetylation, methylation, phosphorylation and ubiquitination, can be added to each of the histone proteins (Bannister and Kouzarides, 2011). Primarily these occur on their highly conserved N-terminal tail regions, which protrude from the nucleosome core. Such modifications can influence the charge and structure of the tails altering their interactions with DNA and intra- and inter-nucleosomal interactions (Morales and Richard-Foy, 2000). This in turn can alter the compaction of chromatin and the stability and dynamics of nucleosome binding. Histone tails also act as platforms for the recruitment of factors to chromatin, including transcription factors, histone modifying enzymes and other chromatin remodelers (Yun et al., 2011). Particular combinations of modifications act to recruit

specific factors. As modifications occur with specific patterns along TUs, the location of factor recruitment can be controlled (Petty and Pillus, 2013; Rando, 2007; Smolle and Workman, 2013).

1.5 Transcription initiation

The promoter of a TU directs the recruitment of a network of factors necessary for initiation, known as the pre-initiation complex (PIC). The core of this promoter contains a "TATA box" sequence with consensus sequence TATA(A/T)A(A/T)(A/G), which provides the site for initial PIC formation. Although 80-90 % of PCG TU promoters have been termed "TATA less" promoters, 99 % of these have a sequence with two or less mismatches from this consensus (Rhee and Pugh, 2012a).

The general transcription factors (GTFs) TFIIA, TFIIB, TFIIE, TFIIF, TFIIH and the TATA box binding protein (TBP), either as part of the TFIID at "TATA less" promoters or the SAGA complex at TATA box promoters, are assembled on the promoter, centred on the TATA box (Basehoar et al., 2004). The TATA box is bound directly by TBP, which binds in the minor groove, inducing a large bend in the DNA (Kim et al., 1993). This weakens the DNA duplex structure, assisting later DNA melting. TFIIA and TFIIB bind DNA on either side of TBP, upstream and downstream respectively, stabilising the TBP-DNA interaction (Buratowski et al., 1989). Pol II is recruited along with the GTFs but is positioned in the PIC so that it does not directly contact promoter DNA (Murakami et al., 2013). The CTD of Rpb1 is largely unphosphorylated during recruitment (Phatnani and Greenleaf, 2006).

Once the PIC is assembled, several reactions occur within the complex to trigger transcription initiation. The activity of the ATP-dependent helicase component of TFIIH, Ssl2, acts to unwind promoter DNA and position it within the Pol II active site (Fishburn et al., 2015). In this transient, open complex, Pol II scans along the template strand and initiates transcription in a region from 40 bp to 120 bp downstream of the TATA box to

form the initially transcribing complex (ITC) (Cheung et al., 2011). Selection of the TSS is determined by DNA sequences upstream of the TSS and GTFs, including TFIIB, which inserts an N-terminal region termed the "B-finger" into the RNA exit channel of Pol II, contacting the DNA (Goel et al., 2012; Kostrewa et al., 2009; Lubliner et al., 2013). The position of the TSS can vary, leading to the production of different transcript isoforms from the same TU (Pelechano et al., 2013). The ITC needs to synthesise ~8 nt of RNA to form a stable transcription complex and can go through several rounds of abortive initiation before reaching this length (Liu et al., 2011b). The B-finger, which stabilises the formation of the initial 5 nt of RNA, sterically hinders further transcript extension and needs to be evicted (Liu et al., 2010; Saunders et al., 2006). A number of phosphorylations also occur during this initiation process within the PIC and to Pol II, including phosphorylation of Ser5 and 7 residues within the Rpb1 CTD by the TFIIH subunit, Kin28 (Akhtar et al., 2009). These alter the conformation of the PIC, assisting Pol II in breaking its contacts with the GTFs and processing away from the promoter, termed promoter clearance (Wong et al., 2014). Meanwhile the GTFs remain bound to the promoter, potentially enabling recruitment of another polymerase and re-initiation (Yudkovsky et al., 2000).

PIC recruitment, promoter DNA melting, ITC formation and stabilisation, promoter clearance and transcription re-initiation are all regulated to control the level of transcription initiating at a TU. Key components of this regulation are DNA-binding transcription factors, which are recruited predominantly upstream of the core promoter by specific sequences known as *cis*-regulatory elements (CREs) (Harbison et al., 2004). The sequence of each promoter varies greatly with different combinations and spacing of CREs, ensuring specific levels and patterns of binding (Segal and Widom, 2009). Factors can have transcriptionally activatory or repressive effects by influencing GTF recruitment and function. This is mediated through either binding the PIC directly or indirectly via Mediator, a multi-subunit complex, which wraps around the PIC (Hahn

and Young, 2011; Plaschka et al., 2015). These transcription factors can also recruit chromatin remodelling enzymes and histone modifying enzymes. By altering the positions of the +1 and -1 nucleosomes, the size of the NFR can be altered and CREs and core promoter elements can become exposed or occluded (Cairns, 2009). This movement can also alter the strength of binding of nucleosomes to DNA, altering the ability of Pol II to scan along DNA. The nucleosomes also act as platforms, which can be modified by enzymes to modulate PIC recruitment (Rhee and Pugh, 2012a; Venters and Pugh, 2009). Environmental signals can impact the level and binding properties of transcription factors, enabling adaptation of the transcriptional response (López-Maury et al., 2008; Morano et al., 2012).

1.6 Transcription elongation

Once the first ~25 nt of DNA have been transcribed, Pol II transitions into the elongation phase of transcription. This is characterised by the binding of many elongation factors, with functions in mRNA processing, chromatin modification and assisting progression through chromatin, to Pol II to form the transcription elongation complex (TEC). The composition of the TEC varies as Pol II progresses from the 5' to the 3' end of the TU dependent on the functions that need to be performed.

1.6.1 Post-translational modification of Rpb1

One mechanism controlling TEC composition is the occurrence of post-translational modifications (PTMs) within the complex. The most highly studied PTMs occur on the CTD of Rpb1. Tyr1, Ser2, 5 and 7 and Thr4 of this domain can all be reversibly phosphorylated while Pro3 and 6 can undergo *cis-trans* isomerisations (Buratowski, 2003; Chapman et al., 2007; Heidemann and Eick, 2012). The CTD is largely unstructured but is thought to transition from a more compact state when unphosphorylated to an extended state with increased phosphorylation (Meinhart et al., 2005). Regions are likely to become more structured upon binding by an induced fit

mechanism (Noble et al., 2005). Combinations of modifications promote the binding of certain factors while repressing the binding of others, ensuring particular factors bind the TEC at particular points along the TU (Eick and Geyer, 2013).

Using ChIP followed by hybridisation to microarrays, multiple groups have attempted to map the position of Rpb1 CTD PTMs genome-wide (Bataille et al., 2012; Kim et al., 2010; Mayer et al., 2010, 2012; Tietjen et al., 2010). To map these PTMs, several different antibodies have been used. The context in terms of surrounding modifications has been shown to influence the specificity and binding strength of particular antibodies (Chapman et al., 2007). Moreover, differences in the ChIP protocols used result in different profiles being obtained, leading to confusion over the distribution of certain PTMs. Ser2 (Ser2P) and Ser5 phosphorylation (Ser5P) have been most highly studied with Ser5P peaking at the start of the TSS before decreasing approximately 150 bp downstream of the TSS irrespective of the length of the gene. Ser2P increases from the start of the gene to around 600-1000 bp downstream of the TSS where the level stabilises before decreasing towards the end of the TU (Mayer et al., 2010). Discrepancies lie in the degree of Ser5P in the middle of the TU and the point at which Ser2P decreases at the 3' end of the TU. The 3E8 antibody detects Ser5P until just before the poly-adenylation site (PAS) (a site involved in transcription termination - see below), whilst the H14 antibody detects no Ser5P after the 5' end of the TSS (Bataille et al., 2012). The H5 antibody shows Ser2P decreases just before the PAS, whilst 3E10 shows Ser2P increases at this point before decreasing after the PAS (Bataille et al., 2012). The Ser7P profile obtained using the 4E12 antibody varies depending on the level of antibody used. High concentrations give a profile similar to the 3E8 Ser5P specific antibody, whilst lower concentrations give a level Ser7P profile across the TU (Bataille et al., 2012; Kim et al., 2010; Mayer et al., 2010). The Tyr1P profile obtained using the 3D12 antibody matches the H5 antibody obtained Ser2P profile, decreasing approximately 180 bp upstream of the PAS (Mayer et al., 2012) (Figure 1.2).

Discrepancies also lie in whether PTM profiles differ for different gene types. Both Mayer et al., (2010) and Bataille et al., (2012) show the profiles are uniform across all gene types, whilst Kim et al., (2010) show Ser7P is enriched on non-coding TUs.

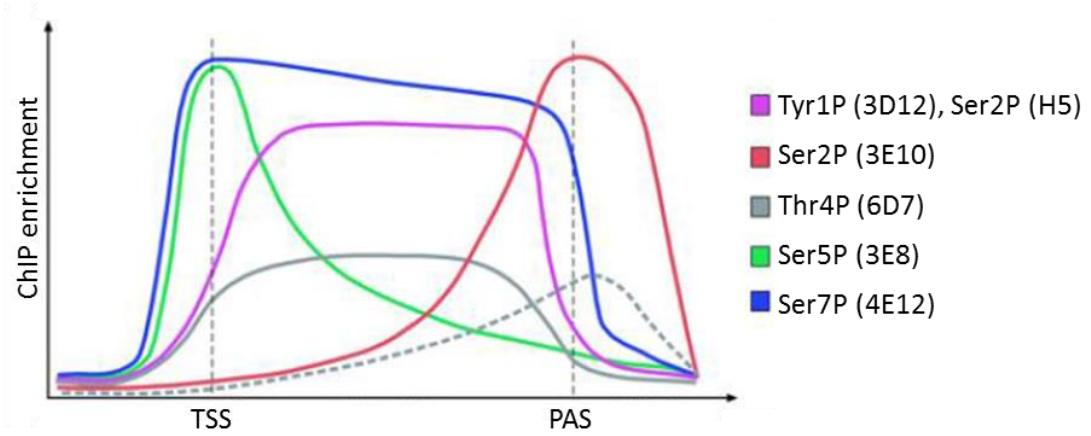


Figure 1.2. The profiles of Rpb1 CTD PTMs across a PCG TU. Schematic showing the metagene ChIP profiles obtained using the antibodies shown in brackets for the different CTD PTMs. The Ser7P profile shown is obtained when low 4E12 antibody concentrations are used. The grey dashed line is obtained for Thr4P in mammalian cells. Figure adapted from Eick and Geyer (2013).

The kinases and phosphatases responsible for these marks have also been identified and mapped. Kin28 phosphorylates Ser5 and Ser7 as Pol II initiates and maps to the promoter of the TU (Akhtar et al., 2009). Bur1 and Ctk1 are the main Ser2 kinases (Bowman and Kelly, 2014), although both are able to phosphorylate other residues *in vitro* and to some extent *in vivo* (Bataille et al., 2012). ChIP shows Bur1 starts to bind TUs just after the TSS and reaches maximal occupancy at around 200-300 bp downstream of the TSS, with Ctk1 having a similar profile whose maximal occupancy is downstream of Bur1 (Mayer et al., 2010). The CTD interacting domain (CID) of Bur1 recognises CTD repeats phosphorylated on Ser5 and preferentially phosphorylates Ser2 in such repeats (Qiu et al., 2009). Another role of Bur1 is in the phosphorylation of Ser1 of the C-terminal repeat domain (CTR) of the elongation factor Spt5 (Qiu et al., 2012). This factor binds to the Pol II clamp domain independent of the CTD of Rpb1 and its

CTR, consisting of 15 repeats, has a similar role to the CTD of Rpb1, also acting as a recruitment platform for factors to the TEC (Hartzog and Fu, 2013). Rtr1 is enriched at the transition point where Ser5P and Ser2P decrease and increase, respectively, and is responsible for Ser5 dephosphorylation at this point (Mosley et al., 2009). Ssu72 dephosphorylates both Ser5 and Ser7 and binds TUs around the TSS and PAS (Zhang et al., 2012). The activity of this enzyme is stimulated by the proline isomerase Ess1 and the crystal structure shows it bound to Pro6 in the minority *cis*-conformation (Krishnamurthy et al., 2009; Werner-Allen et al., 2011). Ess1 mutants show high Ser5P and Ser7P until the end of the TU (Bataille et al., 2012). Fcp1 dephosphorylates Ser2 and is detected bound to TUs just after Ctk1 starts to bind with an increase in binding around the PAS (Zhang et al., 2012). Evidence suggests that Fcp1 can only dephosphorylate Ser2 in repeats without other serine phosphorylation marks and *in vitro* the dephosphorylation of Ser5 and/or Ser7 by Ssu72 is required before Ser2 can be dephosphorylated in all repeats (Bataille et al., 2012). This may be the reason for the different H5 and 3E10 Ser2P antibody profiles observed around the PAS. The action of Ssu72 removing Ser5P in Ser2 phosphorylated repeats may lead to the decrease in H5 signal, which can only detect Ser2P in the vicinity of Ser5P (Chapman et al., 2007). After Ser5 and/or Ser7 has been dephosphorylated by Ssu72 in Ser2 phosphorylated repeats, Fcp1 is able to act on the remaining Ser2 phosphorylated repeats, leading to the decrease in 3E10 signal, which can detect Ser2P in more contexts, after the PAS. The kinase responsible for Tyr1P has not been identified but Glc7, which forms a complex with Ssu72, dephosphorylates this mark around the PAS (Shreiek et al., 2014).

1.6.2 The roles and recruitment profiles of elongation factors

1.6.2.1 RNA capping factors

The 5' ends of RNAs are capped by a sequence of reactions, which occur co-transcriptionally to the 5' nucleotide. First the terminal ɣ-phosphate is removed by Cet1, then a guanylyl group is added by Ceg1, and this guanine is then methylated by Abd1

(Cowling, 2010; Ghosh and Lima, 2010). The ability to form an RNA cap is essential and deletion of any of these enzymes is lethal (Mao et al., 1995; Shibagaki et al., 1992; Tsukamoto et al., 1997). Cet1 and Ceg1 form a stable complex and show near identical ChIP-chip occupancy profiles, both with sharp peaks in occupancy 20 bp downstream of the TSS. The Cet1-Ceg1 complex binds the Ser5 phosphorylated CTD (Cho et al., 1997; Gu et al., 2010), the unphosphorylated CTR of Spt5 (Lidschreiber et al., 2013) and to a binding site on the body of Pol II spanning the RNA exit channel (Martinez-Rucobo et al., 2015). This latter interaction is stabilised when RNA with a 5' triphosphate end protrudes from the exit channel, which starts to occur once 17 nt have been transcribed (Martinez-Rucobo et al., 2015). The model proposed is that these factors are first recruited to the TEC by Ser5P and Spt5 during promoter escape. They then show enhanced occupancy by interacting with the RNA exit channel as RNA begins to protrude from it. It has been shown in *Schizosaccharomyces pombe* (*S. pombe*) that the main function of Ser5P is to recruit the capping enzymes, as artificially tethering the human capping enzyme, which contains both Cet1 and Ceg1 functionality, to Rpb1 was sufficient to bypass the lethal effect of mutating the Ser5 residues in all CTD repeats to alanine (Schwer and Shuman, 2011). Abd1 is also recruited by Ser5P and the unphosphorylated CTR of Spt5 but shows a peak in occupancy a further 90 bp downstream of Cet1 and Ceg1 and thus may only show maximal recruitment once the guanylyl group is added (Lidschreiber et al., 2013; McCracken et al., 1997).

The cap binding complex (CBC) associates with capped RNA. This interaction requires the methyl group on the 5' guanine and so can only occur once the cap is completed (Topisirovic et al., 2011). Consequently CBC proteins start to bind TUs after the peak in Abd1 binding, showing maximal occupancy 200-300 bp downstream of the TSS. The CBC is involved in Bur1 and Ctk1 recruitment to TUs at this position (Gonatopoulos-Pournatzis and Cowling, 2014; Lidschreiber et al., 2013) and it is thought that this creates a checkpoint whereby RNA needs to be correctly capped before Ser2 can be

phosphorylated and in turn downstream factors can be recruited. Both removal of Ser5P by Rtr1 and Bur1 phosphorylation of the CTR and CTD assist in disrupting the interaction of the capping enzymes with the TEC (Lidschreiber et al., 2013).

1.6.2.2 Histone chaperones

In vitro, Pol II can transcribe naked DNA at similar rates to those observed *in vivo* but inefficiently transcribes a chromatin template (Izban and Luse, 1992). For this it needs a number of accessory factors including the Spt4/Spt5, Spt16/Pob3, and Spt6/Iws1 complexes (Endoh et al., 2004).

1.6.2.2.1 Spt4/5

Spt5 forms a complex with Spt4 and is conserved across all domains of life. As mentioned, it contains a C-terminal repeat domain (CTR) which acts as a recruitment platform much like the CTD of Rpb1. This complex competes with the TFIIE initiation factor for the same Pol II clamp domain binding site. Therefore Spt5 only associates with Pol II once it has cleared the promoter (Grohmann et al., 2011; Martinez-Rucobo et al., 2011).

1.6.2.2.2 Spt6/Iws1

Spt6 forms a complex with Iws1 and has been linked to roles in transcription initiation, elongation, termination and mRNA export (Duina, 2011; Mayer et al., 2012; Yoh et al., 2007). It has a non-canonical CTD interacting domain (CID), consisting of two tandem SH2 domains (tSH2). This domain has been shown *in vitro* to interact with Tyr1, Ser2, Ser5 and Ser7 phosphorylated, but not unphosphorylated, CTD peptides, showing the strongest affinity for Tyr1P (Liu et al., 2011a). Phosphorylation of tandem CTD repeats or doubly phosphorylated repeats also increases its affinity. Deletion of its CID, does not completely abolish the Spt6-Pol II interaction suggesting the protein can interact with other sites on Pol II (Mayer et al., 2010). Spt6 is a histone chaperone that can interact with H3 and H2A-H2B histone subunits with a preference for H3 and promote histone

assembly *in vitro* (Bortvin and Winston, 1996). A key function of Spt6 is thought to be in disrupting nucleosomes in front of the processing polymerase while reassembling them behind the polymerase, leading to a faster elongation rate (Ardehali et al., 2009). Mutations in Spt6 lead to depletion of histones from chromatin, enabling transcription to initiate from cryptic promoters in the body of genes (Kaplan et al., 2003).

1.6.2.2.3 Spt16/Pob3

Spt16 interacts with Pob3 to form the FACT (Facilitates Chromatin Transcription) complex. Like Spt6, FACT is able to interact with H2A-H2B dimers and H3-H4 tetramers although its preference is for H2A-H2B dimers (Belotserkovskaya et al., 2003; Hondele et al., 2013). By interacting with nucleosomes it can disrupt nucleosome-DNA contacts, possibly resulting in loss of one H2A-H2B dimer (Xin et al., 2009). As a result of this functionality, it can behave much like Spt6, assisting in disrupting and reassembling nucleosomes as Pol II processes. The exact mechanism of its recruitment to the TEC is poorly understood although it is present on chromatin from the TSS to just before the PAS (Mayer et al., 2010)

1.6.2.3 Histone modifying complexes

A large number of histone modifications occur across a TU with several of the modifying enzymes responsible being recruited via interactions with the TEC. This leads to stereotypical patterns of modifications across genes (Smolle and Workman, 2013). This thesis will focus on modifications brought about by Set2 and the Paf1 complex.

1.6.2.3.1 Set2

Set2 is the methyl-transferase responsible for H3K36 mono- (me1), di- (me2) and tri- (me3) methylation (Strahl et al., 2002). The protein binds Rpb1 via its CID, which recognises Ser2, Ser5 doubly phosphorylated Rpb1 CTD repeats (Krogan et al., 2003a; Vojnic et al., 2006). Although H3K36me2 is found on active TUs (Rao et al., 2005), only H3K36me3 is dependent on Set2 recruitment to the TEC (Youdell et al., 2008). The level

of H3K36me3 correlates with the level of transcription and rises across the TU peaking before the PAS (Pokholok et al., 2005). H3K36me recruits the histone deacetylase Rpd3S, which reduces histone acetylation in the body of TUs, repressing transcription initiation from cryptic promoters (Keogh et al., 2005). It has also been suggested that this mark stabilises nucleosomes, preventing exchange with new acetylated histones (Venkatesh et al., 2012); however, when H3K36me is blocked by mutating lysine 36 to arginine, high levels of histone exchange are not observed (Ferrari and Strubin, 2015).

1.6.2.3.2 The Paf1 complex

Paf1 combines with Ctr9, Cdc73, Leo1, and Rtf1 to form the Paf1 complex (Paf1C). Cdc73, Rtf1 and, to a lesser extent, Ctr9 all have Rpb1 CIDs that have a strong preference for CTD peptides with a di-phosphorylation mark of either Ser2P and Ser5P or Ser5P and Ser7P (Qiu et al., 2012). Paf1C is also recruited to the TEC by binding the Ser1 phosphorylated CTR of Spt5 via its Rtf1 subunit (Mayekar et al., 2013; Qiu et al., 2012). The complex has been detected at gene promoters (Krogan et al., 2003b; Wood et al., 2003) and has a broad ChIP occupancy profile across the body of the TU, increasing just after the TSS, reaching maximal occupancy after approximately 500-600 bp and decreasing just before the PAS (Mayer et al., 2010).

Loss of Paf1C disrupts Ctk1 recruitment, resulting in a large reduction in Ser2P, which in turn decreases the recruitment of the termination factor Pcf11 (see below), Set2 and Spt6. This leads to shorter polyA mRNA tails and decreased H3K36me3 (Dronamraju and Strahl, 2014; Mueller et al., 2004).

Paf1C, and in particular the Rtf1 subunit, is essential for the activity of the Rad6/Bre1 complex, which is responsible for the mono-ubiquitination of H2B on residue K123 (H2Bub1) (Ng et al., 2003a; Piro et al., 2012; Wood et al., 2003). Rad6/Bre1 is recruited to TU promoters independent of Paf1C, whereas its recruitment to TU bodies is Paf1C dependent (Xiao et al., 2005). The genome-wide metagene profile of H2Bub1 closely

matches that of Paf1, showing maximal occupancy downstream of the TSS (Schulze et al., 2011). Rad6 and H2Bub1 levels transiently increase at promoters upon gene induction, whilst showing generally lower levels at constitutively active genes (Kao et al., 2004). Removal of H2Bub1 is carried out by the ubiquitinases, Ubp8, which removes H2Bub1 near the TSS, and Ubp10, which acts further downstream (Schulze et al., 2011).

H2Bub1 stabilises nucleosomes and promotes transcription elongation by stimulating the activity of the histone chaperone Spt16 in reassembling nucleosomes as Pol II processes (Chandrasekharan et al., 2009; Pavri et al., 2006). This increases nucleosome occupancy across the TU, inhibiting transcription initiation from both the promoter as well as cryptic initiation sites in the TU body (Batta et al., 2011; Fleming et al., 2008).

H2Bub1 also plays a role in promoting other histone modifications as blocking H2Bub1 prevents H3K4 di- (H3K4me2) and tri- methylation (H3K4me3) by the Set1-containing COMPASS complex and methylation of H3K79 by Dot1 (Kim et al., 2013; Shilatifard, 2006). Paf1C is also involved in Set1 recruitment to Pol II as deletion of Ctr9, Paf1 and Rtf1 but not Rad6 disrupts this recruitment (Krogan et al., 2003b; Ng et al., 2003b). Set1 recruitment has been reported to be restricted to the 5' end of TUs using ChIP qPCR. However, a recent higher resolution study using ChIP-seq suggests Set1 shows maximum recruitment at ~500 bp downstream of the TSS (Kuang et al., 2014) similar to Paf1.

Paf1C may have additional roles in gene expression downstream of transcription in RNA export. Supporting this, Paf1C has been shown to interact with the THO/TREX RNA export complex via Hpr1 and an Δ*paf1* Δ*hpr1* double mutant is lethal (Chang et al., 1999; Mueller et al., 2004). In addition, Δ*cdc73* is synthetically lethal with Δ*apq12*, a factor involved in RNA export (Baker et al., 2004). H2Bub1 also has a role in promoting RNA export (Vitaliano-Prunier et al., 2012).

1.7 Transcription termination

Transcription termination involves cleavage, processing and release of the nascent RNA along with disengagement of the polymerase from the DNA template. There are two main pathways through which Pol II transcription is terminated in *S. cerevisiae*. Both are coordinated by factors associating with the TEC.

1.7.1 The PAS dependent pathway

The PAS pathway is the canonical pathway at PCG TUs, although it also occurs to some extent at some non-coding TUs, particularly those for SUTs and XUTs (Grzechnik et al., 2015; Marquardt et al., 2011). Occurring at the 3' end of PCG TUs, the PAS in humans is a highly conserved AAUAAA sequence. In yeast, it is more complex and degenerate consisting of a $(C/T)(A)_n$ sequence surrounded by U rich elements. This occurs 10-30 bp downstream of a positioning element with consensus sequence AA(A/T)AAA, which is in turn downstream of an efficiency element with consensus sequence TA(A/T)(C/G)TA (Guo and Sherman, 1996; Mischo and Proudfoot, 2013; Zhao et al., 1999). Once these sequences are transcribed into the nascent RNA they are bound by specific subunits of the cleavage factor (CF) and cleavage and polyadenylation factor (CPF) complexes.

These complexes are also recruited through subunits interacting with Pol II, including the essential CF subunit Pcf11. Pcf11 interacts preferentially with the CTD of Rpb1 phosphorylated only at Ser2 and shows cooperative binding in the presence of multiple sequential Ser2P CTD repeats (Barillà et al., 2001; Lunde et al., 2010; Meinhart and Cramer, 2004). An interaction also occurs between Rpb2 and Pcf11 in complex with the CF subunit Clp1 (Pearson and Moore, 2014). Pcf11 can be detected at low levels at the start of TUs, suggesting the termination machinery is partially recruited to Pol II as it initiates. Alternatively, as genes form transcription-dependent loops in which the TU start interacts with the TU end, this signal may result from factor binding to the TU end also being detected at its start (O'Sullivan et al., 2004). The level of Pcf11 increases

across the body of TUs before showing a peak in binding around the PAS (Mayer et al., 2010). The Pcf11-CTD interaction is incompatible with Tyr1P and so this may restrict high Pcf11 binding to the end of TUs (Mayer et al., 2012).

Recruitment of the complexes at the 3' end of TUs leads to cleavage of the nascent RNA by the CPF endoribonuclease Ysh1 at the $(C/T)(A)_n$ PAS sequence. The CPF polyA polymerase (Pap1) then adds a ~70 nt polyA tail to the 3' end of the RNA and this is usually bound by the polyA binding protein Pab1 (Porrua and Libri, 2015; Schmid et al., 2012).

Pol II continues to transcribe a further ~200 bp downstream of the PAS before it terminates transcription. Nonetheless this termination is closely linked to PAS processing as mutation of CPF or CF factor components or PAS sequences causes termination defects. The exact mechanism by which Pol II is terminated is poorly understood and it is likely that several different pathways are involved. One proposal, termed the torpedo model, involves degradation of the RNA downstream of the site of cleavage by the 5'-3' exonuclease Rat1. This nuclease then collides with Pol II, disrupting the stability of the RNA:DNA hybrid at its active site (Kim et al., 2004). Rat1 may also be recruited to Pol II by another mechanism, as Rat1 and Pcf11 enhance each other's recruitment (Luo et al., 2006), while Rtt103, a factor which also interacts with Ser2P CTD repeats, also recruits Rat1 (Lunde et al., 2010). The allosteric model proposes that conformational changes occur in the TEC after cleavage, leading to the loss of elongation factors near the PAS. This destabilises Pol II and decreases its elongation rate. The high presence of Pcf11 at the end of TUs may also directly cause termination, as, *in vitro*, the protein alone can dismantle the TEC through interactions with the CTD and RNA (Zhang et al., 2005).

1.7.2 The Nrd1-Nab3-Sen1 dependent (NNS) pathway

The NNS pathway is mediated by Nrd1, Nab3 and Sen1, which form the NRD complex. Depletion of Nrd1 and Sen1 from the nucleus leads to defects in termination at non-coding RNA TUs including snoRNAs, snRNAs, CUTs (Schaughency et al., 2014) and to a lesser extent SUTs and XUTs (Schulz et al., 2013), while termination at most PCG TUs is unaffected. Some termination defects are seen at short PCG TUs in *sen1* and *nrd1* mutants (Steinmetz et al., 2006; Tudek et al., 2014). It has also been shown that large termination defects at PCG TUs occur in *nrd1* and *sen1* mutant strains in a *rat1* mutant background. This suggests the NNS pathway may act as a failsafe termination pathway at TUs that usually terminate using the PAS pathway (Rondón et al., 2009).

Nrd1 possesses a CID that binds preferentially to Ser5P over Ser2P CTD peptides but has highest affinity for Ser2, Ser5 doubly phosphorylated peptides (Vasiljeva et al., 2008). Tyr1P inhibits this CID-CTD interaction (Mayer et al., 2012). Nrd1 binding to Ser5P requires Pro6 to be in the *cis*-conformation like Ssu72 (Kubicek et al., 2012) with the proline isomerase Ess1 required for the NNS pathway (Singh et al., 2009). Nrd1 and Nab3 also possess RNA recognition domains, which recognise the RNA consensus sequences GUA(G/A) and UCUUG respectively (Creamer et al., 2011).

The recruitment of the NRD complex is mediated by both CTD and RNA interactions. ChIP occupancy profiles show a peak in Nrd1 at the 5' end of PCG TUs with similarity to the Ser5P profile as well as another smaller peak around the PAS. Nrd1 shows the highest occupancy relative to Pol II levels on CUTs, SUTs and snoRNAs (Kim et al., 2010; Mayer et al., 2012; Tudek et al., 2014). This is likely to be due to the enrichment for NRD complex RNA-binding motifs in these ncRNA trancripts and depletion of these motifs in coding transcripts (Schulz et al., 2013). Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) data shows Nrd1 and Sen1 cross-link to CUTs, SUTs and snoRNAs. It also shows Sen1 and Nrd1 bind to the 3' end of some

mRNA transcripts (Creamer et al., 2011; Jamonnak et al., 2011), consistent with a role for the NRD complex in failsafe termination at these TUs.

Once recruited, Sen1 is proposed to use ATP to translocate towards and then collide with the transcribing Pol II, leading to disruption of the TEC (Porrua and Libri, 2013). Once Pol II has disengaged from the DNA, the CID of Nrd1 is released from its CTD interaction. This enables the CID to interact with the TRAMP complex subunit Trf4 via a domain that mimics the CTD and this targets the RNA for processing or degradation by the nuclear exosome (Tudek et al., 2014).

In addition to the NRD complex, some components of the CPF and CF complexes play a role in transcription termination by the NNS pathway at ncRNA TUs. This includes the APT subcomplex of CPF with mutations in components Glc7, Ssu72, Pta1 and Ref2 showing snoRNA termination defects (Kim et al., 2006; Nedea et al., 2003, 2008). *pcf11* mutants with disrupted CTD binding show termination defects at the majority of SUTs, CUTs, XUTs and snoRNAs. Pcf11 is recruited downstream of Nrd1 at TUs terminated by the NNS pathway and may promote termination by stimulating Sen1 activity. *pcf11* mutants that disrupt CPF cleavage and polyadenylation activity also show termination defects at many CUTs, SUTs and XUTs and a few snoRNAs, suggesting that these TUs are also sometimes terminated by the PAS pathway (Grzechnik et al., 2015).

1.8 Splicing

Splicing is the process by which an intron is removed from RNA and the 5' and 3' exons either side of it are joined together. Less than 5 % of PCG TUs contain introns, however, 101 of these are within ribosomal protein genes, which are amongst the most highly transcribed genes in the yeast genome. This means that, contrary to popular perception, splicing is abundant in *S. cerevisiae*, with approximately 40 % of all mRNAs molecules in the cell having been spliced (Warner, 1999).

RNA transcribed from intron-containing PCG TUs contains a 5' splice site (ss) with a consensus sequence GUAAGU and 3' ss with sequence AG at the 5' and 3' ends of the intron, respectively, as well as a branch point sequence (BPS) near the 3' ss with consensus UACUUAC (Lacadie and Rosbash, 2005). Splicing is mediated by the spliceosome, which is a megadalton complex consisting of 5 snRNAs (U1, U2, U4, U5 and U6) each associated with several proteins to form small nuclear ribonucleoproteins (snRNPs). The spliceosome also consists of many non-snRNP proteins (Wahl et al., 2009). This spliceosome is sequentially assembled on RNA with the U1 snRNP binding to the 5'ss and U2 snRNP binding to the BPS (complex A, Figure 1.3) (Görnemann et al., 2005). The other snRNPs then associate (complex B). The complex is then rearranged, leading to the loss of U1 and U4 snRNPs and the formation of a catalytically active spliceosome (complex B*). Two sequential *trans*-esterification reactions are then carried out in which the branch point attacks the 5' ss releasing the 5' exon, which remains tethered to the complex (complex C). The 3' OH of this 5' exon is then ligated to the 3' ss releasing the intron as a lariat (Wahl et al., 2009). Splicing can be regulated so that some introns are not always spliced and, in genes with multiple exons, particular exons can be included or excluded from the final product. This gives rise to different transcript isoforms coding for different protein isoforms (Kawashima et al., 2014; Kelemen et al., 2013).

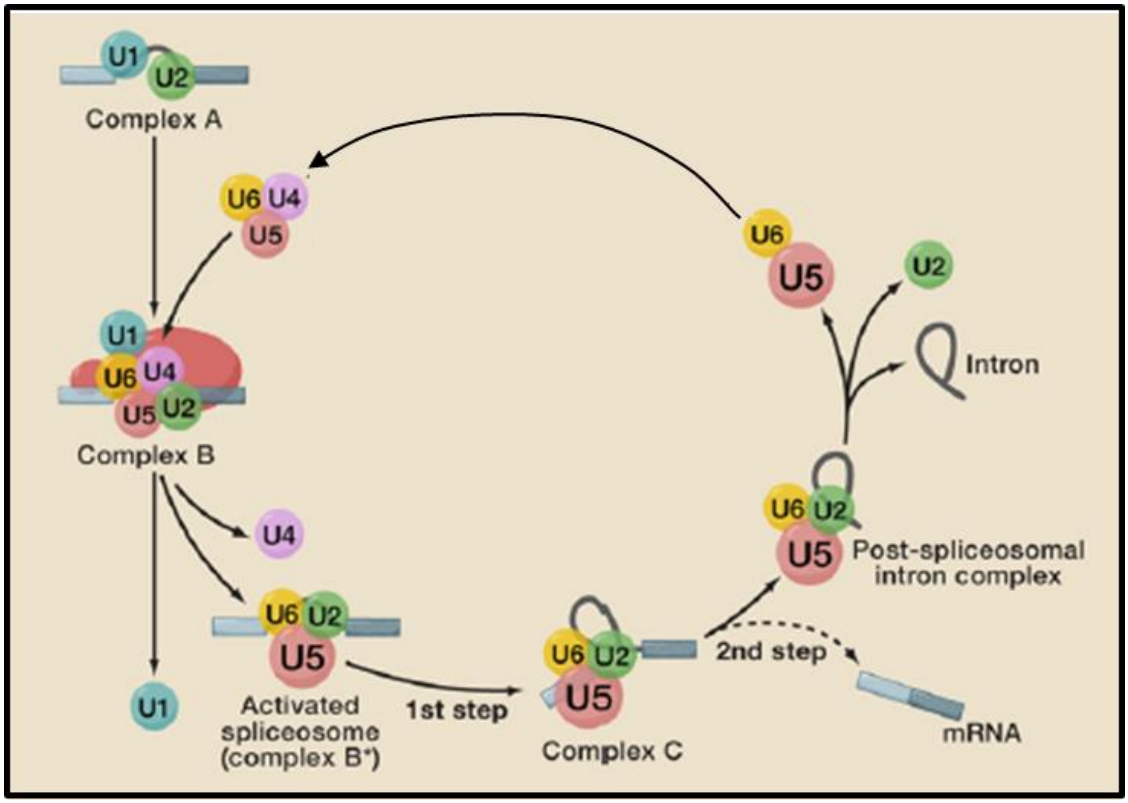


Figure 1.3. Pre-mRNA splicing by the spliceosome. Schematic showing the assembly and disassembly of snRNPs during splicing of pre-mRNA. Figure adapted from Wahl et al., (2009).

While splicing can occur post-transcriptionally, after cleavage of the nascent transcript at the PAS, it has also been observed occurring co-transcriptionally in many organisms (Singh and Padgett, 2009; Wada et al., 2009), genome-wide (Brugiolo et al., 2013). Co-transcriptional splicing has even been visualised by electron microscopy (Beyer and Osheim, 1988). Splicing can only start once the 3' ss has been transcribed. Using artificial constructs in *S. cerevisiae*, Tardiff et al., (2006) showed that, when the 3' exon is shorter than 600 bp, splicing does not occur co-transcriptionally, but when it is longer than 1200 bp, it does. As the length of 3' exons in *S. cerevisiae* are generally short (median length < 600 bp), it has been suggested that most splicing in this organism occurs post-transcriptionally. Supporting this, ChIP shows U1 and U2 snRNPs assemble on chromatin downstream of the 5' ss and 3' ss, respectively; however, the level of U2

snRNP ChIP signal is lower on short genes, suggesting U2 snRNP assembly has generally not occurred before transcription has finished at these genes (Tardiff et al., 2006). In conflict with this, analysis of nascent RNA associated with chromatin suggests 50 % of *S. cerevisiae* introns are greater than 74 % spliced during transcription (Brugiolo et al., 2013; Carrillo Oesterreich et al., 2010). To provide sufficient time for splicing to occur on short genes, they suggest, using Pol II ChIP-seq data, that Pol II pauses 250 bp upstream of the PAS on genes with short 3' exons; however, Pol II pausing at this site is not observed in NET-seq data (Churchman and Weissman, 2011). Overestimation of co-transcriptional splicing may have occurred as a result of post-transcriptional RNAs co-purifying with chromatin.

The functional significance of coupling transcription and splicing is poorly understood, although it may enable these processes to regulate each other. In support of this, controlling the rate of transcription elongation has been shown to regulate whether an exon is included (Perales and Bentley, 2009). Some spliceosome components also associate with the CTD, which may assist the recruitment of the complex to the nascent RNA (Eick and Geyer, 2013); however, no difference in splicing efficiency has been observed between co- and post-transcriptional splicing (Tardiff et al., 2006). Using a variety of techniques, Pol II has been shown to slow down or pause at different positions associated with splicing, such as at the 5' end of exons (Kwak et al., 2013), around splice sites (Mayer et al., 2015), over each exon (Nojima et al., 2015) and over introns (Chathoth et al., 2014). The CTD of Rpb1 may also be post-translationally modified at these pauses to regulate elongation and splicing (Chathoth et al., 2014; Nojima et al., 2015). Although species-specific differences may account for some of the discrepancies between these studies, it is clear that further investigation is required into how transcription elongation and splicing regulate each other.

1.9 RNA fate

Throughout the lifetime of an RNA, it is bound by numerous proteins to form a ribonucleoprotein particle (RNP). These proteins impact on the fate of the RNA, including its degradation and its subcellular localisation. Many of these proteins are loaded onto the RNA during transcription elongation and termination.

1.9.1 Transcript release and RNA export

Custódio et al., (1999) showed β-globin transcripts in mammalian cells accumulate in a nuclear focus at the site of their transcription. Upon inhibition of transcription, this nuclear focus disappears, presumably due to release of the transcript and no further transcript accumulation. However, if splicing or 3' end formation is disrupted, the nuclear focus remains and the transcripts are retained at the site of transcription. This is proposed to be a quality control mechanism, ensuring RNAs are correctly processed before their release. In *S. cerevisiae*, terminating a transcript by ribozyme cleavage leads to retention of the transcript at the site of transcription, while encoding a stretch of at least 48 adenosines before the ribozyme cleavage site leads to release of the transcript (Dower et al., 2004). Furthermore, terminating transcripts with a variety of different 3' UTRs modulates the decay rates of these transcription site-associated foci upon transcription inactivation (Abruzzi et al., 2006). This suggests the efficiency with which transcripts are polyadenylated is an important regulator of transcript release. Transcripts without or with short polyA tails are retained and transcripts with longer tails are released.

Multiple processes regulate polyA tail length. Transcripts terminated by the NNS pathway are oligoadenylated by the TRAMP complex, which results in the production of a shorter polyA tail than transcripts terminated by the PAS dependent pathway (Tudek et al., 2014). Variation between different PAS sequences may also regulate polyadenylation efficiency. The action of polyadenylating enzymes can be reversed by

the activity of the nuclear exosome. This multisubunit complex consists of a channel leading towards the 3'-5' exonuclease and endoribonuclease Dis3 and 3'-5' exonuclease subunit Rrp6 (Bonneau et al., 2009; Lebreton et al., 2008). Deletion of *RRP6* leads to an increase in transcripts with long polyA tails and a reduction in the persistence of transcription site-associated transcript foci after transcription inactivation (Vodala et al., 2008).

The proposed mechanism by which increased polyA tail length causes increased transcript release is that it increases transcript binding by polyA tail binding proteins such as Pab1. These proteins shuttle between the nucleus and the cytoplasm and, by binding RNAs, promote their export out of the nucleus away from nuclear foci (Brune et al., 2005; Dunn et al., 2005). In support of this, transcripts terminated by the NNS pathway, such as CUTs and to a lesser extent SUTs, are bound less by Pab1, probably as a result of their shorter polyA tails, show lower association with nuclear pore components and show greater nuclear retention (Tuck and Tollervey, 2013).

In addition to polyA binding proteins, many other proteins associate with the transcript to form an RNP and are involved in nuclear export regulation. Many of these, including the essential TREX complex, associate with RNA during transcription. The TREX complex consists of Yra1, Sub2, the THO complex (THOc), consisting of subunits Tho2, Hpr1, Mft1, Thp2, and Tex1, and additional RNA-binding proteins. THOc and Yra1 have been shown to interact with Ser2, Ser5 doubly phosphorylated CTD repeat peptides and this leads to binding of TREX to the TEC (MacKellar and Greenleaf, 2011; Meinel et al., 2013; Zenklusen et al., 2002). Additional interactions between TREX subunits and the nascent RNA, and between Yra1 and Pcf11 may strengthen this interaction (Johnson et al., 2009). ChIP detects TREX subunits associated with TUs increasing from the TSS and reaching maximal occupancy just before the PAS (Meinel et al., 2013). Yra1 is proposed to be transferred to the transcript upon transcript cleavage and act as an adapter for the

binding of the Mex67-Mtr2 export receptor complex to the RNA (Zenklusen et al., 2001). Mex67 interacts directly with nucleoporins, leading to transport of the RNA through the nuclear pore complex (NPC) into the cytoplasm (Segref et al., 1997).

Disruption of various export factors, including TREX subunits or Mex67, leads to nuclear accumulation of bulk polyadenylated RNA. This creates the impression that nuclear export is a linear process and is the same for all RNAs. However, RNAs show differential binding to different export factors, including Yra1, Mex67 and polyA binding proteins Nab2 and Pab1 (Baejen et al., 2014; Hieronymus and Silver, 2003; Hurt et al., 2004; Iglesias et al., 2010; Tuck and Tollervey, 2013). In addition, post-translational modification of particular export factors, including the sumoylation of Hpr1, affects the export of a subset of RNAs (Bretes et al., 2014). This suggests there are multiple parallel export pathways, with the export of particular RNAs occurring at different rates and being regulated by association with particular combinations of factors.

The position of genes within the nucleus is another important yet poorly understood process in the regulation of nuclear export. Transcriptionally silenced loci, including telomeres and mating type loci, typically associate with the nuclear periphery. Visualisation of the nuclear periphery in mammalian cells shows it is generally associated with heterochromatin. However, interspersed within heterochromatin are patches of euchromatin in proximity to NPCs (Akhtar and Gasser, 2007). This observation supports the gene gating hypothesis, which proposes that genes translocate from the nucleoplasm or repressive regions of the nuclear periphery to the nuclear pore upon transcriptional activation to enable efficient export of the transcript being produced (Blobel, 1985). Evidence from yeast supports this with highly transcribed genes showing increased association with nuclear pore components and many but not all inducible genes localising to the nuclear periphery upon activation (Brickner and Walter, 2004; Casolari et al., 2004; Taddei et al., 2006). Whether this promotes nuclear

export is contentious because disrupting gene association with the nuclear pore does not always disrupt export (Green et al., 2012).

Multiple factors regulate the recruitment of genes to the NPC. During transcription initiation and elongation, interactions form between promoter- and TEC-associated factors and nuclear pore components (Dieppois and Stutz, 2010; Schneider et al., 2015; Sood and Brickner, 2014). Even after transcription inactivation some genes continue to associate with the nuclear pore. This correlates with the length of time the transcription site-associated transcript foci persists (Abruzzi et al., 2006). This suggests the RNA retained at the site of transcription may also act as a tether recruiting the gene to the nuclear pore, possibly via interactions between RNA-associated proteins and nuclear pore components.

1.9.2 RNA degradation

RNA export is a regulator of RNA stability, with different RNA degradation pathways compartmentalised within the nucleus and cytoplasm. In the nucleus, RNAs can be degraded 3' to 5' by the nuclear exosome or they can be decapped by Dcp2 and degraded from their 5' end by Rat1 (Geisler et al., 2012). Pre-mRNAs that fail to complete certain processing steps, such as 5' capping, splicing, cleavage and polyadenylation and correct mRNP assembly, are retained at the site of transcription and degraded by these pathways (Houseley and Tollervey, 2009). As discussed, termination of ncRNAs by the NNS pathway results in transcripts with short polyA tails, which leads to some retention of these transcripts within the nucleus. Additionally, this short polyA tail creates a short unstructured extension, which targets these transcripts for nuclear exosome activity, as they can be efficiently translocated through the exosome channel (Chlebowski et al., 2013). Of the ncRNAs, CUTs are bound the most by the TRAMP complex and are generally rapidly and completely degraded by the nuclear exosome. SUTs and XUTs also show more TRAMP complex binding than pre-mRNAs leading to some nuclear exosome degradation (Tuck and Tollervey, 2013; Tudek et al.,

2014; Xu et al., 2009). For snRNAs and snoRNAs, which are also terminated by the NNS pathway, additional structures prevent complete exosome degradation instead resulting in 3' end trimming and subsequent transcript release (Lykke-Andersen and Jensen, 2007).

Once in the cytoplasm there are two main pathways of RNA decay. Either the RNA is degraded by the cytoplasmic exosome, which lacks the Rrp6 subunit or they are decapped by Dcp2 and degraded 5' to 3' by Xrn1. Both degradation routes require prior shortening of the polyA tail by the Pan2/Pan3 or Ccr4/Pop2/Not deadenylation complexes (Houseley and Tollervey, 2009). There are also additional degradation pathways for RNAs which are aberrantly translated, including those for RNAs which contain premature stop codons termed non-sense mediated decay (NMD), contain no stop codons or cause excessive ribosomal stalling (Parker, 2012). Despite generally being retained in the nucleus, CUTs and SUTs show 3.6- and 7.6-fold respective increases in average levels on deletion of the predominantly cytoplasmic Xrn1 (van Dijk et al., 2011) and sometimes associate with ribosomes (Smith et al., 2014), confirming some level of export. Those which are exported are also subjected to NMD due to their lack of long open reading frames (ORFs) (Malabat et al., 2015; Marquardt et al., 2011).

mRNAs show great variability in half-lives, ranging from a minimum of 1 min, median of 8 min and maximum of over 2 h (Miller et al., 2011). Major determinants of RNA stability are the sequence and structure of the RNA, which influence the binding of factors that can promote or repress degradation-associated activities (Parker, 2012). RNA stability also varies in different environmental conditions and phases of the cell cycle (Eser et al., 2014; Zid and O'Shea, 2014). This is partly controlled by altering the expression levels of certain degradation components, for example Rrp6 is degraded on entering meiosis leading to the stabilisation of some ncRNAs (Lardenois et al., 2011). Degradation rate can also be controlled by the promoter, with promoter swap

experiments altering transcript stability and localisation within the cytoplasm (Bregman et al., 2011; Trcek et al., 2011; Zid and O'Shea, 2014). This suggests a mechanism whereby factors recruited to the promoter can influence mRNP composition. The promoter can also influence which TSS and PAS are used, leading to different transcript isoforms, which have been shown to form different structures with different stabilities (Geisberg et al., 2014; Pelechano et al., 2013).

1.10 The role of non-coding transcription

The pervasive transcription carried out by Pol II at intergenic and antisense regions has raised the question as to what, if any, is the role of this non-coding transcription, especially when the transcripts produced are often rapidly degraded.

The majority of SUT and CUT TUs initiate divergently from the 5' NDRs found at the promoters of other (usually protein-coding) TUs. Similarly, 48 % of the promoters of PCG TUs share an NDR with the promoter of another TU (Xu et al., 2009). This shows bidirectional transcription from an NDR is a common feature of the yeast genome. This has led some to suggest that most non-coding TUs are a non-functional by-product of PCG transcription; the opening up of chromatin that PCG transcription requires resulting in Pol II occasionally being recruited and initiating in the wrong direction/place (Struhl, 2007). Evidence against this is that there is only a modest correlation between non-coding and coding TUs expressed from the same divergent promoter (Churchman and Weissman, 2011). Therefore the opening up of chromatin by high levels of PCG transcription does not automatically result in high levels of non-coding divergent transcription. ChIP on PIC components has also shown that there is a distinct PIC for each divergently transcribed TU suggesting some degree of independent regulation (Murray et al., 2012; Rhee and Pugh, 2012a).

The compact nature of the yeast genome and this bidirectional transcription often gives rise to overlapping TUs. TUs transcribed in the same direction arranged tandemly can

result in transcription of the upstream TU terminating in or overlapping with the promoter of the downstream TU. TUs can also be transcribed convergently. A specific example of this is the sense/antisense pair (SAP) whereby transcription can initiate in the 3' NDR of a TU (potentially shared with the 5' NDR of a downstream tandem TU, forming a bidirectional promoter) and transcribe the antisense strand of this TU, possibly terminating near its promoter (Mavrich et al., 2008). Overlapping transcription provides a possible role for non-coding transcription in regulating the transcription of neighbouring TUs. Transcription of one TU into the promoter of a second TU from either an antisense or tandem orientation can cause repression of the second TU in a process called transcriptional interference. Specific examples of this include *SRG1* non-coding transcription into the promoter of the downstream *SER3* PCG TU and *RME2* antisense transcription into the promoter of *IME4* (Hongay et al., 2006; Martens et al., 2004). Transcriptional repression by transcriptional interference can be mediated by several non-mutually exclusive mechanisms. These include disruption of transcription factor binding, occlusion of the promoter by deposition of nucleosomes by TEC-associated histone chaperones Spt6 and Spt16, and co-transcriptional H3K4 and H3K36 di- and tri-methylation across promoters, leading to removal of activatory histone acetylation marks by the Rpd3, Hda1 and Set3 histone deacetylase complexes (Castelnuovo and Stutz, 2015; Castelnuovo et al., 2013; Houseley et al., 2008; Pinskaya et al., 2009; Thebault et al., 2011). Convergently transcribed TUs may also lead to collision between polymerases, as both strands cannot be transcribed simultaneously, potentially acting as another mechanism of transcriptional repression (Hobson et al., 2012). Overlapping transcription can alternatively have activatory roles (Crisucci and Arndt, 2012; Uhler et al., 2007). The mechanism of this is poorly understood but may result from transcription removing repressors or disrupting nucleosome positioning, enhancing PIC recruitment.

In order to establish a more general role for non-coding transcription, studies have focused on the role of antisense transcription at SAPs. Consistent with both activatory

and repressive functions of non-coding transcription, levels of sense and antisense transcription at individual genes show no correlation (Murray et al., 2012, 2015). Antisense transcription is however associated with genes with higher gene expression variability (Xu et al., 2011). It has recently been shown that antisense transcription reading into the sense promoter correlates at the promoter and 5' end of the sense TU with increased levels of nucleosome occupancy, histone H3 acetylation, chromatin remodelling enzymes and nucleosome turnover and reduced levels of H3K36 and H3K79 methylation and H2Bub1 associated with more stable nucleosomes (Murray et al., 2015). One hypothesis is that the increased nucleosome turnover as a result of antisense transcription causes greater variability in nucleosome positioning, giving rise to both repressive and activatory effects and may result in a noisier mode of transcription in which the sense TU switches between an off state and rapid bursts of transcription (Murray et al., 2015).

In addition to a role for non-coding transcription, there could also be a role for the non-coding transcripts produced. Despite often being rapidly degraded, such transcripts could potentially act in the short time period before their degradation or be able to function at low steady-state levels. Evidence and mechanisms for possible roles for such transcripts will be discussed in greater detail in Chapter 6.

From studies on non-coding transcription, it is clear that overlapping pervasive transcription can have a range of different effects on neighbouring genes. The exact effect on transcriptional regulation is probably dependent on the environmental conditions and the transcription levels and orientations of the genes within a locus.

1.11 Aims

This thesis aims to investigate different aspects of transcription. To provide additional insight into the role of factors which associate with Pol II, a main aim will be to map the location of factors associated with the TEC with strand-specificity and greater resolution

than has previously been achieved. How transcription changes in response to a change in conditions will also be analysed with additional investigations into the role of non-coding transcription and non-coding transcripts in regulating transcriptional changes. More detailed aims are provided before each results section.

Chapter 2.

Methods

2 Methods

2.1 Yeast culture

Strains were streaked from glycerol stocks onto 2 % agar YPD (1 % yeast extract (Difco), 1 % bactopeptone, 2 % glucose) plates and grown (48 h, 30 ^{o}C). Cells were then grown (overnight, 30 ^{o}C) in 5 ml YPD. This culture was used to inoculate a 25 ml YPD culture at OD_{600} 0.2, which was grown (30 ^{o}C, 200 rpm) to OD_{600} 0.6-0.7. To assess growth using a drop plate assay, overnight cultures were normalised for cell density and then 2 µl of five 10-fold serial dilutions were pipetted onto 2 % agar YPD plates and incubated at 30 ^{o}C for 2 days.

2.2 Yeast strains

All strains used in this study are listed in Table 2.1. Genetic manipulation of strains was carried out using the homologous recombination method described by Longtine et al., (1998). For gene deletion strains, PCR products were made 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 3xFLAG-tagged strains, first a plasmid was created from plasmid pFA6a-GFP(S65T)-His3MX6 (Longtine et al., 1998) with the 3xFLAG sequence, amplified from strain YSC001 by PCR, inserted in place of GFP. PCR products were made using this template, consisting of a 40bp sequence homologous to the first 40 bp upstream of the stop codon of the gene to be tagged followed by the FLAG sequence, His5 selection marker and 40 bp of sequence homologous to a region downstream of the gene to be tagged.

Cells to be transformed were grown to log phase, pelleted (3 min, 5000 rpm), resuspended in 450 µl 0.1 M LiAc/TE and incubated (> 1 h, 4 ^{o}C). 100 µl of cell suspension, 10 µl of gel extracted PCR product, 10 µl calf thymus DNA (Sigma), 700 µl 0.1 M LiAc/TE, 40 % PEG were incubated (30 min, 30 ^{o}C) then heat-shocked (20 min, 42

°C). Cells were pelleted (5 min, 7000 rpm), resuspended in H_20 and plated onto appropriate selection media. When the kanamycin 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.

Name	**Parent strain**	**Genotype**	**Origin**
BY4741	S288C	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*	Euroscarf
BY4742	S288C	*MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*	Euroscarf
YSC001	BY4741	*MATa; his3Δ0; leu2Δ0; met15Δ0; ura3Δ0; RPB3-3xFLAG-NAT*	S. Churchman
BY4741 *RPB3-3xFLAG*	BY4741	*RPB3-3xFLAG-His3MX6*	This study
BY4741 *SPT6-3xFLAG*	BY4741	*SPT6-3xFLAG-His3MX6*	This study
BY4741 *SPT16-3xFLAG*	BY4741	*SPT16-3xFLAG-His3MX6*	This study
BY4741 *PAF1-3xFLAG*	BY4741	*PAF1-3xFLAG-His3MX6*	This study
BY4741 *PCF11-3xFLAG*	BY4741	*PCF11-3xFLAG-His3MX6*	This study
BY4741 *NRD1-3xFLAG*	BY4741	*NRD1-3xFLAG-His3MX6*	This study
BY4741 *CET1-3xFLAG*	BY4741	*CET1-3xFLAG-His3MX6*	This study
BY4741 *ISW1-3xFLAG*	BY4741	*ISW1-3xFLAG-His3MX6*	This study
BY4741 *YRA1-3xFLAG*	BY4741	*YRA1-3xFLAG-His3MX6*	This study
BY4742 *SET2-3xFLAG*	BY4742	*SET2-3xFLAG-KanMX4*	B. Strahl
BY4741 *FCP1-3xFLAG*	BY4741	*FCP1-3xFLAG-His3MX6*	This study
BY4741 *RTR1-3xFLAG*	BY4741	*RTR1-3xFLAG-His3MX6*	This study
BY4741 *SSU72-3xFLAG*	BY4741	*SSU72-3xFLAG-His3MX6*	This study
BY4741 *KIN28-3xFLAG*	BY4741	*KIN28-3xFLAG-His3MX6*	This study
BY4741 *CTK1-3xFLAG*	BY4741	*CTK1-3xFLAG-His3MX6*	This study
BY4741 *BUR1-3xFLAG*	BY4741	*BUR1-3xFLAG-His3MX6*	This study
Δxrn1	BY4741	*xrn1::KanMX6*	Euroscarf
Δxrn1 RPB3-3FLAG	BY4741 *RPB3-3FLAG*	*xrn1::LEU2*	This study
Δpaf1	BY4741	*paf1::KanMX6*	This study
Δcmk2	BY4741	*cmk2::KanMX6*	Open Biosystems
HMS2:ADH1t	BY4741	*HMS2 Δ(651-660)::^{T}ADH1.kanMX6*	D. Brown

Table 2.1. Details of strains used in this study.

2.3 DNA extraction

Pelleted cells (>2mg) were vortexed (10 min) with 200 µl DNA extraction buffer (100 mM NaCl, 100 mM Tris-Cl (pH 8.0), 1 mM EDTA (pH 8.0), 1 % SDS, 2 % Triton X-100), 200 µl acid-washed glass beads and 200 µl phenol:chloroform:isoamyl alcohol. After spinning (16,000 g, 10 min, 4 ^{o}C), the upper layer was combined with 500 µl of a 6 parts ethanol, 1 part 7.5 M ammonium acetate solution. DNA precipitate was pelleted (16,000 g, 10 min, 4 ^{o}C) and resuspended in 200 µl H_20.

2.4 RNA extraction

25 ml OD_{600} 0.6 culture was pelleted (1,000 g, 3 min), resuspended in 400 µl T.E.S. (100 mM Tris Cl (pH 7.5), 100 mM EDTA (pH 8.0), 0.5 % SDS) and 400 µl phenol:chloroform (pH 4.7) and incubated (65 ^{o}C, 20 min, 1400 rpm). After spinning (16,000 g, 20 min, 4 ^{o}C), the upper layer was mixed with 40 µl 3 M NaOAc (pH 5.5) and 1 ml ethanol and incubated (-80 ^{o}C, 30 min). RNA precipitate was pelleted (16,000 g, 10 min, 4 ^{o}C) and resuspended in 100 µl H_20. Concentration was determined using a NanoDrop spectrophotometer.

2.5 Reverse transcription

10 µl of extracted RNA was incubated (50 ^{o}C, 1 h) with 2 µl DNase I buffer, 1 µl RNAseOUT (Invitrogen), 6 µl H_20 and 1 µl DNase I (Roche) and then heated (70 ^{o}C, 15 min). 5 µg RNA was reverse transcribed by incubating (65 ^{o}C, 5 min) with 0.6 µl random primers, 1 µl dNTPs in a 13 µl final volume and then adding 4 µl First-Strand buffer, 1 µl 0.1 M DTT, 1 µl RNAseOUT, 1 µl Superscript III (Invitrogen) (replaced with 1 µl H_20 in control reactions) and incubating at 50 ^{o}C (1 h) and then 70 ^{o}C (15 min). Complementary DNA (cDNA) levels for each gene were analysed by Real Time quantitative PCR (qPCR) using a RotorGene (Corbett), SYBR Green mix and primers from Table 2.2.

Gene	Forward primer sequence	Reverse primer sequence
YPR156C/TPO3	GTGGATCTAAATCCAACATG	ACGGGGGCTTTGGAGCTCAC
YGR138C/TPO2	CTCGTGCAAATCAACTTAAG	GGCGCATTGATATCAGGCAC
YOL016C/CMK2	TGTTCATAGGGATTTGAAAC	CTTCAATTGTTTAGCTATAC
YOL086C/ADH1	GCCACTGACGGTGGTGCTCA	GGCACCAGCTGGCATACCGA
YDR399W/HPT1	GTCGGCAAACGATAAGCAAT	ATGAAACCACCACCACCAAT
YAL035W/FUN12	TGATGTCCCAGGTCTTTTGG	AACCAAAATTGCGATGTTAC
YDL052C/SLC1	ATTGACACCTTGAATAAAGG	ACATTGTCAGCTCACTCGTG
YGR033C/TIM21	ACGTCACTGCTTAGAACAAG	TTGAGGCCATAATGGCTTAG

Table 2.2. Reverse transcription primers used in this study

2.6 Western blotting

Protein samples were separated by gel electrophoresis on 7.5 % or 10 % SDS polyacrylamide gels and transferred to nitrocellulose membranes. Membranes were incubated with 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, washed 3x 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 chemiluminescence (Pierce) and x-ray film.

Antibody Clone Name	Epitope recognised	Species reactivity	Dilution for Western blot
3E8 (Millipore)	Ser2P RPB1 CTD	Rat	1:1000
3E10 (Millipore)	Ser5P RPB1 CTD	Rat	1:2000
4E12 (Millipore)	Ser7P RPB1 CTD	Rat	1:2000
3D12 (Millipore)	Tyr1P RPB1 CTD	Rat	1:250
σ-80 (Santa cruz)	Residues 1-80 from the N-terminus of RPB1	Rabbit	7:1000
M2 (Sigma)	FLAG tag sequence (DYKDDDDK)	Mouse	1:5000

Table 2.3. Details of antibodies used in this study.

2.7 Nuclear Extraction

50 ml OD_{600} 0.6 culture was pelleted (1,000 g, 3 min), and incubated (<5min, 30 ^{o}C) in 0.5 ml spheroplasting buffer A (1 M sorbitol, 0.5 mM 2-mercaptoethanol, 20 mg/ml zymolase 20T (MP Biomedicals)). Once >70 % of cells were spheroplasted (monitored using a light microscope with 40x magnification), cells were pelleted (16,000 g, 10 s, 4 ^{o}C), washed carefully twice with 1 ml 1 M sorbitol, resuspended in 100 µl 18 % Ficoll and then mixed quickly with 1 ml 9 % Ficoll. Nuclei were pelleted (16,000 g, 5 min, 4 ^{o}C) and washed twice with 1 ml 1 M sorbitol. RNA was extracted from nuclei as detailed in section 2.4.

2.8 *In vitro* transcription

DNA sequences were synthesised by PCR with the T7 RNA polymerase promoter sequence (TAATACGACTCACTATAGGG) attached to the 5' end of the sense strand. Transcription mixtures (30 µl) containing 0.05 µM PCR product, 2 µl T7 polymerase, 3 µl 10x T7 polymerase transcription buffer (Ambion), 0.5 mM each of GTP, UTP, CTP and ATP were incubated (37 ^{o}C, 1 h), followed by incubation (37 ^{o}C, 15 min) with 1 µl of Turbo DNAse (Ambion). RNA extraction was carried out as described but with 200 µl T.E.S. and phenol:chloroform, resuspending the pellet in 30 µl H_20.

2.9 Triple-stranded structure formation assay

Cy5-Cy3 end-labelled DNA (1 µM) was incubated (5 min, 22 ^{o}C) with 5 µM RNA in TBM buffer (90 mM Tris-HCl (pH 7.2), 90 mM boric acid, 10 mM $MgCl_2$). Products were separated by electrophoresis on 6% polyacrylamide 0.5x TBM gels. Gels were scanned for both Cy5 and Cy3 fluorescence using a Typhoon FLA-9000 Biomolecular Imager (GE Healthcare). Images were analysed, including fluorescent signal intensity quantification, using Aida Image Analyzer software. Gels were then stained with methylene blue to show all nucleic acids.

2.10 Restriction enzyme digestion protection assay

After triple-stranded structure formation, reactions were incubated (2 h, optimum restriction enzyme temperature) with optimum restriction enzyme buffer (NEB) and 15 U restriction enzyme in a final volume of 10 µl. Products were separated and analysed as before.

2.11 Complementary single stranded DNA competition assay

5 µM RNA was incubated (5 min, 22 ^{o}C) with short 25-30 nt single stranded DNA (ssDNA) complementary to the RNA in TBM buffer. Cy5-Cy3 labelled DNA (1 µM) was added and incubated (5 min, 22 ^{o}C). Products were separated and analysed as before.

2.12 Northern blotting

RNA samples (20 µg) were separated by gel electrophoresis on 20 mM MOPS (pH 7.0), 1.1 % formaldehyde, 5 mM NaOAc, 1 mM EDTA, 20 mM MOPS (pH 7.0), EtBr gels and transferred to nylon membranes (Magma) by wet blotting using 20 x saline-sodium citrate (SSC) buffer as described by Sambrook et al., (1989). 18S and 26S rRNA levels were determined by imaging gels under UV light, serving as a control for RNA loading and quality. Membranes were fixed by incubating at 80 ^{o}C (2 h) and blocked by incubating (65 ^{o}C, 2h) with PerfectHyb (Sigma) hybridisation buffer. *In vitro* transcription in which UTP was replaced with α-32^{P} UTP was used to generate strand-specific radiolabelled RNA probes. Filtration through G350 sepharose was used to remove excess nucleotides. Probes were added and allowed to hybridise overnight at 65 ^{o}C. Membranes were washed by incubating (65 ^{o}C, 20 min) 2x with (1x SSC), 1x (0.2x SSC), 1x (0.1x SSC) and then 1x with (0.05x SSC). Probe binding was visualised by exposure to x-ray film from 24 hours to 7 days.

2.13 RNA fluorescence in situ hybridisation (FISH)

50 ml OD_{600} 0.6 culture was pelleted (1,000 g, 3 min), and fixed by incubating (45 min, 80 rpm) with 35 ml PBS and 4.8 ml 32 % paraformaldehyde. Fixed cells were washed twice with 10 ml FISH buffer A (1.2 M sorbitol, 0.1 M $KHPO_4$ (pH 7.5)) and resuspended in 1 ml FISH buffer B (FISH buffer A, 20 mM ribonucleoside vanadyl-complex (VRC), 20 µM 2-mercaptoethanol). The mixture was incubated at 30 °C with 15 U of lyticase until >70 % of cells were spheroplasted (~25 min). Cells were pelleted (1000 g, 3 min, 4 °C) and washed with and then resuspended in 1 ml FISH buffer B. 150 µl of cells were left to settle (30 min, 4 °C) on poly-L-lysine treated coverslips. These were washed with 2 ml FISH buffer A to remove unattached cells, incubated (-20 °C, >3 h) in 2 ml 70 % ethanol and then rehydrated for 5 min in 2 ml FISH wash buffer (10 % formamide, 2x SSC). For hybridisation, coverslips were placed cell-coated side down on a 48 ul drop containing 50 nM probes (Stellaris), 0.1 g/ml dextran sulphate, 1 mg/ml *E. coli* tRNA, 2 mM VRC, 20 µg/ml BSA, 2x SSC, 10 % formamide and incubated (30 °C, 24 h) in a sealed Parafilm chamber. Coverslips were twice incubated (30 °C, 30 min) in 2 ml FISH wash buffer, washed with H_20, air dried, placed cell-coated side down on a drop of ProLong Diamond Antifade Mountant with DAPI (Life technologies), allowed to polymerise for 24 h in the dark and then sealed with nail varnish. Cells were imaged using a DeltaVision Core wide-field fluorescence deconvolution microscope. A 100x objective lens was used to image 31 “z-stacks” using an exposure time of 0.5 s for DAPI and 1 s for Cy3 and Cy5 filters for each stack. Images were deconvolved and compressed to 2D images displaying the maximum intensity for each pixel across all z-stacks. Image quantification was carried out using a bespoke image analysis pipeline within ImageJ.

2.14 NET-seq

2.14.1 Yeast culture

Cells were grown in 25 ml YPD overnight. This was used to inoculate 1 L of YPD in a 2 L flask to give a starting OD_{600} of 0.2. 2 L of cells were grown (30 °C, 170 rpm) to OD_{600} of 0.65 and harvested by filtration onto a 0.45 µm pore size, 90 mm diameter nitrocellulose filter paper (Whatman). Cells were scraped off the filter paper with a spatula pre-cooled in liquid nitrogen and flash frozen in liquid nitrogen. Cells were ground in 6 cycles of 3 min at a 15 Hz shaking frequency in a 50 ml grinding jar (Retch) with a 25 mm stainless steel ball using a Retsch MM400 mixer mill (3 min, 15 Hz). The grinding chamber was cooled in liquid nitrogen in between cycles. 1 g of yeast grindate was stored at -80 °C.

For induction in YPG, cells were first grown in YPD to an appropriate OD_{600} so that OD_{600} after induction would be in the range 0.6-0.7. Cells were spun (3 min, 5000 rpm) in 500 ml pots in a JA-10 rotor, resuspended in 1 L YPG pre-warmed to 30 °C and returned to a rinsed 2 L flask. After growth (30 °C, 170 rpm) for the induction time, cells were harvested by filtration as above.

2.14.2 Immunoprecipitation

1 g of yeast grindate was resuspended in 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 (complete, EDTA-free (Roche)), 50 U/ml SUPERase.In RNase inhibitors (Invitrogen)). Resuspended grindate was incubated (4 °C, 20 min) with 660 U of DNase I (Promega). Insoluble cell debris was pelleted by spinning (16,000 g, 4 °C, 10 min). Supernatants were combined (a 20 µl input sample was taken and combined with 20 µl 2 x SDS loading buffer) and added to 0.5 ml of anti-FLAG M2 affinity agarose beads (Sigma) pre-washed with 2 x 10 ml Lysis buffer A (without SUPERase.In). Beads and supernatant were incubated (4 °C, 2.5 h) on a nutator, before spinning (1,000 g, 4 °C, 2 min). A 20 µl unbound sample was taken and combined with 20 µl 2 x SDS loading buffer. Beads were

washed four times 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). Beads were incubated (30 min, 4 ^{o}C) twice with 300 µl of Elution Buffer (Lysis Buffer A with 1 mg/ml 3 x FLAG peptide (Sigma)). The elution supernatants were combined (a 20 µl elute sample was taken and combined with 20 µl 2 x SDS loading buffer) and RNAs were extracted using a miRNeasy kit (Qiagen). Input, unbound and elute samples were analysed by Western blot to monitor the immunoprecipitation efficiency.

2.14.3 Library generation

2.14.3.1 Linker ligation and RNA fragmentation

Immunoprecipitated RNA (3 µg in 30 µl 10 mM Tris-HCl, pH 7.0) was denatured (2 min, 80 ^{o}C) and placed on ice. A 5' adenylated (to ligate to the 3' OH of RNA), 3' blocked with a dideoxy C base (to prevent self-ligation) cloning linker 5rApp/CTGTAGGCACCATCAAT/3ddC (Integrated DNA Technologies) was ligated to the 3' ends of RNAs by first dividing the RNA into three microfuge tubes and then adding 10 µl ligation reaction mix to give final concentrations of 50 ng µl^{-1} cloning linker 1, 12 % PEG 8000, 1 x T4 Rnl2 (truncated) ligation buffer, 10 U µl^{-1} T4 Rnl2 (truncated) (NEB) and incubating for 3 h at 37 ^{o}C. Ligated RNA was incubated (95 ^{o}C, 35 min) with 20 µl Alkaline Fragmentation Buffer (100 mM $NaCO_3$ (pH 9.2), 2 mM EDTA) to fragment linker-ligated RNA to a narrow size range to reduce size bias of future steps. RNA was precipitated by incubating (30 min, -20 ^{o}C) with ice-cold 500 µl H_20, 60 µl 3M NaOAc (pH 5.5), 2 µl 15 mg ml^{-1} GlycoBlue (Ambion) and 0.75 ml isopropanol and then spinning (16,000 g, 4 ^{o}C, 30 min). Pellets were washed with 0.75 ml 80 % ethanol, dried (10 min, 22 ^{o}C) and resuspended sequentially in the same 10 µl 10 mM Tris-HCl (pH 7.0).

2.14.3.2 Size selection

Ligated and fragmented RNA was mixed with 10 µl 2x TBE-urea loading dye (89 mM Tris-Cl, 89 mM Boric acid, 2 mM EDTA, 12 % Ficoll, 7 M Urea, 2.5 mg/ml Orange G),

denatured (2 min, 80 °C) and run on a 10 well 10 % TBE-urea gel (Biorad) (200 V, 32 min). The gel was stained with sybrGold (Invitrogen) and the region containing 38-95 nt fragments excised. This gel piece was divided between three microfuge tubes and physically disrupted. Each was incubated (70 °C, 10 min) in 200 µl H_20. The tubes were pooled and gel debris was removed using a Costar-Spin-X column (Corning). RNA was precipitated by adding 60 µl 3M NaOAc (pH 5.5), 2 µl 15 mg ml^{-1} GlycoBlue and 0.75 ml isopropanol, incubating (30 min, -20 °C) and then spinning (16,000 g, 4 °C, 30 min). Pellets were washed with 0.75 ml 80 % ethanol, dried (10 min, 22 °C) and resuspended in 10 µl 10 mM Tris-HCl (pH 7.0).

2.14.3.3 Reverse transcription and circularisation

Reverse transcription (RT) was carried out by adding 3.28 µl 5 x First-Strand buffer, 0.82 µl dNTPs (10 mM each), 0.5 µl 100 µM RT primer (This is phosphorylated at the 5' end (5 Phos) for later circularisation and contains two 18 carbon spacer sequences (iSp18) to add flexibility, facilitating circularisation) (Table 2.4), denaturing (80 °C, 2 min) and incubating (48 °C, 30 min) with 0.5 µl SUPERase.In, 0.82 µl 0.1 M DTT and 0.82 µl Superscript III (ThermoScientific). 1.8 µl 1 M NaOH was added and RNA was degraded (98 °C, 20 min). To neutralise, 1.8 µl 1 M HCl was added.

cDNA was mixed with 20 µl 2 x TBE-urea loading dye, denatured (3 min, 95 °C), and run (loaded in 2 wells, 20 µl per well) on a 10 well 10 % TBE-urea gel (Biorad) (200 V, 50 min). The gel was stained with sybrGold (Invitrogen) and the regions containing the RT product excised. Both gel pieces were physically disrupted and incubated (70 °C, 10 min) in 200 µl H_20 in separate tubes. Gel debris was removed and tubes were pooled. cDNA was precipitated by adding 25 µl 3 M NaCl, 2 µl 15 mg ml^{-1} GlycoBlue and 0.75 ml isopropanol, incubating (-20 °C, 30 min) and then spinning (16,000 g, 4 °C, 30 min). The pellet was washed with 0.75 ml 80 % ethanol, dried (10 min, 22 °C) and resuspended in 15 µl 10 mM Tris-HCl (pH 8.0).

The RT product was circularised by incubating (60 °C, 60 min) with 2 µl 10 x CircLigase buffer, 1 µl 1 mM ATP, 1 µl 50 mM $MnCl_2$, 1 µl CircLigase (Epicentre). The enzyme was heat-inactivated at 80 °C, 10 min.

2.14.3.4 PCR amplification

16.7 µl 5 x HF Phusion buffer, 1.7 µl dNTPs (10 mM), 0.4 µl 100 µM Barcoded primer (A, B, C or D) (Table 2.4), 0.4 µl 100 µM Primer 1 (Table 2.4), 59.2 µl H_20, 0.8 µl Phusion polymerase (NEB) was added to 5 µl circularised DNA. This was divided (16.7 µl per tube) between five 0.2 ml tubes. PCR reactions were heated (98 °C, 30 s) and then submitted to 7 temperature cycles (98 °C, 10 s; 60 °C, 10 s; 72 °C, 10 s). At the end of cycle 3 and each subsequent cycle one tube was removed and placed on ice. 3.4 µl DNA loading dye (1.5 g Ficoll 400, 25 mg Orange G in 10 ml H_20) was added to each reaction and run on an 8 % TBE gel (Invitrogen) (180 V, 55 min). The gel was stained with sybrGold and the PCR product excised from the PCR reaction with the highest unsaturated signal without higher molecular weight products. The gel piece was physically disrupted and incubated (22 °C, overnight, with agitation) in 0.67 ml DNA soaking buffer (0.3 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA). DNA was precipitated by adding 2 µl 15 mg ml^{-1} GlycoBlue and 0.68 ml isopropanol, incubating (-20 °C, 30 min) and spinning (16,000 g, 4 °C, 30 min). The pellet was washed with 0.75 ml 80 % ethanol, dried (10 min, 22 °C) and resuspended in 10 µl 10 mM Tris-HCl (pH 8.0).

2.14.4 Sequencing

Samples were submitted to Harvard Biopolymers Facility, Massachusetts, USA. The DNA quality and quantity was determined using the TapeStation (Agilent) and qPCR and then samples were sequenced on an Illumina HiSeq 2000 machine. 50 nt were sequenced from one end using the sequencing primer. Barcoded samples were pooled so that 2-3 samples were multiplexed per lane giving ~40-60 million reads per sample.

Primer	Sequence
RT primer	5Phos/ATCTCGTATGCCGTCTTCTGCTTG/iSp18/CACTCA/iSp18/ TCCGACGATCATTGATGGTGCCTACAG
Barcoded primer 1	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACG TCTGAACTCCAGTCACATGCCATCCGACGATCATTGATGG
Barcoded primer 2	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACG TCTGAACTCCAGTCACTGCATCTCCGACGATCATTGATGG
Barcoded primer 3	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACG TCTGAACTCCAGTCACTTAGGCTCCGACGATCATTGATGG
Barcoded primer 4	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACG TCTGAACTCCAGTCACTGACCATCCGACGATCATTGATGG
Primer 1	CAAGCAGAAGACGGCATACGA
Sequencing primer	TCCGACGATCATTGATGGTGCCTACAG

Table 2.4. Primers used for library generation and sequencing. Barcoding sequences are underlined. iSp18, 18 carbon spacer; 5Phos, 5' phosphorylated end.

2.15 Bioinformatics

2.15.1 NET-seq/TEF-seq data processing

2.15.1.1 Alignment of sequencing reads

FastQ files with sanger (+33) encoded quality scores were uploaded to Galaxy (https://usegalaxy.org) (Blankenberg et al., 2010; Giardine et al., 2005; Goecks et al., 2010). Adapter sequences were clipped and reads shorter than 15 nt after clipping were discarded. Reads were aligned to the *S. cerevisiae* genome (2011 assembly) using Bowtie for Illumina (Langmead, 2010). Parameters were chosen so non-uniquely aligned reads were discarded (-n 1, -e 70, -l 28, -k 1, -m 1). The SAM file output was converted to a BAM file and downloaded (Li et al., 2009).

2.15.1.2 Narrowing reads and data visualisation

Using R/Bioconductor, aligned reads were narrowed to their 3' nucleotide and used to create BigWig files for data visualisation in IGV (Robinson et al., 2011; Thorvaldsdóttir et al., 2013).

2.15.1.3 Subtraction of data from non-FLAG-tagged strains

When NET-seq is carried out on a strain lacking a FLAG-tagged protein there is a background signal. This needs to be subtracted from the signal for a FLAG-tagged strain. For strains with non-FLAG-tagged data available, the Pol III-transcribed *SCR1* gene was divided into 10 bp bins exclusive of the first 20 bp and final 50 bp of the gene (position chrV:442007-442458). The ratio of counts (FLAG/non-FLAG) for each bin was calculated. The mean ratio was taken excluding infinite ratios and ratios greater than 5 standard deviations above the mean. 1/(mean *SCR1* ratio) x non-FLAG-tagged reads were subtracted from FLAG-tagged reads genome-wide. For visualisation, an 11 bp running mean was calculated except the running mean window did not overlap gene TSSs or TTSs or the 3' bp of exons or introns.

2.15.2 mRNA and ncRNA annotations

The TSS and PAS positions of TUs for all PCGs, SUTs, CUTs and XUTs analysed were taken from Pelechano et al., (2013), selecting first for the most abundant transcript isoforms and, of these, the longest, to obtain unique annotations for 5579 PCGs, 612 SUTs, 344 CUTs and 761 XUTs. The annotations for the start and end of coding sequences, introns, snRNAs, snoRNAs and other ncRNAs were taken from the Saccharomyces Genome Database (SGD) (www.yeastgenome.org).

2.15.3 Metagene analysis

2.15.3.1 Standard average occupancy profiles

PCG TUs were selected whose TSS or PAS was not within 100 bp on either strand of another PCG TU TSS or PAS or snRNA/snoRNA. The aim of this was to minimise potentially confounding effects from signal from neighbouring genes, seen especially with the non-strand-specific ChIP-chip data. For plots of the 5' region of TUs, TUs were aligned by their TSS. The final 200 bp upstream of the PAS were excluded to avoid transitions occurring at the 3' end of TUs influencing these profiles. The region from the TSS -100 bp to +1000 bp was split into 10 bp bins and the number of counts in each bin

was calculated from ChIP-chip data taken from Mayer et al., 2010 and 2012 and TEF-seq data from this study. The mean number of counts for each bin at each position was then calculated excluding the top 5 % and bottom 5 % of bins. These were excluded to prevent highly and lowly transcribed TUs from skewing the mean. Mean counts were plotted on a linear scale from the minimum value to the maximum value. For plots of the 3' region of TUs, TUs were aligned by their PAS, TUs shorter than 601 bp and the region from the TSS to +500 bp of TUs were excluded to avoid transitions occurring at the 5' end of TUs influencing these profiles. The region from the PAS -400 bp to +100 bp was split into 10 bp bins and again the mean number of counts for each bin was calculated and plotted in the same manner.

2.15.3.2 Rpb3-normalised average occupancy profiles

For these plots, all TUs, including overlapping TUs, were included. For plots of the 5' and 3' end of TUs, TUs were aligned and the same regions were excluded as before. For plots of the transition across the 3' end of the ORF, TUs were aligned by this 3' end. Regions downstream of the PAS were excluded. For plots of transitions relative to nucleosomes +2 to +5, the position of nucleosomes was taken from Brogaard et al., (2012). This paper uses chemical modification of histones to give a base pair resolution genome-wide map of the position of the centre of each nucleosome. The region from -100 bp to + 100 bp around each of up to the first four nucleosomes after the TSS +50 bp until the PAS were taken from each TU. Again each region was split into 10 bp bins and the number of counts from TEF-seq for each elongation factor divided by the number of counts for Rpb3 was calculated. These were multiplied by a factor specific scaling factor (see Table 2.5) to enable plotting on the same arbitrary linear scale. This is necessary as differences in sequencing depth and the efficiency with which different factors co-immunoprecipitate Pol II, leads to large differences in the level of counts for each factor. The mean (scaling factor x elongation factor)/Rpb3 ratio was then calculated for each bin after first excluding undefined ratios and the top 10 % and bottom 10 % of ratios.

These were excluded to prevent anomalous single base pair spikes in the data seen with TEF-seq from skewing the mean.

Elongation Factor	Scaling Factor
Cet1	1
Nrd1	0.75
Paf1	2.2
Pcf11	2.7
Set2	6
Spt6	0.47
Spt16	1.04
Ssu72	8

Table 2.5. Scaling factors used for plotting the Rpb3-normalised average elongation factor occupancy profiles.

2.15.4 Differential occupancy analysis

The 5579 PCG TUs were taken. TUs were excluded that overlapped with another PCG TU on the same strand. To attempt to remove some misannotated TUs, TUs longer than 750 bp, with higher counts per base over the whole TU than in the first 500 bp after the TSS in both the Cet1 and Rpb3 TEF-seq combined data were excluded. TUs with higher than 1.5x the counts over the region from the TSS -100 bp than the region from the TSS +100 bp in the Rpb3 TEF-seq combined data were excluded. Lowly expressed TUs, including those with anomalous spikes in signal possibly resulting in false positives, were also excluded. To do this, TUs were split into 100 bp bins. The counts per base of each bin were calculated after removing anomalous spikes in the data by discarding the top and bottom 2 counts from each bin. TUs with a maximum bin count per base of less than 0.5 were excluded. These exclusions left 4072 TUs. For each TEF-seq repeat, total counts from the TSS to 200 bp upstream of the PAS (PAS-200 bp) for Spt6 and Spt16 and from the TSS +400 bp to PAS-200 bp for Paf1 were calculated. Rpb3 counts for each region and each repeat were also calculated. These were rounded down to the nearest integer and TUs with negative integers were excluded. The DESeq algorithim within the DESeq2 package (Love et al., 2014) was then applied to each set of factor and Rpb3 counts. TUs

were selected as being significantly enriched or depleted for a particular factor on Pol II if the p value, adjusted for multiple testing using the Benjamini-Hochberg method, was less than 0.05.

2.15.5 Differential transcription analysis

Rpb3 NET-seq counts for TUs were taken for each condition. The DESeq algorithim was applied to counts between two conditions, including replicates where available. TUs were selected as being significantly differentially transcribed or 2-fold differentially transcribed if the p value, adjusted for multiple testing using the Benjamini-Hochberg method, was less than 0.05.

Chapter 3.

Mapping the location of transcription elongation complex components genome-wide

3 Mapping the location of transcription elongation complex components genome-wide

3.1 Introduction

Numerous studies have been carried out to map the location of components of the TEC bound to chromatin. These have proved highly informative in defining the mechanism and positions with which different factors are recruited to chromatin at specific positions along TUs. This in turn has been useful for determining functions for such factors. The large majority of these studies have used the technique of chromatin immunoprecipitation (ChIP) (Kim et al., 2010; Mayer et al., 2010, 2012; Meinel et al., 2013).

3.1.1 Chromatin immunoprecipitation

ChIP involves first crosslinking chromatin, which most commonly is carried out by exposing cells to the cross-linking agent formaldehyde, creating heat-reversible protein-DNA and protein-protein cross-links. Cells are then lysed, followed by DNA fragmentation, either by sonication or restriction enzyme digestion, to give fragment sizes of approximately 300-500 bp. Immunoprecipitation using antibodies specific to the factor of interest (or epitope for epitope-tagged factors) is then carried out, enriching for the DNA fragments cross-linked to this factor. After washing and removal of the cross-linked proteins, the DNA is purified. The selective enrichment of particular DNA fragments can then be detected either by quantitative PCR (qPCR) or using a genome-wide method of either hybridising fragments to a genome-wide tiling array (ChIP-chip) or generating libraries from the fragments followed by high-throughput sequencing (ChIP-seq) (Mahony and Pugh, 2015).

3.1.2 Problems with ChIP

Despite the power of ChIP, it has a number of problems. A key issue lies in its low spatial resolution. The use of high-density tiling arrays has increased resolution with several

recent studies using arrays with 25 nt probes spaced at 4 nt intervals (Mayer et al., 2010). ChIP-seq further increases resolution (Ho et al., 2011), however, data for most TEC components in *S. cerevisiae* using ChIP-seq is lacking. An additional constraint on the technique's resolution is caused by the size of the immunoprecipitated DNA fragments. Increasing the extent of DNA fragmentation to an average length of 200 bp increases resolution (Fan et al., 2008). A modification of ChIP in which immunoprecipitated DNA is digested with an exonuclease up to the point of crosslinking (ChIP-exo) further increases resolution (Rhee and Pugh, 2012b).

Another issue is caused by lack of sensitivity both in detecting lowly bound regions above background noise and small changes in binding due to limits in the dynamic range of detection. This is dependent on the efficiency with which bound fragments are enriched above input DNA and the sensitivity of the detection method. Again ChIP-seq improves on ChIP-chip by increasing the dynamic range (Ho et al., 2011).

Studies using ChIP to map TEC component locations often make the assumption that the factor solely interacts with chromatin via its interaction with the TEC. This assumption may be flawed as many elongation factors also interact with other chromatin components, for example both Spt6 and Spt16 interact with histones (Duina, 2011). Indeed some factors can still be detected bound to chromatin once the domain by which they interact with Pol II has been removed (Mayer et al., 2010; Tudek et al., 2014). Moreover, ChIP is unable to distinguish between active and inactive TECs. ChIP's lack of strand-specificity also means it is not possible to distinguish between factors interacting with the TEC transcribing the sense strand and that transcribing the antisense strand of DNA. This is particularly problematic in highly compact, pervasively transcribed genomes, such as that of *S. cerevisiae*, where many TUs overlap on both strands.

The use of a cross-linking step has made it possible to detect even weakly bound chromatin factors. This step may, however, create non-native interactions. Additionally,

the higher order structure of chromatin and the occurrence of transcription-dependent gene loops have the potential to create misleading binding patterns for certain factors.

3.1.3 Improving on ChIP

A number of techniques have been developed to improve upon ChIP in mapping Pol II genome-wide in several model organisms.

3.1.3.1 Run-on sequencing

Nuclear run-on is a technique, which first involves permeabilisation of cells, often by adding sarcosyl, or isolation of cell nuclei. Transcription is then reinitiated in the presence of radiolabelled UTP, leading to the labelling of nascent transcripts (Smale, 2009). By hybridising these transcripts to tiling arrays, transcription rates can be determined genome-wide (García-Martínez et al., 2004). Genome run-on sequencing (GRO-seq) is an adaptation of these techniques in which transcription is reinitiated in isolated nuclei using a ribonucleotide analog (5' bromouridine 5' triphosphate (BrUTP)) (Core et al., 2008). Labelled nascent RNAs are then immunoprecipitated and used to construct cDNA libraries for sequencing. This technique gives strand-specific information on the position of transcriptionally active Pol II complexes. The distance travelled by Pol II after transcription reinitiation cannot be determined, limiting the spatial resolution to approximately 50 bp (Core et al., 2012). Precision run-on sequencing (PRO-seq) has since been developed, which limits the distance travelled by the reinitiated polymerase to a few nucleotides (Kwak et al., 2013), giving near single nucleotide resolution. With all these techniques, the ability of Pol II to reinitiate in non-native conditions is being assayed. This reinitiation may not be possible in all circumstances, potentially biasing results. The manipulation of cells necessary to isolate nuclei or permeabilise cells may also cause stress, potentially altering transcription. So far GRO-seq and PRO-seq have not been adapted for *S. cerevisiae*. The presence of a cell wall makes isolating nuclei difficult, which may prevent the use of these techniques in yeast.

3.1.3.2 Native elongating transcript sequencing

Native elongating transcript sequencing (NET-seq) is a technique in which Pol II is immunoprecipitated from non-cross-linked cell lysate by adding a 3x FLAG epitope tag to its Rpb3 subunit. After washing, nascent RNA bound to the active site of Pol II is purified and used to construct cDNA libraries so that, upon sequencing, the sequence of the 3' ends of these RNAs is obtained (Figure 3.1). This gives a strand-specific nucleotide resolution map of the positions of actively transcribing Pol II genome-wide (Churchman and Weissman, 2011). As cells are immediately flash frozen and cryogenically lysed after growth, the potential for stress-dependent transcriptional changes to occur is minimised.

3.1.4 Aim

Due to its higher resolution and strand-specificity, NET-seq is an improvement on ChIP in mapping active Pol II genome-wide. The aim of this chapter was to develop a technique to map with strand-specificity and higher resolution the location of components of the TEC genome-wide. To do this, the standard NET-seq protocol was first established in the laboratory. Then this standard technique was adapted so that the 3x FLAG epitope tag was added to an elongation factor instead of Rpb3. This allowed immunoprecipitation of this elongation factor and co-immunoprecipitation of associated Pol II. By then purifying the nascent RNAs from the active site of Pol II and carrying out the same library generation and sequencing steps, it was possible to obtain a strand-specific nucleotide resolution genome-wide map of the locations of the TECs bound by this elongation factor.

This adapted NET-seq technique, termed transcription elongation factor sequencing (TEF-seq), was used to map the location of several different TEC components. The high resolution afforded by this new technique revealed increased levels of complexity within the transcription cycle due to additional transitions in the exchange of factors within the TEC as well as periodic binding of certain factors. The sensitivity of the

technique, along with its strand-specificity, made it possible to resolve whether the series of transitions the TEC undergoes is uniform for all TUs. This revealed a possible role for increased Paf1 binding to Pol II in increasing the nuclear export of the RNA being transcribed. TEF-seq data also show that splicing intermediates co-immunoprecipitate with the TEC, suggesting that splicing occurs to some extent co-transcriptionally at many intron-containing genes.

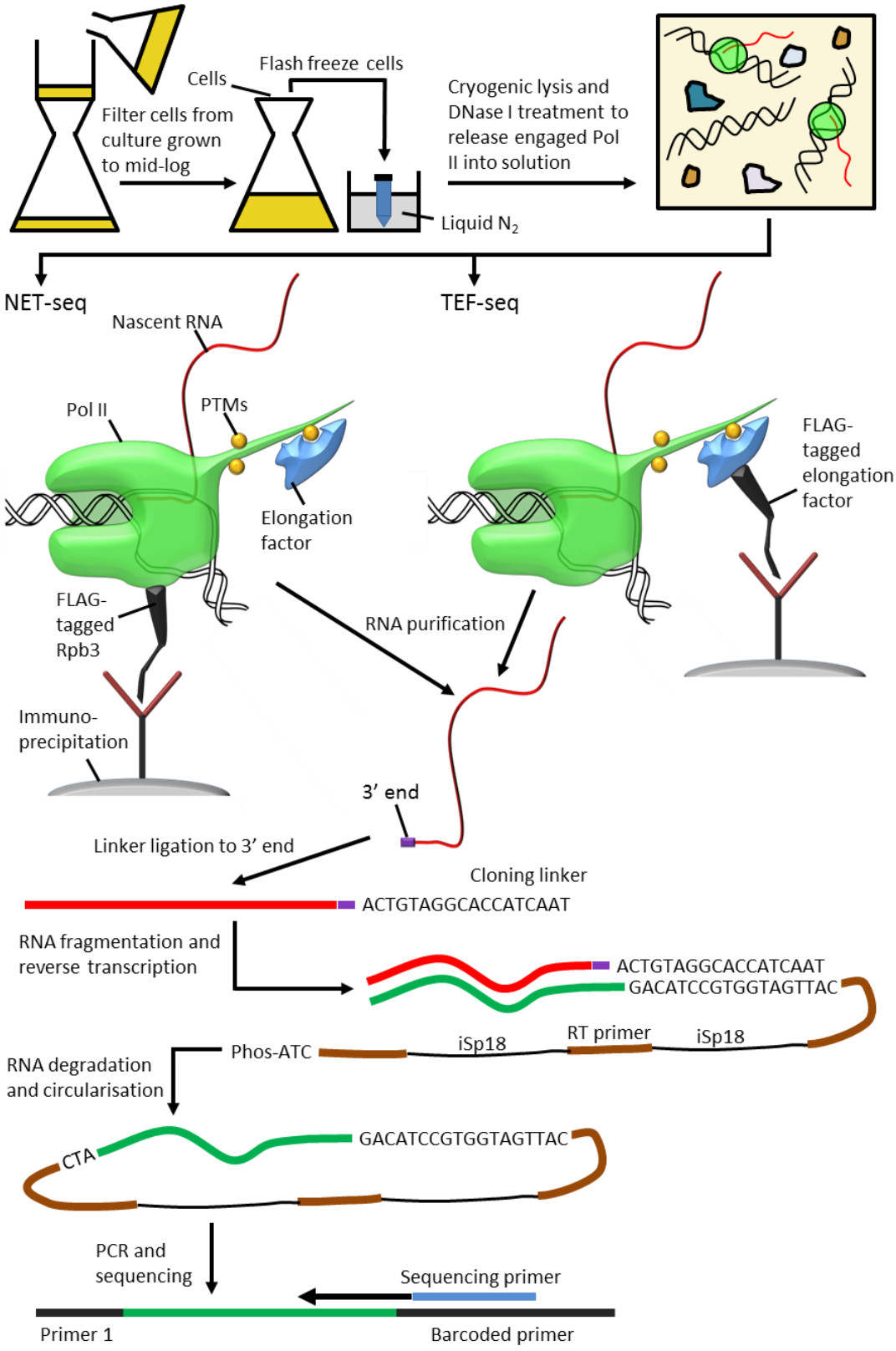


Figure 3.1. Native elongating transcript-sequencing (NET-seq) and transcription elongation factor sequencing (TEF-seq). Schematic outlining the NET-seq technique and its adaptation for TEF-seq.

3.2 Results

3.2.1 Establishing the standard NET-seq technique

3.2.1.1 Immunoprecipitation of Pol II

The standard NET-seq protocol was carried out as detailed in the Methods (Figure 3.1). Briefly, 2 L of BY4741 *S. cerevisiae* culture grown to mid-log (OD_{600} of 0.6-0.7) in rich media (YPD) was flash frozen and cryogenically lysed. Lysate was resuspended in lysis buffer and extensively treated with DNase I to maximise release of transcriptionally engaged Pol II into the soluble fraction. After clearing the lysate, Pol II was immunoprecipitated, via its endogenously expressed FLAG-tagged Rpb3 subunit, from this soluble input fraction using anti-FLAG antibody conjugated agarose beads.

During immunoprecipitation of Pol II, 20 µl samples were taken from the input, unbound and eluted fractions. These were assayed for the levels of FLAG-tagged Rpb3 (Rpb3-3xFLAG) and Rpb1 by Western blotting (Figure 3.2A and B) to assess immunoprecipitation efficiency. Rpb3 strongly binds to the beads, showed by the absence of FLAG signal in the unbound sample. The Rpb3 signal in the eluted sample is greater than the input sample and the final eluted volume is approximately 1/10 the input volume. Therefore the efficiency of Rpb3 immunoprecipitation from the input is greater than 10 %. The Rpb1 western using the σ-80 antibody specific to the N-terminus of Rpb1 (Santa cruz) shows this subunit is co-immunoprecipitated with Rpb3 in the eluted sample, suggesting that the entire Pol II complex has been immunoprecipitated. There are both high (~235 kDa) and low (~220 kDa) molecular weight bands, which correspond to the hyper-phosphorylated and hypo-phosphorylated forms of Rpb1, respectively. The hyper-phosphorylated form is created by phosphorylation of the CTD of Rpb1 during transcription initiation and elongation (Eick and Geyer, 2013). The signal of the hyper-phosphorylated band is stronger than the hypo-phosphorylated band, showing that the majority of Pol II that has been immunoprecipitated is of the

transcriptionally active form, suggesting the lysis and DNase I steps have effectively released this form for immunoprecipitation.

RNA was purified from the eluted fraction using a miRNeasy kit (Qiagen) giving a yield of approximately 3-5 μg. RNA in the process of being transcribed by Pol II is estimated to make up 0.34 % of the total RNA in the cell with the immunoprecipitation step estimated to enrich this fraction approximately 100-fold (Churchman and Weissman, 2011). This means that, despite the enrichment, approximately 66 % of the RNA purified consists of mature transcripts. Analysis of the purified RNA by denaturing TBE-urea polyacrylamide gel electrophoresis shows prominent bands confirming the presence of mature transcripts, including the 5S rRNA at 170 bp (Figure 3.2C).

3.2.1.2 Library generation

To generate libraries for sequencing from the 3' ends of purified RNAs, a 20 bp DNA linker was ligated to the 3' hydroxyl group at the 3' end of RNAs. RNA was then fragmented and linker ligated RNA fragments from 38-95 nt were selected by gel electrophoresis to minimise length bias in downstream enzymatic reactions (Figure 3.2D). These were reverse transcribed using a long primer with sequence complementary to the DNA linker sequence and, following RNA degradation, the product was purified by gel electrophoresis (Figure 3.2E). This was circularised and 4-6 cycles of PCR were carried out to give double-stranded DNA with barcoded adapters specific for Illumina high-throughput single read 50 nt sequencing. This PCR product was purified by gel electrophoresis and submitted for sequencing (Figure 3.2F). Sequencing was carried out by the Harvard Biopolymers Facility.

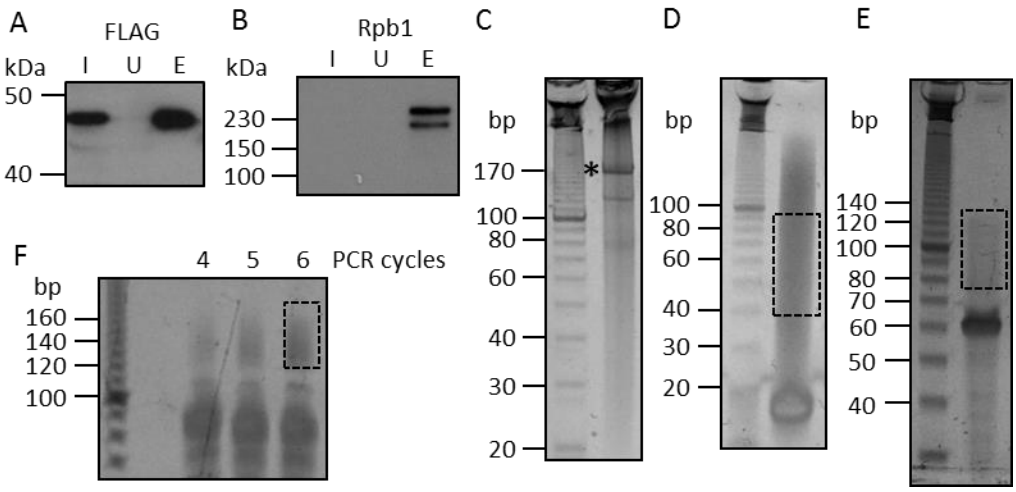


Figure 3.2. Immunoprecipitation and library generation steps of standard NET-seq protocol. A and B) Western blots showing the levels of FLAG-tagged Rpb3 (A) and Rpb1 (B) in input (I), unbound (U) and eluted (E) samples from the NET-seq immunoprecipitation step. C) TBE-urea gel showing RNA purified from the eluted fraction. 5S rRNA is marked with an asterisk. D) TBE-urea gel showing size selection of linker-ligated RNA fragments from 18-95 nt. E) TBE-urea gel showing size selection of reverse transcribed product. F) TBE gel showing the final barcoded products following 4, 5 and 6 PCR cycles. All TBE-urea gels are stained with sybrGold. Dotted boxes show the excised region.

3.2.1.3 Data analysis

Sequencing reads were aligned and narrowed to their 3' nucleotide as explained in the Methods. Total reads aligning between the TSS and PAS of PCG TUs from two repeats and from the published NET-seq data from a wild type BY4741 strain (Churchman and Weissman, 2011) were counted. Annotations used were taken from Pelechano et al. (2013) as explained in the Methods. A regularised log transformation (rlog) was applied to the counts using the DESeq2 package (Love et al., 2014) in R/Bioconductor. This effectively produces $\log_2$ transformed data normalised for library size for genes with medium to high counts while shrinking differences at genes with low counts. A standard $\log_2$ transformation exaggerates relative differences at low count genes at which random noise likely dominates any meaningful signal. Transformed data from two repeats strongly correlated (spearman correlation coefficient greater than 0.98) with

the published data, showing that the technique had been effectively replicated (Figure 3.3).

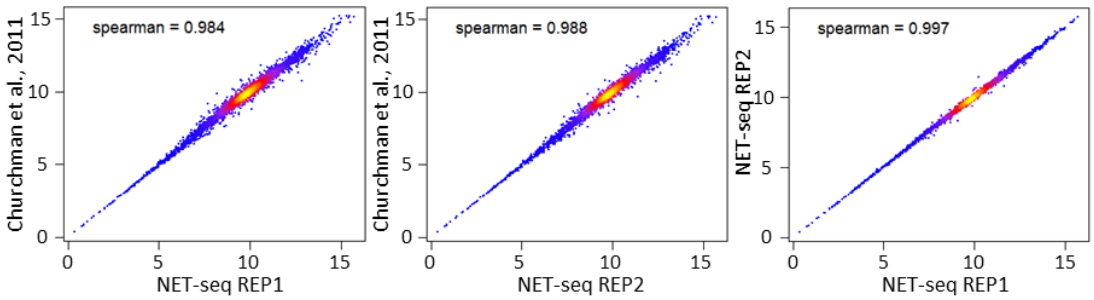


Figure 3.3. Correlation between published NET-seq data from Churchman and Weissman (2011) and two NET-seq repeats (REP1 and REP2) from this study. Reads aligning to each PCG TU have been counted and correlated after applying an rlog transformation. Scales show rlog(read counts). Scatter plots are colour-coded from low density (blue) to high density (yellow).

3.2.2 Improving the NET-seq technique

As highlighted above, the majority of the RNA that is purified and subsequently sequenced using NET-seq is not RNA in the process of being transcribed by Pol II. Instead the majority of sequencing reads map to the Pol I- and Pol III-transcribed rDNA locus. Consequently these are discarded, as this region is repetitive and only uniquely aligning reads are accepted. Analysis of the Pol III transcribed 0.5 kb *SCR1* gene shows NET-seq reads aligning to it in data from this study and the published data (Figure 3.4A). As sequencing occurs from the 3' end of transcripts, if the signal at the *SCR1* locus came solely from mature full-length *SCR1* transcripts, reads should only be seen at the 3' end of the gene. However, despite there being a peak of reads at the 3' end, there are also many reads aligned across the body of this gene. This suggests that shorter *SCR1* transcripts have also been purified, possibly resulting from degradation of full-length transcripts. If there is a signal at this Pol III-transcribed gene from RNA that is not being actively transcribed by Pol II, it is possible that a similar signal is also present at Pol II-transcribed genes.

To assay whether this was the case, the NET-seq protocol was applied to wild-type BY4741 lacking the epitope tag on Rpb3. In two repeats, approximately 3 µg of RNA were purified each time. The data generated also show a similar pattern of reads aligning across *SCR1* (Figure 3.4A). There are also reads aligning to Pol II transcribed genes. These observations demonstrate that in the standard NET-seq protocol there is a background signal that is not specific to the immunoprecipitation of Rpb3-3xFLAG. At intron-containing genes, this signal is higher over exons than introns (Figure 3.4B). This suggests the signal results from mature spliced transcripts rather than signal from non-specific binding of actively transcribing Pol II to the anti-FLAG antibody-conjugated agarose beads. These mature transcripts must have been degraded to give the signal from their 3' ends across the body of the gene. mRNAs are generally either degraded from their 5' end by the 5' to 3' exonuclease Xrn1 following decapping or from the 3' end by the exosome. Only degradation from the 3' end would give rise to the range of 3' ends observed. Alternatively this degraded mRNA signal may be generated after cell lysis by the release of ribonucleases; however, as the immunoprecipitation is carried out at 4 ^{o}C in the presence of ribonuclease inhibitors, their activity should be minimised.

All current publications using NET-seq (Churchman and Weissman, 2011; Marquardt et al., 2014; Nguyen et al., 2014; Terzi et al., 2011) do not carry out these controls in non-tagged Rpb3 strains and make no attempt to remove or account for the background signal present in their data sets. As the level of enrichment of the Pol II specific signal is generally several times greater than the background signal, many of the conclusions drawn from these papers may still be correct. Nonetheless, any observations from these data attributed to transcription could potentially be confounded by this background mature transcript signal, especially for mutants that strongly alter transcript stability. Ideally, all NET-seq experiments should be carried out with a non-tagged strain control.

To subtract the background signal, the *SCR1* gene was used to obtain a scaling factor. This was firstly because all reads aligning to *SCR1* were assumed to result from the background signal in both strains because this gene is not transcribed by Pol II and secondly because the signal level at this gene is high, reducing the effect of fluctuations due to noise. The reads from the two non-tagged strain replicates were combined. The *SCR1* gene was then divided into 10 bp bins from position chrV:442007 to chrV:442458. This excludes the first 20 bp, which have few reads, as reads need to be longer than 18 bp to align, and the final 50 bp of the gene, which have a peak in reads from mature full-length *SCR1* transcripts and tend to have greater variability between repeats. The ratio of counts (FLAG-tagged/non-FLAG-tagged) for each bin was calculated. The standard deviation in the ratios for all bins was less than one suggesting the background signal was almost constant across *SCR1* for the different strains. The mean ratio was taken and 1/(mean *SCR1* ratio) x non-FLAG-tagged reads were subtracted from FLAG-tagged reads for each nucleotide genome-wide. This process will be termed no-tag normalisation and is demonstrated for the *YDR064W* locus (Figure 3.4B). For visualisation, the no-tag-normalised data was smoothed by calculating an 11 bp running mean except the running mean window did not overlap TU TSSs or PASs or the 3' nt of exons or introns (Figure 3.8A-F show example loci).

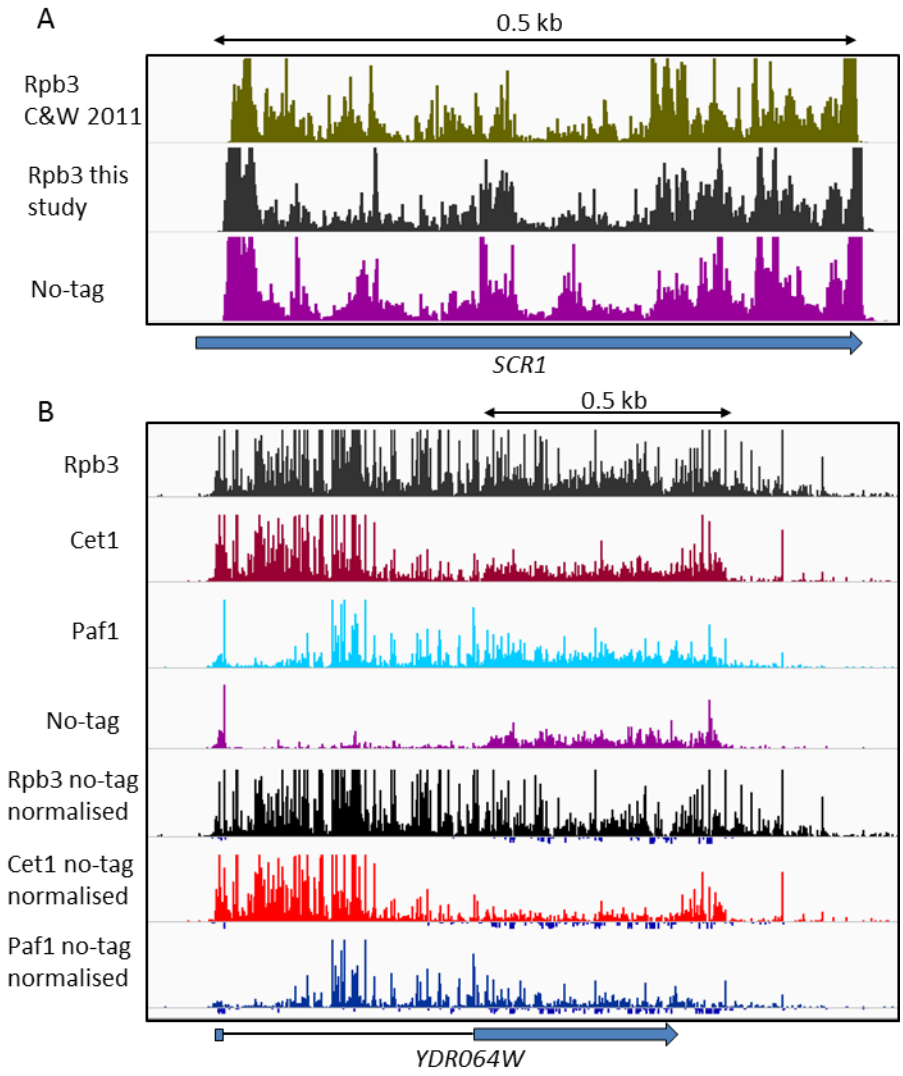


Figure 3.4. The normalisation process. A) Reads aligning to Pol III-transcribed *SCR1* from published NET-seq data (Churchman and Weissman, 2011 (C&W 2011)), NET-seq data from this study and using the NET-seq protocol on a non-tagged BY4741 strain. B) Tracks showing data for *YDR064W* from Rpb3, Paf1 and Cet1 TEF-seq pre- and post-normalisation by subtraction of reads from the non-tagged strain. Reads over *SCR1* are used to determine a scaling factor for tagged factor TEF-seq reads relative to no-tag TEF-seq reads before subtraction. Pre-normalisation tracks are scaled relative to the no-tag track according to this scaling factor. No-tag normalised tracks are not smoothed and are on the same scale as their pre-normalised tracks.

3.2.3 Co-immunoprecipitation of Pol II with elongation factors

The positions of elongation factors have so far been mapped using ChIP. Due to its advantages over ChIP, adapting NET-seq to map the locations of several elongation factors within transcriptionally active TECs would yield higher resolution, strand-specific data. To do this, strains in which elongation factors had been C-terminally FLAG-

tagged at their endogenous genomic location were generated. The elongation factors chosen were: Cet1, Isw1, Nrd1, Paf1, Pcf11, Spt6, Spt16 and Yra1 (see Chapter 1 for details of these factors). These were tested for their ability to co-immunoprecipitate Pol II using a scaled down version of the standard NET-seq immunoprecipitation protocol in which 1/4 of all materials was used and the final 150 µl eluted sample was mixed with 150 µl 2x SDS loading buffer for Western blotting. Levels of FLAG-tagged protein in the eluted fraction show each factor was effectively immunoprecipitated with ~10 % or greater efficiency (Figure 3.5).

The levels of total Rpb1 (σ-80) and Tyr1P (3D12), Ser2P (3E10), Ser5P (3E8) and Ser7P (4E12) CTD forms in the eluted fraction were assayed (antibody clone names are shown in brackets) (Figure 3.6A and B). Western blots were then quantified by measuring the signal intensity of each band using ImageJ (Figure 3.6C). The level of Ser2P and Ser5P CTD forms co-immunoprecipitated from Yra1- and Isw1-tagged strains are low and very similar to a non-tagged strain suggesting these factors do not specifically enrich for Pol II. As Yra1 contains a CID, has been shown to interact with CTD peptides with Ser2, Ser5 doubly phosphorylated repeats and maps to TUs in ChIP experiments it is surprising that it does not co-immunoprecipitate Pol II (MacKellar and Greenleaf, 2011; Meinel et al., 2013). This may be because the Yra1-TEC interaction is weak and particular immunoprecipitation conditions or cross-linking is required to detect it. Isw1 is an ATP-dependent chromatin remodelling enzyme, which has not been shown to interact with Pol II, but does interact with H3K36me3 (Smolle et al., 2012). This, therefore, acts as a control to show that not all chromatin associated proteins co-immunoprecipitate Pol II. This suggests co-immunoprecipitation of Pol II by other factors occurs through a direct interaction with the TEC rather than indirectly, by interacting with other chromatin components.

Not surprisingly, Rpb3 co-immunoprecipitates more total Rpb1 than any of the elongation factors. It is also the only factor to co-immunoprecipitate the hypo-phosphorylated form, suggesting the other factors only bind Pol II when it is actively transcribing. The total Rpb1 level co-immunoprecipitated by Spt6 is at least twice that of any other elongation factor. Spt6 Ser2P and Ser7P levels are similar to Rpb3 suggesting this factor co-immunoprecipitates these forms as efficiently as the Pol II subunit. Tyr1P levels, however, are lower than Rpb3, which is surprising as this mark is specifically bound by the CID of Spt6 (Mayer et al., 2012). Both Spt6 and Rpb3 co-immunoprecipitate a form of Rpb1 that runs at a size between the hypo- and hyper-phosphorylated forms. This can be detected by all the phospho-specific antibodies, possibly representing a form of Pol II with an intermediate level of phosphorylation. Spt16 co-immunoprecipitates total Rpb1 and Ser2P at ~2.5x lower levels than Spt6 and Tyr1P, Ser5P and Ser7P at even lower levels. Cet1 co-immunoprecipitates Ser2P and Tyr1P at much lower levels than Ser5P and Ser7P forms relative to the other factors. Nrd1 and Paf1 co-immunoprecipitate Rpb1 and its different phosphorylated forms at similar levels. Nrd1 does not co-immunoprecipitate Ser5P very strongly, although it has previously been reported to co-immunoprecipitate the Ser5P and not the Ser2P form (Vasiljeva et al., 2008). That study, however, used antibodies H5 (Ser2P specific) and H14 (Ser5P specific) whose binding specificities are influenced by modification to nearby residues (Chapman et al., 2007). That study also showed Nrd1 binds preferentially to Ser2, Ser5 doubly phosphorylated CTD peptides than to Ser5P peptides. Pcf11 shows similar levels of co-immunoprecipitation of Ser2P and lower levels of Ser5P and Ser7P compared to Paf1 and Nrd1. It also shows slightly higher levels of co-immunoprecipitation of Tyr1P compared to these factors, a mark which has previously been shown to inhibit Pcf11 binding *in vitro* (Mayer et al., 2012).

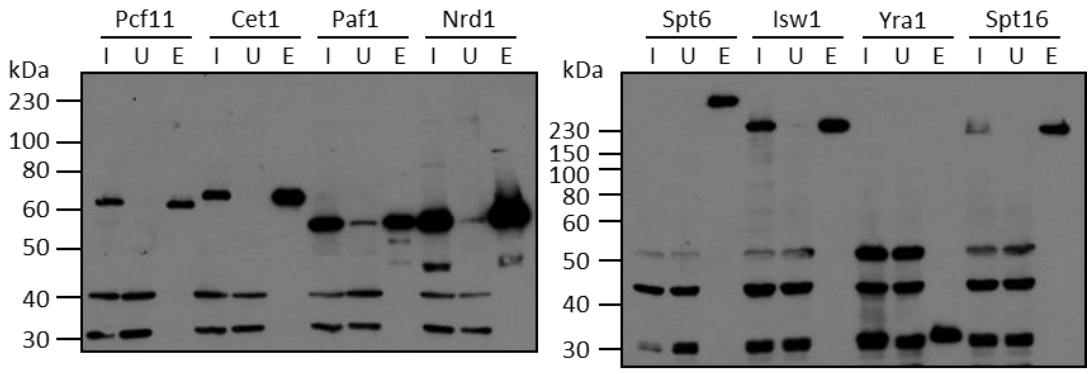


Figure 3.5. Immunoprecipitation of FLAG-tagged factors. Western blots showing the levels of FLAG-tagged factors in input (I), unbound (U) and eluted (E) samples using the standard NET-seq immunoprecipitation protocol.

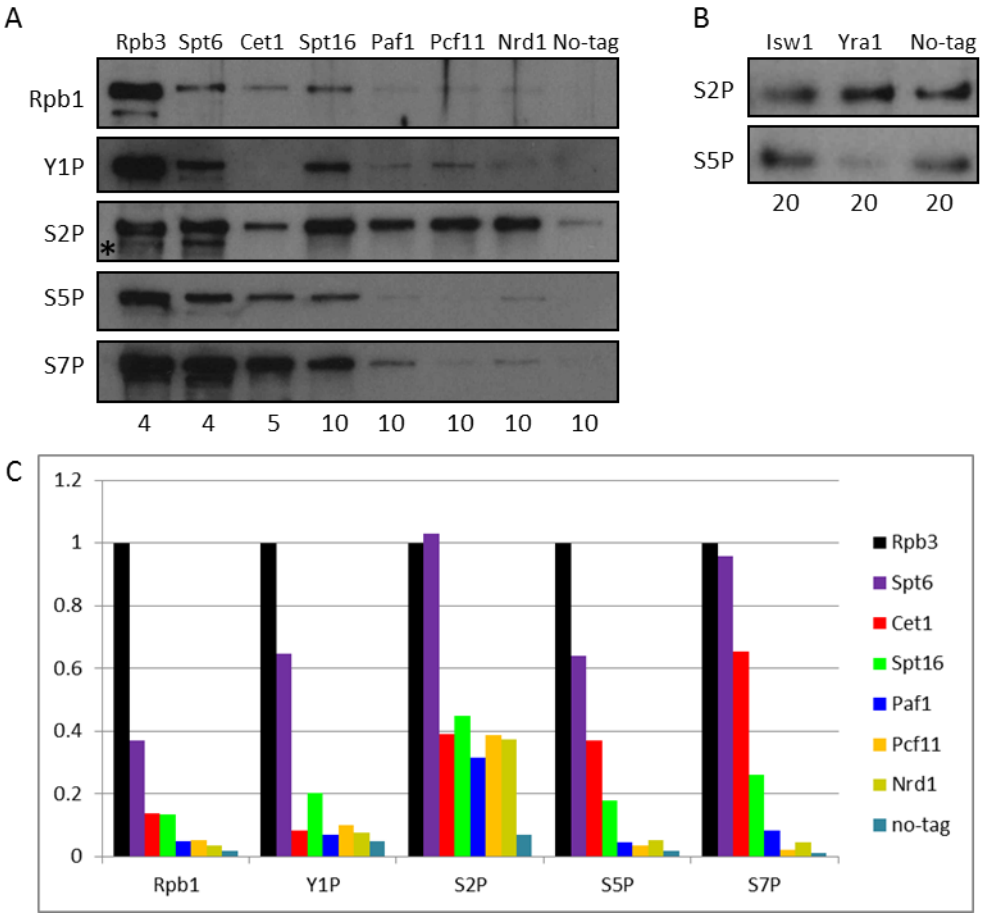


Figure 3.6. Co-immunoprecipitation of Rpb1 by FLAG-tagged factors. A and B) Western blots showing the levels of total Rpb1 and its Tyr1P, Ser2P, Ser5P and Ser7P phosphorylated forms co-immunoprecipitated with each FLAG-tagged factor in the eluted fraction using the standard NET-seq immunoprecipitation protocol. Numbers show the volume loaded in µl of the eluted fraction. Asterisk shows form of Pol II with intermediate level of phosphorylation. C) Bar chart showing quantification of Western blots in A using ImageJ. Signal for each factor for each blot is first adjusted for volume loaded and then plotted as a ratio of the Rpb3 signal within that blot.

3.2.4 Applying the adapted NET-seq technique to elongation factors

The NET-seq protocol was applied to the FLAG-tagged elongation factor strains. The background signal was removed using no-tag normalisation as detailed previously (see Appendix for total read counts before and after no-tag normalisation). This data processing is demonstrated in Figure 3.4B showing pre- and post-no-tag normalisation tracks at the intron-containing *YDR064W* gene for Paf1- and Cet1-FLAG-tagged strains. No-tag normalisation is especially necessary for tagged elongation factor strains as the

efficiency with which Pol II is immunoprecipitated is generally lower than with Rpb3-3xFLAG leading to a higher proportion of background signal relative to FLAG-tagged factor-specific signal. This technique will be termed transcription elongation factor sequencing (TEF-seq).

For the strains *YRA1-3xFLAG* and *ISW1-3xFLAG*, following no-tag normalisation there is no obvious signal above the noise confirming these factors do not efficiently co-immunoprecipitate Pol II. As suggested, the Yra1-TEC interaction may be weak. Consequently the use of TEF-seq, due to it lacking a cross-linking step, may only be possible with factors that interact strongly with the TEC. Data from these two strains were not further analysed.

3.2.5 Replicates show adapted NET-seq is highly reproducible

Two TEF-seq replicates (with no-tag normalisation) were obtained for Rpb3-, Cet1-, Nrd1-, Paf1-, Pcf11-, Spt6- and Spt16-tagged strains. The Nrd1 data were not included in the following analysis because it contains a confounding signal, as explained below, which could distort the analysis. For each replicate, reads mapping to each PCG TU, from the TSS to the PAS, using annotations specified in the methods were counted and rounded down to give integer values. An rlog transformation was applied to these counts. Correlating the transformed data for the replicates shows the high reproducibility of the data (Figure 3.7A), with spearman correlation coefficients for all factors not lower than 0.91.

A heatmap of the Euclidean distance between all replicates and all factors shows that the replicates cluster together (Figure 3.7B). This can also be visualised using principal component analysis (PCA) in which the samples are projected into the 2-dimensional directions that explain the most differences between them (Figure 3.7C). Again the replicates generally cluster closely together. The heatmap and PCA suggest the Rpb3,

Spt6 and Spt16 data are very similar to each other, quite similar to the Cet1 data and least similar to the Pcf11 and Paf1 data.

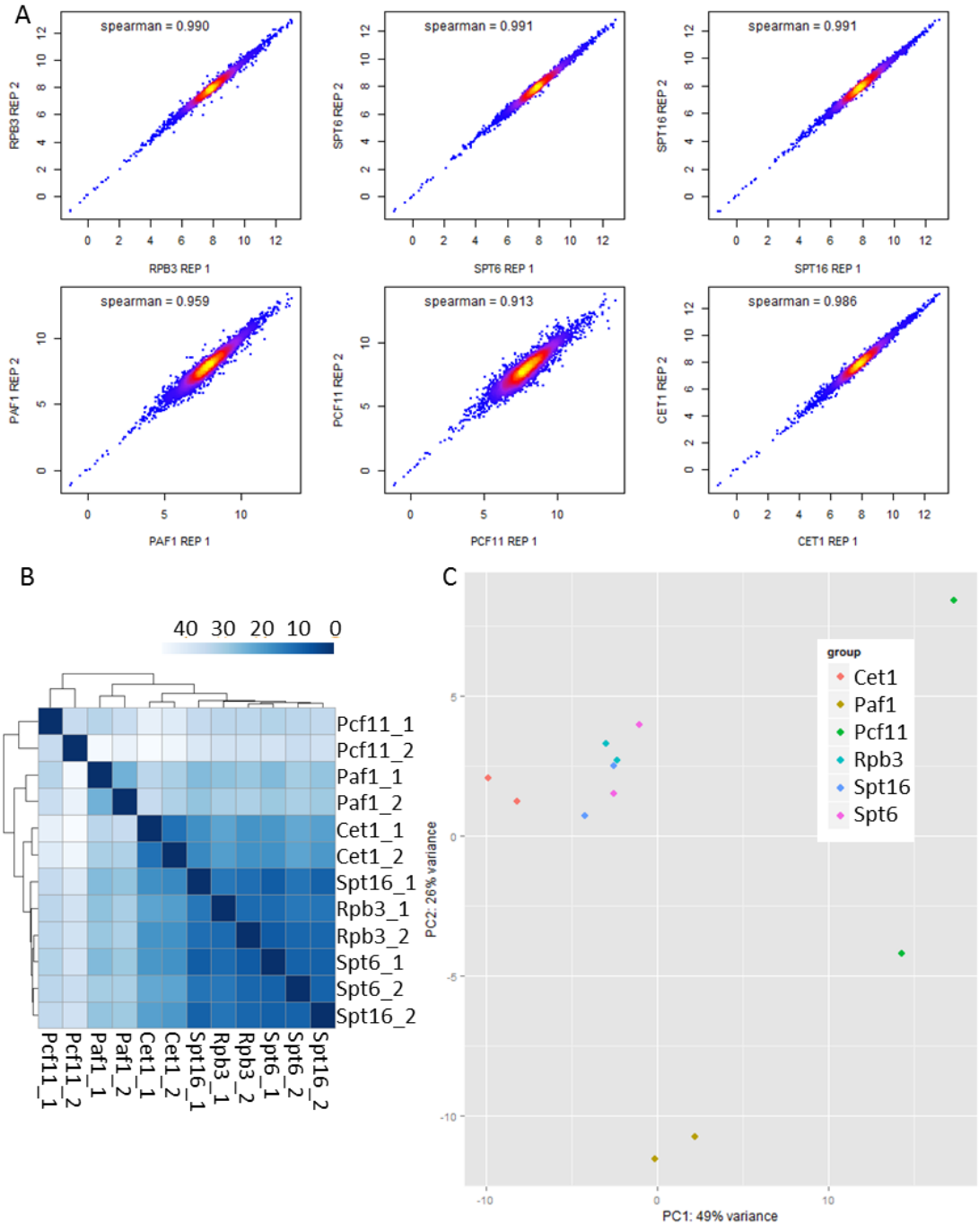


Figure 3.7. TEF-seq is highly reproducible. A) Scatter plot showing the correlation between two TEF-seq biological replicates of rlog transformed total counts at all PCG TUs for six factors. B) Heatmap showing the Euclidean distance between replicates for all six factors. C) Principal component analysis showing the similarity between factors and replicates

3.2.6 Analysis of factor occupancy profiles shows the advantages and disadvantages compared to ChIP

3.2.6.1 Individual loci

Due to the high correlation between replicates the raw reads were added together and no-tag normalised. Rpb3, Cet1, Nrd1, Paf1, Pcf11, Spt6 and Spt16 combined data are shown for six loci following smoothing as before (Figure 3.8A-F). Annotations shown are for the most abundant transcript isoform for each TU (Pelechano et al., 2013). A different arbitrary scale is used for each track (Table 3.1), which is consistent across all loci, allowing relative changes in factor binding to Pol II to be visualised.

Factor	Maximum value of scale in Figure 3.8A-F/TEF-seq data counts
Rpb3	11
Spt6	20
Spt16	9
Cet1	11
Paf1	4
Pcf11	4
Nrd1	15

Table 3.1. Maximum value of the scale for each factor TEF-seq track in Figure 3.8A-F.

Observation of these loci suggests Spt6 and Spt16 closely match the occupancy profile of Rpb3. Cet1 shows maximum occupancy in the first ~300 bp of each TU in agreement with its role in capping. Paf1 and Pcf11 show lower occupancy at the start of TUs rising towards the 3' end. This is generally consistent with published conclusions from ChIP data, although Paf1 has previously been described as having a similar occupancy profile to Spt6 and Spt16 (Mayer et al., 2010). Comparison of factor occupancy profiles from the ChIP-chip experiments from Mayer et al, (2010, 2012) with the profiles from TEF-seq shown for all factors at the loci in Figure 3.8A-D highlights the advantages of this new technique. The strand-specificity and higher resolution enables clear visualisation of the start and end points of TUs. Factors that associate with Pol II transcribing a TU on one

strand can be reliably distinguished from those associating with Pol II transcribing the opposite strand or nearby TUs.

The Nrd1 TEF-seq occupancy profile shows a sharp peak at the 5' end of many TUs. For TUs with an intron within ~50 bp of the TSS there is a peak in signal over the first exon and the 5' end of the second exon with lower signal over the intron (Figure 3.9). This suggests this signal is from spliced RNA rather than Nrd1 binding to the transcribing Pol II at these positions, especially as it is not seen in Nrd1 ChIP-chip data (Mayer et al., 2012) (data not shown). Much like the signal from the non-FLAG-tagged strain, this cannot be full-length spliced RNA otherwise the range of 3' end reads near the 5' end of the RNA would not be seen. It may be that Nrd1 binds the short, possibly degraded, RNA species, which are present as a background signal in all TEF-seq data. This binding would have to be specific to the RNAs mapped to the 5' end of TUs because these are the ones which appear to be specifically enriched. In support of this model, TUs with a high 5' peak in signal in the Nrd1 data sets also have a high signal across the whole TU in the non-tagged strain data sets. Nrd1 has also been shown to preferentially bind to the 5' end of mRNA transcripts and associates with the 5' RNA cap binding complex (Vasiljeva and Buratowski, 2006; Webb et al., 2014). As shown here for Nrd1, for factors with RNA binding properties, caution must be taken when interpreting the data, which is one disadvantage of the technique compared to ChIP.

In the Nrd1 data, there is also likely to be signal from Nrd1 associated with Pol II, as Pol II co-immunoprecipitates with the factor (Figure 3.6). Nrd1 has been shown by ChIP to be enriched at the 5' end of TUs, which is thought to be due to it binding specifically to Ser5P Rpb1 CTD localised to this position (Vasiljeva et al., 2008). This is not particularly apparent at individual loci in TEF-seq data, although a 5' peak in signal may be partially masked at some TUs by the non-Pol II-specific sharp 5' peak in signal. Instead there is a

broader signal extending into the body of TUs, consistent with the association of Nrd1 with Ser2P Rpb1 CTD seen by Western blot (Figure 3.6).

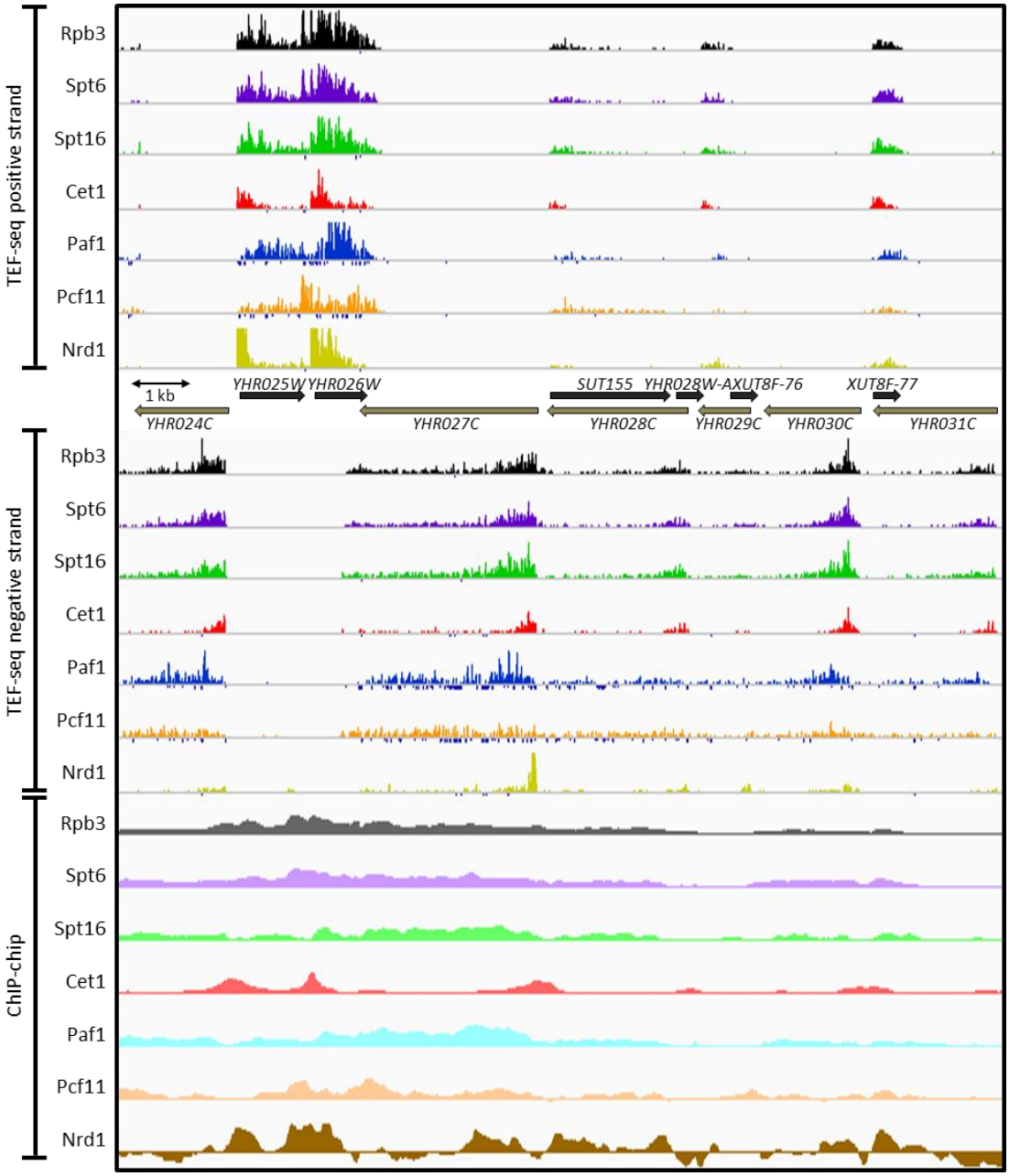


Figure 3.8A. Locus showing TEF-seq data and ChIP-chip data for elongation factors. A different arbitrary scale is used for each TEF-seq track (Table 3.1), which is consistent across all loci, allowing relative changes in factor binding to Pol II to be visualised. Data has not been normalised for differences in library size or the efficiency of immunoprecipitation. ChIP-chip data is from Mayer et al., 2010 and 2012.

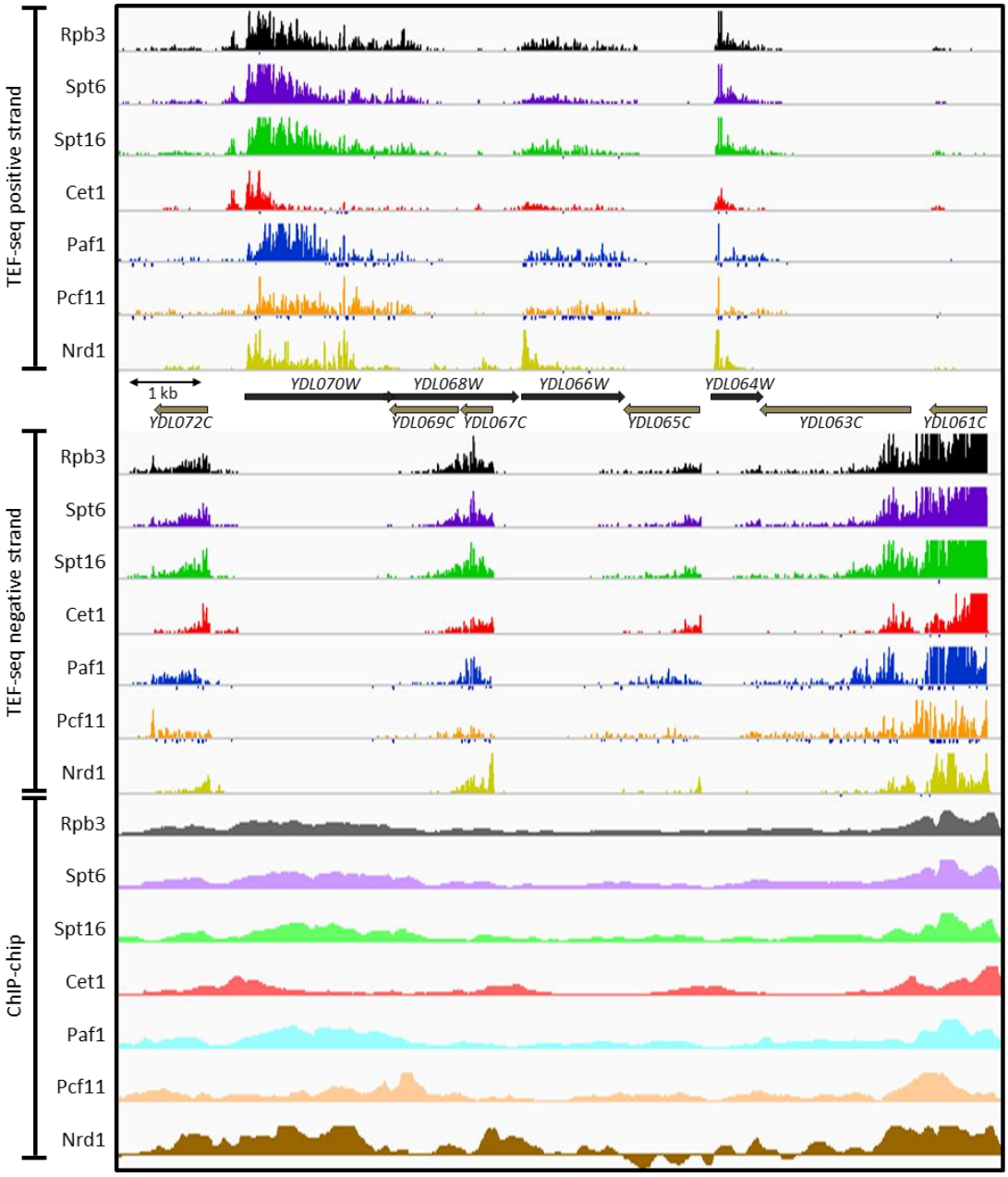


Figure 3.8B. Locus showing TEF-seq data and ChIP-chip data for elongation factors.

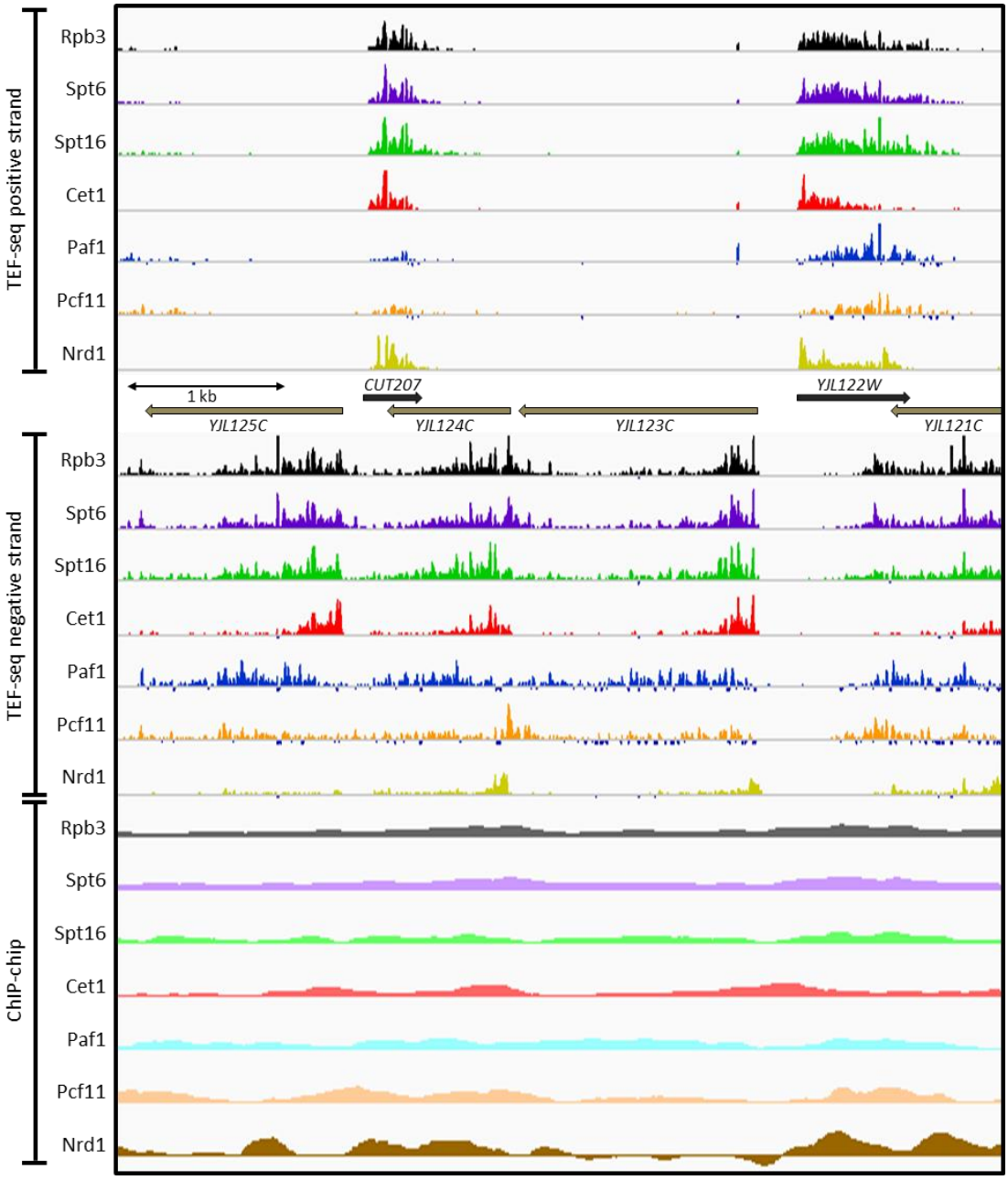


Figure 3.8C. Locus showing TEF-seq data and ChIP-chip data for elongation factors.

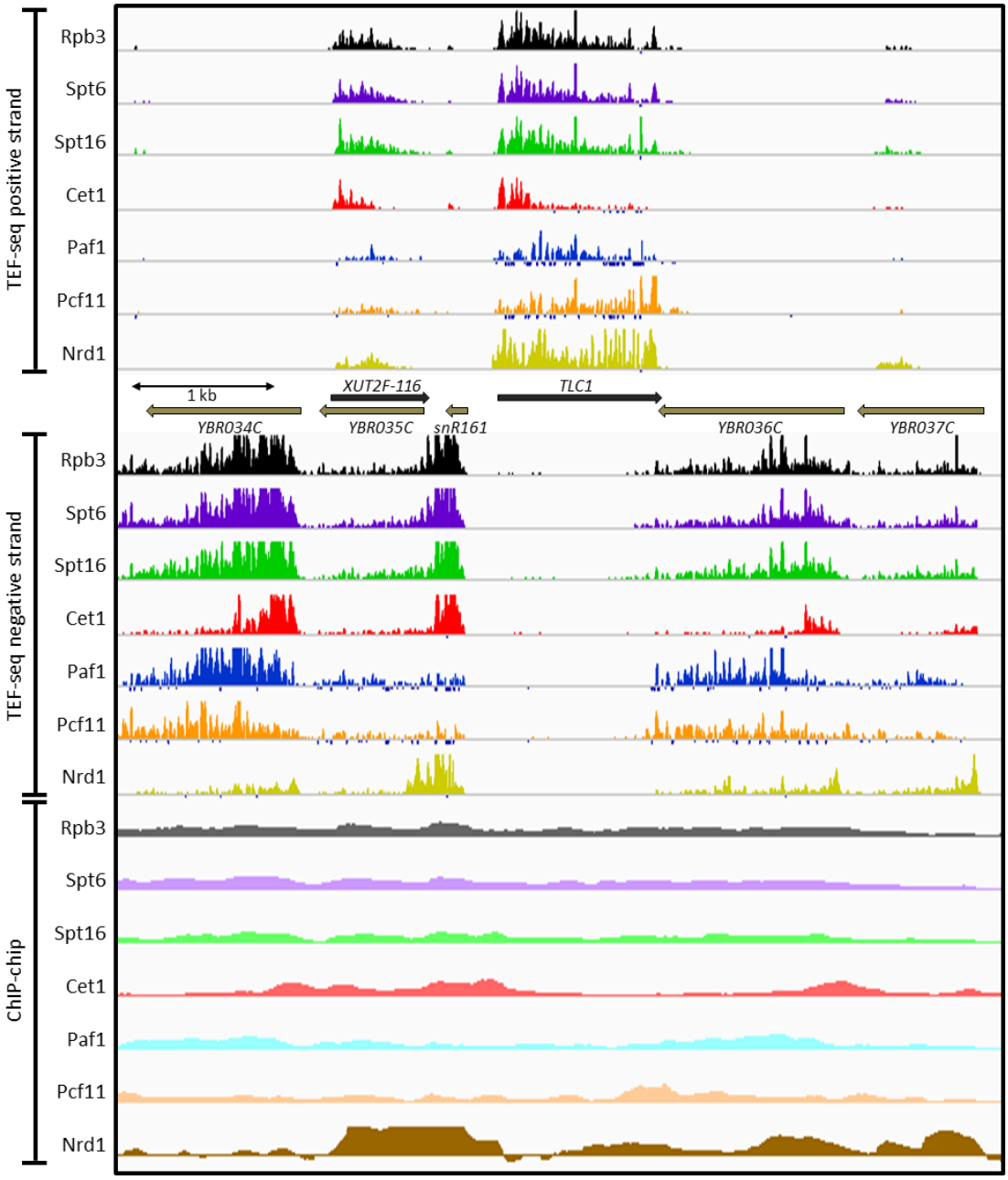


Figure 3.8D. Locus showing TEF-seq data and ChIP-chip data for elongation factors.

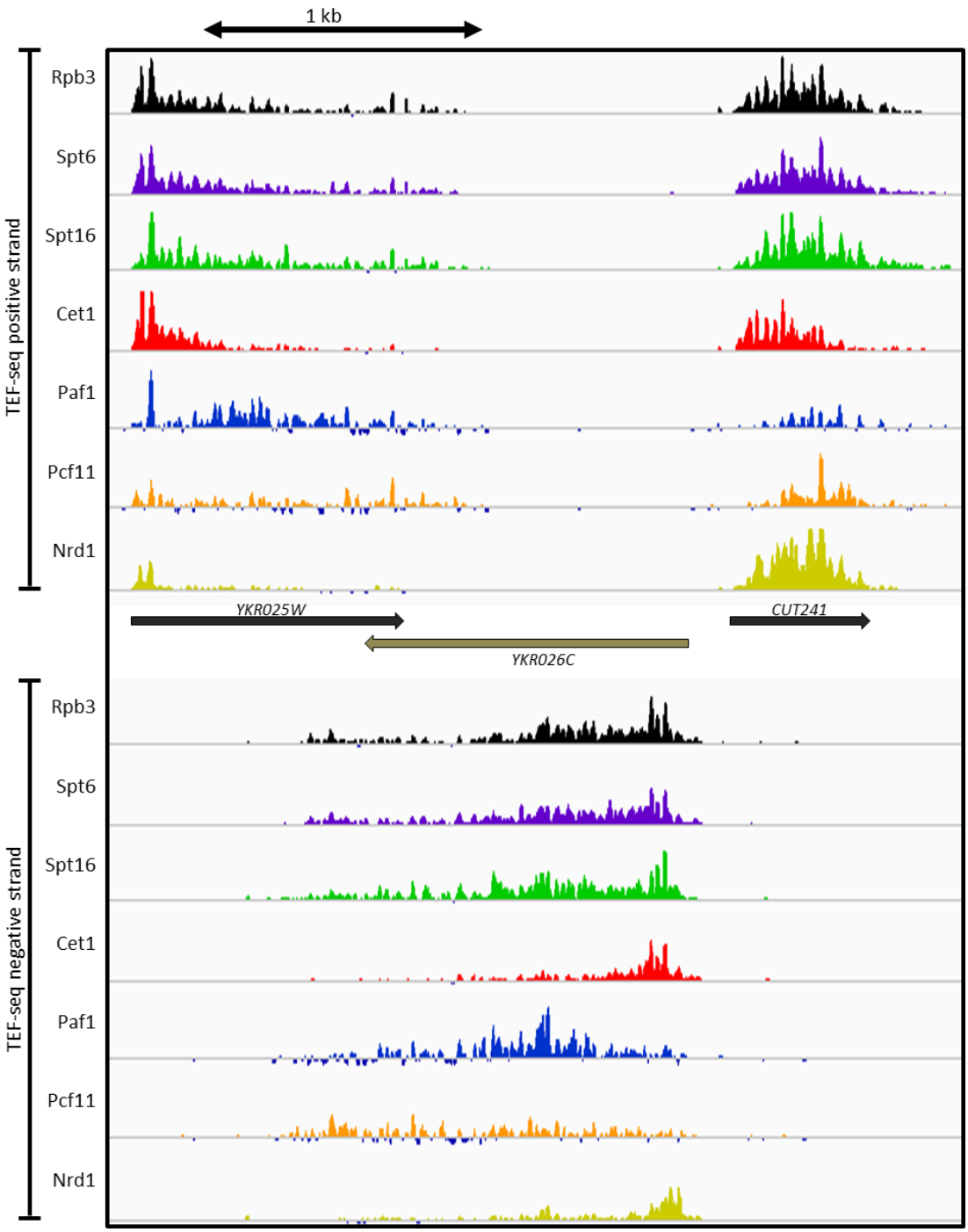


Figure 3.8E. Locus showing TEF-seq data for elongation factors.

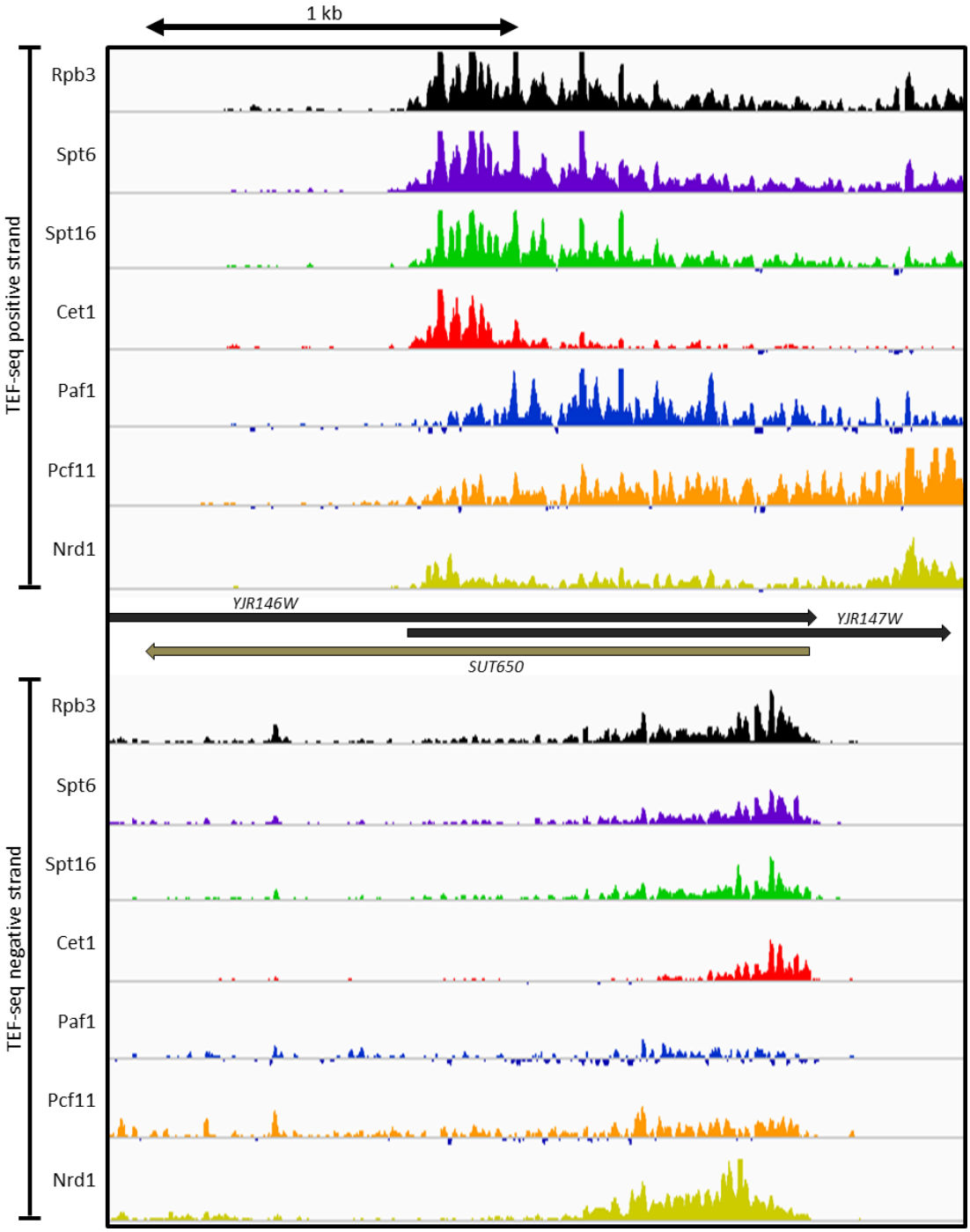


Figure 3.8F. Locus showing TEF-seq data for elongation factors.

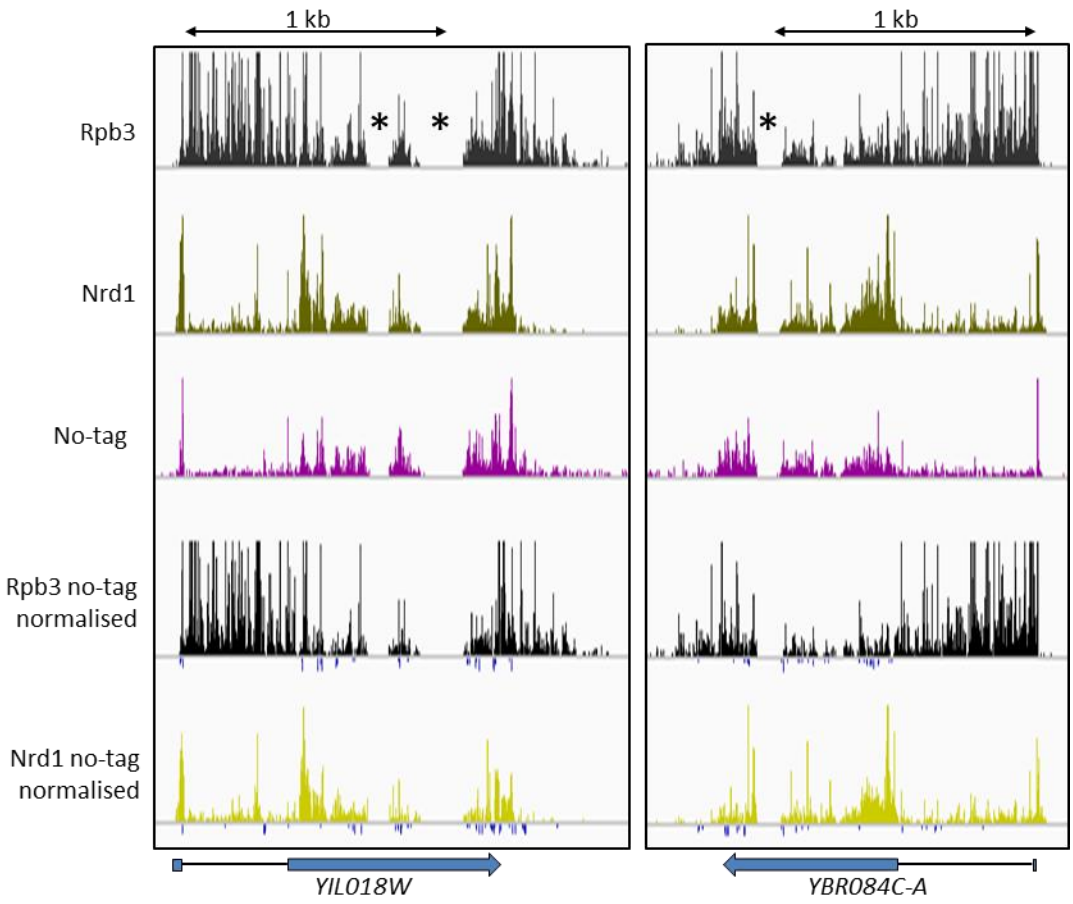


Figure 3.9. The Nrd1 data show a sharp peak in signal over the first exon and start of the second exon for PCG TUs with early introns. Nrd1 and Rpb3 TEF-seq data for two representative PCG TU with an early intron (*YIL018W* and *YBR084C-A*). Gaps in the data within the second exon (marked with asterisks) are the result of discarding non-uniquely aligned reads, which possibly map to these non-unique sites. Pre-normalisation tracks are scaled relative to the No-tag track according to their *SCR1* scaling factor. No-tag normalised tracks are not smoothed and are on the same scale as their pre-normalisation tracks.

3.2.6.2 Metagene profiles

PCG TU averaged occupancy profiles for each factor were plotted for TEF-seq data (TS) and published ChIP-chip data (CC) (see Methods) (Figure 3.10). The Rpb3 occupancy profile is similar for both techniques, showing the highest occupancy just downstream of the TSS and a gradual decline in occupancy towards the PAS. This may be due to a slower transcription speed, increased pausing and/or early transcription termination at the 5' end of TUs. CC shows higher occupancy upstream of the TSS probably because this

technique can detect Pol II that has transcribed less than 18 nt of RNA, including PIC bound Pol II. TS shows a small increase in Rpb3 occupancy just before the PAS, suggesting Pol II pauses just before it reaches this site. This increase at the end of the TU is greater and broader in CC. This possibly results from signal from nearby TUs, due to the lower resolution of ChIP, or cross-linking of 5' TU bound Rpb3 to the 3' end, due to gene looping.

The Cet1 occupancy profile is similar for both techniques with a high peak in occupancy at the 5' end of TUs, consistent with its role in co-transcriptionally capping the 5' end of transcripts shortly after they emerge from the RNA exit channel of Pol II. This is also consistent with its interaction with Ser5P and Ser7P Rpb1 (Figure 3.6), which both show high levels in this region (Mayer et al., 2010). As with Rpb3, the Cet1 CC profile shows a signal upstream of the TSS and a higher signal at the 3' end of the TU. The reasons for these differences are likely to be similar to those for Rpb3.

The Nrd1 profile was plotted excluding the first 300 bp from the TSS so as to avoid the spike in signal in this region. TS and CC profiles closely match, although TS shows a clearly defined peak in binding immediately before the PAS, whereas in the CC profile this is much broader. As mentioned, Nrd1 is often stated as binding strongest to the 5' end of TUs. These profiles show that this is indeed the case; however, the profiles more closely resemble those of Rpb3 than Cet1 suggesting Nrd1 is bound to Pol II constantly across the whole TU. In support of this Nrd1 ChIP profiles normalised to Rpb3 ChIP profiles have been recently shown to give a largely flat profile across TUs (Tudek et al., 2014).

Spt6 and Spt16 TS profiles are similar to the Rpb3 TS profile with some differences around the PAS, which will be discussed later. Spt6 and Spt16 CC occupancy profiles are higher in the middle and 3' end of the TU. ChIP detects all interactions with chromatin rather than just interactions with Pol II. This higher signal may therefore reflect an

interaction with another chromatin component such as nucleosomes with which they have been shown to interact (Duina, 2011). Paf1 has similar CC and TS profiles to those of Spt6 and Spt16 but starts to bind the TEC and shows maximum occupancy downstream of these factors. This conflicts with the previous interpretation of the ChIP-chip data (Mayer et al., 2010), which suggested Paf1 starts to bind with Spt16, and is caused by differences in how the profiles are plotted relative to each other. The Paf1C is required for Set1 recruitment to Pol II (Krogan et al., 2003b). Discrepancies lie in the position of Set1 recruitment with some suggesting its recruitment is restricted to the 5' end of a TU (Ng et al., 2003b). More recent reports suggest a broader recruitment profile across TUs with a peak in occupancy ~+500 bp (Kuang et al., 2014). If Paf1 is responsible, the TS and CC profiles of Paf1 would support the +500 bp peak in Set1 occupancy. This raises the question as to how H3K4me3 deposited by Set1 is restricted to the 5' end of TUs. Paf1 is required for Rad6/Bre1 recruitment, which deposits H2Bub1, a mark which is required for H3K4me3. Both H3K4me3 and H2Bub1 show transient increases at the 5' end of TUs upon transcription induction before decreasing to lower levels after activation (Kao et al., 2004; Ng et al., 2003b). One possibility is that Paf1 and Set1 show different occupancies on TUs during induction.

Pcf11 shows higher occupancy in the middle of the TU in the TS profile compared to the CC profile; however, both profiles show highest occupancy at the 3' end of the TU. This peak in binding can be clearly seen to occur downstream of the PAS for the TS profile. Due to its lower resolution, the CC profile shows a less precisely defined peak.

In addition to the factors mentioned, TEF-seq was also applied to Set2 and to enzymes, which phosphorylate or dephosphorylate the CTD of Rpb1 (see Chapter 1 for details of these factors). For strains in which Bur1, Ctk1, Kin28, Rtr1 and Fcp1 were FLAG-tagged, following no-tag normalisation there was no obvious signal above the noise showing these factors do not efficiently co-immunoprecipitate Pol II (see Appendix). For Set2 and

Ssu72 there was a high enough signal above noise and the gene averaged profiles for these factors were plotted. The Set2 occupancy profile closely resembles Paf1. The Ssu72 profile has similarities to the Pcf11 profile but with a peak in occupancy just before the PAS rather than after it.

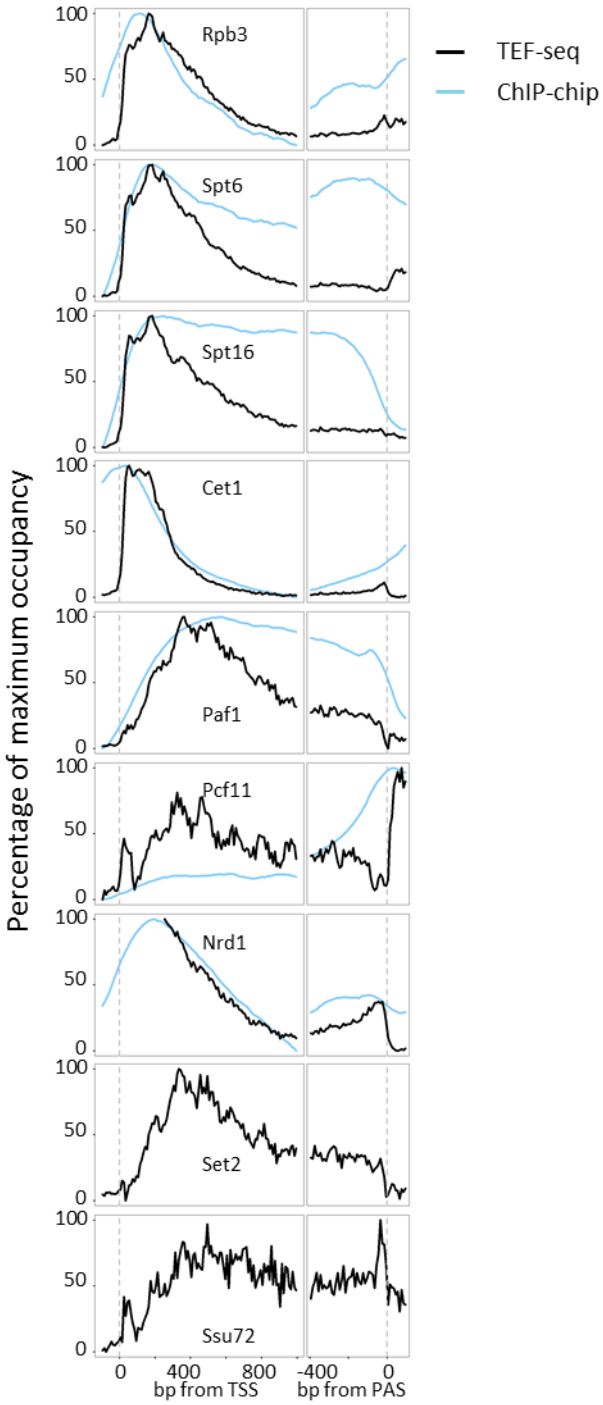


Figure 3.10. PCG TU averaged occupancy profiles for TEF-seq and ChIP-chip data. Profiles are plotted as a percentage of maximum occupancy. TEF-seq, Black; ChIP-chip, Blue. These profiles are not normalised to Pol II levels.

3.2.7 Pol II-normalised elongation factor average profiles reveal periodic factor binding and an increased number of transitions in factor binding

3.2.7.1 Periodic factor binding and transitions at the 5' end of TUs

Normalising elongation factor TEF-seq data to Rpb3 TEF-seq data enables visualisation of factor transitions on Pol II. PCG TU averaged elongation factor/Rpb3 (EF/Rpb3) ratio profiles were plotted for each elongation factor for specific regions of the TU. To do this, each EF/Rpb3 ratio is multiplied by its own arbitrary scaling factor and has not been normalised for differences in sequencing depth. This enables clear comparison of factor binding transitions but not absolute binding levels to the TEC. These profiles were first plotted from the TSS to +1400 bp. The 200 bp upstream of the PAS were excluded to avoid transitions occurring at the 3' end of TUs influencing these profiles (see Methods) (Figure 3.11A-C). Figure 3.11B shows these profiles for Paf1, Spt6 and Spt16. In the first ~800 bp after the TSS these factors have profiles with an oscillating pattern, although Paf1 generally increases over this region. This means that the binding affinity of these factors for the TEC changes periodically as Pol II transcribes. This oscillating pattern can also be seen in the non-Pol II-normalised average profiles (Figure 3.10). The period of each oscillation is ~180 bp, closely matching the distance between nucleosomes. To analyse this relationship, the PCG TU averaged EF/Rpb3 ratio profiles for the 100 bp upstream and downstream of the central bp (centre) of each of the first 4 nucleosomes after the TSS +50 bp (nucleosomes +2 to +5) were plotted (see Methods; Figure 3.11D). As the levels of the factors change over this region a scale has been chosen to focus on the dynamics of Spt6 and Spt16 for nucleosomes +2, +3 and +4 and Paf1 and Spt16 for nucleosome +5. This shows that Spt16 and Paf1 bind strongest to the TEC as Pol II transcribes through the centre of each nucleosome, whilst Spt6 binds strongest as Pol II transcribes 50 bp downstream of this. A one-tailed Student's t-test was used to show the increase in binding from the nucleosome centre (bin from centre to +10 bp) to 50 bp downstream of the centre (bin from +50 bp to +60 bp) was significant for Spt6 for nucleosomes +2 to +4 and the decrease in binding over the same region was significant

for Spt16 for nucleosomes +2 to +5 and for Paf1 for nucleosomes +4 and +5 (see Table 3.2 for p values). These changes in occupancy are in addition to the changes in occupancy seen for Pol II around nucleosomes (Brogaard et al., 2012; Churchman and Weissman, 2011) because the oscillating pattern is seen after normalisation to Rpb3.

Nucleosome	**+2**	**+3**	**+4**	**+5**
Spt6	8.7×10^{-19}	2.3×10^{-14}	1.0×10^{-12}	0.89
Spt16	2.3×10^{-9}	1.6×10^{-16}	1.1×10^{-4}	2.9×10^{-7}
Paf1	0.997	6.5×10^{-2}	8.9×10^{-4}	3.6×10^{-5}

Table 3.2. P values from a one-tailed Student's t-test testing, for each nucleosome, whether there is a significant mean increase in binding from the nucleosome centre to +50 bp for Spt6 and a mean decrease in binding for Spt16 and Paf1 across the same region.

Spt6 and Spt16 have been proposed to interact with histone proteins enabling partial disassembly and reassembly of nucleosomes as Pol II transcribes through them (Duina, 2011). Both proteins associate with all core histones but show preferential affinities for H3 and H2A-H2B dimers, respectively (Bortvin and Winston, 1996; Hondele et al., 2013). This difference in binding preference is the likely cause of the offset pattern of binding of these factors. In the nucleosome structure, H3 predominantly contacts the DNA in a ~50 bp region centred on the nucleosome centre, whilst H2A-H2B dimers contact the DNA at regions downstream and upstream of this (Luger et al., 1997). One possible explanation is that, as Pol II transcribes into the region bound by H2A-H2B dimers, they compete with Pol II for Spt16 binding, leading to decreased occupancy of Spt16 on Pol II. Then, as Pol II progresses into the central region of the nucleosome bound by the H3-H4 tetramer, this competition is reduced, leading to increased Spt16 occupancy on Pol II. Transcription into the region downstream of the centre would again lead to increased competition and reduced occupancy. Applying such a model to Spt6 would predict a peak in Spt6 occupancy on Pol II either side of the centre of the nucleosome rather than that seen on just the downstream side of the nucleosome. This

suggests other mechanisms are involved. As Pol II transcribes through nucleosomes it has previously been shown to pause before the centre of each nucleosome (Brogaard et al., 2012; Churchman and Weissman, 2011). This probably results from DNA being more tightly wrapped around the histone as Pol II transcribes towards, rather than away from the nucleosome centre. The asymmetrical pattern of binding by Spt6 could therefore be due to the protein showing decreased binding to Pol II as it is required to bind nucleosomes and destabilise the tightly wrapped structure before the nucleosome centre but to a lesser extent after it. In addition, the pausing of Pol II may result in changes occurring within the TEC and possible changes in the modification of the Rpb1 CTD, which may affect the binding affinities of factors. Paf1 has been shown to physically and genetically interact with Spt16 which may be why the positioning of their oscillating patterns matches (Squazzo et al., 2002).

The EF/Rpb3 ratio profiles for this region for Pcf11, Paf1, Set2 and Ssu72 (Figure 3.11C) and Cet1 and Nrd1 (Figure 3.11A) also show some oscillating patterns to a lesser extent. This becomes clearer upon smoothing the profiles using a running mean average (Figure 3.16A). These figures reveal several transitions in factor binding (Table 3.3 and Figure 3.13). Firstly there is a transition (T1) at ~+100 bp at which Pcf11, Paf1, Ssu72 and Set2 start to increase. Pcf11 shows a peak in binding at the TSS decreasing up to this transition, whereas the other factors show very little change in binding before this point. In addition, Cet1 occupancy decreases at T1 from its high level at the start of the TU. It continues to decrease until ~+450 bp (Transition 3 (T3)). Cet1 is recruited in complex with Ceg1 by Kin28-mediated CTD Ser5P, which occurs concomitantly with promoter escape (Akhtar et al., 2009; Cho et al., 1997). ChIP-chip data has shown that Cet1 and Ceg1 are the first RNA capping factors to be recruited, while Abd1, the cap methyl-transferase, is recruited a further 90 bp downstream. Once the cap is methylated, the CBC can bind, which assists in the recruitment of Bur1 and Ctk1 (Lidschreiber et al., 2013). Bur1 phosphorylates the Spt5 CTR at Ser1 and CTD repeats at Ser2, particularly

targeting those already phosphorylated at Ser5 (Qiu et al., 2009). Both modifications are bound by Paf1, which also assists in Ctk1 recruitment (Qiu et al., 2012). Set2, which also binds Ser2, Ser5 doubly phosphorylated marks (Li et al., 2005), increases with Paf1. Bur1- and Ctk1-mediated CTD Ser2P recruits Pcf11. During this period, Rtr1 removes most of the Ser5P from CTD repeats (Mosley et al., 2009) leading to the loss of Cet1. CTR phosphorylation also disrupts Cet1 binding (Lidschreiber et al., 2013). The presence of the nascent RNA and the fact that Ser5P is not lost from all repeats (Bataille et al., 2012) may prevent the complete loss of Cet1 from the TEC. This cascade of events could therefore explain transitions T1 and T3.

At ~+650 bp (Transition 4 (T4)) Paf1 and Spt16 reach maximum occupancy. This supports a possible link between these factors, already shown to have matching oscillating patterns, in their mechanism of recruitment to the TEC at some regions of the TU. There may be an indirect link between these factors as the H2Bub1 modification, which is dependent upon Paf1, enhances the interaction between Spt16 and H2B (Fleming et al., 2008). As Ser5P and Ser7P decrease while Ser2P increases from the TSS to +650 bp, this increase up to +650 bp for Spt16 that is not seen with Spt6 may explain the higher binding of Spt16 to Ser2P than Ser5P and Ser7P, relative to total Rpb1 binding, compared to Spt6 (Figure 3.6). Both Spt6 and Spt16 decrease after reaching peaks at ~+250 bp (Transition 2 (T2)) and +650 bp, respectively. These decreases may be due to the transfer of these factors from the TEC to nucleosomes, leading to the higher ChIP signal than TEF-seq signal at the 3' end of TUs seen in Figure 3.10. Pcf11, Ssu72 and Set2 occupancies continue to increase past +650 bp, with Ssu72 and Set2 no longer increasing past ~+1000 bp and Pcf11 no longer increasing past ~+1200 bp (Transition 5 (T5)). The Ser2, Ser5 double phosphorylation mark, bound by Paf1C, is thought to be limited to a few repeats of the CTD as Rtr1 dephosphorylates Ser5P in most repeats. This may limit any further increase in Paf1 binding after +650 bp. The Ser2 single phosphorylation mark may be able to occur at more repeats enabling further

increases in Pcf11 binding past this +650 bp position. The CID of Set2 has been shown to bind more strongly to Ser2, Ser5 doubly phosphorylated repeats than only Ser2P or Ser5P repeats (Li et al., 2005). At +650 bp the rate at which Set2 occupancy is increasing decreases. This may reflect preferential binding to the Ser2, Ser5 double phosphorylation mark like Paf1 up to +650 bp. Binding to other marks with lower affinity may enable the further increases in occupancy observed past this point.

The Nrd1/Rpb3 ratio profile across this region shows a sharp 5' peak in signal, which corresponds to the non-Pol II-specific signal that has previously been discussed. After this, the profile is flat, suggesting the binding of Nrd1 to Pol II does not change. This is consistent with the matching non-Pol II-normalised average Nrd1 and Rpb3 occupancy profiles in this region (Figure 3.10).

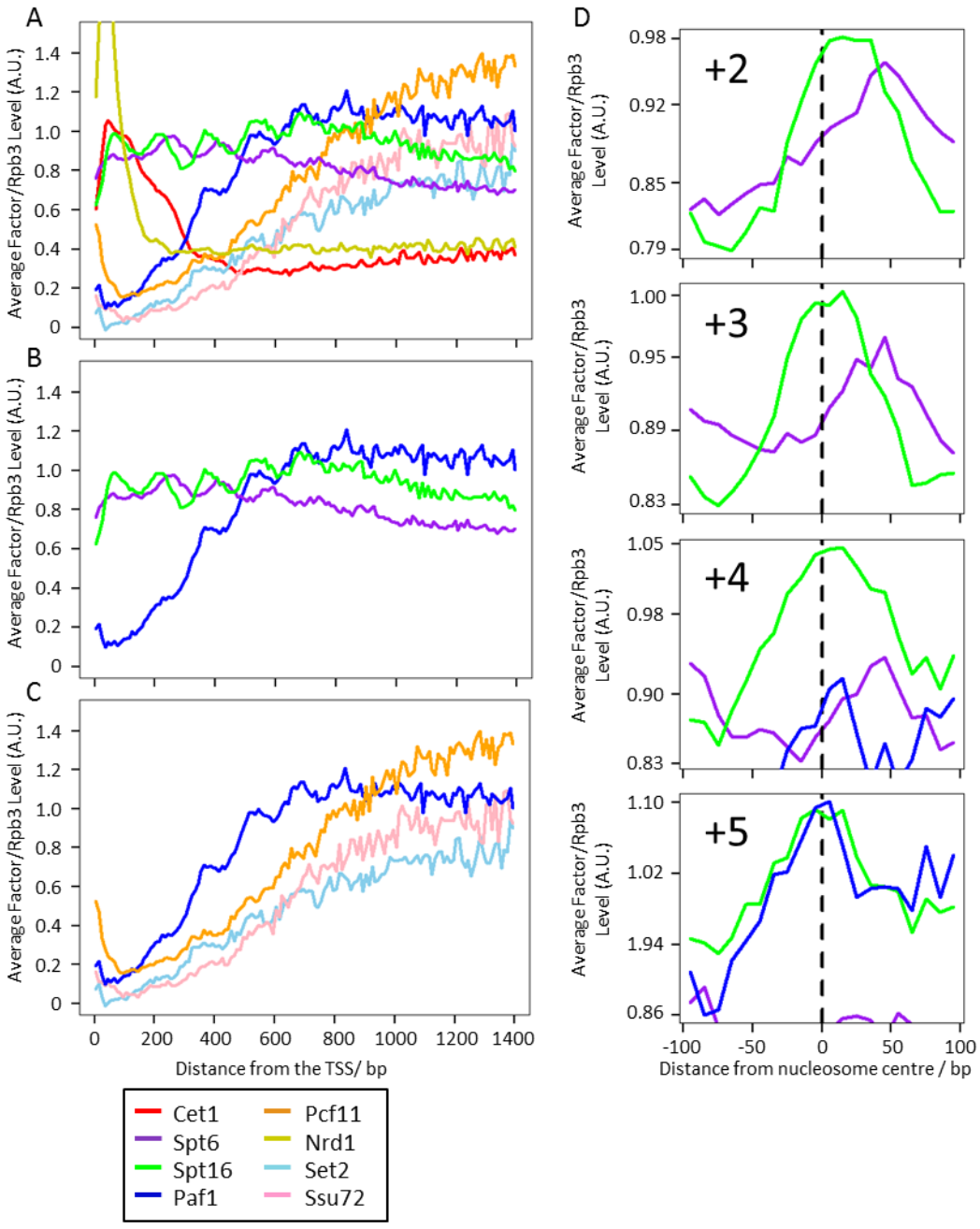


Figure 3.11. Transitions in factor binding at the 5' end of TUs. A, B and C) PCG TU averaged TEF-seq EF/Rpb3 ratios from the TSS to +1400 bp. Each EF/Rpb3 ratio is multiplied by its own arbitrary scaling factor and has not been normalised for differences in sequencing depth. D) PCG TU averaged TEF-seq factor/Rpb3 ratios for the first 4 nucleosomes after the TSS +50 bp (nucleosomes +2 to +5).

3.2.7.2 Transitions at the 3' end of TUs

PCG TUs were aligned by the 3' end of their open reading frame (ORF) and the EF/Rpb3 ratio profiles were plotted for each elongation factor from this position -100 bp to +300 bp (Figure 3.12A). These have been split into two separate figures to aid visualisation. TUs shorter than 601 bp (390 TUs) were excluded to avoid transitions occurring at the 5' end of TUs influencing these profiles. Signal downstream of the PAS was also excluded. These figures show that, while Nrd1 and Cet1 do not show a change in binding at this position, the other factors all show decreased binding to the TEC as it passes the 3' end of ORF (Transition 6 (T6)). This means this position, which is important for translation, provides a signal, which is recognised during transcription. Recognition of this position may mean that the stop codon of a transcript is marked during transcription, possibly assisting later translation. This transition may also prime the TEC for later termination. As the distance between the stop codon and termination is not uniform, other signals are clearly required for determining the position of termination.

PCG TUs were then aligned by their PAS and the EF/Rpb3 ratio profiles were plotted for each elongation factor from this position -400 bp to +100 bp (Figure 3.12B). These have again been split into two separate figures to aid visualisation. TUs shorter than 601 bp (390 TUs) were again excluded. ~150 bp upstream of the PAS, Set2, Pcf11, Paf1 and Spt6 show decreased binding to the TEC (Transition 7 (T7)). Spt16 shows a more gradual decrease in binding. This may overlap with transition T6 occurring at the end of the coding sequence.

Nrd1 increases at this transition reaching a peak at ~80 bp upstream of the PAS (Transition 8 (T8)) before decreasing back to its original level at the PAS. Across this same region, Ssu72 shows little change or even a slight increase in binding before also decreasing ~80 bp before the PAS. The noise of this Ssu72 profile makes an increase hard to distinguish although, upon smoothing the profile using a running mean, a slight

increase is more apparent (Figure 3.21A). Both Ssu72 and Nrd1 bind preferentially to the CTD in which Pro6 is in the *cis*-conformation (Kubicek et al., 2012; Werner-Allen et al., 2011). In addition, the proline isomerase Ess1 is required for termination by the NRD complex (Singh et al., 2009) and stimulates Ssu72 activity (Krishnamurthy et al., 2009). It may be that Ess1 acts 150 bp upstream of the PAS to increase the number of CTD repeats with Pro6 in the *cis*-conformation. The CTD would then recruit Ssu72 and Nrd1. Ssu72 may be flatter across this region rather than showing a noticeable increase due to an averaging effect from the decrease seen at the end of the ORF. This isomerisation and the dephosphorylation of remaining Ser5P and Ser7P by Ssu72 may disrupt the binding of other factors. Supporting this, Pcf11 binds preferentially to CTD repeats with prolines in the *trans*-conformation (Noble et al., 2005). Ssu72 is also recruited as part of the APT complex containing the Tyr1P dephosphorylating enzyme Glc7 (Schreieck et al., 2014). ChIP profiles for Tyr1P show this mark decreases ~180 bp before the PAS, which is likely to cause a decrease in binding of Spt6 which recognises this mark (Mayer et al., 2012). There is also an increase in TEF-seq Rpb3 signal just before the PAS (Figure 3.10). This may relate to a slowing in the speed of Pol II, possibly due to the decrease in Spt6 binding, as Spt6 has been linked to increased elongation rates (Ardehali et al., 2009).

At the PAS (Transition 9 (T9)), Pcf11 and Spt6 show increased binding while all other factors show decreased binding. This could be due to the increased presence of Ser2P around the PAS detected by the 3E10 antibody (Bataille et al., 2012). Loss of other factors may also reduce the competition for binding sites on the CTD of Rpb1. Pcf11 is also recruited by factors of the CF/CPF complexes, which recognise RNA sequences in the region of the PAS. The cleavage and release of the nascently transcribed RNA at the PAS may decrease the affinity of Cet1 and Nrd1, which both interact with RNA, for the TEC. ~50 bp downstream of the PAS (Transition 10 (T10)), Pcf11 decreases, possibly due to the removal of the remaining Ser2P by Fcp1 (Bataille et al., 2012). Spt6 remains

bound to Pol II at least until it disengages from DNA. See Table 3.3 and Figure 3.13 for a summary of all transitions and the possible TEC modifications regulating these transitions.

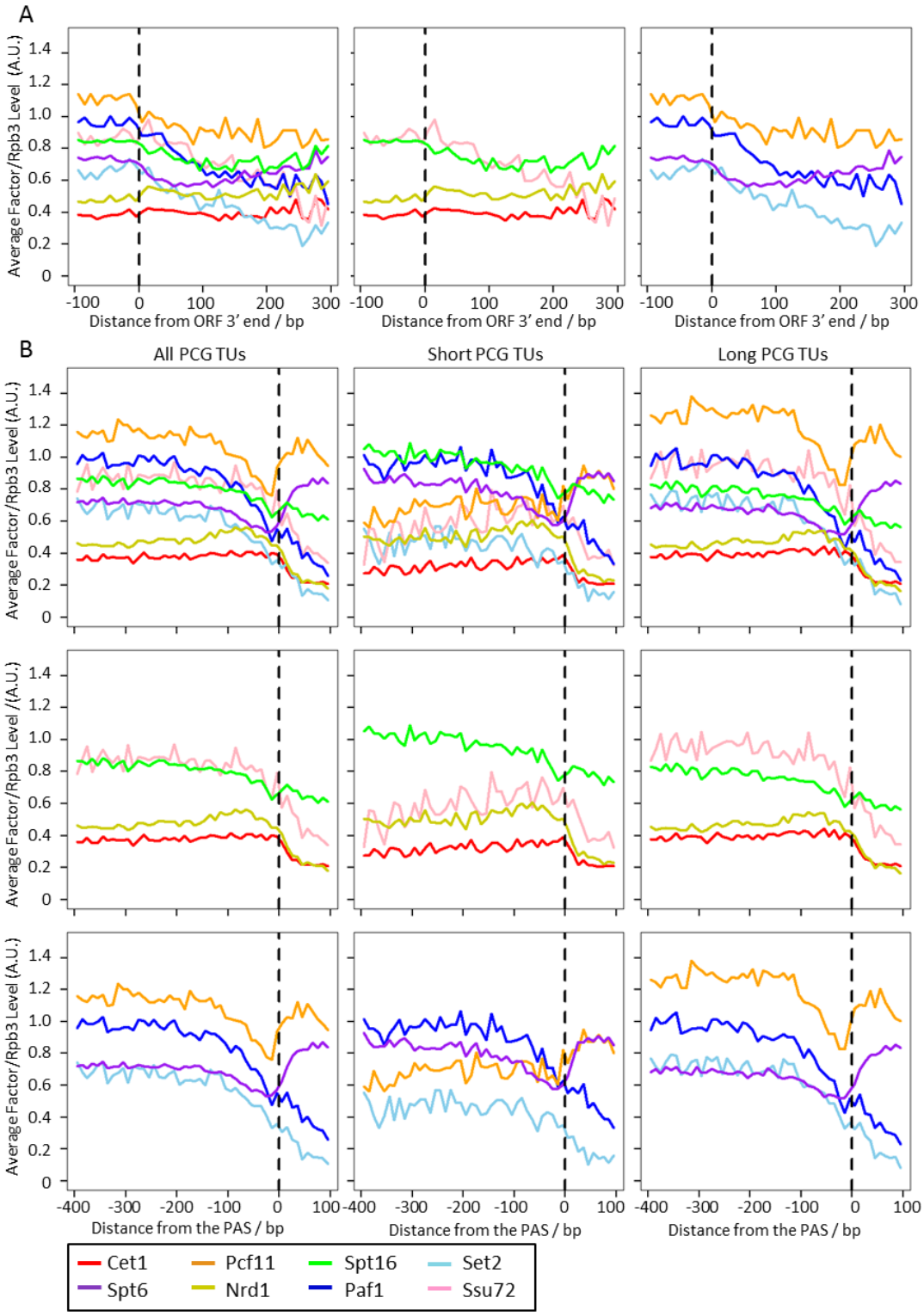


Figure 3.12. Transitions in factor binding at the 3' end of TUs. A) PCG TU averaged TEF-seq EF/Rpb3 ratios from the3' end of the ORF -100 bp to +300 bp excluding regions downstream of the PAS. Profiles for all factors (left) are split between two panels (middle and right) to aid visualisation. B) PCG TU averaged TEF-seq factor/Rpb3 ratios from the polyA site -400 bp to +100bp. These are for all PCG TUs and after sorting TUs by length into long and short TU groups. Profiles for all factors (top) are split between two panels (middle and bottom) to aid visualisation.

Transition	Position/ bp	Description	Possible TEC modification
T1	TSS+100	Cet1 decreases; Paf1, Pcf11, Set2 and Ssu72 increase	Loss of Ser5P; Gain of Ser2P; Spt5 CTR Ser1 phosphorylation
T2	TSS +250	Spt6 decreases	Unknown
T3	TSS +450	Cet1 stops decreasing	Limit to the loss of Ser5P; Binding to nascent RNA
T4	TSS +650	Paf1 stops increasing; Spt16 decreases; Set2 increases at a slower rate	Limit to the number of Ser2P, Ser5P doubly phosphorylated repeats
T5	TSS +1000 to +1200	Pcf11, Set2 and Ssu72 stop increasing	Limit to the number of Ser2P singly phosphorylated repeats
T6	ORF end	Paf1, Pcf11, Spt6, Spt16, Ssu72 and Set2 decrease	Unknown
T7	pA -150	Paf1, Pcf11, Spt6, Spt16 and Set2 decrease; Ssu72 and Nrd1 increase (may overlap with T6)	Pro6 isomerisation to cis form; Loss of Ser5P, Ser7P and Tyr1P
T8	pA -80	Nrd1 and Ssu72 decrease	Loss of Pro6 isomerisation
T9	pA site	Cet1 decreases; Nrd1, Paf1, Set2, Spt16, Ssu72 continue to decrease; Pcf11 and Spt6 increase	Loss of nascent RNA; Continued presence of Ser2P
T10	pA +50	Pcf11 decreases	Loss of Ser2P

Table 3.3. Summary of the transitions in elongation factor binding to Pol II across PCG TUs

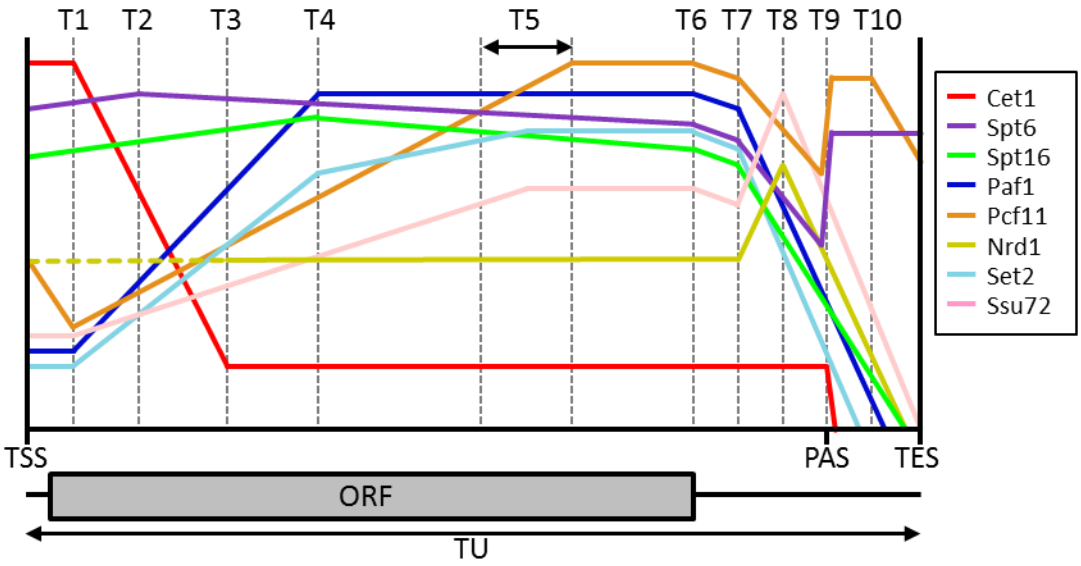


Figure 3.13.Summary of the transitions in elongation factor binding to Pol II across PCG TUs. The Nrd1 profile is shown as a dashed line before transition 3 (T3) showing the predicted level on Pol II. The true signal of Nrd1 associated with Pol II in this region is masked by a signal from Nrd1 binding to mature degraded RNA.

3.2.8 TU length does not affect the position of factor binding transitions but affects the level of factors at 3' ends

TUs were split into two groups of short (less than 1100 bp; 1802 TUs; median length 795 bp) and long (greater than or equal to 1100 bp; 3777 TUs; median length 1849 bp) PCGs. The elongation factor/Rpb3 ratios for each group were plotted from the TSS to +800 bp (Figure 3.14) and, after aligning by the PAS, from the PAS -400 bp to +100 bp (Figure 3.12B and 3.14). These have again been split into two separate figures to aid visualisation (Figure 3.12B). Figure 3.14 shows the position of factor transitions T1 to T5 do not change relative to the TSS with changes in gene length. This means these transitions are dependent on the transcription of x number of bases rather than x percent of a TU. It also means that for TUs shorter than ~1200 bp, Pcf11 occupancy does not reach the point at which it plateaus. This leads to a lower Pcf11 occupancy at the 3' end of short TUs even after the PAS (Figure 3.12B). Only short PCG TUs have been shown to have termination defects upon Nrd1 disruption (Tudek et al., 2014). Nrd1 possibly acts in a transcription termination fail-safe pathway (Rondón et al., 2009),

which may be more important at these short TUs due to their lower Pcf11 levels. Set2 and Ssu72 also do not reach the point at which they plateau on short TUs, also leading to lower levels at the 3' end of these TUs. The decrease in Spt6 and Spt16 at transitions T2 and T4, respectively, means that for long TUs these two factors are lower at the 3' end of the TU.

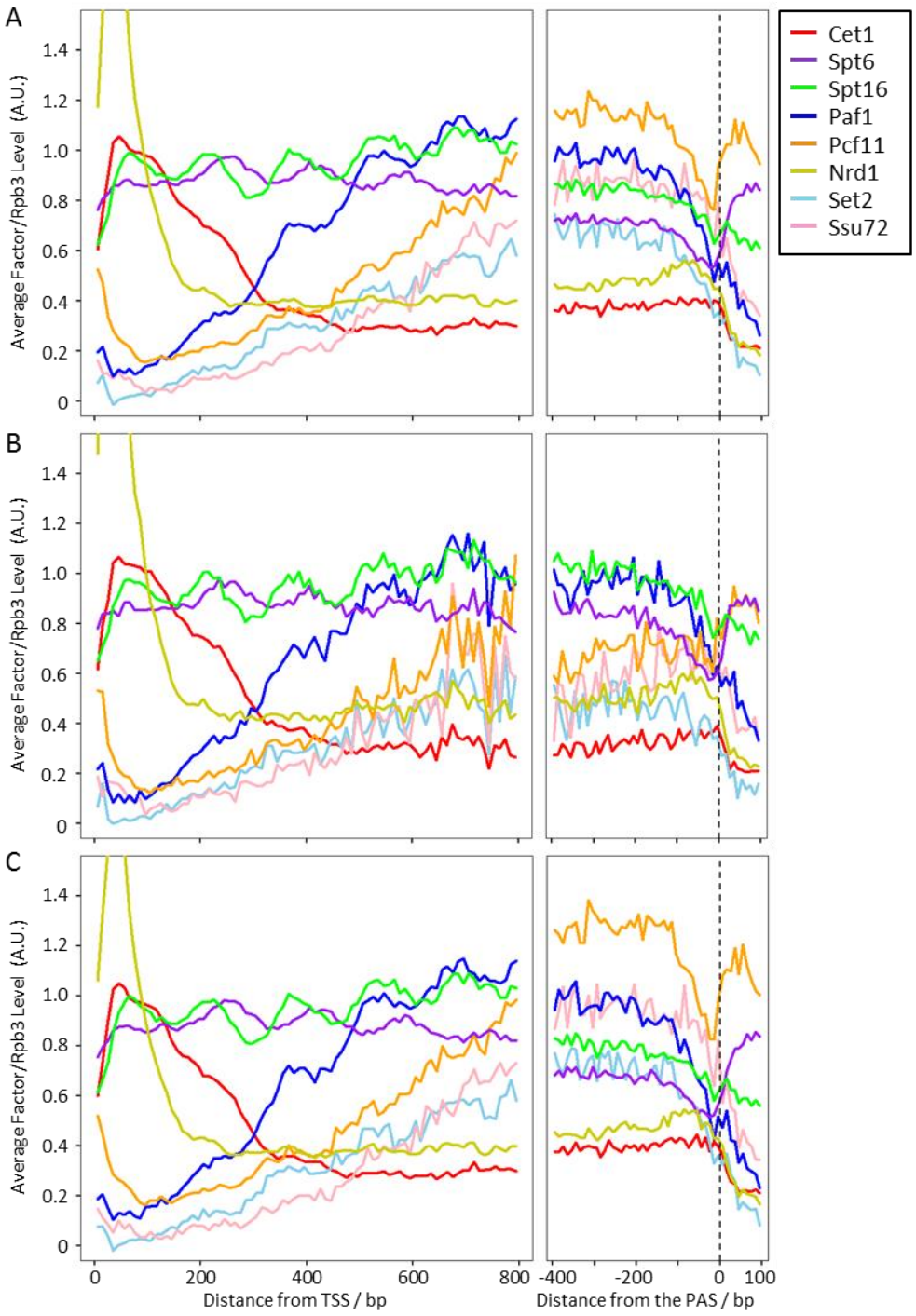


Figure 3.14. The effect of TU length on the levels of factors and factor binding transitions. PCG TU averaged TEF-seq EF/Rpb3 ratios from the TSS to +1400 bp and from the polyA site -400 bp to +100 bp for all PCG TUs (A) and after sorting by length for TUs shorter than 1100 bp (B) and longer or equal to 1100 bp (C).

3.2.9 Splicing intermediates co-immunoprecipiate with Pol II

TEF-seq data show no general differences between the profiles of factors over intron-containing and intronless PCG TUs (compare Figure 3.15A and 3.16A with Figures 3.8A-F). However, a previously reported feature of NET-seq data is that it shows a large peak in reads aligning to the terminal nucleotide of the first exon and first intron of some intron-containing TUs (Churchman and Weissman, 2011). This has been proposed to result from splicing intermediates (the first exon and excised intron) associating and co-immunoprecipitating with the transcribing Pol II. Their 3' hydroxyl termini allow them to be incorporated into libraries and sequenced from their 3' ends (Figure 3.17A). This has been cited as evidence of co-transcriptional splicing (Churchman and Weissman, 2011). Reads from a non-tagged strain, however, also have peaks at these terminal nucleotide positions (Figure 3.4B), calling into question this conclusion and highlighting the importance of carrying out the no-tag control. Nevertheless, after no-tag normalisation, peaks at these positions are still present, showing that these intermediates are specifically enriched above the no-tag signal, by association with Pol II (Figure 3.15B and 3.16B).

The Rpb3 TEF-seq signals at the terminal nucleotides of the first exon and first intron of 263 intron-containing TUs were counted. A threshold was then set by taking the second maximum signal from the 10 nt upstream and 10 nt downstream of each terminal nucleotide position. If this was below 10, a threshold of 10 was set. Each terminal nucleotide position was then assigned as resulting from a splicing intermediate if it was two times greater than this threshold. Positions that were also assigned as splicing intermediates using Isw1 or Yra1 TEF-seq data were excluded. This resulted in a total of 148 TUs showing a splicing intermediate resulting from the first exon and 67 resulting from the excised intron lariat (Figure 3.17C). This suggests that, for the majority of intron-containing transcripts, splicing has at least started to occur while the transcript is still associated with Pol II. This may be while the transcript is being transcribed (Figure

3.17A) or the transcript may remain associated with Pol II after cleavage at the PAS or even after transcription termination (Figure 3.17B).

It has previously been shown that co-transcriptional splicing in *S. cerevisiae* occurs preferentially when the 3' exon is longer (Tardiff et al., 2006). Comparison of the Rpb3 TEF-seq signal for splicing intermediates for a TU (*YFL039C* (Figure 3.15)) with a long 3' exon (1219 bp) shows a higher proportion of excised intron signal compared to first exon signal (first exon = 308 counts; excised intron = 209 counts; excised intron/first exon = 0.68) than a TU (*YOR069W* (Figure 3.16)) with a short 3' exon (518 bp) (first exon = 224 counts; excised intron = 47counts; excised intron/first exon = 0.21). This proportion was calculated for all 148 TUs assigned as having a first exon signal and correlated with the distance from the 3' ss to the PAS, showing a weak, but significant, positive correlation (spearman correlation coefficient = 0.305, p value = 0.00016). The first exon and excised intron signals result from the first and second steps of splicing, respectively. An explanation of this result could be that there is less time to complete both splicing reactions during transcription on TUs with a shorter second exon, giving rise to a higher proportion of the intermediate from the first step of splicing, than the second step, associating with Pol II. The reason why this correlation is not stronger may be due to noise within the data or differences in elongation rate on particular intron-containing TUs (Carrillo Oesterreich et al., 2010).

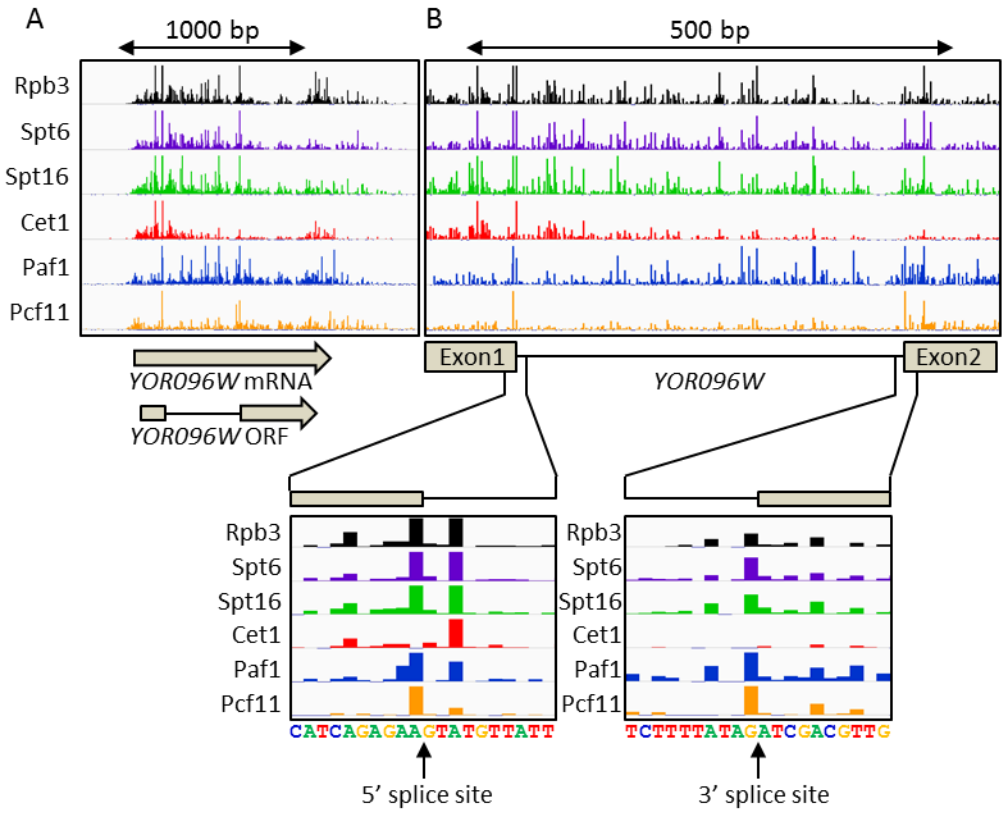


Figure 3.15. *YOR096W* shows peaks at the terminal nucleotide of its first exon and intron in TEF-seq data. A) TEF-seq data for six factors across *YOR096W*. B) Increased resolution of this data across the region around the *YOR096W* intron and further increased resolution over the ends of the intron. Tracks are not smoothed and are on the same relative scales as in Figure 3.8. The maximum values of these scales shown in Table 3.1 have been multiplied by 9 due to the high level of transcription on this TU.

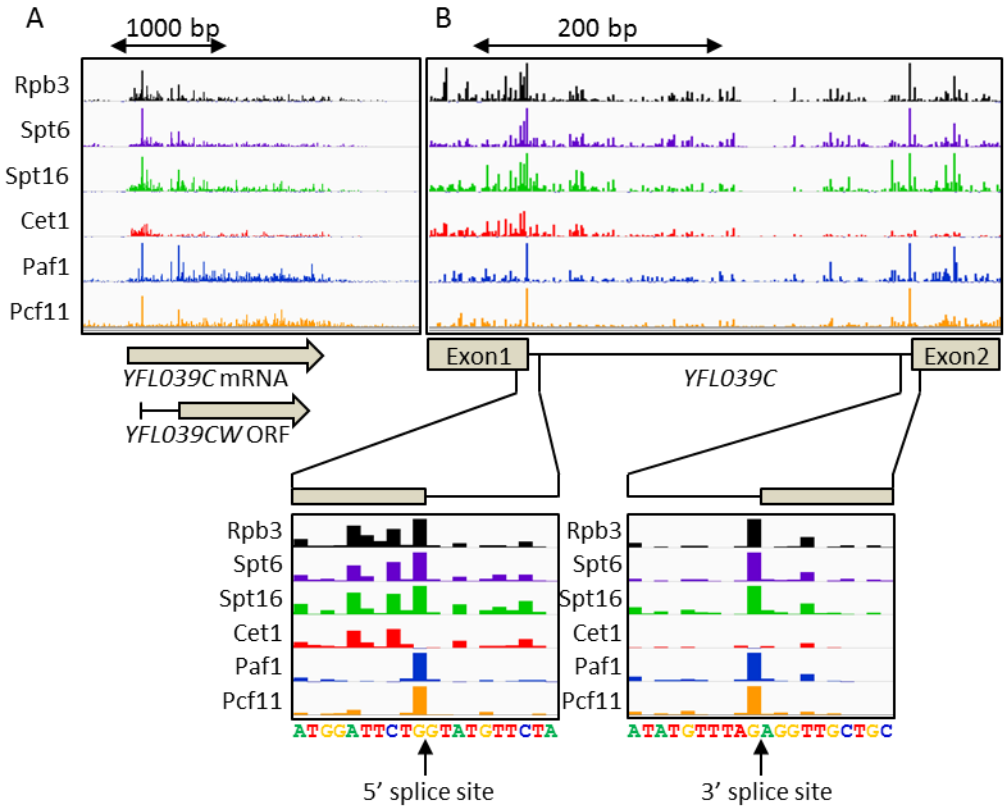


Figure 3.16. *YFL039C* shows peaks at the terminal nucleotide of its first exon and intron in TEF-seq data. A and B) Same as Figure 3.15.

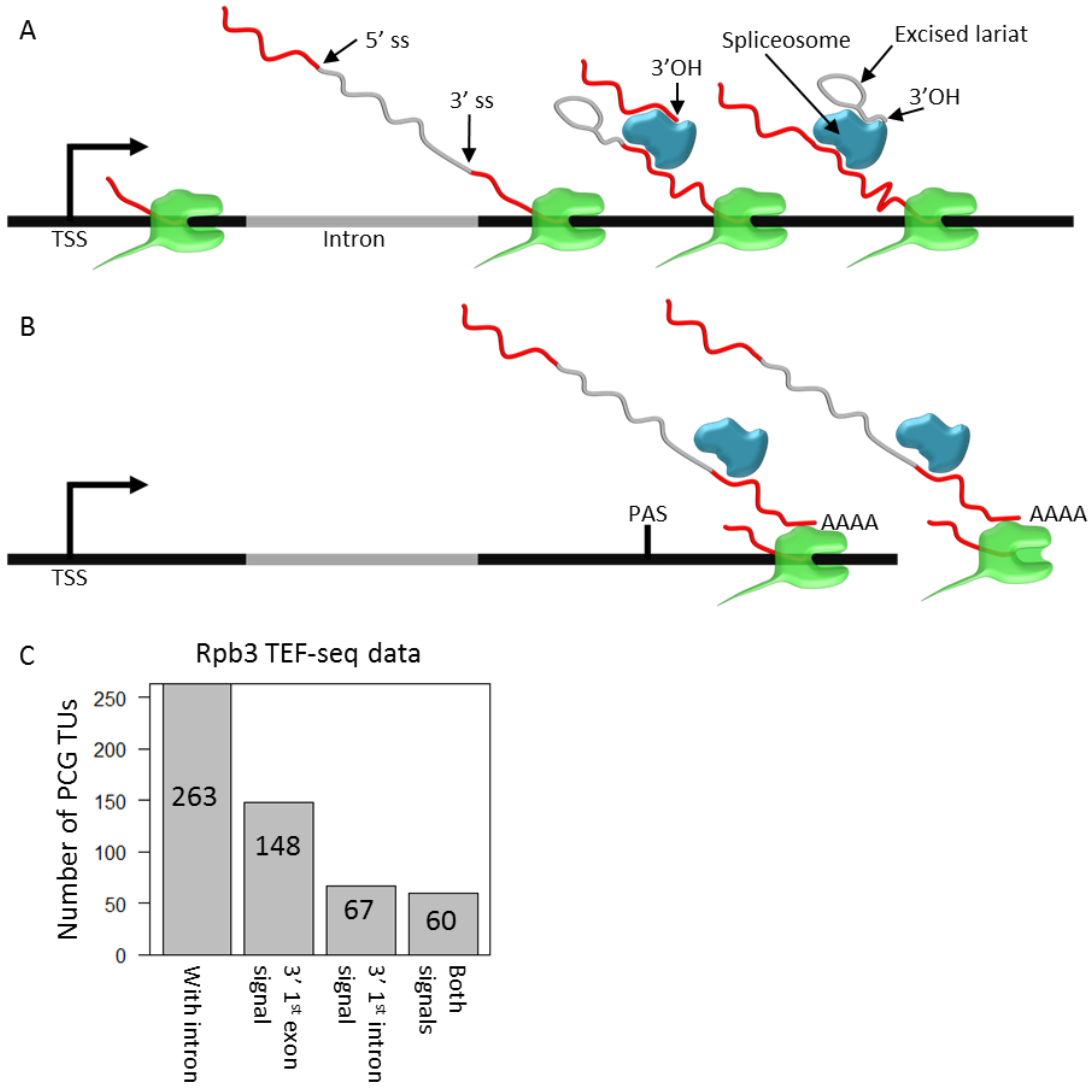


Figure 3.17. Splicing intermediates co-immunoprecipitate with Pol II at many intron-containing TUs. A) Schematic illustrating how splicing intermediates consisting of the first exon and excised intron lariat from the first and second step of the splicing reaction, mediated by the spliceosome, co-immunoprecipitate with Pol II either during their transcription (A) or after cleavage of the transcribed RNA or even after transcription termination (B). These intermediates are sequenced from their terminal nucleotide, giving rise to peaks at the terminal nucleotide of the first exon and first intron in TEF-seq data. C) Bar chart showing the number of intron-containing PCG TUs showing splicing intermediate peaks either at the terminal nucleotide of the first exon or first intron.

3.2.10 Cet1 co-immunoprecipitates a TEC without splicing intermediates

The presence of these splicing intermediates in TEF-seq data from Spt6-, Spt16-, Cet1-, Paf1- and Pcf11-tagged strains was then analysed (Figure 3.15 and 3.16). Nrd1 data was not analysed due to the high peaks in signal from binding to the start of mature degraded RNA typically occurring close to splicing intermediate signal positions. Terminal nucleotide signals for the first exon and intron were counted and assigned as splicing intermediates as before. The number of TUs showing a splicing intermediate at these terminal nucleotide positions was calculated out of those showing this signal in the Rpb3 TEF-seq data (Figure 3.18A-B). Remarkably, Cet1 was the only factor to show almost no signal for either splicing intermediate. An explanation for this could be due to lower levels of Cet1 over the 3' ends of TUs where splicing begins (Figure 3.18C). This may also reflect differential association between factors within the TEC; while some factors may interact more strongly with Pol II, others may interact more strongly with the spliceosome (Figure 3.18D). Cet1 shows greatly reduced affinity for its binding site over the RNA exit channel of Pol II once it has removed the terminal phosphate of the 5' nucleotide of the RNA (Martinez-Rucobo et al., 2015). The lack of Cet1 association with splicing intermediates may therefore reflect a control mechanism, which ensures that splicing is restricted to RNAs that have been capped. In support of this, the cap binding complex is required for splicing but can only associate with the TEC once the RNA has been fully capped (Gonatopoulos-Pournatzis and Cowling, 2014; Görnemann et al., 2005; Lidschreiber et al., 2013).

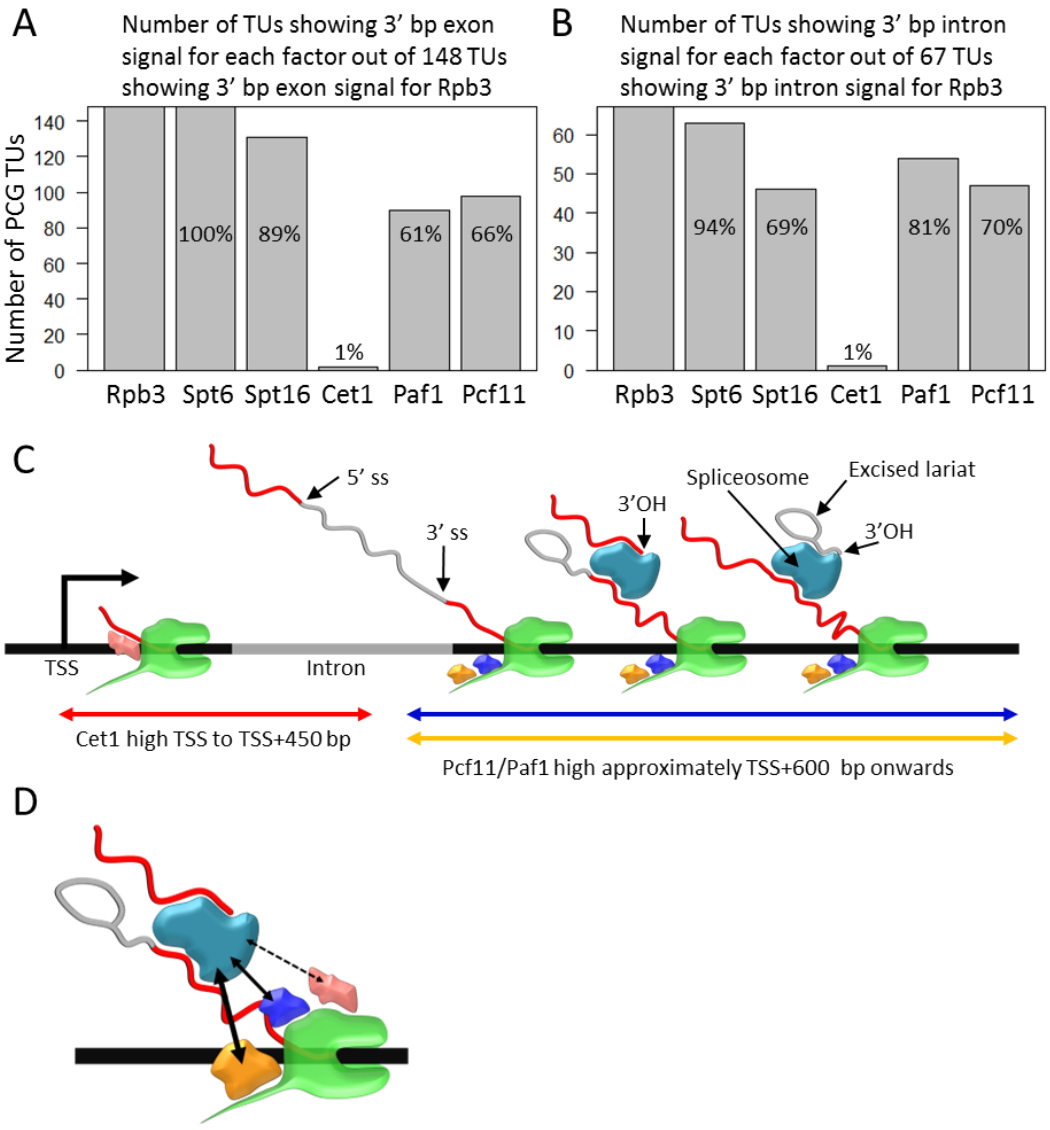


Figure 3.18. Cet1 co-immunoprecipitates a TEC without splicing intermediates. A and B) Bar charts showing the number of TUs assigned as showing a splicing intermediate for the first exon (A) or first intron (B) in Rpb3 TEF-seq data also assigned as showing these intermediates in TEF-seq data for other factors. C and D) Possible models explaining how Cet1 co-immunoprecipiates a TEC without splicing intermediates while other factors do. C) Cet1 is lower over the 3' exon of TUs where splicing occurs co-transcriptionally. D) Different factors may associate with the spliceosome within the TEC with different affinities.

3.2.11 Differential factor binding to non-coding and protein- coding gene TUs

Several studies using ChIP-chip on elongation factors and Rpb1 CTD modifications have attempted to determine whether the same levels and profiles of factor binding and modifications occur at every TU. Some conclude that the levels of elongation factors

scale uniformly with the level of Pol II across all TUs (Bataille et al., 2012; Mayer et al., 2010) with others concluding that some non-coding TUs, including CUT TUs, have different levels of particular modifications (Kim et al., 2010). Ultimately its low resolution and lack of strand-specificity, when many non-coding TUs are transcribed from the antisense strand of PCG TUs, means ChIP cannot adequately answer this question.

To address this question using TEF-seq, TU averaged EF/Rpb3 ratios were plotted for SUT, CUT and PCG TUs from the TSS to +400 bp (Figure 3.19A). The non-coding TU profiles are noisier because the average ratios for each 10 bp bin are calculated over fewer TUs than for PCGs. Therefore all profiles have been smoothed by plotting a running mean average of these average ratios using a moving window consisting of five 10 bp bins. XUT TUs were not analysed as they largely overlap with SUTs. The Nrd1 occupancy on CUT TUs lacks the same high 5' peak coming from the binding of mature transcripts as these ncRNAs are generally less stable than mRNAs and so the level of mature transcripts is low. Across the body of CUT TUs and to a lesser extent SUT TUs, Nrd1 occupancy is significantly higher than on PCG TUs (p values = 2.5×10^{-14} and 2.6×10^{-23}, respectively). These p values were calculated using a Student's t-test to test the difference in the means for these groups for the bin 350-360 bp after the TSS. This high occupancy can also be seen for individual SUTs and CUTs (e.g. *CUT207*, Figure 3.8C, *CUT241* Figure 3.8E, *SUT650* Figure 3.8F). Nrd1 is likely to be targeted to these transcripts, during their transcription by Pol II, because they are enriched for Nrd1 and Nab3 (an Nrd1 binding partner) binding motifs (Schulz et al., 2013). Nrd1 recruitment leads to the termination of these transcripts and targets them for rapid degradation by the exosome (Tudek et al., 2014). The significantly higher level (p value = 6.1×10^{-7} (calculated as above)) on CUT TUs than SUT TUs is likely to lead to more efficient targeting of CUTs for exosome degradation. These non-coding TUs have been split into two groups determined by whether (SUTs) or not (CUTs) their steady state level can be

detected above a threshold using RNA-seq or microarrays in a wild type strain. The steady state level is determined by the balance between their transcription and degradation rates. It is likely that there is a gradual transition between the two groups. Higher transcription and/or lower binding by Nrd1, leading to less efficient targeting for degradation, is more likely to lead to detection and classification as a SUT.

Paf1 and Set2 occupancies are reduced on both SUT and CUT TUs compared to PCG TUs. This reduction of Paf1 can be seen for individual SUT and CUT TUs (e.g. *CUT207*, Figure 3.8C, *CUT241* Figure 3.8E, *SUT650* Figure 3.8F). Antisense transcription at PCG TUs often includes CUT and SUT TUs. This is therefore consistent with the lower level of the histone modifications, H2Bub1 and H3K36me3, dependent upon these factors, seen at PCG TUs with high levels of antisense transcription (Murray et al., 2015).

CUTs are generally short. Therefore to assay for differential binding of factors downstream of +400 bp, the EF/Rpb3 ratios were plotted for combined CUT and SUT TUs compared to PCG TUs from the TSS to +800 bp (Figure 3.19B). These p values were calculated, using a Student's t-test, to test the difference in the means for these groups for the bin 350-360 bp after the TSS. The profiles have again been smoothed using the same running mean average as above (Figure 3.19A). A Student's t-test was used to calculate p values for the difference between the two groups for each factor for the bin 700-710 bp after the TSS (Table 3.4). The levels of Paf1 and Set2 are reduced by approximately half from the levels of occupancy seen on PCG TUs. Spt16 also no longer shows the increase in occupancy from the TSS to ~+650 bp seen for PCG TUs. These observations suggest the number of Ser2, Ser5 doubly phosphorylated repeats on the CTD of Rpb1 is reduced. Spt6 levels are also slightly reduced. This is possibly a result of the lower levels of Paf1, which is involved in recruiting Spt6 and Ctk1 to Pol II (Dronamraju and Strahl, 2014). In contrast, Pcf11 shows similar levels on CUT and SUT TUs compared to PCG TUs, which is surprising given that *PAF1* deletion results in loss of

Pcf11 from TUs (Mueller et al., 2004). The effect of lower Paf1 levels on Pcf11 levels may be counteracted by the higher levels of Nrd1, which plays a role in Pcf11 recruitment (Grzechnik et al., 2015).

Factor	p value
Cet1	0.42
Paf1	$1.7x10^{-17}$
Pcf11	0.22
Set2	$1.7x10^{-8}$
Spt6	0.0024
Spt16	$2.1x10^{-18}$
Nrd1	$1.4x10^{-8}$
Ssu72	0.61

Table 3.4. P values from a Student's t-test testing whether there is a significant difference in the mean occupancy of each factor relative to Rpb3 on PCG TUs compared to non-coding (SUT and CUT) TUs. These were calculated for values at bin 700-710 bp after the TSS.

Figures 3.20 and 3.21 show extended versions of these profiles from the TSS to + 1400 bp and from the PAS -400 bp to + 100 bp before and after smoothing, respectively. The transitions occurring at the PAS also differ for CUT and SUT TUs. These differences may be due to these TUs having less precisely defined and more variable termination sites. Consequently any transitions are also likely to be less precisely defined. Nonetheless, no decrease is seen in factor binding upstream of the PAS. This may be because there is no coding sequence to trigger transition T6 seen at PCG TUs. If this is a mechanism by which protein-coding transcripts are marked, this may be a mechanism by which non-coding transcripts are distinguished from coding transcripts in the cytoplasm and targeted for degradation. Both Nrd1 and Ssu72 rise and have a peak in occupancy before the PAS. This supports the proposed role of the NRD and APT complexes in terminating non-coding transcription (Kim et al., 2006; Nedea et al., 2003, 2008; Schaughency et al., 2014).

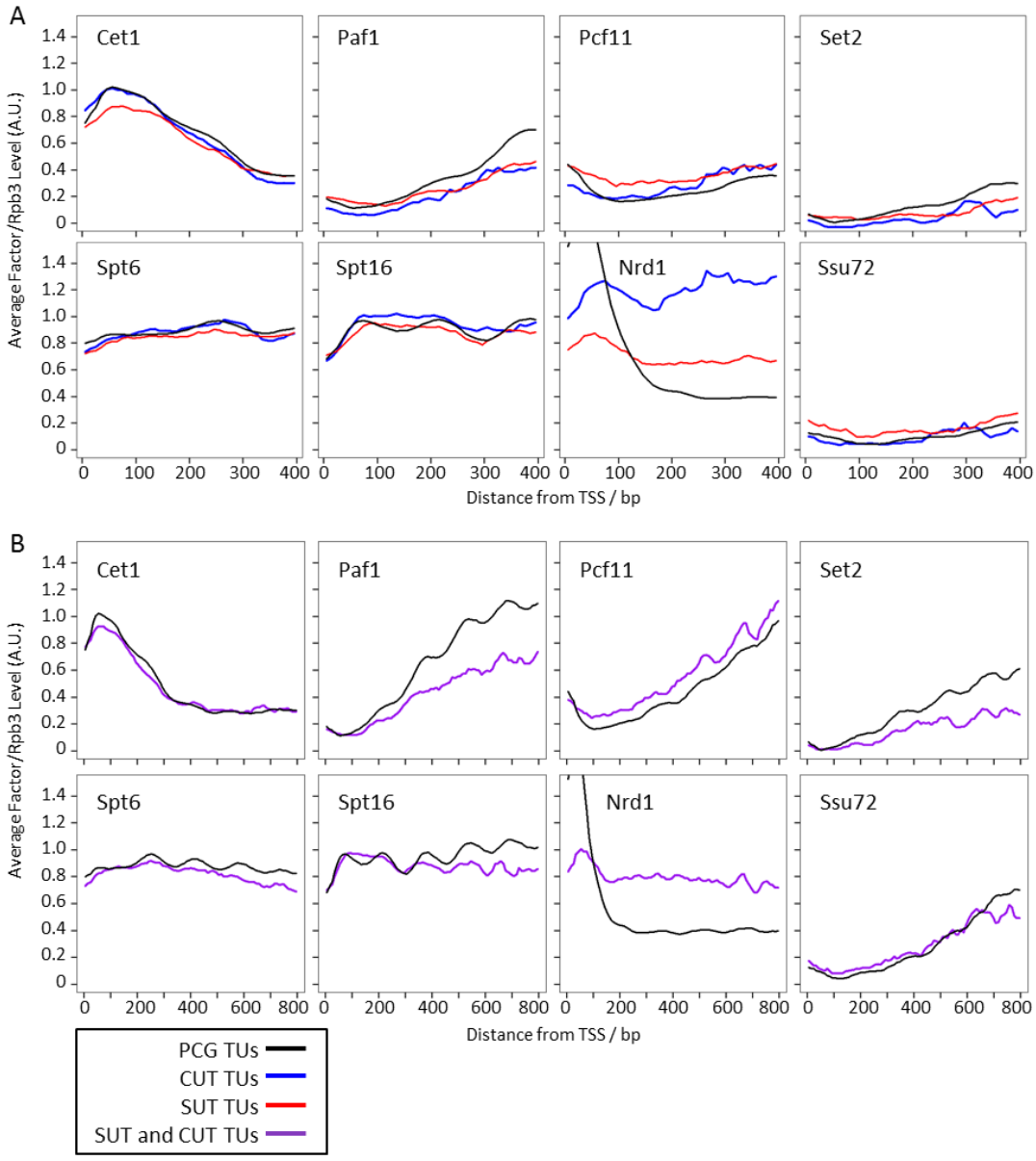


Figure 3.19. Some factors show differential occupancy between non-coding and protein coding TUs (TSS to TSS +800 bp). A) Average TEF-seq EF/Rpb3 ratios from the TSS to +400 bp for TUs grouped into PCGs (Black), CUTs (Blue) and SUTs (Red). B) Average TEF-seq EF/Rpb3 ratios from the TSS to +800 bp for TUs grouped into PCGs (Black), and CUTs and SUTs grouped together (Purple). A running mean average of has been calculated to smooth all profiles using a moving window of five 10 bp bins.

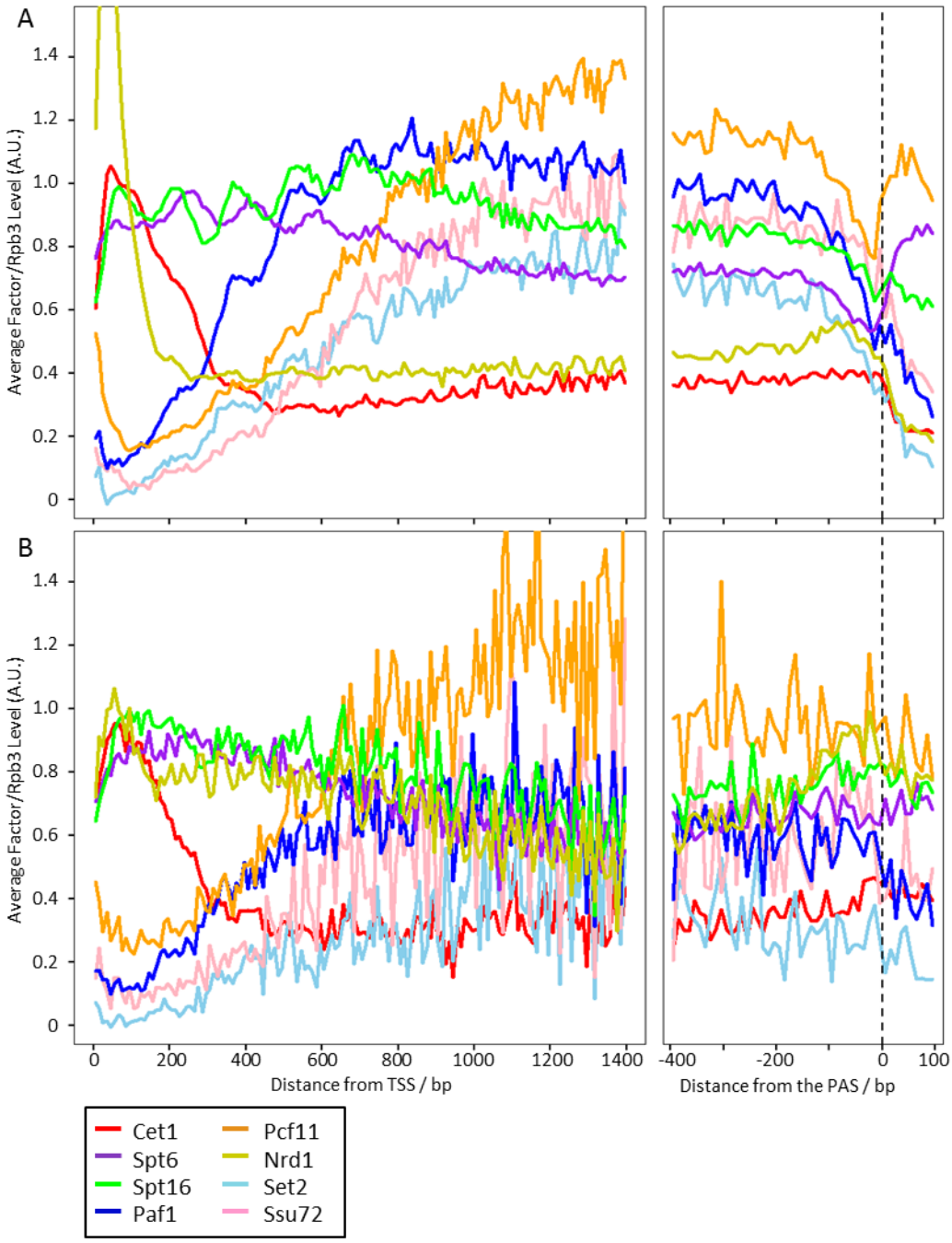


Figure 3.20. Some factors have differential occupancy between non-coding and PCG TUs (whole TU). A) PCG TU averaged TEF-seq EF/Rpb3 ratios from the TSS to +1400 bp and from the polyA site -400 bp to +100 bp averaged across PCG TUs (A) and grouped CUTs and SUTs (B). High noise in B particular towards the 3' end is due to the average being calculated across fewer TUs.

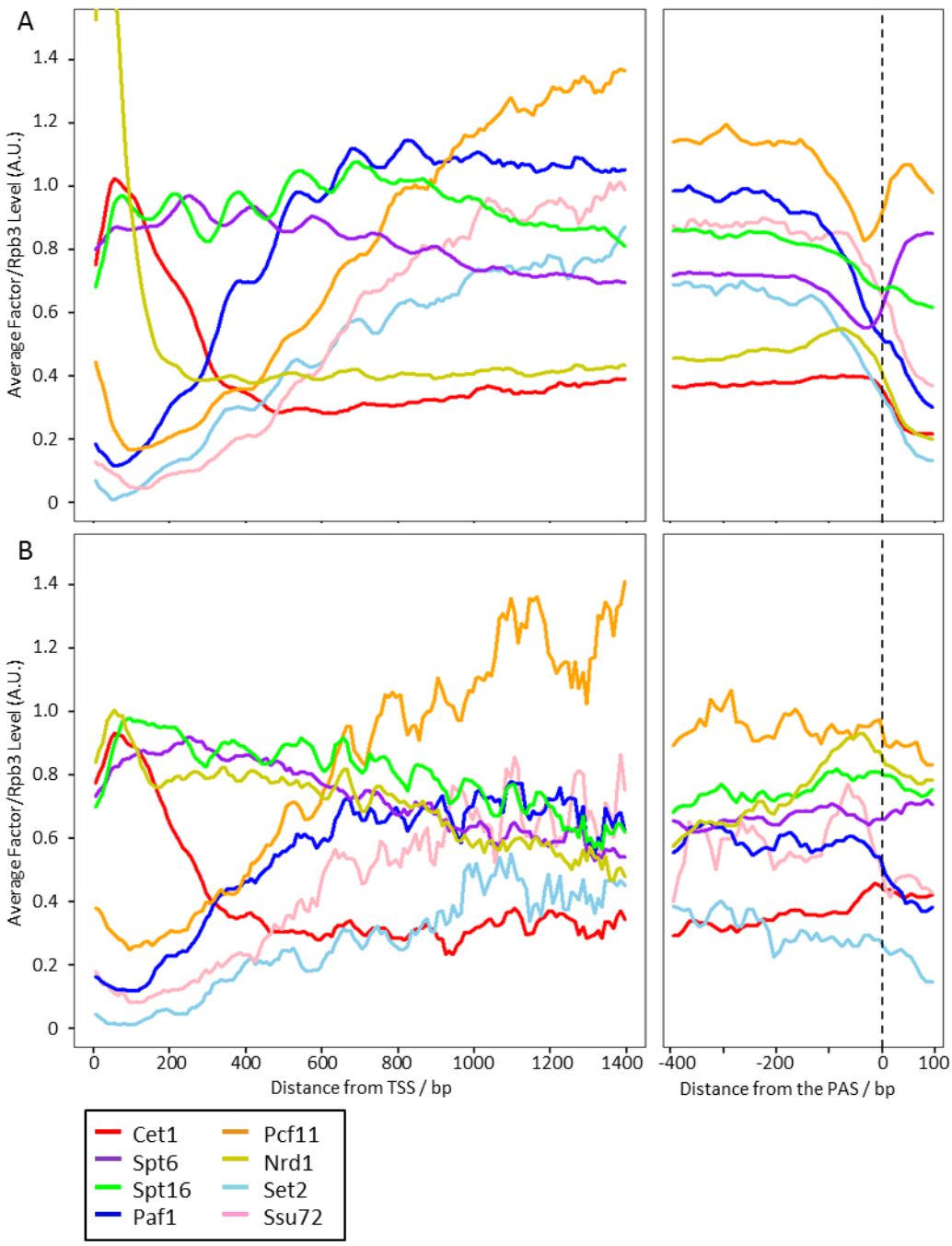


Figure 3.21. Smoothed profiles showing some factors have differential occupancy between non-coding and protein coding TUs (whole TU). A) PCG TU averaged TEF-seq EF/Rpb3 ratios from the TSS to +1400 bp and from the PAS -400 bp to +100 bp averaged across PCG TUs (A) and grouped CUTs and SUTs (B). A running mean average has been calculated to smooth all profiles using a moving window of five 10 bp bins.

3.2.12 Differential binding of factors relative to Pol II to particular protein coding TUs

As Spt6, Spt16 and Paf1 have lower levels relative to Pol II on non-coding TUs, the level of these factors relative to Pol II across PCG TUs was also analysed. By comparing counts for particular regions of PCG TUs from the two replicates for each factor to the two Rpb3 replicates, differential occupancy was assayed using the differential expression function within the DESeq2 package (see Methods). For Spt6 and Spt16, counts were from the TSS to 200 bp upstream of the PAS (PAS -200 bp) while for Paf1 counts were from the TSS +400 bp to PAS -200 bp, as Paf1 is low near the TSS. The final 200 bp before the PAS are excluded to avoid influence from factor transitions occurring in this region. Figure 3.22 shows scatter plots of the base mean occupancy against the relative $\log_2$ fold change in occupancy for each factor relative to Rpb3 occupancy. The base mean occupancy at each TU is the mean occupancy across both factor and Rpb3 replicates after scaling by a size factor calculated by the DESeq algorithm. PCG TUs showing a change in factor occupancy relative to Rpb3 with a p value adjusted for multiple testing between 0.1 and 0.05 are shown in yellow, while those less than 0.05 are shown in red. Most of the PCG TUs show no significant change in occupancy for any factor, showing that, in general, the level of each elongation factor is linearly proportional to the level of transcribing Pol II. Paf1 shows the greatest number of TUs with a significant relative enrichment or depletion over Rpb3. These will be termed Paf1-enriched or Paf1-depleted TUs, respectively. Paf1 also shows the greatest range in relative occupancy levels, followed by Spt16 and then Spt6. Among the Paf1-enriched TUs are a group shown by triangles. These are 5 PCG TUs (*YER130C*/*COM2*, *YIR035C*, *YPR156C*/*TPO3*, *YPR157W*/*TDA6*, *YGR138C*/*TPO2*) from a group of 10, which generally encode membrane proteins and are rapidly induced by the Haa1 transcription factor in response to acid stress (the remaining 5 genes were too lowly transcribed to be selected for differential occupancy analysis) (Fernandes et al., 2005; Keller et al., 2001). Four of these 5 PCG TUs, not including *YIR035C*, are also significantly enriched for Spt16. Of

these four genes only *TPO2* is significantly enriched for Spt6, although the other genes show a trend towards enrichment. This shows enrichment of these three factors on particular TUs is linked and could be determined by transcription factor binding to the promoter, at least for this group of TUs.

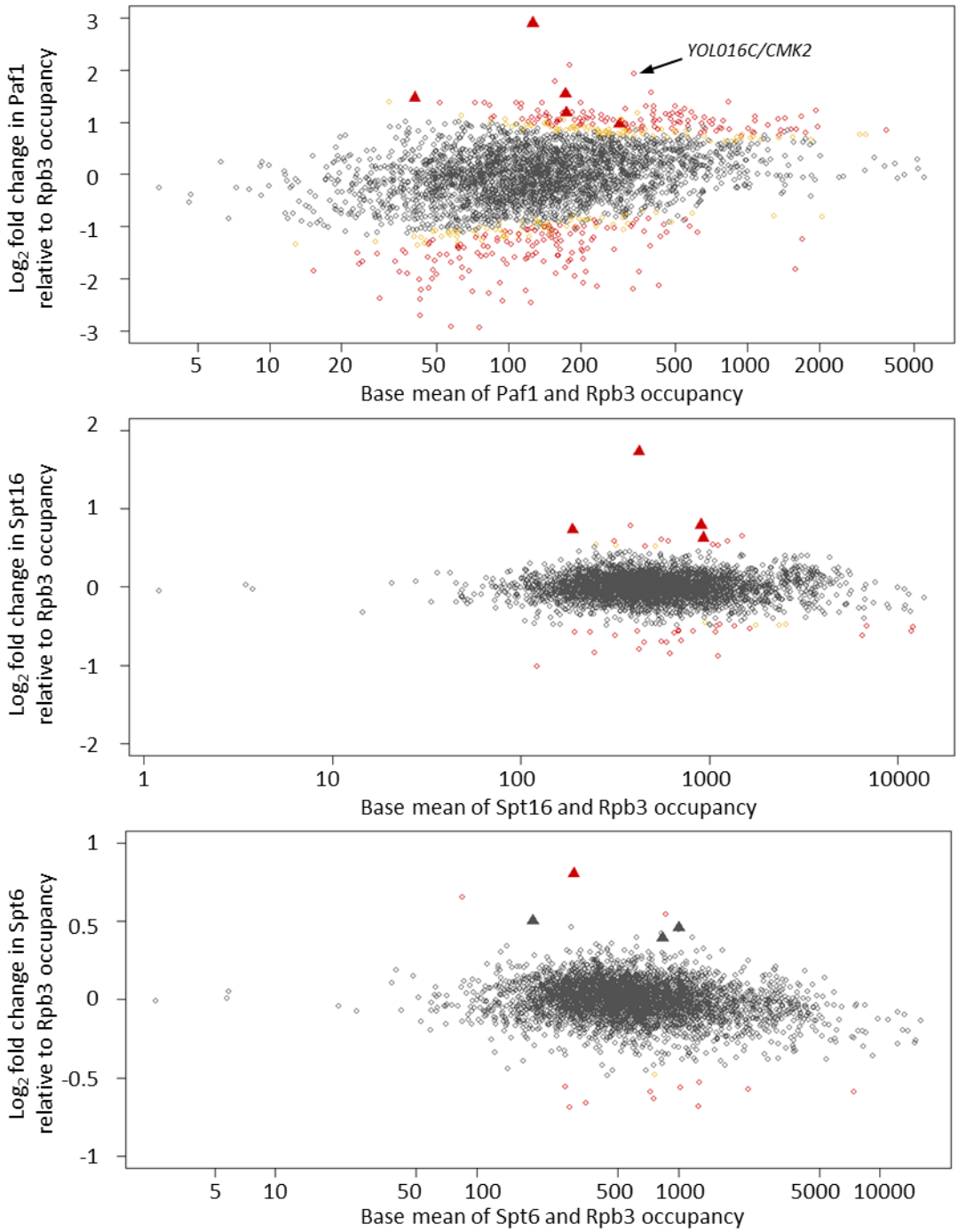


Figure 3.22. Differential factor occupancy at PCG TUs. Scatter plots showing the base mean occupancy against the relative $\log_2$ fold change in occupancy for each factor relative to Rpb3 occupancy for each PCG TU. Yellow points $0.05 \leq$ p value < 0.1. Red points p value < 0.05. Triangles show TUs known to be regulated by the Haa1 transcription factor.

3.2.13 FLAG-tagged strains show no growth or expression differences

Different occupancy levels of particular factors could be due to differences in transcription between the different FLAG-tagged strains. None of the strains show a growth defect, suggesting the presence of the FLAG-tag on any of the factors has no impact on global expression levels (Figure 3.23). RNA levels measured using quantitative RT-PCR from Paf1-enriched (*TPO3*, *CMK2*, *TPO2*), Paf1-depleted (*SLC1*) and no significant difference in Paf1 (*HPT1*) TUs. This was determined by first normalising to the level of *ADH1* RNA in each sample. Then the average RNA level from five replicates was calculated for each of the five RNAs between the Paf1- and Rpb3-FLAG-tagged strains (Figure 3.24). This demonstrates that the changes in occupancy seen for particular factors are genuine.

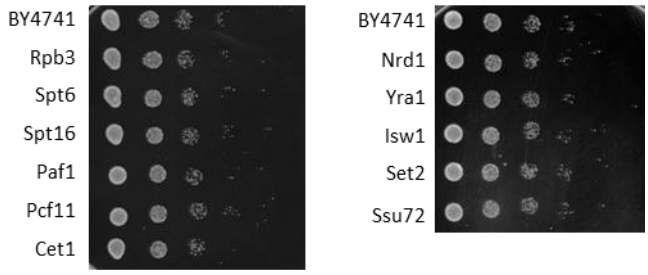


Figure 3.23. No growth differences between wild type and FLAG-tagged factor strains. Drop plates showing the growth of different FLAG-tagged factor strains and the wild-type untagged BY4741 background strain following serial 10x dilution on YPD agar.

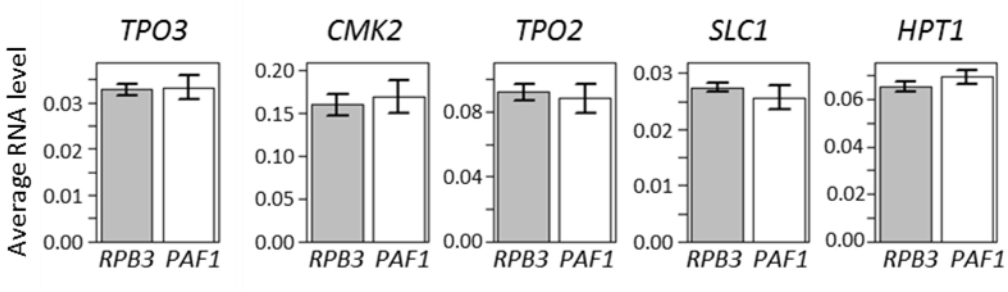


Figure 3.24. No difference in RNA levels between *PAF1-3xFLAG* and *RPB3-3xFLAG* strains. Bar charts showing the average RNA level in *RPB3-3xFLAG* (*RPB3*) and *PAF1-3xFLAG* (*PAF1*) strains from five replicates for Paf1-enriched (*TPO3, CMK2, TPO2*), Paf1-depleted (*SLC1*) and no significant difference in Paf1 (*HPT1*) TUs. Error bars show the standard error of the mean.

3.2.14 mRNAs synthesised from enriched Paf1 TUs are enriched for Yra1 and Mex67 binding

To establish a role for Paf1 enrichment relative to Rpb3 levels on particular TUs, PAR-CLIP data showing the level of binding of two factors involved in RNA export, Mex67 and Yra1, were used (Baejen et al., 2014). PCG TUs were split into three groups of Paf1-enriched, depleted or no significant change. Figure 3.25 shows the average counts per bp for Yra1 and Mex67 for each group, with the raw data having already been normalised for RNA level. A Welch's one-tailed t-test was used to show mRNAs from the Paf1-enriched group showed significantly increased binding by both factors compared

to the no significant change group. mRNAs from the Paf1-depleted group showed significantly decreased binding by Mex67 compared to the no significant change group.

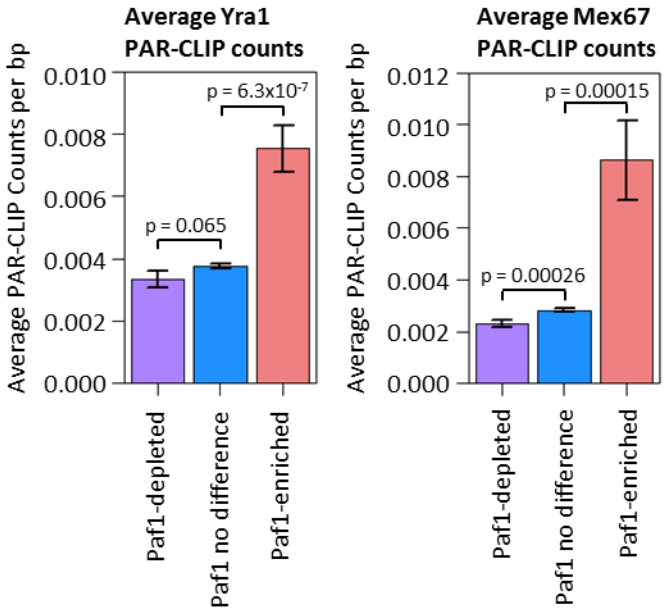


Figure 3.25. RNA from Paf1-enriched TUs show increased binding to Yra1 and Mex67. TUs were split into Paf1-enriched (red), Paf1-depleted (purple) and no significant difference in Paf1 (blue) groups. Bar chart showing the average counts per base normalised to RNA level for Yra1 and Mex67 bound to RNAs from these groups calculated using PAR-CLIP data from Baejen et al., (2014). Error bars show the standard error of the mean. P values show the significance of the increase or decrease in the mean of the enriched or the depleted group relative to the no significant difference group, respectively.

3.2.15 mRNAs synthesised from Paf1-enriched TUs accumulate in the nucleus in a *Δpaf1* strain

mRNAs from Paf1-enriched TUs seem to be specifically targeted for binding by these two export factors which may mean these RNAs are targeted for more efficient or rapid export from the nucleus. Two Paf1-depleted TUs (*YDL052C*/*SLC1*, *YGR033C*/*TIM21*), two TUs showing no change in Paf1 (*YOL086C*/*ADH1, YDR399W*/*HPT1*) and four Paf1-enriched TUs (*YAL035W*/*FUN12*, *YPR156C*/*TPO3*, *YGR138C*/*TPO2*, *YOL016C*/*CMK2*) were selected for further analysis. Figure 3.26 shows the relative level of Rpb3 and Paf1 across each of these TUs. The level of Paf1 for each TU is multiplied by the same

constant factor used for the Paf1/Rpb3 ratio profiles. The level of Mex67 and Yra1 bound by RNAs from these TUs from the PAR-CLIP data is listed in Table 3.4. To test the effect of Paf1 on nuclear export, a *Δpaf1* strain was generated. This strain was slow growing and has previously been demonstrated to have a growth rate 0.59x that of a wild-type strain as well as large changes in RNA levels (O'Duibhir et al., 2014). RNA was purified from a nuclear extract and whole cell extract from a BY4741 wild-type (wt) and the *Δpaf1* strain. Quantitative RT-PCR was used to measure the level of RNA in each of these four samples. All RNA levels were normalised to the level of *HPT1* RNA in each sample. The *HPT1* TU shows no significant enrichment or depletion of Paf1 (Figure 3.26 from 400 bp onwards). The proportion of each RNA in the nucleus in the *Δpaf1* strain relative to the wt strain controlled for changes in total RNA levels was then calculated using the formula: (nuclear *Δpaf1* level / whole cell *Δpaf1* level)/(nuclear wt level/ whole cell wt level). Figure 3.27 shows the mean from five independent repeats for each RNA. A Welch's one-tailed t-test was used to test the significance of any increase relative to *ADH1* (Table 3.5). Both the RNAs from Paf1-depleted TUs show no significant increase relative to *ADH1* in their nuclear levels. Three of the four RNAs from the Paf1-enriched TUs (*TPO3, TPO2* and *CMK2*) all become significantly enriched in the nucleus after deletion of *PAF1*, while *FUN12* shows a trend towards this enrichment (p value = 0.077). These data suggest the enrichment of Paf1 within the TEC during elongation is important for the export of the RNA being synthesised, possibly by targeting the RNA for binding by the RNA export factors Yra1 and Mex67 (Figure 3.28).

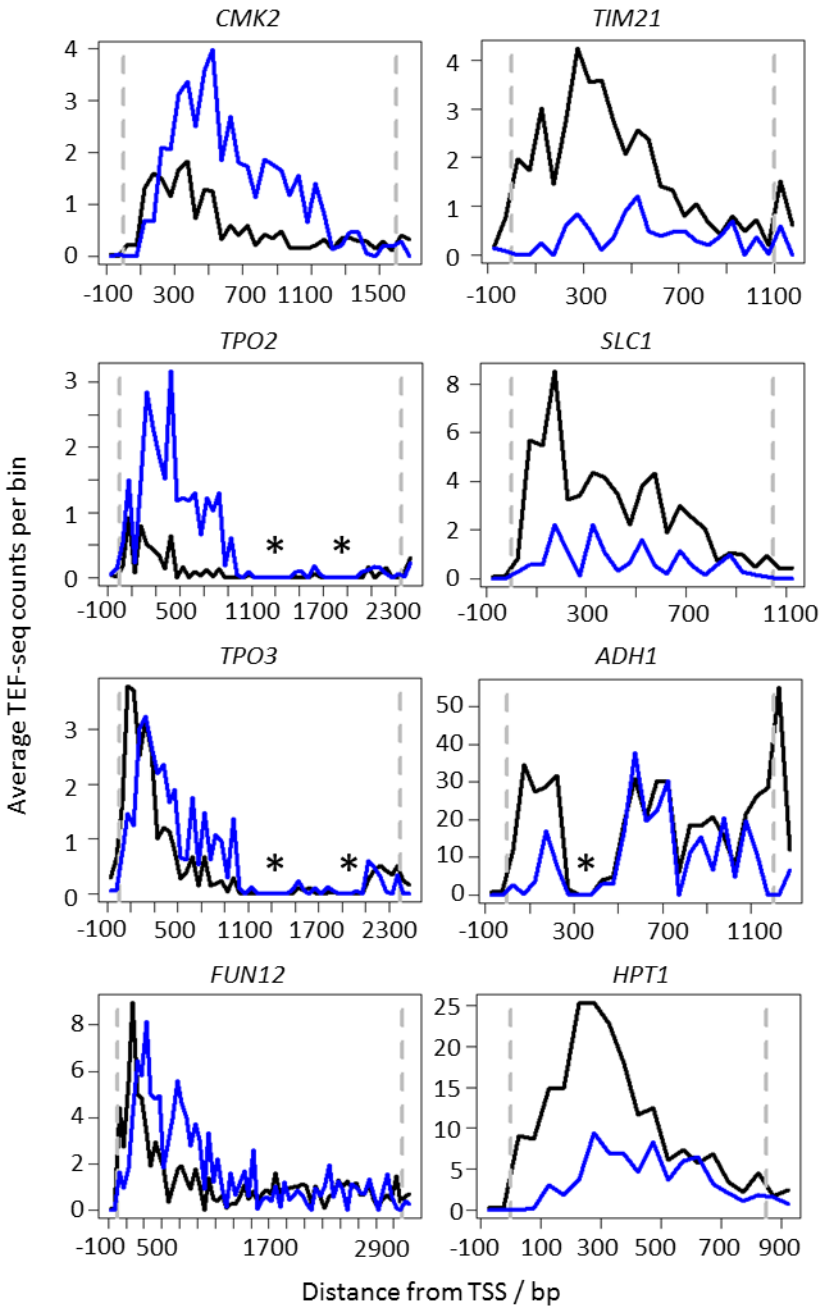


Figure 3.26. Paf1 and Rpb3 occupancy on the TUs analysed. Profiles showing average Paf1 (Blue) and Rpb3 (Black) occupancy in 50 bp bins from the TSS -100 bp to the PAS +100 bp on TUs under analysis. Paf1 and Rpb3 occupancy have not been normalised for differences in sequencing depth and Paf1 levels have been multiplied by a constant factor to enable plotting on the same scale. These profiles therefore show relative, rather than absolute occupancy differences. TSSs and PASs are shown by dashed grey lines. Gaps in the profiles, marked by asterisks, particularly apparent for *ADH1*, result from discarding non-uniquely aligned reads. Paf1 only reaches maximal levels by TSS +650 bp onwards, hence the difference seen between Rpb3 and Paf1 levels at the start of TUs particularly apparent for *HPT1*.

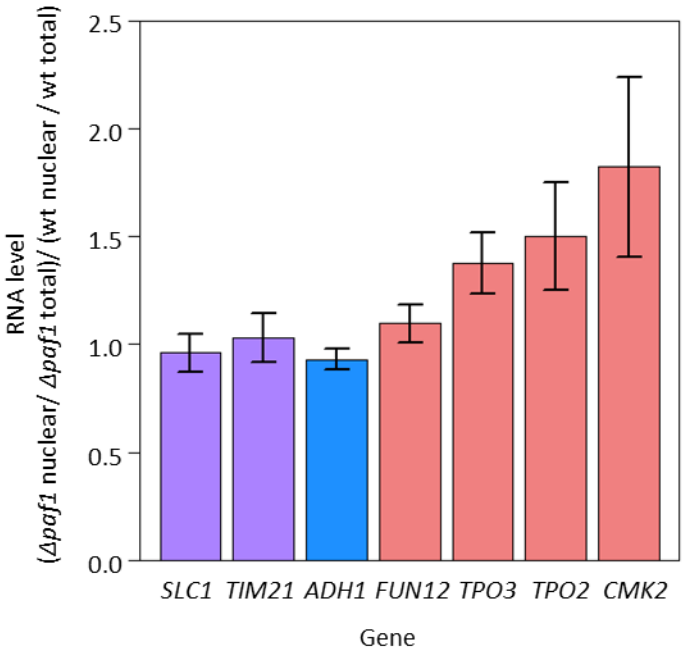


Figure 3.27. RNAs from Paf1-enriched TUs accumulate in the nucleus upon deletion of *paf1*. Bar chart showing the mean ratio from five replicates of the proportion of the total of each RNA in the nucleus of a Δ*paf1* strain compared to a wild type strain. Paf1-enriched TUs (red), Paf1-depleted TUs (purple), TUs showing no significant difference in Paf1 (blue). Error bars show the standard error of the mean.

Gene	Alias	Increase from *ADH1* mean	p value	Mex67 PAR-CLIP	Yra1 PAR-CLIP
YDL052C	*SLC1*	0.03	0.3910	0.001897	0.002193
YGR033C	*TIM21*	0.10	0.2225	0.000987	0.000618
YOL086C	*ADH1*	0	N/A	0.001340	0.001511
YDR399W	*HPT1*	N/A	N/A	0.006397	0.004978
YAL035W	*FUN12*	0.17	0.0774	0.002022	0.005138
YPR156C	*TPO3*	0.45	0.0160	0.026412	0.012426
YGR138C	*TPO2*	0.57	0.0423	0.072042	0.014801
YOL016C	*CMK2*	0.89	0.0496	0.121836	0.027809

Table 3.5. For each RNA, the increase relative to *ADH1* in its proportion in the nucleus in a Δ*paf1* compared to a wt strain is shown. P values from a t-test on this increase and Mex67 and Yra1 PAR-CLIP counts per base are also shown.

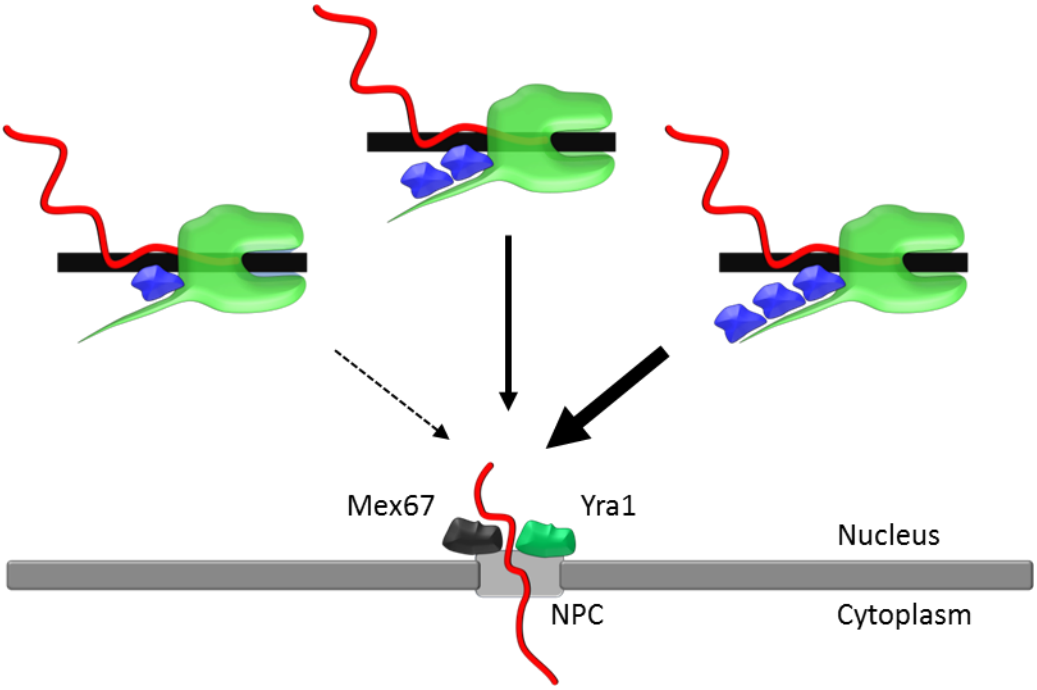


Figure 3.28. The link between Paf1 and nuclear export. Increased Paf1 (Blue) levels on Pol II possibly leads to increased binding of the nascent RNA by export factors Yra1 and Mex67, resulting in more efficient RNA export through the nuclear pore complex (NPC).

3.3 Discussion

Here, the recently developed technique of NET-seq has been replicated. The technique has then been improved by removing a previously unidentified background signal present in all NET-seq data. This improved NET-seq technique has then been adapted for mapping factors associated with Pol II within the TEC. This new technique, termed TEF-seq, gives strand-specific high-resolution maps of the positions of factors bound to the actively transcribing Pol II, giving it clear advantages over the non-strand-specific, lower resolution studies carried out using ChIP.

A drawback of TEF-seq results from the presence of the background signal. This means samples have to be sequenced to high depth in order to obtain sufficient factor-specific signal above the background. Reducing the background signal, caused by the non-specific association of RNA with the anti-FLAG beads during immunoprecipitation, would greatly decrease the number of sequencing reads required and the cost. The FLAG peptide sequence (DYKDDDDK) consists predominantly of negatively charged aspartate amino acids. The anti-FLAG antibody may therefore have a preference for binding to negatively charged molecules, including some binding to RNA. Switching to a different epitope tag with fewer charged amino acids, such as an HA tag, may help reduce the background signal. The FLAG epitope is, however, very efficiently immunoprecipitated and so switching to a different tag may also reduce this efficiency, resulting in no net increase in signal over background. An alternative strategy could be to block the beads prior to immunoprecipitation with RNA. The RNA used would have to be chemically modified so as to prevent the DNA linker ligating to its 3' end and thus prevent it from being incorporated into the sequencing library.

The resolution afforded by TEF-seq has demonstrated for the first time that, as Pol II transcribes through nucleosome-bound DNA, factors change in the strength of their association with the TEC. It has also enabled the identification of ten transitions in the

binding of factors to Pol II across TUs. These transitions affect multiple processes including: ensuring co-transcriptional RNA capping occurs correctly; determining the speed of transcription; determining the location of transcription termination and histone modifications; possibly marking the end of the ORF of a transcript. The higher number of transitions implicates increased regulation through an increased repertoire of patterns of PTMs within the TEC, in particular specific combinations of reversible phosphorylation and isomerisation of CTD residues. Further experiments carrying out TEF-seq on other Pol II-associated factors are likely to add to this complexity. The use of mutants in TEF-seq experiments, particularly those which affect CTD binding, including mutants in which the CID of a factor has been swapped with that of another factor (Heo et al., 2013), should provide information on the mechanisms by which these factors are recruited. This should also provide additional information on the roles of these factors in transcription. It would also be interesting to test whether the recruitment profiles of factors change in different environmental conditions or upon gene induction.

Of particular interest is the effect of gene length on the level of particular factors at the 3' end of TUs and how this may impact on transcription termination. In *S. cerevisiae* there are two main Pol II termination pathways – the NNS pathway and the PAS pathway (Porrua and Libri, 2015). A popular model for explaining how the choice of termination pathway is mediated stems from the observation by ChIP that Nrd1 is enriched at the 5' end of TUs (Vasiljeva et al., 2008). Therefore for short TUs that generally include most sn/snoRNAs and CUTs as well as shorter PCG TUs, which show termination defects in NNS pathway mutants (Steinmetz et al., 2006; Tudek et al., 2014), Nrd1 would still be enriched by the end of the TU and so the NNS pathway is used. Observations here show Nrd1 is not enriched relative to Pol II at the 5' end of PCG TUs. Additionally, Nrd1 shows a peak in occupancy just before the PAS at SUT, CUT and PCG TUs and this is irrespective of PCG TU length. In contrast, Pcf11 shows lower levels at the end of short PCG TUs as it requires ~1200 bp to reach maximum occupancy. The

NNS pathway may therefore be active at all TUs but may become increasingly redundant as the PAS pathway becomes more active with increasing levels of its essential Pcf11 component. This dependence of the PAS pathway on length may have had an evolutionary impact on the length of TUs. For instance, the ORFs of the 137 ribosomal protein genes in *S. cerevisiae* are often short. The majority of the TUs of these genes have introns either in their 5' UTRs or shortly downstream of their start codons. One effect of this is to lengthen the TU, which leads to higher levels of Pcf11 at the 3' end of these TUs. These TUs are amongst the most highly transcribed in YPD (Warner, 1999) and so may require more Pcf11 and a more active PAS pathway to terminate effectively.

The strand-specificity and sensitivity of TEF-seq has reliably demonstrated differential occupancy of particular factors on TUs. The data suggest that Paf1 is enriched on a specific group of PCG TUs, resulting in increased binding of RNAs from these TUs by the nuclear export adaptor, Yra1, and receptor, Mex67. The group of genes targeted include a number of genes that are upregulated in response to particular stresses including *CMK2* and genes upregulated by Haa1 in response to acid stress (Cyert, 2001; Fernandes et al., 2005). These genes need to be expressed rapidly to minimise damage to the cell upon stress. This would be assisted by efficient nuclear export. SUTs and CUTs are generally retained in the nucleus and show decreased binding by Mex67 (Tuck and Tollervey, 2013). Their TUs are depleted for Paf1. RNAs synthesised from PCG TUs depleted for Paf1 also show significantly decreased binding by Mex67. This suggests Paf1 depletion leads to nuclear retention of RNA.

Deletion of *PAF1* caused specific enrichment of RNA from Paf1-enriched TUs within the nucleus. This provides evidence that Paf1 enrichment on TUs causes more efficient nuclear export of RNAs. As deletion of *PAF1* causes large growth defects and large transcriptional changes, the observed effect on export may result from an indirect effect such as alterations in the expression of other genes with roles in export. Transcriptional

changes may also occur to different extents during extraction of cellular and nuclear RNAs, potentially affecting the levels of RNA extracted (discussed in Chapter 4). Further experiments, including analysis of the location of RNAs by FISH upon *PAF1* deletion, or disruption should be carried out to support this causal link between Paf1 and export.

As a mechanism, Paf1 could potentially be transferred from Pol II to the nascent RNA at the 3' end of TUs and then recruit export adaptors to these RNAs. In support of this, Paf1 contains an RNA recognition motif and decreases its occupancy within the TEC just before the RNA is released at the PAS. Paf1 also interacts with Hpr1 (Chang et al., 1999), which in turn interacts with Yra1 and Mex67 within the TREX complex. Paf1 may also regulate the modification of Hpr1. This protein undergoes sumoylation, enabling it to bind specific RNAs, including members of the Haa1 regulon (Bretes et al., 2014). Paf1 levels may also affect polyadenylation efficiency as deletion of Paf1 leads to a reduction in polyA tail length, which is known to affect nuclear export (Dower et al., 2004; Mueller et al., 2004). Another mechanism may link to the role of Paf1 in histone modification. Prevention of H2Bub1 by mutation of H2BK123 or deletion of the E3 ubiquitin ligase *BRE1* leads to accumulation of bulk polyA+ RNA in the nucleus when cells are grown at 39^{o}C (Vitaliano-Prunier et al., 2012). Mutation of H2BK123 also causes a decrease in the association between Mex67 and Yra1 with RNA. This is proposed to result from H2Bub1 stimulating the ubiquitination of the APT complex subunit, Swd2, which in turn causes an association between Mex67 and Swd2 (Vitaliano-Prunier et al., 2008, 2012). Paf1 may therefore also regulate export through this mechanism as Paf1 is required for H2Bub1 by Bre1 (Wood et al., 2003). By recruiting the NPC component, Mex67, to specific TUs during their transcription, Paf1 may also cause the association of TUs with the pore. This may enable fast activation of TUs in response to stress as well as regulating the overall organisation of chromatin within the nucleus (Dieppois and Stutz, 2010; Sood and Brickner, 2014).

A study in *S. pombe* has shown that mutants of Paf1C components are susceptible to PCG silencing by small interfering RNAs (siRNA) expressed *in trans* (Kowalik et al., 2015). The proposed mechanism is that mutation of Paf1C disrupts transcript release from the site of transcription. Base-pairing interactions form between these transcripts and siRNAs bound to the RNA-induced transcriptional silencing complex (RITS), recruiting this complex to the site of transcription, and subsequent heterochromatin formation and silencing. Although *S. cerevisiae* lacks a RITS complex, depletion of Paf1C on Pol II during elongation may be a conserved mechanism of not only retaining transcripts within the nucleus but also at the site of transcription. This implies that low Paf1 on CUT and SUT TUs may help retain these transcripts at the site of their transcription, possibly enabling them to play a regulatory role in this region before their degradation.

What regulates the level of Paf1 on a TU remains to be determined. One possibility is it is regulated by the promoter because all PCG TUs regulated by the Haa1 transcription factor that were tested show Paf1 enrichment. This factor may be able to regulate CTD modification patterns leading to increased Paf1. It has previously been shown that RNA stability and RNA localisation can be regulated by the promoter (Trcek et al., 2011; Zid and O'Shea, 2014). It would be interesting to test whether deleting *HAA1*, mutating Haa1 binding sites or inserting the binding site into other promoters could alter Paf1 recruitment to TUs and influence nuclear export of RNA. In addition to Paf1, Set2, Spt16 and Spt6 are also all depleted from CUT and SUT TUs while Spt6 and Spt16 show some enrichment on TUs regulated by Haa1. This suggests these factors are recruited by the same mechanisms or regulate each other's recruitment with many publications confirming links between these factors (Chu et al., 2007; Dronamraju and Strahl, 2014; Squazzo et al., 2002). The variations in the levels of these key factors involved in chromatin regulation is likely to have effects, in addition to that observed on RNA fate, on many transcriptional processes, including the rate of transcription elongation and

the stability of the chromatin in the vicinity of a TU. Investigations into these effects may provide additional clues into the roles of CUT and SUT transcription.

Chapter 4.

The dynamics and regulation of transcriptional changes upon a change in environmental conditions

4 The dynamics and regulation of transcriptional changes upon a change in environmental conditions

4.1 Introduction

4.1.1 Co-ordinated regulation of the expression of neighbouring genes

Genes within the yeast genome are generally in close proximity to each other. Adjacent genes as well as nearby genes are often functionally related and show positively correlated expression levels, which persist when conditions change (Batada et al., 2007; Cohen et al., 2000; Lee and Sonnhammer, 2003). To some extent this is due to the binding of similar transcription factors to their promoters. This is more likely for the many co-expressed adjacent genes that are divergently expressed on opposite strands from the same bidirectional promoter. However, in mammalian cells adjacent genes no longer show co-expression when separated from each other despite maintaining the same promoter sequences (Raj et al., 2006). In addition, within clusters of several genes, not all of them show correlated expression changes. This suggests there are other mechanisms involved.

4.1.2 Regulation by overlapping transcription units

NET-seq shows that transcription often occurs in intergenic regions as well as antisense to genes (Churchman and Weissman, 2011; Murray et al., 2012). Transcript isoform profiling shows transcription can occur across multiple genes to form multi-cistronic transcripts (Pelechano et al., 2013). As discussed in Chapter 1, this overlap between neighbouring TUs has the potential to affect their transcription. The effects can be either activatory or repressive, with mechanisms including disruption of transcription factor binding or alteration of the positioning and modifications of nucleosomes over promoters (Hongay et al., 2006; Houseley et al., 2008; Martens et al., 2004). Despite these observations at individual loci, a more general, genome-wide role and mechanism

of action for overlapping transcription has not been established. One possibility is that different effects are observed dependent upon the types of TU involved, their orientation with respect to each other and the timing and level of their transcription.

4.1.3 Problems with population studies in steady state conditions

Typically studies in yeast are carried out on cells grown exponentially in standard laboratory conditions of highly aerated, rich media at constant temperature. Cells grown in their natural environment would not normally encounter such stable favourable growth conditions. Instead yeast cells have to endure periods of low nutrient conditions, be able to adapt to metabolise different types of nutrient and increase growth upon an increase in nutrient supply, whilst also enduring changes in temperature, osmolarity and acidity. The role of many of the gene expression regulation mechanisms that have evolved to cope with such variability may be less apparent during steady state conditions. In order to understand them better, it may be more informative to study yeast cells as they undergo a change in conditions.

Analysis of single cells from genetically identical populations has demonstrated heterogeneity in their expression patterns (Castelnuovo and Stutz, 2015). This is partly due to cells undergoing unsynchronised metabolic and cell division cycles in which many genes show cycling expression patterns (Silverman et al., 2010). Studies on unsynchronised populations of cells may give the impression that two neighbouring TUs are expressed contemporaneously in the same cell when in fact their expression is temporally segregated. This is clearly an important consideration when investigating overlapping transcription.

4.1.4 The yeast metabolic cycle

Yeast culture undergoing continuous growth in nutrient-limited conditions becomes metabolically synchronised within a yeast metabolic cycle (YMC), resulting in a robust oscillating pattern of oxygen consumption. Analysis of the transcriptome at time points throughout the YMC has shown that 3922 genes show periodic changes in their RNA

level with the same frequency as the oscillations in oxygen consumption (Kuang et al., 2014; Tu et al., 2005). Particular groups of RNAs and their corresponding proteins, assessed using GFP reporters, show maximal expression in different temporal windows of the YMC. These have been broadly characterised as the oxidative (OX), reductive building (RB) and reductive charging (RC) phases of the YMC (Laxman et al., 2010; Tu et al., 2005). The overall logic is to allow the temporal compartmentalisation of particular processes, ensuring that the conditions for these processes are optimised (Figure 4.1). Although largely characterised in *S. cerevisiae*, such metabolic cycling has also been observed in *S. pombe* and mammalian cells, showing this mechanism of temporal compartmentalisation is highly conserved (Slavov et al., 2012).

The OX phase is a period of high oxygen consumption (HOC) with increased ATP production driven by high levels of aerobic respiration. During this phase, genes associated with growth and, in particular, protein synthesis are upregulated. Thus this energetically demanding process occurs when the ATP supply is maximal. High-temporal resolution profiling of the YMC has revealed further subdivision of the OX phase into two distinct expression phases, OX1 and OX2 (Kuang et al., 2014). High levels of ribosomal assembly and rRNA processing factors (ribosomal biogenesis (ribi) genes) are expressed in OX1, while high levels of ribosomal subunits (RP genes) and translation factors are expressed in OX2. This further subdivision possibly helps maximise the efficiency of ribosome assembly and protein synthesis.

The reductive (RB and RC) phases are both periods of low oxygen consumption (LOC). The RB phase begins as the period of HOC ends and is characterised by the maximal expression of most nuclear-encoded mitochondrial genes and genes required for DNA synthesis, chromatin assembly and cell division. The RB phase can also be further subdivided into three different phases of expression, RB1-3, with cell cycle and DNA replication genes generally expressed in RB1 and mitochondrial genes generally

expressed in RB2 and RB3 (Kuang et al., 2014). It has been proposed that the RB phase enables the segregation of DNA replication from the DNA damaging effects of the high production of reactive oxygen species, which occurs during the OX phase (Cai and Tu, 2012). However, it has been observed that DNA replication does not always occur during this phase (Slavov and Botstein, 2011). Despite the connection between cell cycle gene expression and this phase of the YMC, the cell cycle and YMC are not always linked. Cells can go through a metabolic cycle without dividing and within a metabolically cycling culture there appears to be two sub-populations of dividing and non-dividing cells (Laxman et al., 2010; Slavov et al., 2011). The RC phase follows the RB phase and can also be subdivided into three different sub-phases of expression, RC1-3. The RC phase is characterised by the maximal expression of genes involved in the response to stress and survival including vacuole-, peroxisome-, ubiquitin-proteasome- and autophagy-associated proteins. In addition, genes required for synthesis of acetyl-CoA, NADPH and the storage carbohydrates, trehalose and glycogen, are upregulated, leading to the accumulation of these molecules during the RC phase in preparation for their use in the subsequent OX phase (Shi et al., 2010; Tu et al., 2007).

Using single cell techniques, such as fluorescence in-situ hybridisation, it has been shown that individual cells from asynchronous, exponentially growing cultures also go through a YMC with oscillations in RNA levels (Silverman et al., 2010). Increasing the growth rate of cultures increases the time spent in the OX phase relative to the reductive phases (Slavov et al., 2012). This means that in an asynchronous culture, there is an increased fraction of the total cell population in the OX phase at any particular time point. Consequently, analysis of transcriptomes in asynchronous cultures at different growth rates show a positive correlation between growth rate and levels of RNAs expressed in the OX phase and a negative correlation between growth rate and levels of RNAs expressed in the reductive phases (Brauer et al., 2008).

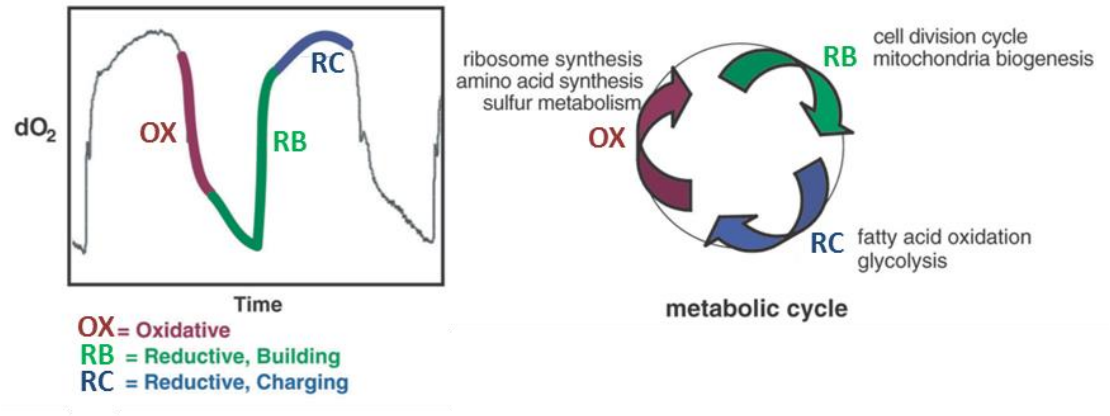


Figure 4.1. The yeast metabolic cycle (YMC). Yeast culture synchronised within the YMC shows an oscillating pattern of oxygen consumption. Each oscillation is split into three phases and each of these is associated with the expression of a particular set of genes leading to temporal compartmentalisation of processes. Figure adapted from Tu et al., (2005).

4.1.5 The environmental stress response

Upon exposure of yeast culture to a wide range of stresses, including the presence of toxic chemicals or changes in temperature, osmolarity or the availability of nutrients, there is a condition-specific response in RNA level changes. Additionally, there is a pattern of RNA level changes that occur common to all stresses, known as the environmental stress response (ESR) (Causton et al., 2001; Gasch et al., 2000; O'Duibhir et al., 2014). There is high correlation between the genes upregulated and downregulated in the ESR and those which are upregulated and downregulated in the reductive phases of the YMC, respectively (Brauer et al., 2008). This has led to the hypothesis that upon exposure of an asynchronous culture to stress, cells in the OX phase transition into the reductive phase and start spending a longer proportion of each YMC in this phase. This leads to an increase in the fraction of cells in this phase and the corresponding increase in the expression of reductive phase genes over oxidative phase genes observed in the ESR (Slavov et al., 2012).

4.1.6 Regulation of the YMC

It is thought that the cell is able to detect the concentration of metabolites indicative of particular phases of the YMC and use this to determine when to drive the expression changes associated with transition into the following phase. DNA-binding transcription factors play a major role. 60 % of them show oscillating expression patterns within the YMC (Tu et al., 2005), while genes which are upregulated at the same point in the YMC tend to have the same transcription factor binding motifs in their promoters (Machné and Murray, 2012). Key signalling pathways, which activate these factors, include the target of rapamycin (TOR) and protein kinase A (PKA) pathways, which sense nitrogen and carbon source availability (Cai and Tu, 2012). In addition, metabolites may have a more direct role, with the high levels of acetyl-CoA at the end of the RC phase acting as a substrate for the specific acetylation of histones at the promoters of OX phase genes by the SAGA complex (Cai et al., 2011). This in turn promotes transcription from these promoters.

4.1.7 Aims

The aim of this chapter was to investigate the gene expression changes that occur on a switch from growth in media containing glucose to that containing galactose, particularly focussing on the changes in nascent transcription using NET-seq. Widespread changes were observed and comparisons were made between expression changes occurring in the YMC. Similar patterns in expression changes were observed at different clusters of TUs, suggesting a possible role for overlap between TUs in regulating these changes. Using a representative locus, this transcriptional overlap was further investigated.

4.2 Results

4.2.1 Transcriptional changes upon a change in carbon source have links to the YMC

To analyse the effect on transcription of an environmental change, the standard NET-seq protocol was applied to the *RPB3-3xFLAG* BY4741 strain after growth to OD_{600} of 0.6 in YP + 2 % glucose (YPD) and after three hours growth in YP + 2 % galactose (YPG). NET-seq measures the density of Pol II at each base pair of a TU across the population. By assuming that the average Pol II elongation rate at each TU is similar, NET-seq can be used to measure the level of transcription. No-tag normalisation was not carried out on these data sets because it is the assumed that the signal is several times greater than the background signal. Counts across all 5579 PCG TUs in the two conditions were summed and differential expression analysis was carried out using the DESeq2 package (Love et al., 2014) within R/Bioconductor. 1054 TUs were shown to have a significant (p value adjusted for multiple testing <0.05) 2-fold decrease (537) or increase (517) in transcription upon the shift to galactose. This is many more than the 14 genes identified as playing a role specific to galactose metabolism or galactose gene transcription (Timson, 2007). To understand why a large number of genes were showing large transcriptional changes as a result of this switch in conditions, the differentially transcribed TUs were compared to genes with oscillating expression within the YMC (Kuang et al., 2014). This showed that the TUs with decreased transcription in galactose are enriched (460, 85.7 %) for OX phase genes while those with increased transcription in galactose are enriched (351, 67.9 %) for RC phase genes (Figure 4.2B). Taking the top 124 genes from these groups and analysing their transcript levels in the YMC confirms that they show peak expression in the OX and RC phases respectively (Figure 4.2D and E) (Kudlicki et al., 2007). The number of genes in each of the subphases of these phases is shown in Figure 4.2C.

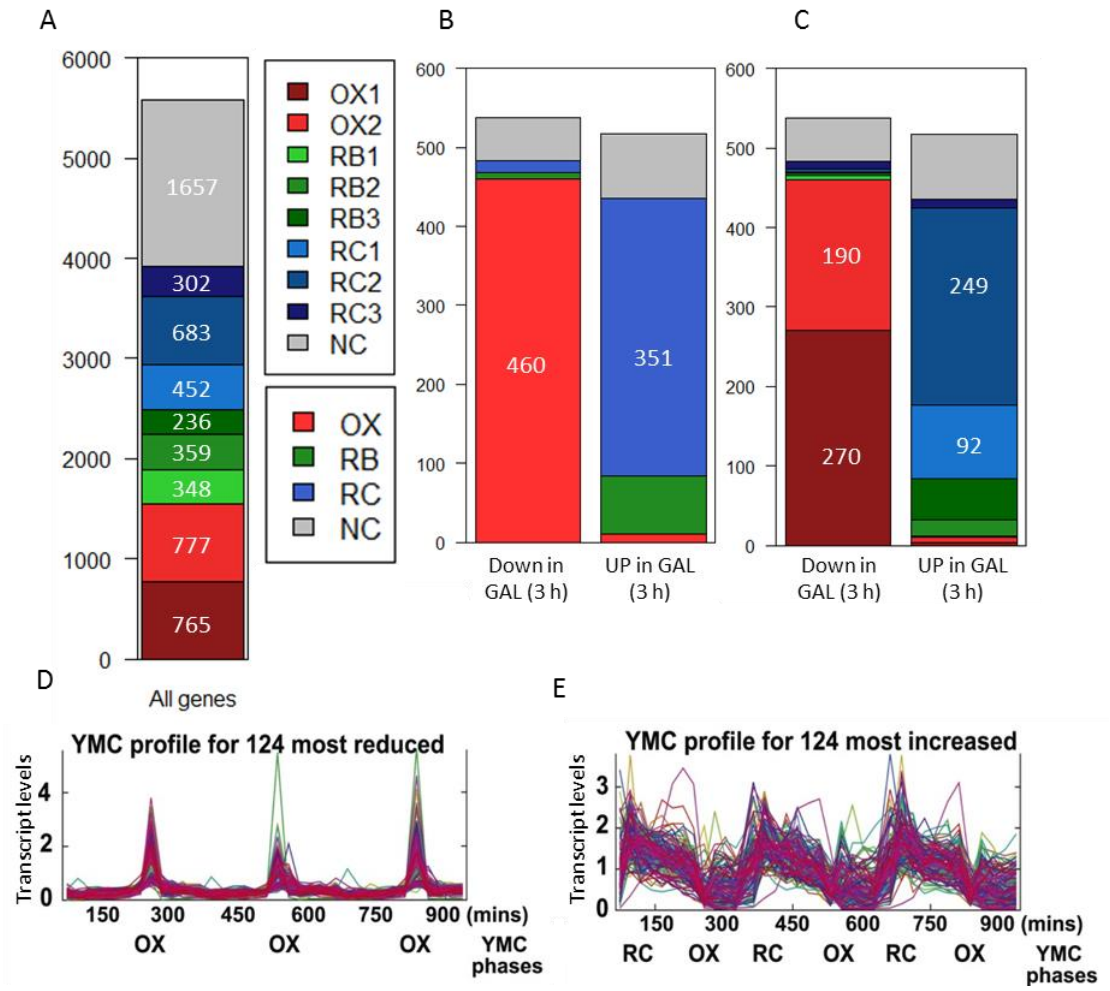


Figure 4.2. Transcriptional changes upon a shift from glucose to 3 h growth in galactose show links to the YMC. A) Number of PCG TUs showing maximum transcript levels in each of the different subphases of the YMC. OX (oxidative), RB (reductive building), RC (reductive charging), NC (non-cycling). Numbers refer to subphases of each phase. B and C) Bar charts showing the number of genes showing two fold decrease or increase in transcription after a shift from growth in glucose to growth in galactose for 3 h and the enrichment of these classes of genes for particular YMC phases (B) or subphases (C) of the YMC. D and E) Transcript expression profiles within the YMC for the 124 PCG TUs showing the largest decrease (D) or increase (E) in transcription (Kudlicki et al., 2007).

An explanation for the changes in transcription seen is as follows. In an asynchronous culture it is thought that individual cells are still cycling metabolically with corresponding oscillating gene expression patterns (Silverman et al., 2010). During growth in glucose, the cells can easily metabolise this carbon source. Consequently, the cells spend a longer proportion of each YMC in the OX phase, expressing genes

associated with translation and biosynthesis, because they are able to release sufficient energy to support these energetically demanding processes. Upon a switch to galactose, the cells are not expressing the necessary genes at high enough levels to efficiently take up and metabolise this new carbon source. Consequently, the cells can no longer release sufficient energy to support the same level of translation and biosynthesis and so the proportion of time spent in the OX phase is reduced. Instead the cells spend a longer proportion of each YMC in the RC phase in which genes associated with survival and response to stress are expressed. Sampling the population of cells at these two time points gives a higher proportion of cells in either the OX phase in glucose or RC phase in galactose. Whether the length of each YMC changes as well as the proportion of time spent in each phase remains to be established.

Differential expression analysis was then applied to non-coding TUs for SUT, CUT and XUT RNAs using the same fold change threshold and significance cut-off. This showed that, in addition to PCG TUs, 15 % of the 1717 ncRNA TUs tested also showed large significant transcriptional changes.

4.2.2 Gene clusters with transcriptional changes upon a shift to galactose share common features

Individual loci at which transcriptional changes were occurring were focused on (Figure 4.3 and Figure 4.4A-D). NET-seq data were analysed alongside high density tiling array data looking at polyA+ purified RNA levels (Nguyen et al., 2014) and published TIF-seq data looking at the diversity of transcript isoforms (Pelechano et al., 2013) in the same two conditions. The expression profiles of PCG transcripts from each locus within the YMC are shown. Many TUs that show transcriptional and transcript level changes upon this change in conditions are clustered together. These changes are occurring at coding and non-coding TUs including CUT and SUT TUs as well as some as yet unannotated TUs. Within each cluster shown, there are TUs which show increased and decreased transcription in galactose. In support of the genome-wide observations, those which

increase show a peak of expression in the RC phase of the YMC and those which decrease show a peak of expression in the OX phase of the YMC. Common features of these clustered genes include the presence of intergenic transcription, with transcription occurring on an upstream gene reading into the promoter and 5' end of a downstream gene (e.g. Figure 4.4B, *SUI2* into the promoter of *YJR008W*; Figure 4.4D, *RPS21B* into the promoter of *GLG2*). There are also high levels of transcription initiating near the 3' end and terminating near the 5' end of some genes within each cluster, forming sense-antisense pairs (SAP) (e.g. Figure 4.4A, *PAN6* and *SUT176*; Figure 4.4B, *YJR008W* and *SUT637*). This level of non-coding transcription is not always represented to the same extent in the microarray data compared to coding transcription, showing that these non-coding transcripts are often rapidly degraded, giving low steady state levels. The TIF-seq data show that, for many of the clustered TUs, there are many different isoforms, including isoforms that overlap with two or even three different genes to give multi-cistronic transcripts.

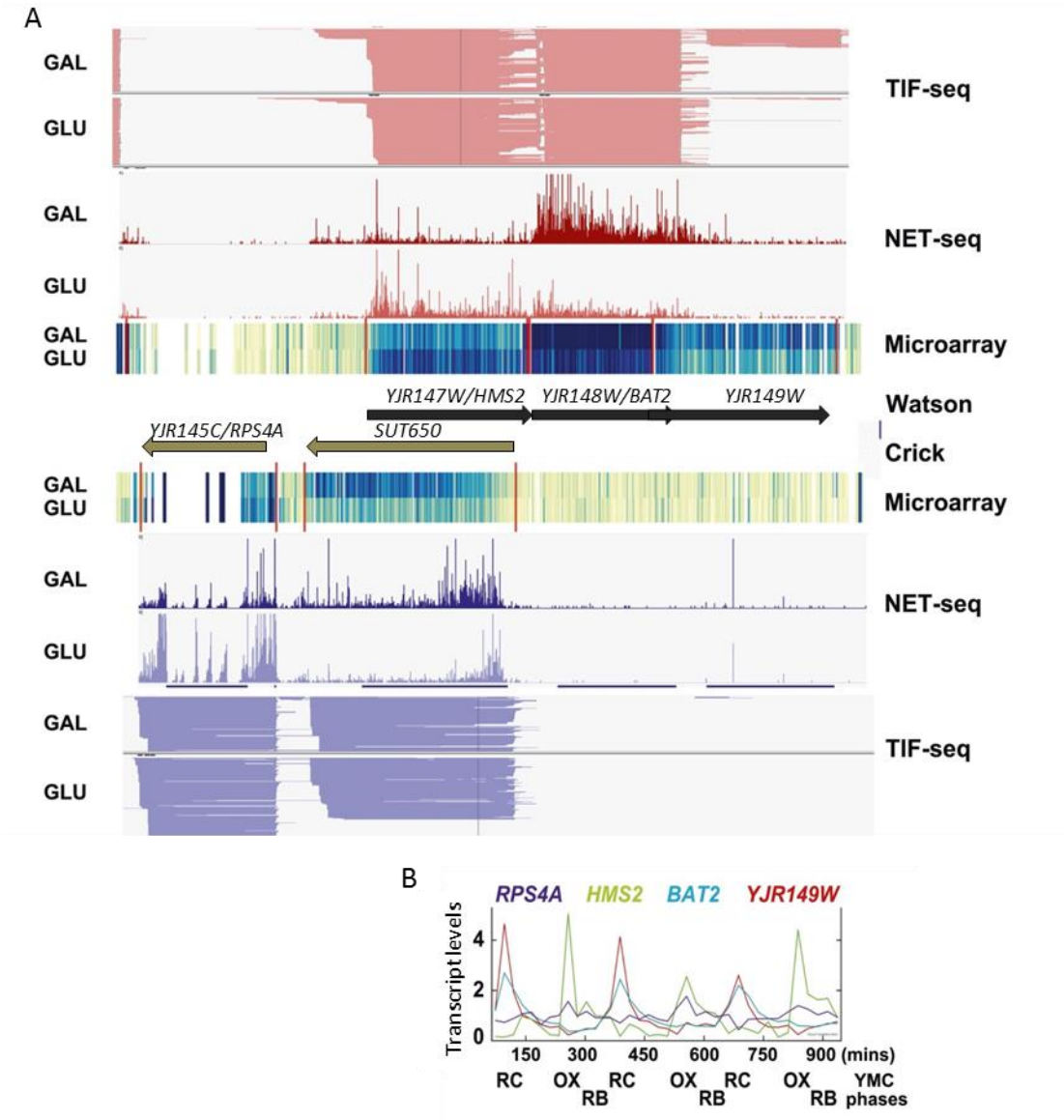


Figure 4.3. Expression changes at an individual locus. A) Changes in transcription (NET-seq), transcripts (microarray) (Nguyen et al., 2014) and transcript isoforms (TIF-seq) (Pelechano et al., 2013) at an individual locus upon a shift from glucose to growth in galactose for 3 h. Microarray scale: yellow to dark blue = low to high transcript levels; Each horizontal line in TIF-seq data represents an alternative transcript isoform. B) Transcript expression profiles of the genes from this locus within the YMC (Kudlicki et al., 2007).

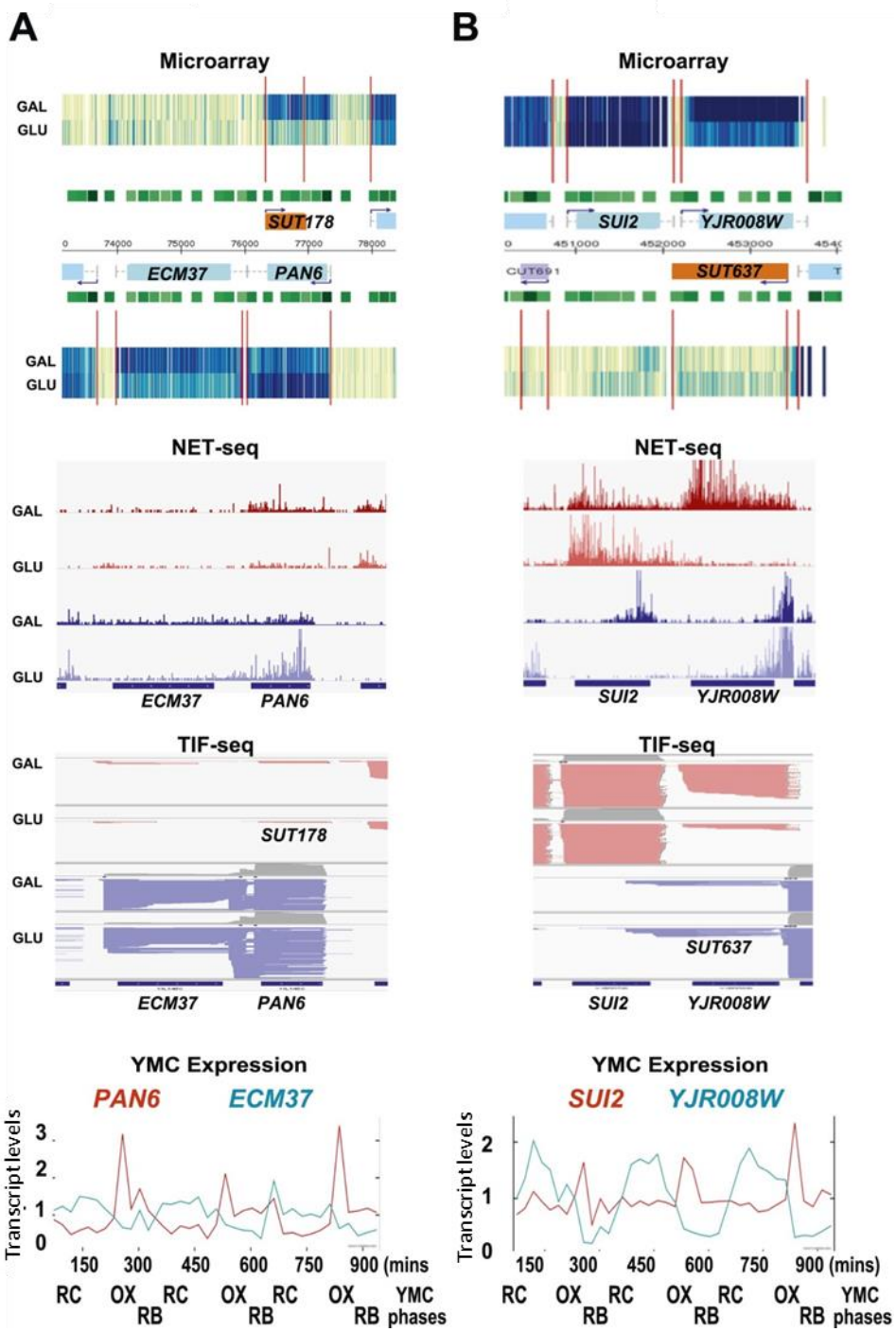


Figure 4.4A and B. Expression changes at individual loci. As Figure 4.3 at two more example loci. Key features include OX genes (*PAN6* and *SUI2*) upstream of RC genes (*ECM37* and *YJR008W*) and SAPs formed between *PAN6* and *SUT178,* between *SUT637* and *YJR008W* and between *SUI2* and an unannotated antisense TU.

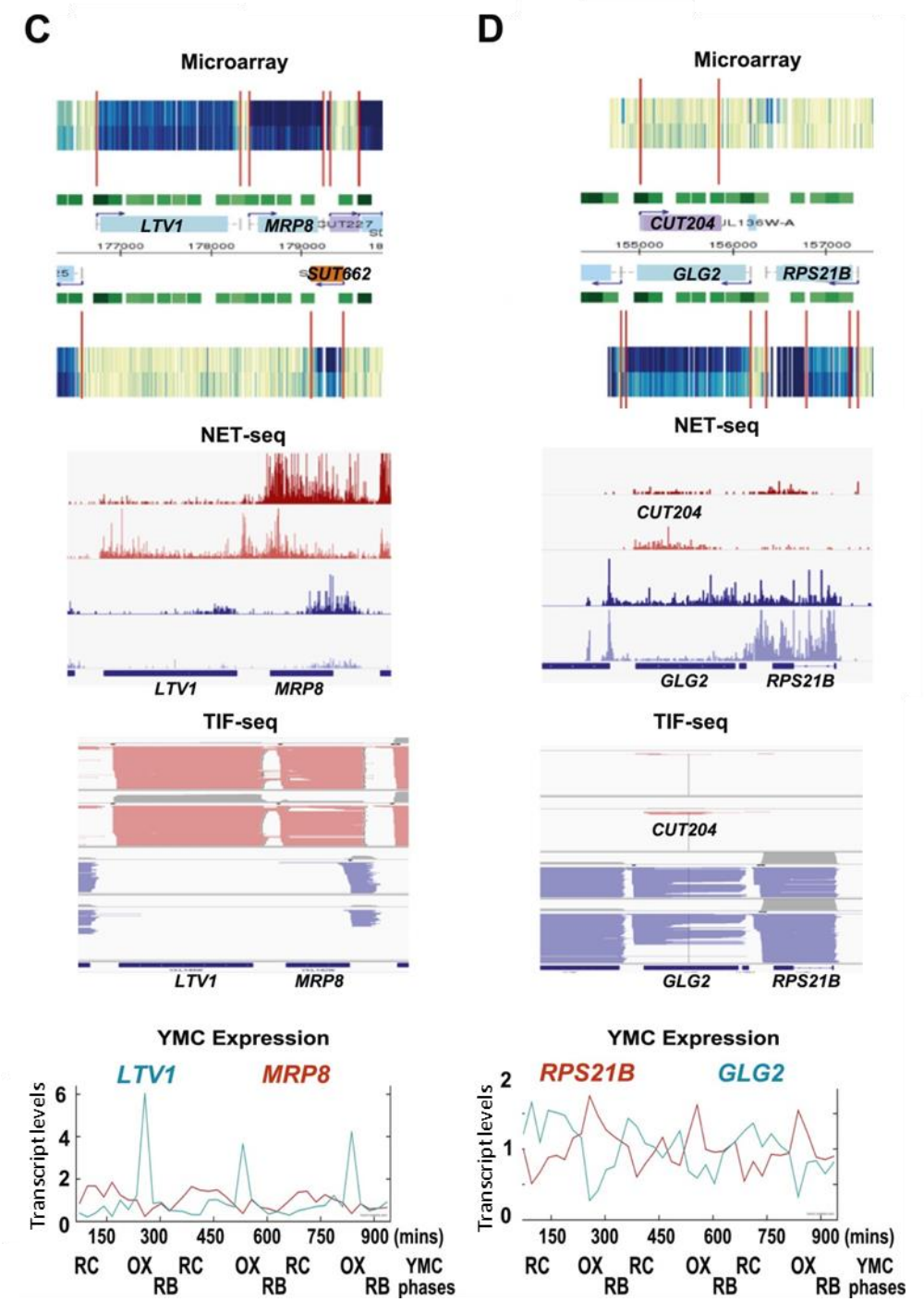


Figure 4.4C and D. Expression changes at individual loci. As Figure 4.3 at two more example loci. Key features include OX genes (*LTV1* and *RPS21B*) upstream of RC genes (*MRP8* and *GLG2*) and SAPs formed between *MRP8* and *SUT662* and between *GLG2* and *CUT204*.

4.2.3 Transcription patterns at the *HMS2-BAT2* locus

To understand the interplay between these clustered TUs, the locus consisting of the PCGs *RPS4A, HMS2, BAT2* and *YJR149W* was analysed (Figure 4.3). The divergent *HMS2*

and *RPS4A* genes show increased expression in glucose and in the OX phase of the YMC, while the tandem *BAT2* and *YJR149W* genes show increased expression in galactose and the RC phase of the YMC. *HMS2* transcription often continues into the promoter of the downstream *BAT2* TU and in glucose the most abundant *HMS2* transcript isoform ends just 43 bp upstream of the start of the most abundant *BAT2* isoform (Pelechano et al., 2013). The non-coding *SUT650* TU, whose most abundant transcript starts 410 bp inside the *HMS2* TU 3' end and ends 312 bp before the start of the *RPS4A* TU, forms an SAP with the *HMS2* TU. *SUT650* shows increased transcription and transcript levels in galactose.

4.2.4 *HMS2* and *SUT650* show temporally segregated expression

Most studies are carried out on exponentially growing bulk cultures. Analysis of transcription and transcripts in these cultures give the impression that all cells in the culture are expressing genes at similar levels. With high levels of transcription occurring on HMS2 and the antisense SUT650 TUs, a likely consequence of this would be for the transcribing polymerases to collide with each other, potentially inhibiting transcription (Hobson et al., 2012). Dual labelled RNA FISH was carried out by Tania Nguyen and Françoise Howe on cells grown in glucose to assess whether these TUs are contemporaneously expressed (Figure 4.5A). Fluorescent dots indicate the location of transcripts. Dots in the nucleus, which are often brighter than cytoplasmic dots, are generally thought to show a cluster of transcripts retained at the site of transcription and transcripts in the process of being transcribed. In multiple analyses there were no cells in which there was signal for both *SUT650* and *HMS2* at the site of transcription, suggesting their transcription is temporally segregated, which limits the possibility of polymerase collision. 35.5 ± 7.9 % of cells contained both transcripts. Where this was the case, generally the two transcripts were located in the cytoplasm, while there was a single dot in the nucleus for one of the transcripts. This probably represents cells that have recently switched from transcribing one TU to the other, with transcripts from

transcription of the first TU yet to have been completely degraded. 22.8 ± 10.5 % of cells lacked either transcript. The remainder of cells showed dots for only one transcript. RNA FISH was also carried out on cells switched from growth in glucose to growth for one hour in galactose. In galactose there were more cells expressing the *SUT650* transcript than the *HMS2* transcript, which was the opposite of that seen in glucose (Figure 4.5B). Thus the increase in expression of *SUT650* in galactose results from an increase in the proportion of cells expressing *SUT650* and a decrease in the proportion of cells expressing *HMS2.* This supports the idea that the effect of the switch to galactose is to increase the proportion of time cells spend in the RC phase of the YMC. It also suggests there may be co-regulation between *SUT650* and *BAT2* and reciprocal regulation between these TUs and *HMS2*.

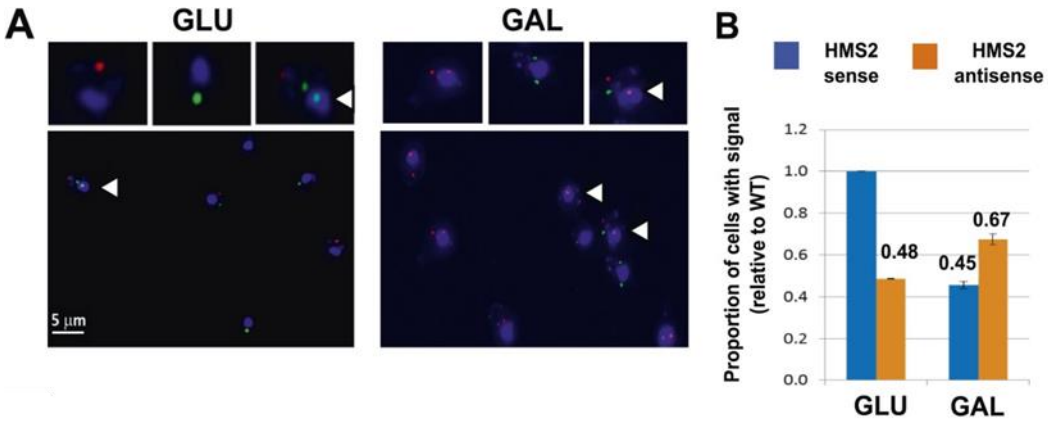


Figure 4.5. *HMS2* and *SUT650* show temporally segregated transcription within the YMC. A) Representative images from dual labelled RNA FISH showing the location and level of *HMS2* (cy5, green dots) and *SUT650* (cy3, red dots) transcripts within individual cells grown in glucose and after 1 h growth in galactose. Cells with nascent transcription events are marked with white arrows. B) Bar charts showing the proportion of cells with signal for each of the transcripts for each of the conditions. RNA FISH and quantification were carried out by Tania Nguyen and Françoise Howe. n = 2; error bars show the standard error of the mean.

4.2.5 *HMS2* transcription into the promoter of *BAT2* represses *BAT2* transcription

To test the nature of this reciprocal regulation, the *ADH1* terminator sequence was inserted into the *HMS2* gene (strain *HMS2:ADH1t*) (Figure 4.6A). This leads to early termination of *HMS2* transcription, separating it from transcription of *BAT2* (Figure 4.6E). Levels of the various transcripts were assayed by Northern blot in both glucose and 1 h in galactose for this strain and a wild type strain (Figure 4.6B-D), quantified in Figure 4.6F. The *HMS2:ADH1t* strain shows increased levels of the *BAT2* transcript in both glucose and galactose. This suggests transcription of *HMS2* into the promoter of *BAT2* inhibits *BAT2* transcription. *SUT650* transcription is also partially terminated by the *ADH1* terminator, leading to lower levels of the longer transcript. The early termination of *SUT650* means that its transcription is prevented from overlapping with the promoter of *HMS2*. The shorter *HMS2* transcript is more highly expressed in this strain, suggesting the longer *SUT650* transcript may inhibit the transcription of *HMS2*. However, this higher level may be due to increased stabilisation as a result of the change in RNA length and altered termination signal.

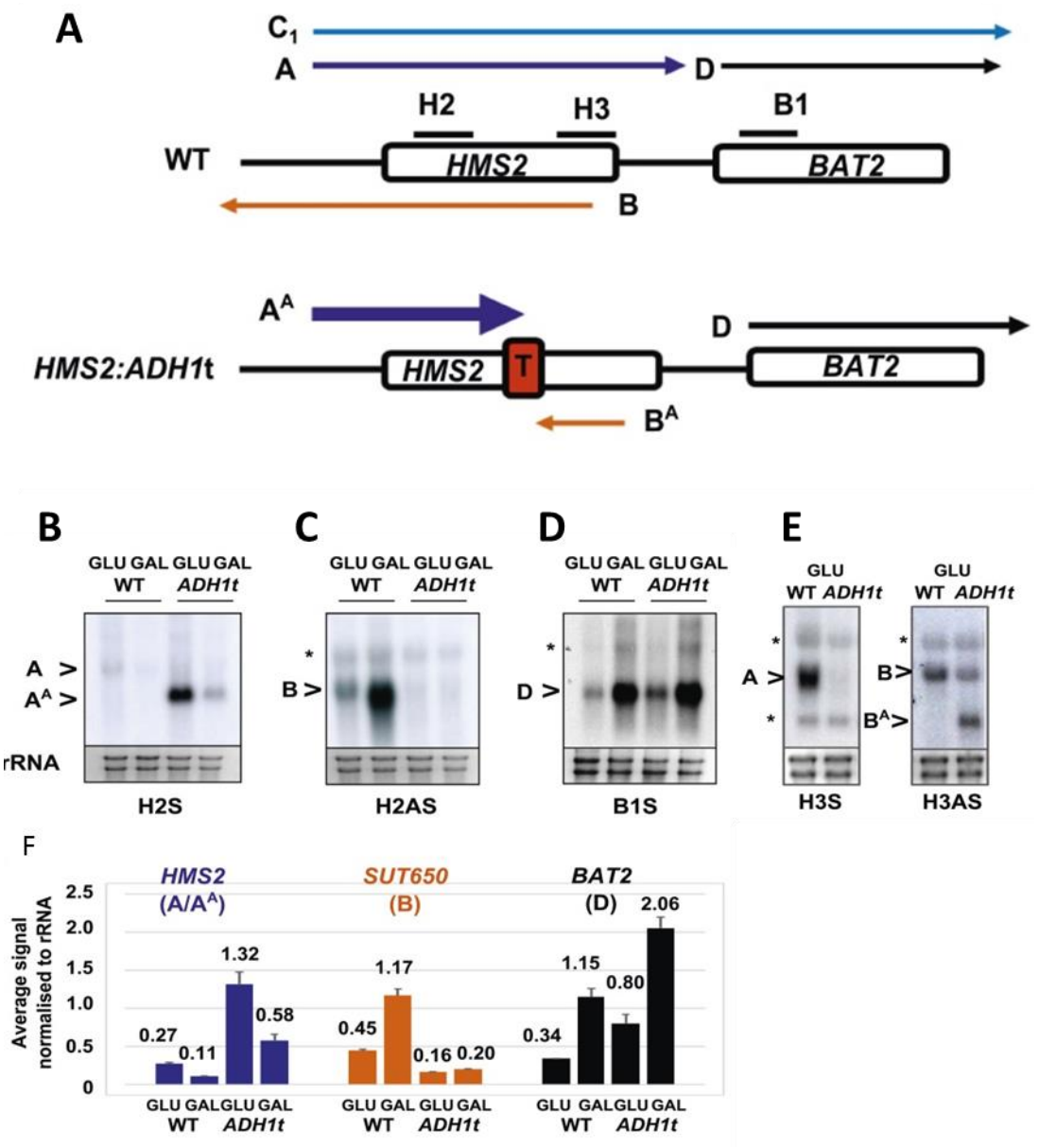


Figure 4.6. Transcriptional interference between overlapping TUs. A) Schematic of the genes and main transcripts in a wild type strain and after insertion of an the *ADH1* terminator sequence into the middle of *HMS2* to create strain *HMS2:ADH1t*. Increased arrow width is used to represent increased transcript levels. The positions of probes H2, H3 and B1 used for Northern blotting are marked. B-D) Northern blots showing the level of transcripts assayed with sense (B) and antisense (C) probes at position H2 and a sense probe at position B1 (D) in both glucose and after 1 h in galactose. E) Northern blot showing the level of sense and antisense transcripts at position H3 in glucose. F) Quantification of levels of transcripts in B, C and D. n= 2, error bars show the standard error of the mean. Northern blotting and quantification were carried out by Tania Nguyen.

4.2.6 A model for transcriptional regulation at the *HMS2-BAT2* locus

Based on these data, a two-state model for regulation of this locus by overlapping TUs was proposed. In the sense-dominant state, which occurs during the OX phase, an input signal leads to the activation of transcription at *HMS2*. Transcription of this TU overlaps with the bidirectional promoter of *SUT650-BAT2*, repressing transcription of these TUs. Transcription of polycistronic transcripts initiating from the *HMS2* promoter overlaps with the promoter of *YJR149W*, repressing *YJR149W* transcription. In the antisense-dominant state, which occurs during the RC phase, the input signal is lost leading to a decrease in *HMS2* transcription. This relieves the repression of the bidirectional promoter, enabling *SUT650* and *BAT2* transcription. *SUT650* transcription overlaps with the promoter of *HMS2* and *RPS4A*, repressing their transcription (Figure 4.7).

This model could also apply to other loci, such as those shown in Figure 4.4. For example, *SUI2* transcription in the OX phase overlaps with and may repress the bidirectional promoter of *YJR008W* and the *SUI2* antisense TU. A decrease in *SUI2* transcription in the reductive phases may relieve this repression, enabling *YJR008W* and *SUI2* antisense transcription. *SUI2* antisense transcription may repress *SUI2* transcription.

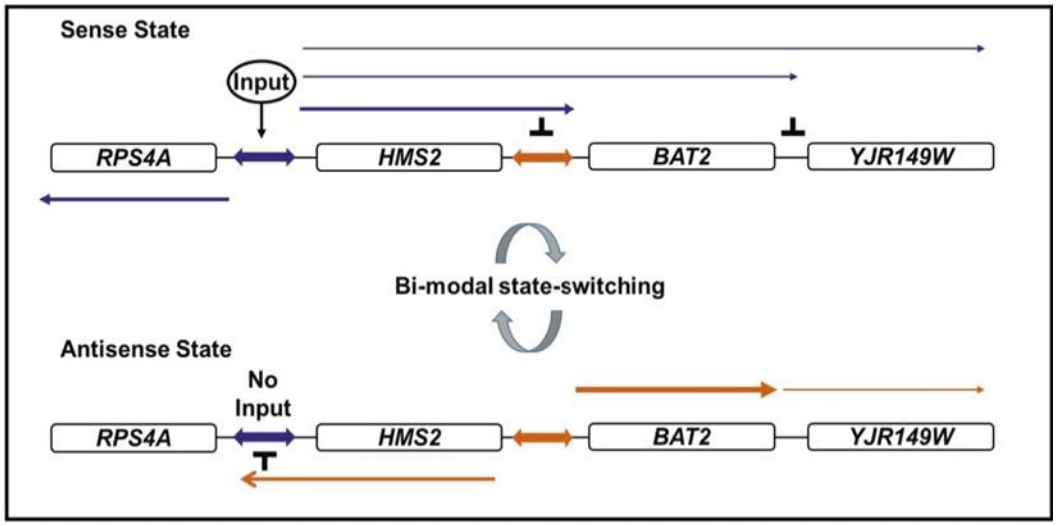


Figure 4.7. Model for transcriptional regulation by overlapping TUs at the *HMS2-BAT2* locus. Cells cycle between a sense-dominant state where *HMS2* transcription represses *HMS2* antisense (*SUT650*), *BAT2* and *YJR149W* transcription and an antisense-dominant state where *SUT650* transcription from the *BAT2-SUT650* bidirectional promoter represses *HMS2* and *RPS4A* transcription.

4.2.7 The dynamics of transcriptional changes

To analyse the dynamics of the transcriptional changes seen upon the switch from glucose to galactose, the standard NET-seq protocol was applied at several earlier time points (5 min (GAL5), 15 min (GAL15) and 60 min (GAL60)) after a switch to galactose. Two repeats were obtained for the 15 min and 60 min time points. All data were analysed alongside the data obtained from growth in YPD (GLU) (two repeats) and after a switch to growth in galactose for 180 min (GAL180). To switch the growth media, cells were pelleted from an exponentially growing culture in YPD by centrifugation at 5000 rpm for 3 min. Then the cells were resuspended in YPG, which had been pre-warmed to 30 ^{o}C. To test whether the centrifugation step affected transcription, a control was carried out in which cells underwent centrifugation and then were resuspended and grown for 15 min in the original YPD media (reGLU).

The reads mapping to each PCG TU were counted and an rlog transformation was applied across all data (Love et al., 2014). Figure 4.8 shows a heatmap of the Euclidean

distance between each time point. The GLU and reGLU data cluster closely, suggesting that if any transcriptional changes occur as a result of the centrifugation step, these are generally much smaller than the large changes occurring upon a switch to galactose. All galactose time points cluster together, showing that these samples show greater similarity to each other than to transcription in glucose. This suggests that many of the transcriptional changes observed at later time points, have occurred or started to occur by 5 minutes. Out of all the time points, the GAL15 replicates show the most dissimilarity to the GLU replicates. This suggest the transcriptional changes occurring at 5 minutes generally become more enhanced by 15 minutes but by 60 minutes there is some recovery back to the pattern of transcription seen in glucose. This is also represented by principal component analysis. 85 % of the variance between samples can be explained by principal component 1 (PC1) with the galactose samples separated from the glucose samples by this component with less variance between GLU and GAL5, GAL60 and GAL180 samples than between GLU and GAL15. The separation of the galactose samples along PC2 shows there is some additional variance between these samples.

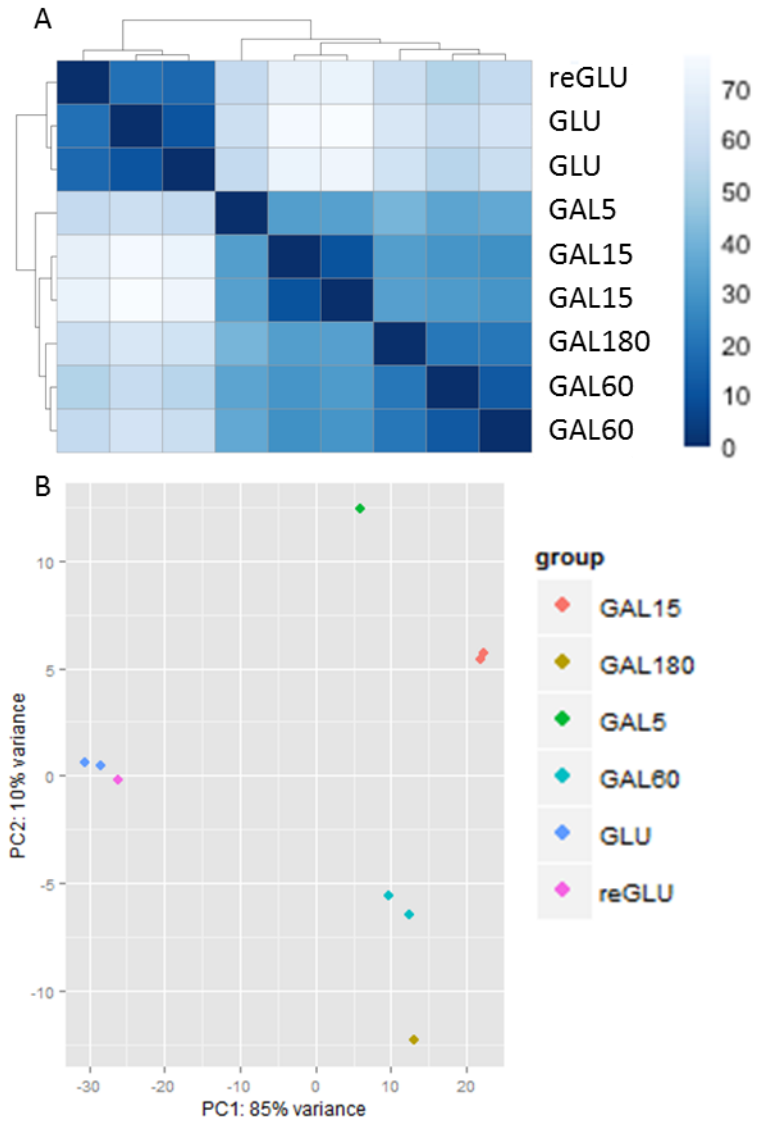


Figure 4.8. Similarity between genome-wide transcription at different time points after a switch from glucose- to galactose-containing media. A) Heatmap showing the Euclidean distance between all PCG TU NET-seq counts after growth in YPD (GLU), centrifugation and resuspension in the same media for 15 min (reGLU), resuspension in YPG for 5 min (GAL5), 15 min (GAL15), 60 min (GAL60), 180 min (GAL180). B) Principal component analysis showing the similarity between replicates and time points.

Differential expression analysis was carried out on the counts mapping to PCG TUs looking for those showing a significant (p value adjusted for multiple testing < 0.05) change between two conditions rather than a two-fold change. Comparing GLU to reGLU shows significant transcriptional changes at 155 genes, with an increase in RC gene transcription (Figure 4.9C). This is possibly due to the centrifugation step inducing a

mild stress response in the cells. Although these changes are small compared to the changes seen for a switch to galactose, it is important to note that such a step, common to many experimental procedures, can have an impact on transcription.

Comparing glucose to each galactose time point, the genes showing an increase are enriched for RB phase genes from subphases RB2 and RB3 and RC phase from subphases RC1 and RC2. The genes showing a decrease are enriched for OX phase genes and subphase RB1 genes. It is thought that not all cells go through a cell division cycle with each YMC. However, cells in glucose divide more rapidly than those which have been switched to growth in galactose, suggesting that a higher proportion of cells go through a cell division cycle with each YMC. In glucose, the cells spend longer in the OX phase expressing protein synthesis genes. Consequently they are more likely to accumulate sufficient quantities of new proteins needed to form a new cell and to pass through the checkpoints required for DNA replication and mitosis and hence spend longer in the RB1 phase in which genes involved in DNA replication are expressed. The 3343 genes showing a transcriptional change after 5 minutes in galactose highlights the speed with which transcriptional changes occur (Figure 4.9A and B). This suggests cells are able to switch rapidly from the OX phase to the RC phase of the YMC.

Next, comparisons were made between each galactose time point and its following time point. From GAL5 to GAL15 the genes with increased transcription are enriched for RB2, RB3, RC1 and RC2 phase genes and those with decreased transcription are enriched for OX phase and RB1 phase genes. This continues the trend seen from GLU to GAL5 suggesting that, while some cells have switched from the OX phase to the RC phase by 5 minutes, a larger fraction have switched by 15 minutes (Figure 4.9D). This trend reverses from GAL15 to GAL60 with the genes showing an increase in transcription enriched for OX phase and RB1 phase genes and those showing a decrease enriched for RB3, RC1 and RC2 phase genes (Figure 4.9E). An explanation for this is that the cells

which have switched to the RC phase by 15 minutes have now started progressing through the YMC with some switching back into the OX phase of the YMC. By 60 minutes there is a small but significant increase in the level of transcription of galactose-specific genes (including *GAL1*, *2*, *3*, *7* and *10*) (Figure 4.10). These cells may be able to metabolise galactose, leading to release of more energy, enabling them to spend longer in the OX phase. Only 267 genes show changes in transcription from GAL60 to GAL180, however, the trends in enrichment for phases of the YMC for these genes reverse from that seen from GAL15 to GAL60 (Figure 4.9F). This is surprising considering the much higher levels of transcription of the galactose-specific genes at GAL180 compared to GAL60 (Figure 4.10). The ability to metabolise galactose should release more energy, leading to longer oxidative phases and a higher fraction of cells expressing OX phase genes. Upon the switch to galactose, cells switch to the RC phase, which may lead to some synchronisation of the cells in the YMC. This synchronisation may be maintained, despite gradually being lost, so that at 60 minutes there is an enrichment of cells synchronised in the OX phase and at 180 minutes these synchronised cells may have progressed back into the RC phase. Higher temporal resolution profiling is required to assess this hypothesis. This increase in transcription of RC phase genes from GLU to GAL5, continuing increase from GAL5 to GAL15, decrease from GAL15 to GAL60 followed by small increase from GAL60 to GAL180 is exemplified by *YAL061W* (Figure 4.11) and *BAT2* (Figure 4.12). The reverse for OX phase genes is exemplified by *HMS2* (Figure 4.12).

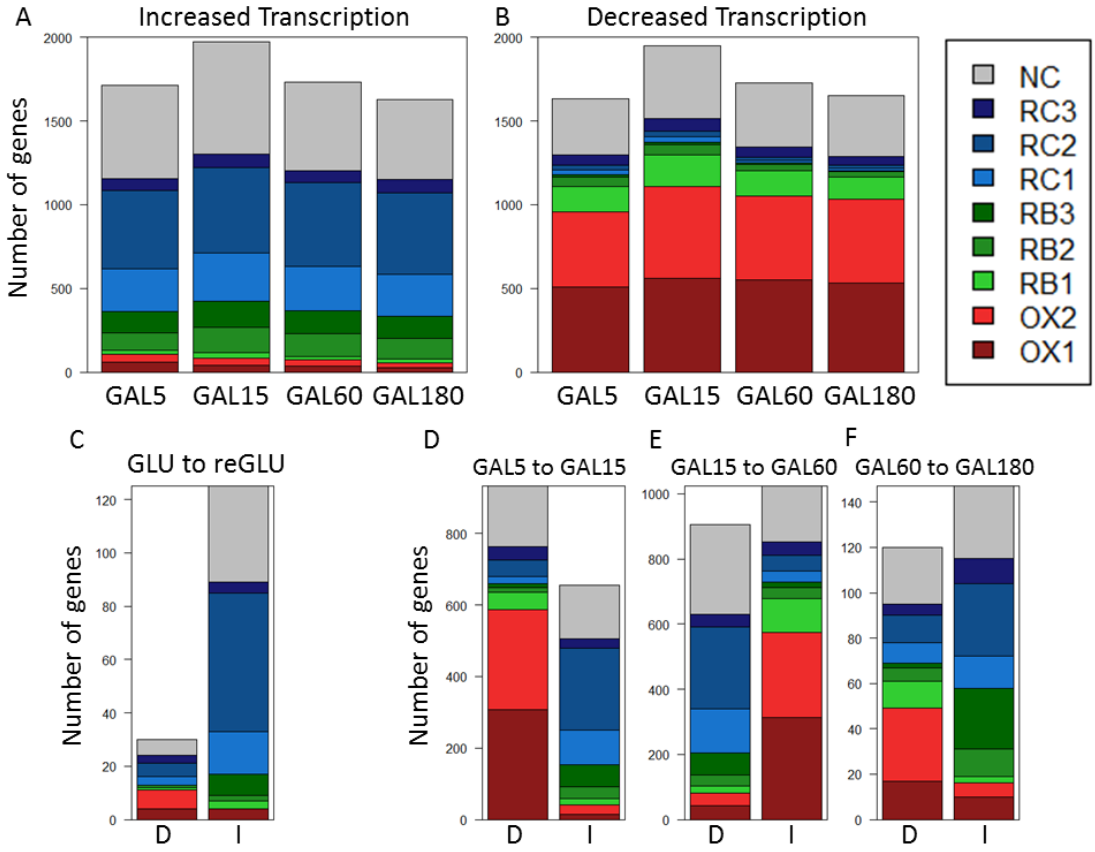


Figure 4.9. Numbers of PCG TUs showing transcriptional changes between different conditions and their connections to the YMC. A and B) Bar charts showing the number of PCG TUs showing an increase (A) or decrease (B) (p value < 0.05) in transcription compared to growth in glucose at each of the 4 time points. C-F) Bar charts showing the number of PCG TUs showing a decrease (D) or an increase (I) (p value < 0.05) in transcription between two time points. PCG TUs are colour-coded according to whether they show cycling transcript levels in the YMC and the phase in which they show maximum transcript levels. OX (oxidative), RB (reductive building), RC (reductive charging), NC (non-cycling).

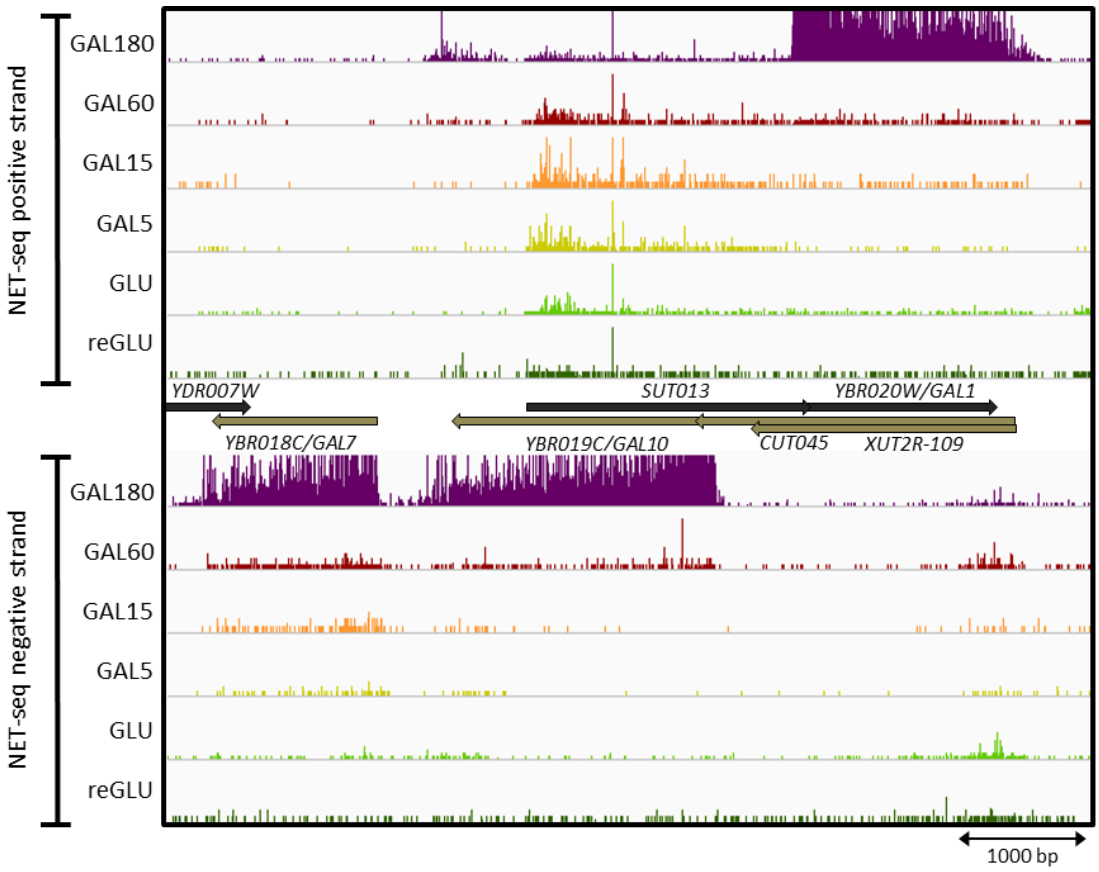


Figure 4.10. Dynamics of the transcriptional changes occurring at the locus containing *GAL1, GAL10* and *GAL7* after a switch from glucose- to galactose-containing media. Tracks are scaled relative to size factors determined by the DESeq2 algorithm. The same scales are used for Figures 4.10-4.12 and 4.15-4.16. Only one replicate is shown for each track.

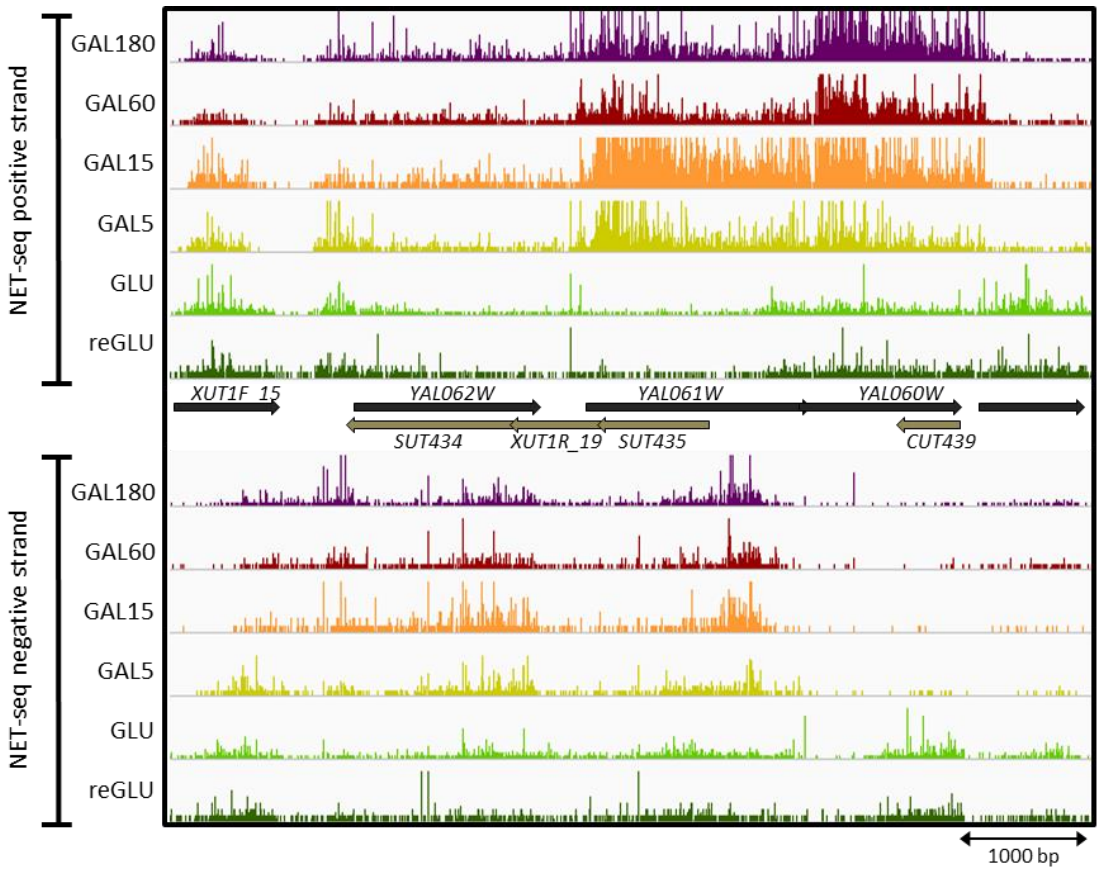


Figure 4.11. Dynamics of the transcriptional changes occurring at the *YAL061W* locus after a switch from glucose- to galactose-containing media.

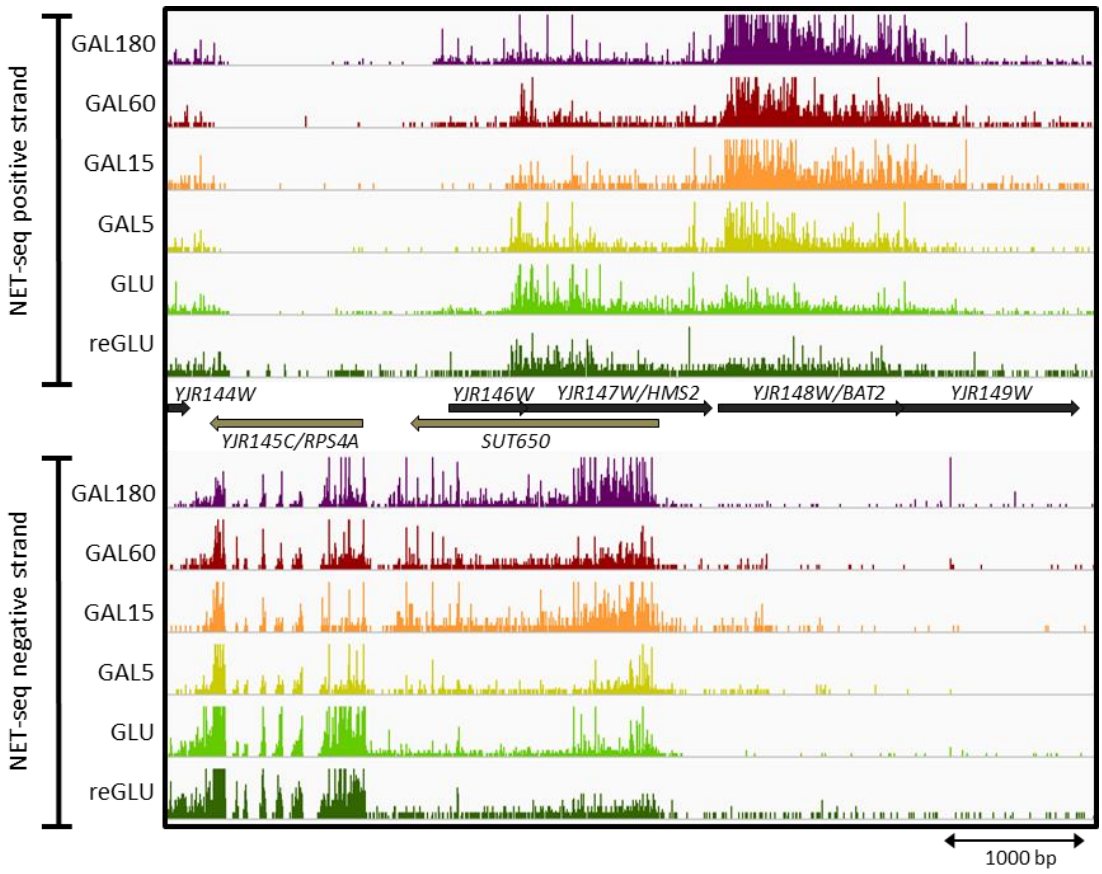


Figure 4.12. Dynamics of the transcriptional changes occurring at the *HMS2-BAT2* locus after a switch from glucose- to galactose-containing media

How genes that show transcriptional changes from GLU to GAL15 change from GAL15 to GAL60 were then analysed. The general trend is that genes that increase in transcription from GLU to GAL15 decrease or show no significant change from GAL15 to GAL60 and are enriched for RC, RB2 and RB3 phase genes, while those that decrease, increase or show no significant change and are enriched for OX and RB1 phase genes as expected (Figure 4.13).

How genes that show transcriptional changes from GLU to GAL5 change from GAL5 to GAL15 were then analysed. The general trend is that genes that increase in transcription from GLU to GAL5 continue to increase or show no significant change from GAL5 to GAL15, and are enriched for OX phase genes, while those that decrease continue to

decrease or show no significant change, and are enriched for RC phase genes (Figure 4.14). There is, however, a group of 221 genes that increase in transcription from GLU to GAL5 before decreasing from GAL5 to GAL15 (for example see Figure 4.15, *YOL016C* and *YOL014W*). Interestingly, several but not all of the genes shown to be particularly Paf1-enriched in Chapter 3 are included in this group such as *YOL016C/CMK2*, *YER130C/COM2* and *YER138C/TPO2.* The link between Paf1 and the NPC component Mex67 demonstrated in Chapter 3 may be involved in localising these TUs close to the NPC, enabling their fast activation. Gene ontology analysis of this group of 221 genes using YeastMine (http://yeastmine.yeastgenome.org/) shows it is enriched for genes associated with a response to a stimulus (p value = 0.0061). These are probably required to be fast activating in order for the rapid upregulation of their corresponding protein, ensuring a quick response to the stress before cell damage occurs. This suggests they may be primed for fast activation. The reason for their subsequent rapid deactivation may be because only small quantities of the protein are required to mediate the appropriate response. The group is also enriched for transcription factors (p value = 0.0462) such as Hap4, a global repressor of OX phase gene transcription (Rao and Pellegrini, 2011), and Rtg1, involved in RC phase gene activation (Ruiz-Roig et al., 2012). These may be required for the initial and rapid switch in transcription from that associated with the OX phase to the RC phase and may no longer be required at high levels once this switch has occurred.

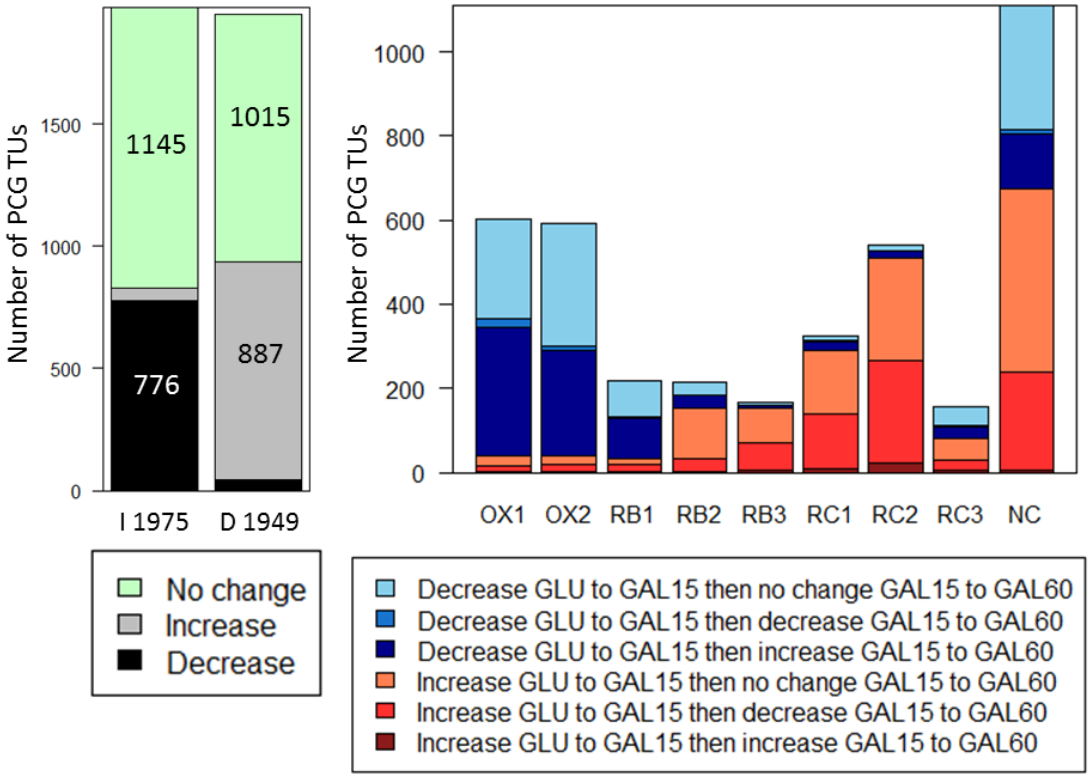


Figure 4.13. Numbers of PCG TUs that show transcriptional changes from GLU to GAL15 showing transcriptional changes from GAL15 to GAL60 and their connections to the YMC. A) Bar chart showing the number of PCG TUs showing an increase (I) or decrease (D) (p value < 0.05) in transcription from GAL15 to GAL60 for those TUs showing a change from GLU to GAL15. B) Bar chart showing whether these PCG TUs, colour-coded according to these changes, show cycling expression in the YMC and the phase in which they show maximum expression.

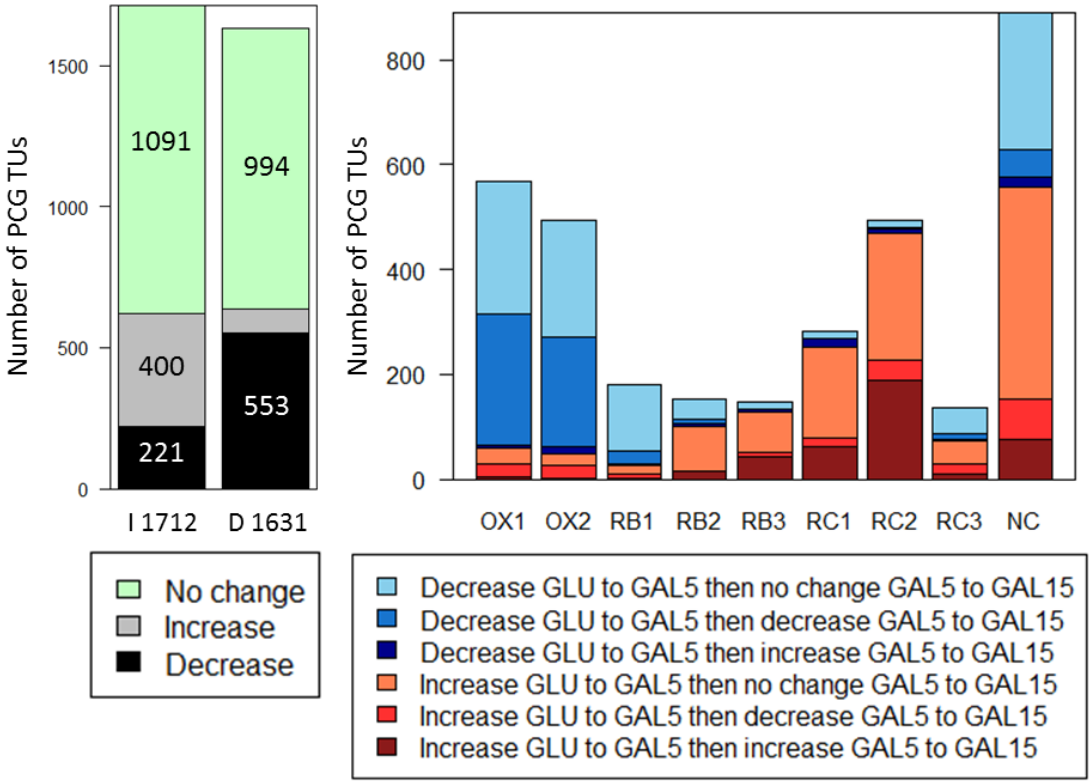


Figure 4.14. Numbers of PCG TUs that show transcriptional changes from GLU to GAL5 showing transcriptional changes from GAL5 to GAL15 and their connections to the YMC. A) Bar chart showing the number of PCG TUs showing an increase (I) or decrease (D) (p value < 0.05) in transcription from GAL5 to GAL15 for those TUs showing a change from GLU to GAL5. B) Bar chart showing whether these PCG TUs, colour-coded according to these changes, show cycling expression in the YMC and the phase in which they show maximum expression.

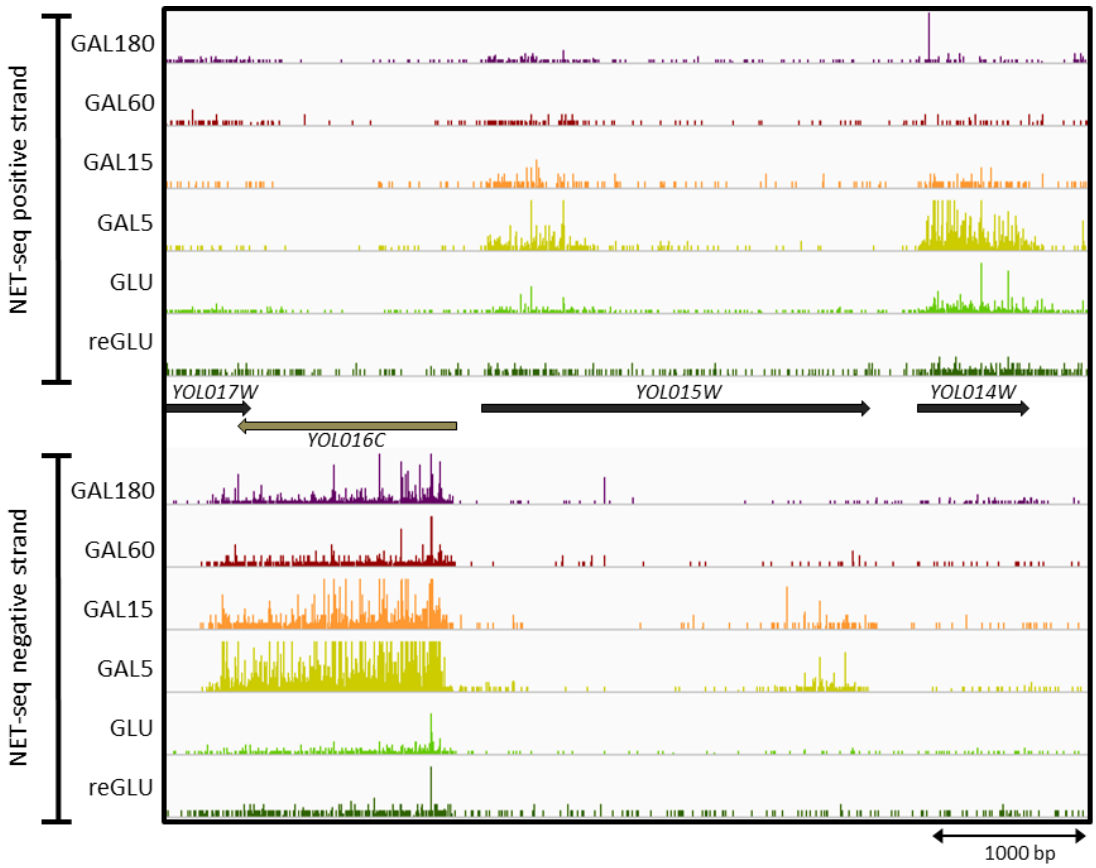


Figure 4.15. TUs showing rapid transcriptional activation from GLU to GAL5 and deactivation from GAL5 to GAL15

In addition, some non-coding TUs show the same pattern of transcriptional changes including an unannotated non-coding TU antisense to *GAL3* (Figure 4.16). This is of particular interest given that Gal3 forms a complex with Gal80 relieving the repressive action of Gal80 on Gal4 bound to the promoters of *GAL* genes, leading to their transcriptional activation. This rapid burst of transcription followed by transcriptional repression may play a role in activating *GAL3* transcription, possibly by disrupting nucleosome positioning or transcription repressor binding at the *GAL3* promoter. The high temporal resolution profiling during this glucose to galactose switch used here also reveals additional non-coding TUs, suggesting additional regulatory mechanisms involved in transcriptional regulation at particular loci. For example, at the *HMS2-BAT2* locus there is a TU starting upstream of the *SUT650* TU inside the 5' end of *BAT2* reading

into the *SUT650* promoter (Figure 4.12). Transcription at this TU does not occur in GLU but does in galactose, particularly at GAL5 and GAL15, and may play a role in disrupting the binding of repressors at the *SUT650-BAT2* bidirectional promoter, enabling its activation. Understanding the role of this non-coding transcription will require further manipulation of the loci at which they occur.

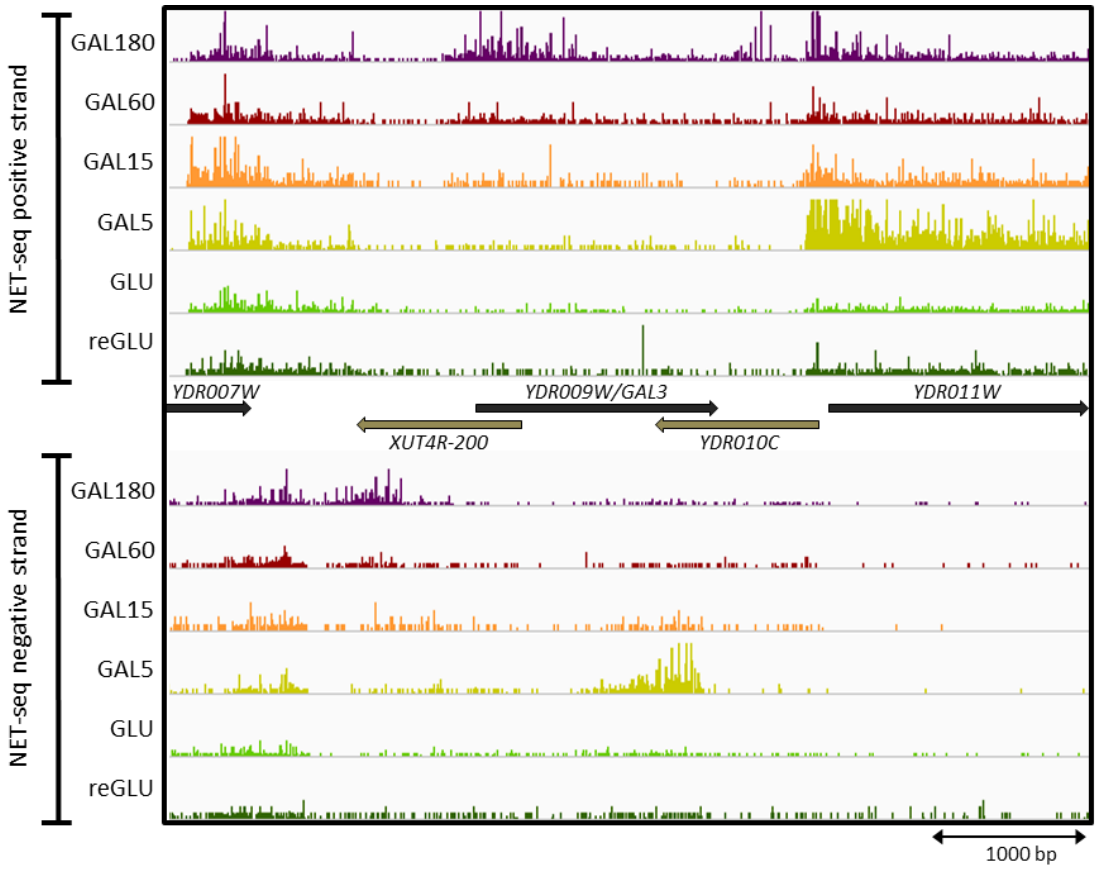


Figure 4.16. *GAL3* antisense non-coding TU showing rapid transcriptional activation from GLU to GAL5 and deactivation from GAL5 to GAL15

4.3 Discussion

Here, measurement of transcription using NET-seq of a population of cells shows that widespread and rapid changes occur at both PCG and non-coding TUs upon shifting cells from growth in glucose- to galactose-containing media. The TUs showing changes are enriched for TUs with periodic expression in the YMC. The model proposed to explain this is that the switch to a less favourable environment induces a stress response causing cells to transition rapidly from the OX and RB1 phases to the RB2, RB3 and RC phases of the YMC. These cells then spend a longer proportion of time in the RB2, RB3 and RC phases of the YMC than they did in glucose, an effect that continues for at least 3 hours. Population measurements at time points after the switch to galactose are carried out on cells in which a higher proportion are in these phases and therefore show higher levels of transcription of TUs expressed in these phases and lower levels of transcription of TUs expressed in the OX and RB1 phases.

The observation that altering conditions can have such a dramatic impact on transcription in a short time period through inducing the ESR and altering the YMC has far reaching implications for many biological experiments. The use of strains in which the function of a particular gene has been disrupted can result in slower growth. This results in an increase in the proportion of cells in the RC phase of the YMC and corresponding alterations in RNA expression levels at the population level (Brauer et al., 2008; O'Duibhir et al., 2014). Temperature sensitive mutants or the use of Degron-tagged proteins have been used to study genes whose deletion is lethal (Dohmen et al., 1994). Alternatively gene function can be tested by expressing genes from inducible promoters or tethering proteins to the ribosome in a technique called anchors away (Haruki et al., 2008). These techniques require a change in conditions, which, in addition to the loss of the functional protein, has the potential to initiate an ESR and lead to cells transitioning to the RC phase of the YMC. Initial analysis of transcriptional changes

which occur upon a shift from glucose- to galactose-containing media in a mutant (*Δxrn1*) strain shows the proportion of cells that have transitioned to the RC phase by 15 min and the extent to which they show some recovery by 60 min is altered compared to a wild type strain (data not shown). This suggests it may not always be possible to control for these effects. The issues highlighted here may mean that effects observed as a result of the loss of protein may be an indirect effect rather than a direct effect attributable to the function of the protein. Harvesting of cells and downstream processing steps may cause additional problems. Common techniques such as ChIP and RNA FISH involve a series of steps including centrifugation, resuspension of cells in a buffer such as PBS and addition of a cross-linking reagent. These all have the potential to cause cells to transition from the OX phase to the RC phase, resulting in genome-wide transcriptional changes before the cells are completely fixed. Attempts to compare results from different techniques such as NET-seq data to ChIP-seq data may be confounded by differences in the timing and procedures used to generate samples.

At a genome-wide level, in multiple species, it has been shown that functionally related genes are often clustered together and show similar patterns of expression. This co-expression is thought to result from chromatin remodelling leading to increased or decreased chromatin compaction spanning across multiple clustered genes (Batada et al., 2007). Not all genes within the same region show this co-expression. As shown here, there are many examples of genes expressed in both the OX and RC phases of the YMC situated in close proximity and showing opposing transcriptional changes upon a switch to galactose. This raises the question as to why this has not been selected against with the opening up of chromatin required to express an OX phase gene increasing the potential of a nearby RC phase gene being expressed out of phase and vice versa with potentially detrimental effects.

Here analysis of the *HMS2-BAT2* locus shows that there are two mutually exclusive sense and antisense transcriptional states at this locus. In one state, associated with the OX phase, *HMS2* is transcribed, while in the other, associated with the RC phase, *BAT2* and the *HMS2* antisense TU, *SUT650*, is transcribed. Regulation of this is to some extent dependent upon the binding of transcription factors as shown by factor binding sites mapping to both the *HMS2* promoter and *BAT2-SUT650* bidirectional promoter and deletion of these factors resulting in altered expression patterns (Tania Nguyen, data not shown). Nevertheless the orientation of these TUs is such that transcription of *HMS2* into the bidirectional promoter and transcription of *SUT650* into the *HMS2* promoter both lead to repression of the promoter. Therefore clustering of genes with reciprocal transcription patterns can be advantageous by helping to provide an additional layer of regulation through overlapping TUs. This helps reinforce the repression of particular TUs, preventing their expression out of phase. Consistent with this, antisense transcription has been shown to have little effect on highly transcribed sense TUs but results in more repression of lowly transcribed sense TUs (Xu et al., 2011).

The mechanism by which transcription overlapping with promoters results in repression is probably a result of a combination of activities characterised at other loci. These include the procession of the Pol II through promoter DNA disrupting transcription factor binding and PIC recruitment and the co-transcriptional deposition of repressive histone modifications over the promoter (Hongay et al., 2006; Houseley et al., 2008; Martens et al., 2004). Supporting this, increased levels of H3K36me3 are present over the promoter of *HMS2* in galactose (Tania Nguyen, data not shown). Gene looping may also play a role. The promoter and terminator of a TU associate with each other in a transcription-dependent manner (El Kaderi et al., 2009; O'Sullivan et al., 2004). The gene loop formed may be mutually exclusive to the formation of other transcriptional loops in the vicinity. For example when the most common *SUT650* TU is

transcribed the *HMS2* promoter would be inside the loop, while the terminator would be outside the loop, which may limit their ability to interact (Figure 4.17).

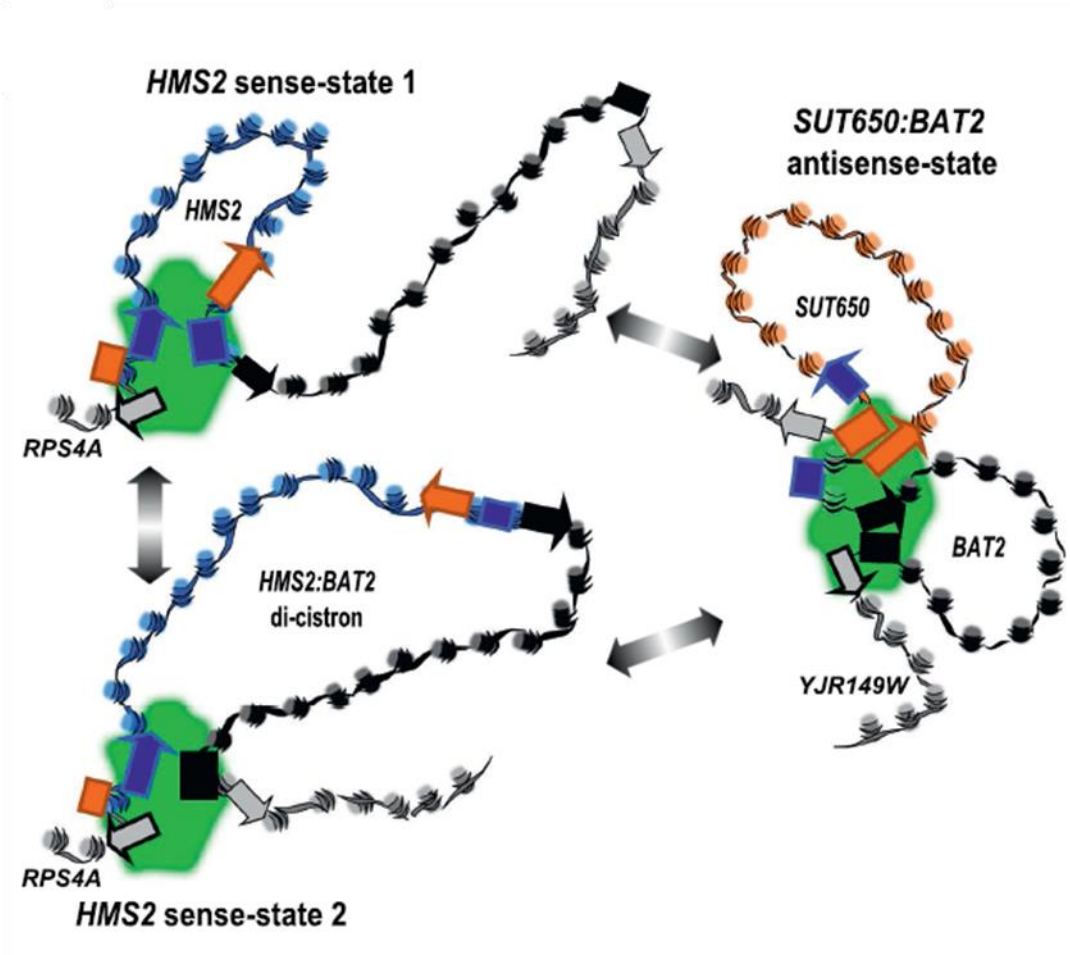


Figure 4.17. Gene loops may prevent other overlapping gene loops from forming. Schematic illustrating how gene loops could form in the sense-state between the *HMS2* promoter (blue arrow) and *HMS2* terminator (blue box) to form the most abundant *HMS2* TU or the *BAT2* terminator (black box) to from a di-cistronic TU. These loops would be mutually exclusive to the formation of gene loops between the promoter and terminator of *SUT650* (orange arrow and box) and the promoter and terminator of *BAT2* (black arrow and box) in the anti-sense state. Green shading indicates regions within which promoters and terminators need to come together for transcription of the TU to occur. Figure adapted from Nguyen et al., (2014).

The switch from the sense to the antisense state can occur within 5 min of the switch to galactose. Rapid activation and then inactivation of several transcription factors occurs in this time frame, suggesting they play a role in mediating this rapid switch. Whether there is also a role for some non-coding TUs as well remains to be established, although

NET-seq reveals many previously unidentified TUs, which could be functional. The repurposing of CRISPR technology has led to the development of a way blocking transcription elongation strand-specifically (Qi et al., 2013) rather than the insertion of terminator sequences used here, which disrupts both sense and antisense transcription. This should be an invaluable tool for further investigations into these non-coding TUs and antisense transcription in general.

Chapter 5.

Assessing the ability of RNA to act by binding double-stranded DNA

5 Assessing the ability of RNA to act by binding double-stranded DNA

5.1 Introduction

5.1.1 RNAs and chromatin

In eukaryotic cells, ncRNAs are enriched in the nucleus and associate with chromatin (Quinodoz and Guttman, 2014; Rodriguez-Campos and Azorin, 2007). Numerous studies suggest that these RNAs play functions in regulating processes occurring within chromatin such as organisation of chromatin structure, transcription, DNA replication and DNA repair (Bernstein and Allis, 2005; Ge and Lin, 2014; Liu and Lu, 2012; Quinodoz and Guttman, 2014; Vance and Ponting, 2014). The precise mechanism through which these RNAs mediate their function is usually poorly understood.

5.1.2 Non-coding transcript function in *S. cerevisiae*

Given the typically short half-lives of many non-coding transcripts in *S. cerevisiae*, a functional role for these transcripts is often overlooked. Regulatory effects of this non-coding transcription are usually attributed to the process of transcription rather than a role for the transcripts produced (Castelnuovo and Stutz, 2015), especially as these effects are usually only observed at regions in close proximity to the non-coding TU (see Chapter 4). Due to their interlinked nature, however, it is difficult to exclude a role for the ncRNA; affecting ncRNA transcription also affects ncRNA production.

Several examples give support to a role for the non-coding transcript. At the *GAL1-GAL10* locus a ncRNA (*GAL10ncRNA,* also known as *SUT013*) is transcribed antisense to *GAL10*, initiating from a promoter towards the 3' end of *GAL10* and transcribing across the bi-directional *GAL1-GAL10* promoter (see Chapter 4, Figure 4.10). In glucose, the transcription of *GAL10ncRNA* is proposed to lead to the co-transcriptional deposition of histone methylation across the *GAL1-GAL10* promoter, recruitment of Rpd3S and histone de-acetylation (Houseley et al., 2008; Pinskaya et al., 2009). This leads to an

additive transcriptional repression effect alongside well-characterised carbon source responsive protein transcription factors, which bind to this bidirectional promoter. An additional role has since been proposed for the *GAL10ncRNA*. Deletion of *XRN1*, the decapping factor *DCP2* or the RNA helicase *DBP2* leads to *GAL10ncRNA*-dependent faster induction of *GAL1* and *GAL10* upon a switch from glucose to galactose media. It is proposed that the ncRNA promotes the transient removal of transcriptional repressors, possibly by interacting with the promoter DNA, through an unknown mechanism. Helicase activity or RNA degradation acts to remove the *GAL10ncRNA* enabling the repressors to re-associate with DNA (Cloutier et al., 2013). Deletion of *XRN1* has also been shown to stabilise a transcript antisense to *TIR1* and cause *TIR1* transcription repression in an antisense transcription-dependent manner (van Dijk et al., 2011). Similarly, at *PHO84,* deletion of *RRP6* leads to stabilisation of the *PHO84* anti-sense transcript and also results in repression of *PHO84* sense transcription (Camblong et al., 2007).

Such studies using decay factor mutants do not, however, definitively demonstrate a role for the ncRNA rather than non-coding transcription in transcriptional regulation. This is because these factors also affect transcription as well as RNA decay, with Rrp6 involved in early termination(Castelnuovo et al., 2013; Fox et al., 2015) and Xrn1 and Dcp2 found within the nucleus, bound to chromatin and shown to influence transcription initiation and elongation and possibly play a role in co-transcriptional RNA decay (Cloutier et al., 2013; Haimovich et al., 2013).

A more effective demonstration of non-coding transcript function comes from studies in which a non-coding transcript expressed from an ectopic location is able to cause a decrease in levels of transcripts from the endogenous locus *in trans* (Berretta et al., 2008; Camblong et al., 2009; Wang et al., 2014). Whether the non-coding transcript travels from its ectopic location to the endogenous locus to mediate this function has not

been established in these studies. Some suggest these *in trans* effects result from a post-transcriptional effect such as base pairing between the non-coding antisense transcript and the affected sense transcript, leading to transcript decay (Sinturel et al., 2015). In addition, cases where an *in trans* effect for a ncRNA cannot be demonstrated (Camblong et al., 2009; Houseley et al., 2008) does not exclude a role for the ncRNA. For example, the RNA may require transcription in close proximity to its site of activity in order to locate this site.

5.1.3 Non-coding transcript function in other organisms

In many eukaryotic cell types, small non-coding RNAs such as microRNA and endogenous short interfering RNAs play roles in transcriptional and post-transcriptional gene regulation. As *S. cerevisiae* lacks the factors necessary to mediate these types of regulation, the focus here will be on long ncRNAs (lncRNAs) classified arbitrarily as ncRNAs longer than 200 nt. Analysis of the transcriptomes of many mammalian cells has identified greater than 10,000 intergenic lncRNAs, which is increasing as more cell types are sequenced at greater depth (Vance and Ponting, 2014). A function for the vast majority of these has not yet been identified. As many lncRNAs show greater conservation in their positioning than their sequence, many suggest, as with *S. cerevisiae*, that the process of their transcription plays a more important role than the lncRNA (Ulitsky et al., 2011). Functions for some lncRNAs are better characterised. For example, lncRNAs *Hotair* and *Xist* bind histone-modifying complexes and direct these to particular locations to mediate transcriptional silencing (Lee, 2011; Tsai et al., 2010). Techniques including Chromatin Isolation by RNA Purification (ChIRP-seq) and capture hybridisation of RNA targets (CHART) have shown that some lncRNAs can be found bound to chromatin at multiple locations, distal from their own transcription site, in sharp peaks and with binding preferences for particular DNA motifs, suggesting a similarity to DNA-binding protein transcription factors (Chu et al., 2011; Simon et al., 2011). It is not clear from these studies whether the RNA binds

directly to the DNA or is recruited indirectly through an interaction with a protein factor or another RNA.

5.1.4 Mechanisms through which ncRNAs may interact with DNA

The interaction of ncRNAs with chromatin is likely to be critical for many of them to function in regulating transcription and other chromatin processes. Although this may occur indirectly through protein factors, the focus here will be on a direct interaction between RNA and DNA. There are several ways in which this may occur. During transcription, RNA is tethered to its TU, and acts as a recruitment platform along with other components of the TEC to mediate the recruitment of factors to this site (see Chapter 3). Also during transcription, the nascent RNA can bind to the template DNA strand upstream of the polymerase active site forming an RNA:DNA hybrid known as an R-loop. This structure may remain upon loss of the polymerase. This process is favoured by the formation of negatively supercoiled DNA behind the polymerase and G rich RNA (Skourti-Stathaki and Proudfoot, 2014), while RNA within the structure is specifically degraded by RNase H (El Hage et al., 2010). The formation of an R-loop between RNA and double-stranded DNA is not thought to occur non-co-transcriptionally under physiological conditions.

Upon release of the nascent RNA from the DNA template, RNA may be able to bind directly to duplex DNA, forming a triplex structure. The classical triplex structure (H-form DNA) forms when an oligonucleotide binds in the major groove of a homopurine:homopyrimidine duplex by sequence-specific Hoogsteen base pairing. The oligonucleotide can bind parallel to the homopurine strand via T-A-T (or U-A-T), C+-G-C (C+ is a protonated cytosine base) and G-G-C base triplets or anti-parallel to the homopurine strand via T-A-T, G-G-C and A-A-T triplets (Buske et al., 2012; Frank-Kamenetskii and Mirkin, 1995). This triplet code ensures specificity, although some mismatched triplets can be tolerated. The parallel form requires an acidic pH to protonate cytosine (Sugimoto et al., 2001), while the anti-parallel form is pH-

independent but RNA oligonucleotides may be excluded from this form (Semerad and Maher, 1994). There are also other less well characterised triplex structures. These include the binding of an oligonucleotide in the duplex minor groove (Szewczak et al., 1998) and the formation of an R-form triplex, proposed as an intermediate in recombination, in which an oligonucleotide binds parallel to its identical duplex sequence in the major groove (Shchyolkina et al., 1994; Zhurkin et al., 1994).

Two papers support the idea that ncRNAs in human cells can regulate chromosomal processes by forming triplexes with double-stranded DNA. Schmitz et al., (2010) show that a ncRNA transcribed from a region upstream of the rDNA long transcript forms a triplex with the promoter of the rDNA long transcript at the binding site of transcription termination factor 1 (TTF-1). This interaction recruits DNA methyltransferase 3b (DNMT3b) to this site whilst inhibiting TTF-1 binding. Martianov et al., (2007) show a ncRNA from the minor promoter of the dihydrofolate reductase (*DHFR*) gene forms a triplex with the major promoter of this gene, possibly disrupting PIC formation and inhibiting transcription. Further evidence supporting *in vivo* triplex formation is provided by immunofluorescent staining of eukaryotic chromosomes with triplex-specific antibodies (Agazie et al., 1996; Lee et al., 1987) and the discovery of proteins capable of interacting specifically with triplexes, and helicases capable of unwinding triplexes (Buske et al., 2011; Jain et al., 2010).

5.1.5 Transcription of a ncRNA regulates recombination at the rDNA locus *in S. cerevisiae*

The ribosomal DNA (rDNA) locus in *S. cerevisiae* consists of 150-200 repeats of a 9.1 kb sequence, occupying around 60 % of chromosome XII. The 9.1 kb sequence contains a Pol I-transcribed 35S rRNA gene and an intergenic spacer region split into two (IGS1 and IGS2) by the Pol III-transcribed 5S rRNA gene (Figure 2A) (Johnston et al., 1997). The repetitive nature of this sequence means that it is highly recombinogenic. Recombination between repeats is used as a mechanism for varying repeat number in

order to control rRNA synthesis levels as well as a mechanism to repair DNA damage within these highly transcribed sequences. This process needs to be tightly controlled as excess recombination can lead to rDNA instability and the formation of extra-chromosomal rDNA circles (ERCs), which accumulate in mother cells leading to replicative senescence and aging (Ganley et al., 2009; Sinclair and Guarente, 1997).

IGS1 contains a replication fork blocking site (RFB), which is bound by Fob1, creating a unidirectional block to the progress of replication forks initiated from the autonomously replicating sequence (ARS) within IGS2 (Kobayashi, 2003). The block can lead to double-strand break (DSB) formation at the RFB, which can then undergo homologous recombination at sister chromatids (Kobayashi et al., 1998). Kobayashi and Ganley (2005) proposed the currently accepted model for regulating this recombination. In this model the sister chromatid rDNA repeats are normally constrained by cohesin, which is particularly concentrated over the cohesion-associated region (CAR) in IGS2. This restricts the DSB so that homologous recombination is only initiated with its aligned repeat of its sister chromatid, leading to equal sister chromatid exchange (ESCE) and no change in rDNA copy number. Pol II-dependent ncRNA transcription from the bidirectional Epro-promoter within IGS1, which is usually repressed by the histone deacetylase Sir2, removes cohesin from the repeats. This enables recombination at alternative repeats of the sister chromatid, leading to unequal sister chromatid exchange (UESCE) and changes in rDNA copy number (Figure 5.1). It also allows intra-chromatid recombination, causing the formation of ERCs.

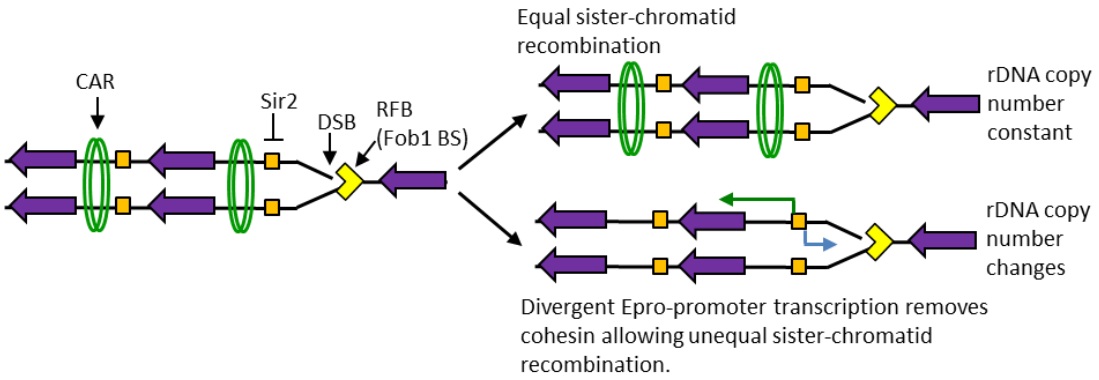


Figure 5.1. The currently accepted model proposed by Kobayashi and Ganley (2005) for rDNA copy number regulation by Pol II transcription from the bidirectional Epro promoter. Normally Sir2 represses Pol II transcription and so cohesin remains bound to the CAR and rDNA repeats remain aligned with sister chromatids. Upon DSB formation the broken end initiates homologous recombination with its aligned repeat leading to no change in rDNA copy number. If Pol II transcription is activated, the transcribing polymerase displaces cohesin from the CAR, allowing the DSB to recombine at alternative repeats leading to changes in copy number. Green rings, cohesin; CAR, cohesin associated region; DSB, double-strand break; yellow arrow head, replication fork blocking site (RFB); orange box, Epro promoter; purple arrow, 35S rRNA gene; blue arrow, IGS1-F transcript; green arrow, IGS1-R transcript.

Work by Houseley et al., (2007) showed that deletion of *TRF4*, a component of the TRAMP4 complex involved in targeting ncRNA to the exosome, leads to stabilisation of the IGS1-F transcript from the Epro-promoter. This mutation also resulted in changes in rDNA copy number without changing cohesin association, suggesting a role for the ncRNA itself, and not just the transcribing Pol II displacing cohesin, in promoting UESCE.

5.1.6 Aims

The aim of this chapter was to determine whether the IGS1-F transcript is able to interact with double-stranded DNA (dsDNA) from the rDNA locus *in vitro* and therefore provide clues as to a mechanism through which this transcript could regulate recombination at this locus. The nature of this interaction was then interrogated. Investigations were also carried out to determine whether ncRNAs from other loci could form similar interactions.

5.2 Results

5.2.1 RNA interacts with double-stranded DNA, sequence-specifically, at the rDNA locus

In order to detect RNA binding to dsDNA a new technique was developed using an electrophoretic mobility shift assay (EMSA). dsDNAs were synthesised by PCR with Cy5 and Cy3 5' end-labelled primers so that one strand was Cy5 labelled and the other Cy3 labelled. These were then incubated with *in vitro*-transcribed RNAs in TBM buffer (90 mM Tris-HCl pH 7.2, 90 mM boric acid, 10 mM $MgCl_2$) and run on polyacrylamide gels containing 0.5x TBM buffer. Gels were then scanned for Cy5 and Cy3 fluorescence, enabling distinction between RNA-dependent mobility shifts of either a single DNA strand with fluorescence for only one cyanine dye, an RNA:DNA heteroduplex, or both DNA strands with both Cy5 and Cy3 fluorescence, an RNA:DNA:DNA triple-stranded structure. Gels were then stained with methylene blue to show all nucleic acids.

Initially three 270 bp Cy5-Cy3-labelled dsDNAs were synthesised, mapping to overlapping sequences across IGS1 (Figure 5.2B). These were incubated with a 585 nt RNA (RNA A) corresponding to the IGS1-F transcript. This RNA interacts with dsDNA Z to form an RNA:DNA:DNA triple-stranded structure but a similar structure does not form with DNAs X and Y, showing that the interaction is DNA sequence-specific (Figure 5.2C). To assess which RNA sequence is required for the interaction, RNAs B, C and D were synthesised as well as an RNA antisense to D, RNA E. RNAs B, D and E all interact to form a triple-stranded structure with DNA Z, whereas RNA C does not (Figure 5.2D). This shows that the interaction is also RNA sequence-specific and occurs over the region between 450 and 585 nt from the 5' end of RNA A. In these interactions, shifts containing only one fluorescent signal also form. These RNA:single-stranded DNA (ssDNA) heteroduplexes may be formed by binding of RNA to low levels of ssDNA produced during PCR, which are not removed by PCR spin column purification. Alternatively the RNA may be displacing ssDNA from the dsDNA. To test this, some DNA

Z was purified by gel extraction to remove ssDNA. This DNA was then incubated separately with RNAs A and D. Triple-stranded structures still formed, whereas heteroduplex interactions did not (Figure 5.2D). The fact that the RNA is unable to displace ssDNA also suggests that RNA is interacting with an intact dsDNA structure rather than two separate single strands to form the triple-stranded structure.

The triple-stranded structures formed were dependent on magnesium ions as they were not detectable when run on 0.5x TBE (90 mM Tris-HCl pH 7.2, 90 mM boric acid, 10 mM EDTA)-containing polyacrylamide gels (data not shown).

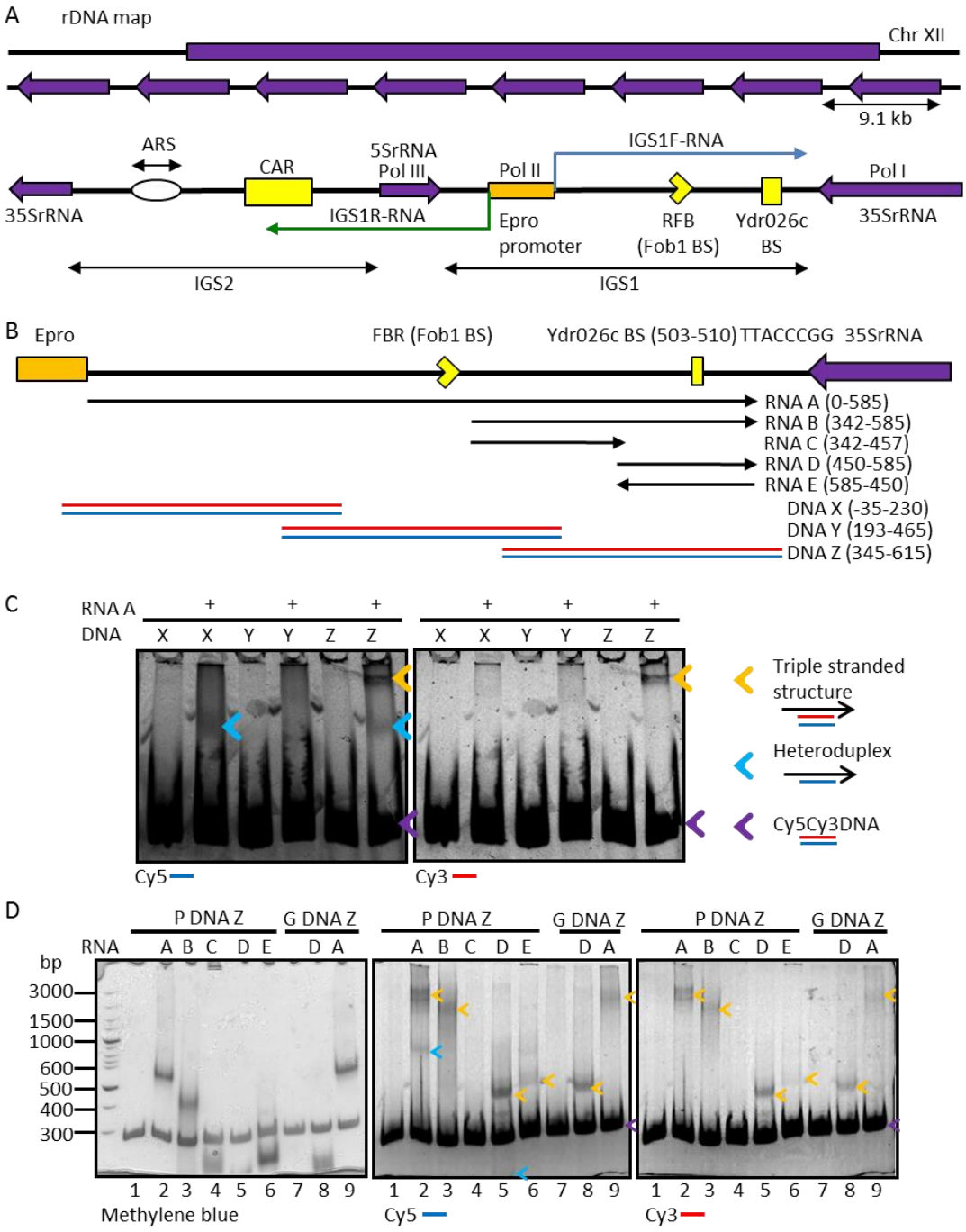


Figure 5.2. RNA interacts with double-stranded DNA, sequence specifically, at the rDNA locus. A) A schematic map of the rDNA locus and the intergenic spacer region. Abbreviations and colour coding as in Figure 5.1. BS, binding site; IGS, intergenic spacer region. B) The location of *in vitro* transcribed RNAs and Cy5-Cy3-labelled DNAs. Numbers are in bp and are relative to the 5' end of RNA A. The Cy5-labelled DNA strand (blue) is sense to RNA A, the Cy3-labelled DNA strand (red) is antisense to RNA A. C) EMSA showing RNA A interacts specifically with DNA Z to form a triple-stranded structure (shift with signal for both labels). D) EMSA showing RNAs A, B, D and E bind PCR spin column-purified Cy5-Cy3-labelled DNA Z (P DNA Z) to form triple-stranded structures (shifts with signal for both labels - lanes 2, 3, 5 and 6). RNAs A, B, D also bind the complementary single Cy5-labelled strand of DNA A to form heteroduplexes (shifts with signal for the Cy5 label only – lanes 2, 3 and 5). RNA C does not bind DNA Z (lane

4). RNAs D and A bind gel-extracted Cy5-Cy3-labelled DNA Z (G DNA Z) to form triple-stranded structures and no heteroduplex structures (shifts with signal for both labels – lanes 8 and 9).

5.2.2 The RNA:DNA interaction is strongest over a short approximately 30 bp region

To narrow down the region of DNA and RNA forming the interaction, six overlapping 25-30 nt ssDNAs were synthesised complementary to RNA A, covering the region 454-581 nt from the 5' end of RNA A (Figure 5.3A). These were incubated with RNA A in separate reactions in order to inhibit the binding of these short regions of the RNA from interacting with dsDNA. Complementary DNA 3 decreases the mobility of RNA A more than the other complementary DNAs (methylene blue-stained gel, Figure 5.3B). DNA Z was then added and products were analysed by EMSA. Only short complementary DNA 4 inhibited the formation of the RNA:DNA:DNA triple-stranded structure, suggesting the 30 nt complementary sequence to this DNA in the RNA is key to binding dsDNA (Figure 5.3B). This sequence is immediately adjacent to the binding site of the Pol I termination factor, Ydr026c, which is a homolog of human TTF-1 (Reiter et al., 2012). The binding site of TTF-1 in humans is also bound by RNA (Schmitz et al., 2010).

RNA A was then incubated with DNA Z and different restriction enzymes that cut at sites along DNA Z, so as to test the ability of the RNA to protect the bound DNA from digestion (Figure 5.3C). The intensity of the Cy5 signal of the triple-stranded structure remaining after digestion by each of the enzymes was measured and normalised to the intensity of this signal for an undigested reaction giving a measure of the undigested DNA in the triple-stranded structure at the end of the reaction (Figure 5.3D). The efficiency of the enzymes is controlled for by the level of digestion of unbound DNA, which is almost completely digested in all reactions. The EcoRI restriction site was most protected, with decreasing levels of protection at restriction sites away from this site. This suggests that the sequence around this site is most strongly bound by the RNA. This is consistent with

the result from the competing short complementary DNAs experiment. The fact that the other restriction sites are also partially protected suggests that these sequences also interact with the RNA and contribute to the formation of the triple-stranded structure. DdeI, which cuts 15 bp from the Cy5-labelled end of the DNA, completely digests the triple-stranded structure, removing the Cy5 signal, showing that this sequence is not protected by RNA A. The remaining Cy3-labelled DNA is still able to bind RNA and form the triple-stranded structure except the size and therefore the mobility shift of the structure is lower.

An RNA corresponding to RNA A was then synthesised in which 40 nt across the region most strongly forming the triple-stranded structure had been deleted (RNA DEL). Its ability to form the triple-stranded structure was 8 times less than RNA A (Figure 5.3E). The fact that the triple-stranded structure still forms supports the result of the restriction enzyme digestion assay in that the regions either side of the key 40 nt region still interact but this interaction is weaker.

A surprising observation from the experiments shown in Figure 5.3B and C is the behavior of the RNA A:ssDNA heteroduplex interaction. This interaction is inhibited more strongly by the presence of complementary DNA 4 like the triple-stranded interaction and the digestion of this interaction by restriction enzymes matches the pattern of digestion observed for the triple-stranded structure. This suggests the RNA A:ssDNA interaction is also strongest over the same short 30-40 bp region as the RNA A:dsDNA interaction. This may be due to the formation of a particular structure within the RNA in this region. Figure 5.2D shows that the triple-stranded structure still forms in the absence of the heteroduplex, showing that their formation is independent.

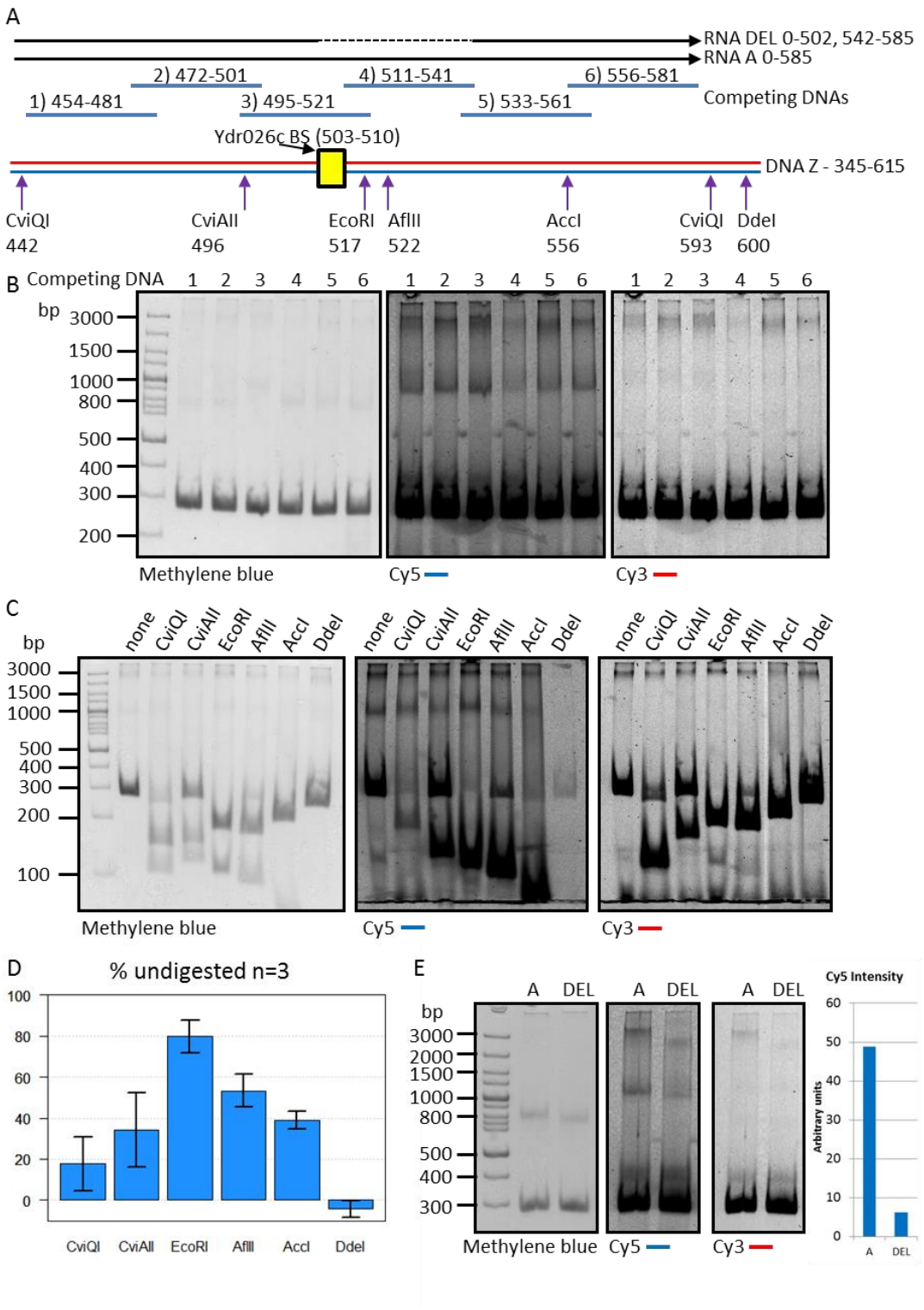


Figure 5.3. The RNA:DNA interaction is strongest over a short, approximately 30 nt and 30 bp, region of the RNA and DNA. A) Schematic showing the positions and lengths of the RNAs and DNAs used in subsequent experiments. The binding site of Ydr026c and restriction sites on DNA Z are shown. All positions are in bp relative to the 5' end of RNA A. B) EMSA showing the effect of incubating RNA A with short, 25-30 nt, competing ssDNAs complementary to RNA A before incubating with DNA Z. Only competing DNA 4 causes a large reduction in triple-stranded structure formation. C) EMSA showing the ability of restriction enzymes to digest DNA at positions along DNA Z

when bound by RNA A. D) Bar chart showing the average percentage of undigested RNA A:DNA Z triple-stranded structure remaining after digestion with each restriction enzyme. This is calculated by measuring the Cy5 intensity of the triple-stranded shift from three repeats of the experiment shown in C. Error bars show the standard error of the mean. E) EMSA showing RNA DEL forms approximately 8 times less triple-stranded structure than RNA A with DNA Z.

5.2.3 There is an optimal DNA length for triple-stranded structure formation

3 additional Cy5-Cy3-labelled dsDNAs were synthesised of different sizes containing the Ydr026c binding site and key triple-stranded structure-forming region (Figure 5.4A). The ability of these to form a triple-stranded structure with RNA A was compared to DNA Z. Both the longer DNA R and shorter DNA S still form triple-stranded structures, although at lower levels to DNA Z, while the shortest DNA (DNA T) does not form this structure (Figure 5.4B). This suggests there is an optimal length of DNA for triple-stranded structure formation. At longer lengths there may be more competing DNA sequence, which may inhibit the RNA interacting with its binding site on DNA. For the shorter DNAs, regions downstream (DNA S) or upstream (DNA T) of the key interacting region have been truncated. As these regions were shown by restriction enzyme digest to be protected by RNA binding it may be that the interaction is weakened by their loss.

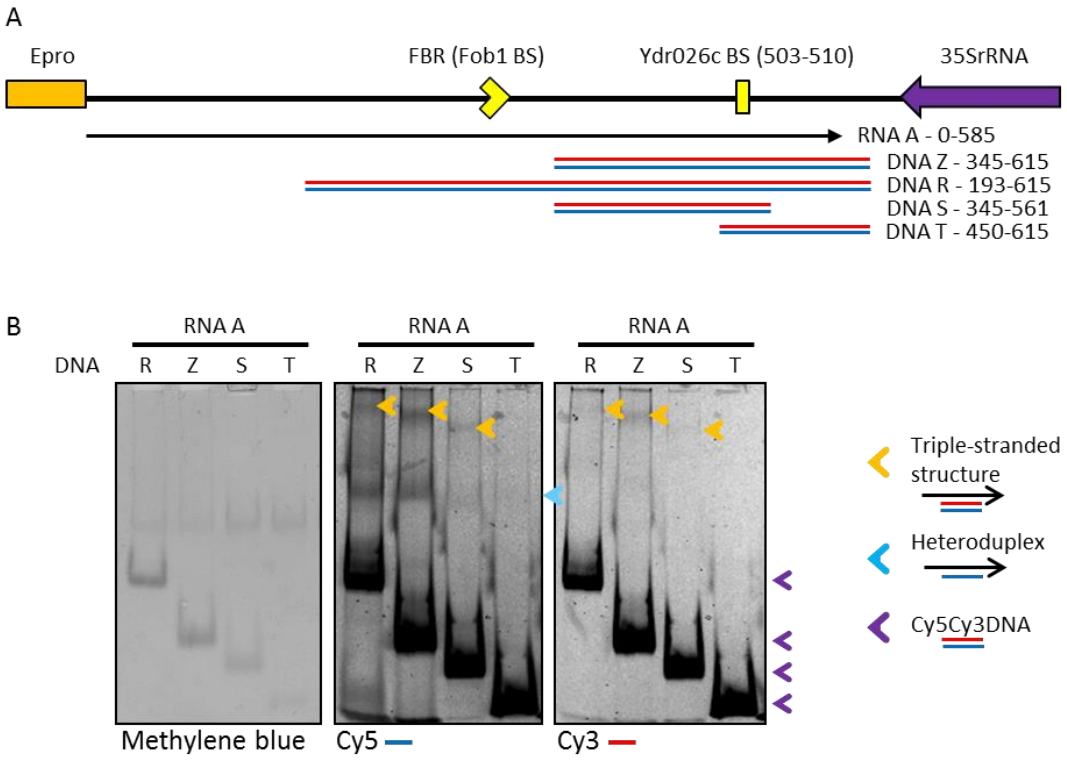


Figure 5.4. There is an optimal DNA length for the RNA:DNA interaction. A) Schematic showing the positions and lengths of the RNA and DNAs used in the EMSA. All positions are in bp relative to the 5' end of RNA A. B) EMSA showing RNA A forms triple-stranded structures with DNAs R, Z and S but not DNA T. Of the three triple-stranded structures formed, the strongest interaction is with the medium length DNA Z.

5.2.4 The interaction of RNA with dsDNA increases logarithmically with increasing RNA concentration

To test the dependence of triple-stranded structure formation on RNA concentration, triple-stranded structure formation between DNA Z (1 μM) and increasing concentrations of RNA D (up to 13.5 μM) was assayed (Figure 5.5B). The amount of DNA Z forming triple-stranded structure at each RNA concentration was determined by measuring the Cy5 fluorescence signal of the shifted band and comparing that to the Cy5 fluorescence intensity from a DNA Z dilution series (Figure 5.5A). Plotting these against the $\log_{10}$ of the corresponding RNA D concentrations shows the interaction increases logarithmically with increasing RNA concentration (Figure 5.5C). When the RNA is in 10-fold excess of the DNA, the proportion of DNA forming the triple-stranded structure

is approximately 6 % suggesting that the interaction is weak. Despite the low proportion of DNA interacting, the interaction is RNA concentration-dependent across the wide range of RNA concentrations tested.

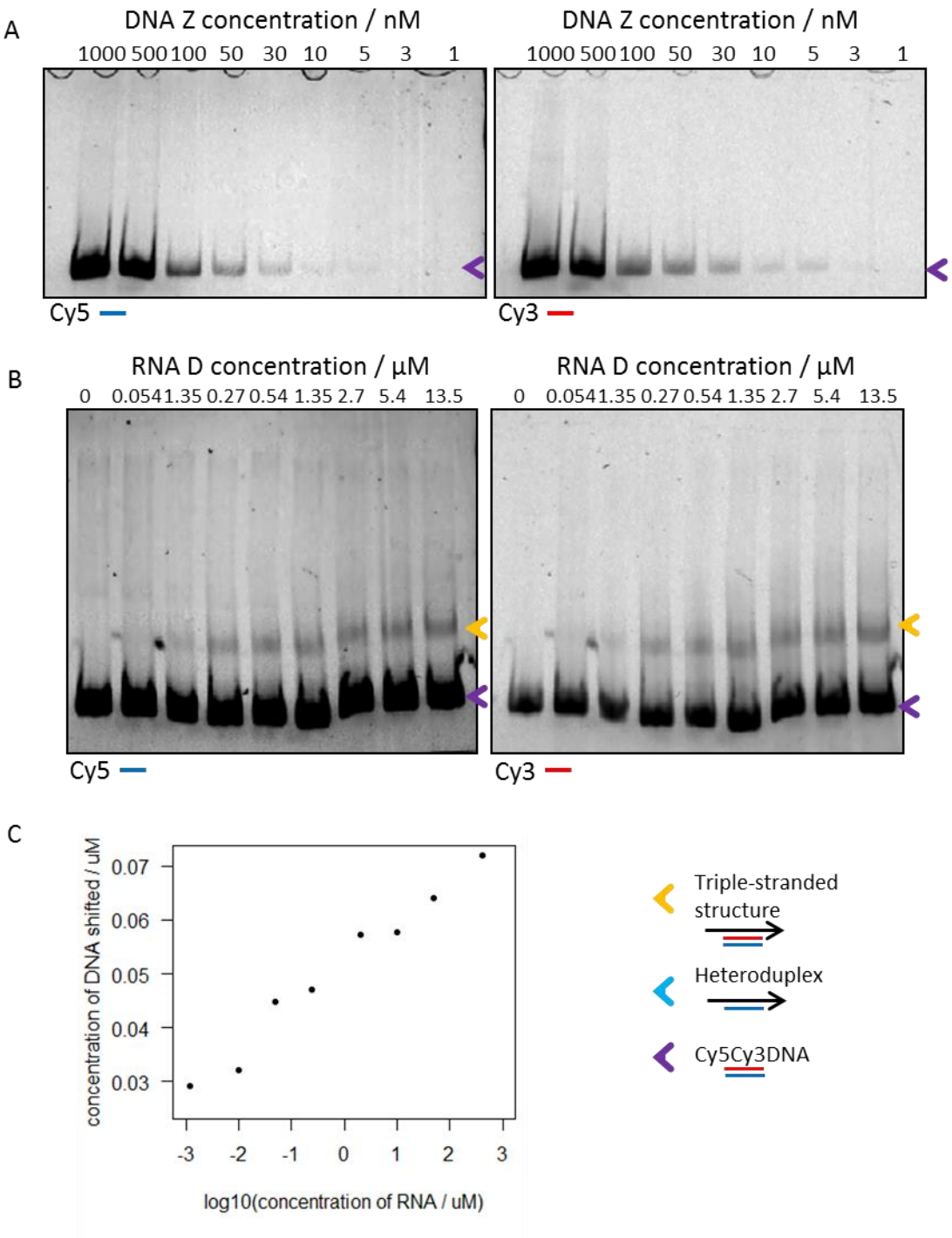


Figure 5.5. The interaction of RNA with dsDNA increases logarithmically with increasing RNA concentration. A) Cy5 and Cy3 fluorescence of DNA Z dilution series used to calculate DNA concentration of shifted bands. B) EMSA showing triple-stranded structure formation between DNA Z and increasing concentrations of RNA D. C) Scatter plot showing the relationship between RNA concentration and the amount of triple-stranded structure formed with DNA Z.

5.2.5 Using RNase H to determine the type of triple-stranded structure formed

RNase H (Figure 5.6) and RNase A (data not shown) were able to degrade both the triple-stranded and heteroduplex structures formed between RNA A and DNA A. RNase H has been shown to specifically degrade ssRNA when hybridized to ssDNA and thus is able to degrade RNA within an R-loop (Donis-Keller, 1979). Its ability to degrade RNA in a triplex has, however, not been determined. Unexpectedly, the RNase H used partially degraded ssRNA when 5 U were added (Figure 5.6, compare lanes 4 and 5) and fully degraded ssRNA when 30 U were added (data not shown). Two separate stocks of the enzyme had the same effect, suggesting that the enzyme may be contaminated or may not be completely specific to degrading RNA in RNA:DNA heteroduplexes in these conditions. Consequently it is not possible to determine what type of triple-stranded structure is being formed. This experiment should be repeated with RNase H from a different supplier.

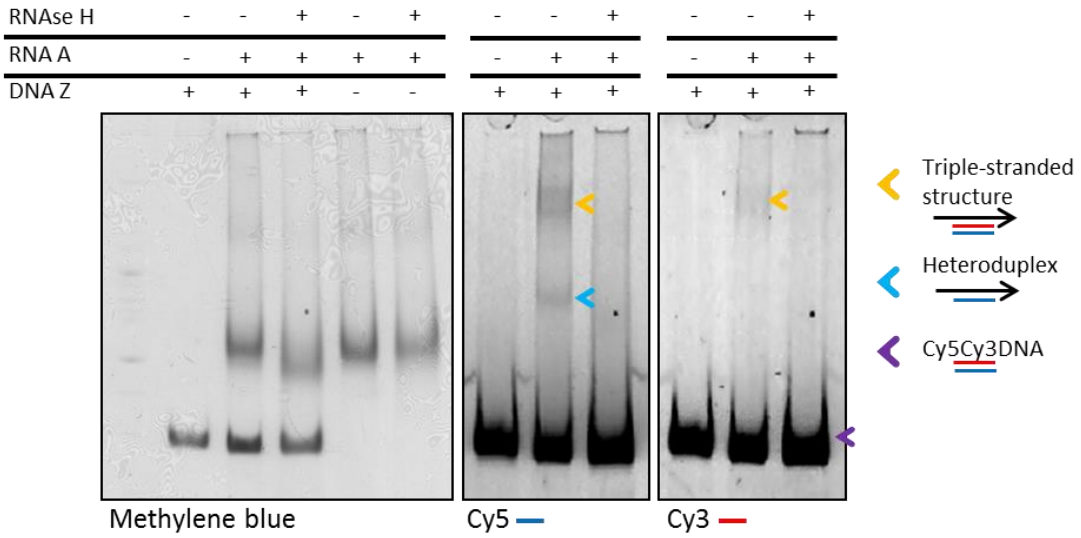


Figure 5.6. Attempt at using RNAse H to determine the type of triple-stranded structure formed. EMSA showing the digestion by RNAse H of RNA within triple-stranded structures and heteroduplex structures and the partial digestion of single stranded RNA.

5.2.6 Triple-stranded structure formation at the *GAL1-10* locus

Non-coding transcription has been shown to occur over the GAL1-GAL10 bi-directional promoter on both strands and shows changes as cells are switched from glucose to galactose (Chapter 4, Figure 4.10). The ability of the ncRNAs produced to interact with DNA from this locus was investigated. Two overlapping 310 bp Cy5-Cy3-labelled DNA molecules (DNA GALiX and Y) and two 572 nt RNAs mapping to each strand from this promoter (RNA GALiF and R) were synthesised (Figure 5.7A). Triple-stranded structure formation between these molecules was investigated. GALiX interacts with both RNA molecules to form triple-stranded structures although high levels of heteroduplex formation masks the Cy5 signal from the X-F triple-stranded structure. GALiY interacts with GALiF and to a lesser extent with GALiR to form triple-stranded structures (Figure 5.7B).

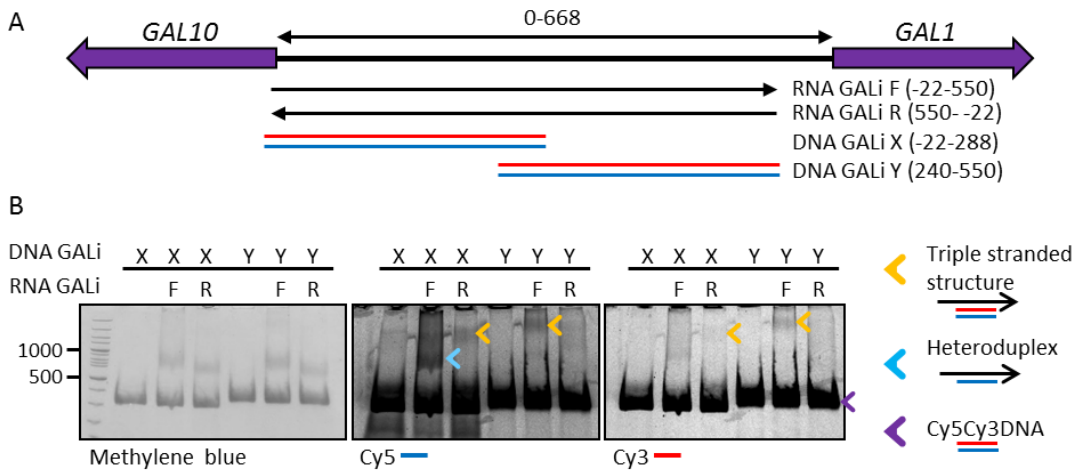


Figure 5.7. RNA and DNA from the *GAL1-10* intergenic region form triple-stranded structures. A) Schematic showing the positions and lengths of the RNA and DNAs used in the EMSA. All positions are in bp relative to the 5' end of *GAL10.* B) EMSA showing RNA F forms triple-stranded structures with both DNAs and RNA R forms a triple-stranded structure with DNA X and DNA Y to a lesser extent. The Cy5 signal of the triple-stranded structure formed between RNA F and DNA X is probably masked by heteroduplex signal.

5.3 Discussion

This study shows that RNA and DNA from the rDNA locus are able to interact to form a triple-stranded structure *in vitro*. This interaction is sequence-specific for both the RNA and DNA and seems to be strongest across an approximately 30-40 bp sequence, although sequences either side of this are involved as shown by the partial protection of these sequences from restriction enzyme digestion. The position of this site immediately next to the binding site of Ydr026c is striking due to the binding site of its homolog in human cells also being bound by RNA. Deletion of *YDR026C* leads to decreased UESCE (Jon Ayling, unpublished data). By binding to this site, RNA may help stabilise Ydr026c at the locus to promote UESCE.

The sequences involved in the interaction lack the polypurine tracts necessary for the formation of a canonical H-form triplex, suggesting that this is an alternative structure. One possibility is this is an alternative triplex structure. The fact that the RNA binds only to duplex DNA containing an identical sequence and the structure is promoted by the presence of magnesium ions fits with observations made for the R-form triplex (Shchyolkina et al., 1994) (Figure 5.8). Another possibility is that this is an R-loop structure with some unwinding of the DNA duplex enabling an interaction between RNA and its complementary strand (Figure 5.8). The PCR reaction could create a small proportion of DNA molecules with some unwound duplex. This should lead to saturation of triple-stranded structure formation rather than the continued increases in formation observed at high RNA concentration. Instead the RNA could promote some unwinding of the DNA duplex. Previous studies to create R-loops *in vitro* have required prior melting of the DNA duplex at higher temperatures than the 22 ^{o}C incubations used here (Thomas et al., 1976) although magnesium ions were not present.

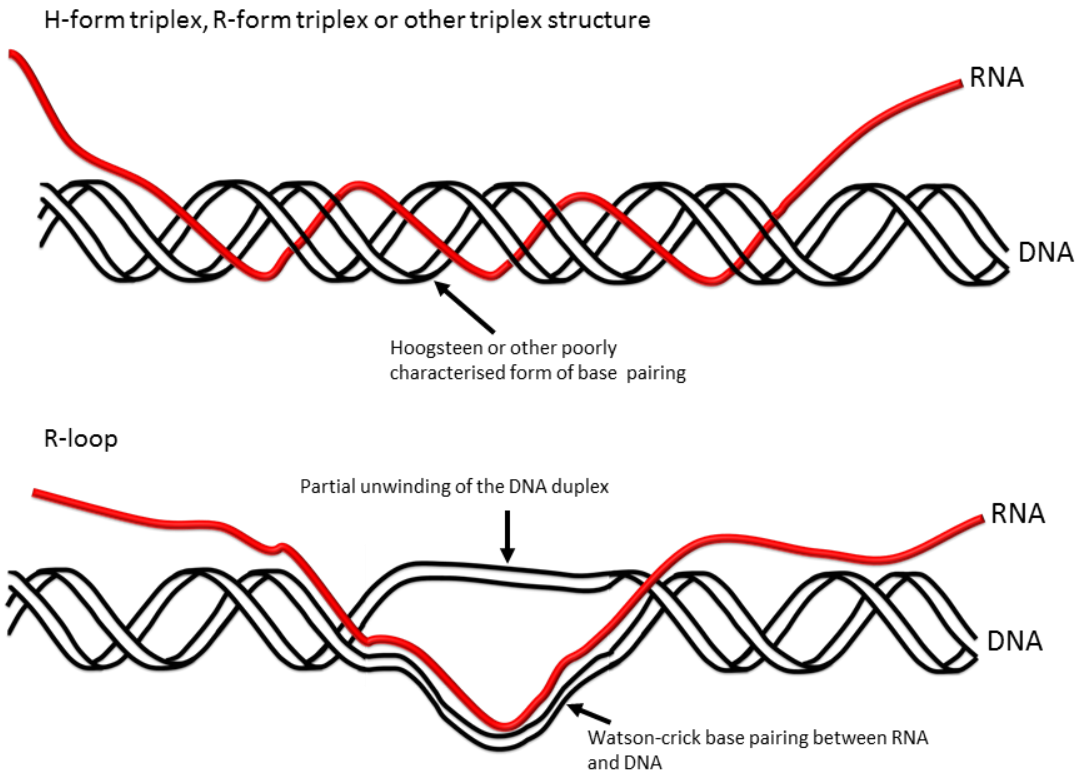


Figure 5.8. Possible triple-stranded structures formed between RNA and double-stranded DNA. RNA binds either to an intact double-stranded DNA molecule to form some form of triplex structure or the DNA is partially unwound enabling RNA to base pair with its complementary DNA strand.

Whatever the structure, the observations here show RNAs can interact with dsDNA *in trans* in physiological conditions. Observation of these structures at both the *GAL* and *rDNA* loci suggests this could be a common mechanism by which ncRNAs could be targeted to DNA and mediate their function. The structure has only been shown to form when the RNA sequence is identical to the DNA sequence and therefore likely restricts it to the site from which the RNA is transcribed. Even after DNA mutation the RNA sequence would still be identical to the DNA sequence, which may explain why lncRNA sequence is not highly conserved. The interaction appears to be fairly weak with only a small fraction of DNA and RNA forming the structure, although this may be different *in vivo*. This may restrict the interaction from forming during a short period immediately after transcription, while the RNA is still in close proximity to the site of transcription.

This may be the reason why expression of most ncRNAs from ectopic locations does not have a phenotype because they are not at high enough concentrations to interact with the DNA at their endogenous loci. The higher Paf1 levels on Pol II transcribing PCG TUs (Chapter 3) may promote movement of the pre-mRNA transcripts away from the site of transcription, preventing them from forming these interactions.

Future experiments would test whether expression of RNA A from an ectopic location was capable of promoting UESCE *in trans*. If this were the case, expression of RNA DEL ectopically would reveal whether triple-stranded structure formation was important for this process. It would also be interesting to express and purify Ydr026c and assay its ability to bind to the triple-stranded structure by assessing if it could cause an additional shift in the mobility of the structure.

Chapter 6.

Discussion

6 Discussion

6.1 Summary

In this thesis investigations were carried out into three main areas.

Firstly, the location of Pol II and factors that associate with the TEC were mapped with strand-specificity and nucleotide resolution genome-wide in cells growing exponentially in glucose-containing media. Due to the high resolution obtained, it was revealed that factors associate and dissociate with Pol II in a previously underappreciated dynamic nature, particularly during passage of Pol II through each nucleosome and at the 3' end of TUs. At many intron-containing TUs it was shown that splicing intermediates co-immunoprecipitate with particular forms of the TEC, which suggests splicing often occurs as the downstream exon is transcribed. The sensitivity and strand-specificity of the mapping revealed that the association of factors with the TEC at the majority of TUs is uniform. However, at some PCG TUs and non-coding TUs some factors associate at different levels, in conflict with the conclusions of previous studies (Bataille et al., 2012; Mayer et al., 2010). This demonstrates that transcription is regulated at this stage and this may affect the stability and export of the transcript produced.

Secondly, how transcription changes upon a change in conditions was analysed by mapping the location of Pol II as cells were switched from glucose- to galactose-containing media. Extensive and rapid transcriptional changes were observed, resulting from cells transitioning within 5 minutes of the change in conditions to a YMC in which they spend a larger proportion of each cycle in the RC phase. OX and RC phase genes are often clustered together at genomic loci with extensive overlap between these TUs and antisense non-coding TUs. Transcription overlapping with promoter regions was shown to act as an additional layer of regulation, possibly used to repress out of phase expression of particular TUs. The mechanism of this repression may result from the transcriptional process but could also involve a role for the transcripts produced.

Thirdly, non-coding transcripts produced at two loci were investigated for their ability to interact with dsDNA. Non-coding transcripts were shown to interact sequence specifically with DNA to form triple-stranded structures. This provides a mechanism by which these RNAs could play a functional role possibly by influencing the binding of protein factors.

6.2 Future directions

By enabling genome-wide analyses, the development of techniques using high-throughput sequencing and tiling arrays have been key to many of the recent advances in gene expression research. The discoveries in Chapter 3 and 4, which were reliant upon NET-seq and its adaptation for transcription elongation factor mapping, TEF-seq, demonstrate the advantages of such techniques. A drawback of these techniques is that experiments need to be carried out on populations of cells. All signal obtained is therefore a population average, which means heterogeneity between cells may be overlooked. This is highlighted in Chapter 4 where an increase in transcription observed by NET-seq was shown by RNA FISH to represent an increase in the proportion of cells in which a TU is transcribed rather than an increase in transcription in all cells. Heterogeneity in gene expression between cells not only results from cells in different phases of the YMC but can also occur between cells within the same phase.

Such heterogeneity could well be present in TEF-seq data; a particular level of signal of elongation factor binding to Pol II along a TU may result from high levels on Pol II in some cells and low levels on Pol II in others. It has been shown for some PCG TUs that a proportion of the transcripts produced lack or have a short polyA tail, causing them to be retained at the site of transcription, potentially assisting in tethering TUs to the NPC (Abruzzi et al., 2004; Vodala et al., 2008) (see Chapter 1). While these may represent transcripts in the process of polyadenylation, another option is that heterogeneity in the association of factors with Pol II stimulates polyadenylation at different levels, giving

rise to variation in polyA tail length with each round of transcription. By also influencing the position of termination, such a process may also have a role in generating the observed heterogeneity in transcript isoforms (Geisberg et al., 2014; Pelechano et al., 2013).

To test this, TEF-seq should be used to map factor binding as cells are switched from glucose- to galactose-containing media, or alternatively, although more technically challenging, after synchronising cells in the YMC. This would reveal whether the ratio of factor binding to Pol II varies in different conditions and phases of the YMC. If this were the case, this would suggest that this is regulated rather than constitutive and that the levels of factor association with Pol II are heterogeneous. Such studies would still require populations of cells and would not provide proof of cell-to-cell differences in factor-Pol II association for cells in the same YMC phase. A complementary approach would be first to determine the regulator controlling levels of particular elongation factors on particular TUs, such as the possible link between the Haa1 transcription factor and levels of Paf1, Spt16 and Spt6 discussed in Chapter 3. Transcripts in single cells could then be followed either by RNA FISH or a live cell imaging approach upon disruption of Haa1, looking for a change in transcript distribution and variation between cells in these distributions.

The development of high-throughput techniques for single cells may provide an alternative strategy for observing heterogeneity. Single cell DNA and RNA sequencing is possible for cells from higher eukaryotes (Farlik et al., 2015; Tang et al., 2009) suggesting that single cell NET-seq and TEF-seq may be possible in the future. Technological challenges due to the presence of a cell wall and the low RNA quantities present have so far prevented single cell RNA sequencing in *S. cerevisiae*. These problems and the inefficiencies of immunoprecipitation and presence of background signal may therefore prove insurmountable hurdles to such technological advances.

6.3 Closing remarks

Conventional theories of transcription propose that transcription is predominantly restricted to PCGs and the level of expression of each gene is largely determined by the level of recruitment of Pol II to gene promoters by DNA-binding transcription factors. Recent discoveries demonstrating that most genomes are pervasively transcribed and gene expression is regulated by many mechanisms at multiple stages have shown this to be far too simplistic. Work presented here adds to these discoveries by suggesting a major role for non-coding transcription and non-coding transcripts in regulating PCG transcription and revealing an additional layer of regulation occurring during transcription elongation to influence transcript fate. The evolution of such complex and intricate regulatory mechanisms demonstrates the importance to cell survival of being able to fine tune gene expression for particular conditions and rapidly alter gene expression when conditions change.

7 Appendix

Tagged factor	Repeat number	Total reads before alignment	Total uniquely aligned reads	Mean *SCR1* ratio	Total counts after no-tag normalisation
Rpb3	1	55715348	8734606	0.72287	5665252
Rpb3	2	52941531	11622648	0.951067	7584353
Rpb3	1 + 2	108656879	20357254	1.673937	13249605
Cet1	1	92672875	8902755	1.154386	4001156
Cet1	2	54572943	3343537	0.464203	1372498
Cet1	1 + 2	147245818	12246292	1.618589	5373653
Nrd1	1	33466014	3549764	0.467231	1565870
Nrd1	2	151184398	19577503	1.933952	11365815
Nrd1	1 + 2	184650412	23127267	2.401183	12931685
Paf1	1	56462163	5221437	0.876595	1499354
Paf1	2	86827099	4303687	0.900999	477984.9
Paf1	1 + 2	143289262	9525124	1.777594	1977339
Pcf11	1	52091129	5101946	0.971322	977647.4
Pcf11	2	70733164	3963217	0.828654	444698.6
Pcf11	1 + 2	122824293	9065163	1.799976	1422346
Spt16	1	78106056	10505280	1.017822	6183538
Spt16	2	80851993	7976597	0.798672	4585382
Spt16	1 + 2	158958049	18481877	1.816495	10768920
Spt6	1	106201956	28635390	1.928179	20448211
Spt6	2	28051242	4440652	0.380375	2825553
Spt6	1 + 2	134253198	33076042	2.308555	23273765
Isw1	1	32726190	1506850	0.471084	-493405
Isw1	2	83878982	3054553	0.792244	-309366
Isw1	1 + 2	116605172	4561403	1.263328	-802771
Yra1	1	63894167	2520891	0.766252	-732666
Yra1	2	95349653	4751769	1.279875	-682667
Yra1	1 + 2	159243820	7272660	2.046127	-1415333
Set2	1	57688679	2489167	0.512015	315117.9
Bur1	1	50155126	1645821	0.402068	-61388.7
Ctk1	1	46440200	1190840	0.297868	-73927.3
Fcp1	1	60748146	1890977	0.383786	261397.2
Kin28	1	87104614	2252370	0.535671	-22123
Rtr1	1	27677406	823707	0.216275	-94612.3
Ssu72	1	67040107	2543032	0.515201	355453.4
No-tag	1	44577031	1808565	N/A	N/A
No-tag	2	64347685	2437502	N/A	N/A
No-tag	1 + 2	108924716	4246067	N/A	N/A

Table 7.1. Total read counts for TEF-seq data before and after no-tag normalisation. Mean *SCR1* ratio is the average ratio of factor to no-tag reads in 10 bp bins across the *SCR1* gene. This ratio is used as a scaling factor in no-tag normalisation.

8 Publications

Murray, S.C., Haenni, S., Howe, F.S., Fischl, H., Chocian, K., Nair, A., and Mellor, J. (2015). Sense and antisense transcription are associated with distinct chromatin architectures across genes. Nucleic Acids Res. *43*, 7823–7837.

Nguyen, T., Fischl, H., Howe, F.S., Woloszczuk, R., Barros, A.S., Xu, Z., Brown, D., Murray, S.C., Haenni, S., Halstead, J.M., et al. (2014). Transcription mediated insulation and interference direct gene cluster expression switches. Elife *3*, 1–21.

9 References

Abruzzi, K.C., Lacadie, S., and Rosbash, M. (2004). Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. EMBO J *23*, 2620–2631.

Abruzzi, K.C., Belostotsky, D. a, Chekanova, J. a, Dower, K., and Rosbash, M. (2006). 3'-end formation signals modulate the association of genes with the nuclear periphery as well as mRNP dot formation. EMBO J. *25*, 4253–4262.

Agazie, Y.M., Burkholder, G.D., and Lee, J.S. (1996). Triplex DNA in the nucleus: direct binding of triplex-specific antibodies and their effect on transcription, replication and cell growth. Biochem J *316 (Pt 2*, 461–466.

Akhtar, a, and Gasser, S.M. (2007). The nuclear envelope and transcriptional control. Nat Rev Genet *8*, 507–517.

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. Mol. Cell *34*, 387–393.

Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., and Lis, J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J. *28*, 1067–1077.

Baejen, C., Torkler, P., Gressel, S., Essig, K., Söding, J., and Cramer, P. (2014). Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. Mol Cell 745–757.

Baker, K.E., Coller, J., and Parker, R. (2004). The yeast Apq12 protein affects nucleocytoplasmic mRNA transport. RNA *10*, 1352–1358.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Barillà, D., Lee, B. a, and Proudfoot, N.J. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. *98*, 445–450.

Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. Cell *116*, 699–709.

Batada, N.N., Urrutia, A.O., and Hurst, L.D. (2007). Chromatin remodelling is a major source of coexpression of linked genes in yeast. Trends Genet. *23*, 480–484.

Bataille, A.R., Jeronimo, C., Jacques, P.E., Laramee, L., Fortin, M.E., Forest, A., Bergeron, M., Hanes, S.D., and Robert, F. (2012). A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Mol Cell *45*, 158–170.

Batta, K., Zhang, Z., Yen, K., Goffman, D.B., and Franklin Pugh, B. (2011). Genome-wide function of H2B ubiquitylation in promoter and genic regions. Genes Dev. *25*, 2254–2265.

Belotserkovskaya, R., Oh, S., Bondarenko, V. a, Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science *301*, 1090–1093.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. Genes Dev. *19*, 1635–1655.

Berretta, J., Pinskaya, M., and Morillon, A. (2008). A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. Genes Dev. *22*, 615–626.

Beyer, a. L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. *2*, 754–765.

Bi, X. (2014). Heterochromatin structure: Lessons from the budding yeast. IUBMB Life *66*, 657–666.

Blankenberg, D., Kuster, G. Von, Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: A web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol.

Blobel, G. (1985). Gene gating: a hypothesis. Proc. Natl. Acad. Sci. U. S. A. *82*, 8527–8529.

Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009). The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. Cell *139*, 547–559.

Bortvin, a, and Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. Science *272*, 1473–1476.

Bowman, E. a, and Kelly, W. (2014). RNA Polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases. Nucleus *5*, 1–13.

Brauer, M.J., Huttenhower, C., Airoldi, E.M., Rosenstein, R., Matese, J.C., Gresham, D., Boer, V.M., Troyanskaya, O.G., and Botstein, D. (2008). Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. Mol. Biol. Cell *19*, 352–367.

Bregman, A., Avraham-Kelbert, M., Barkai, O., Duek, L., Guterman, A., and Choder, M. (2011). Promoter elements regulate cytoplasmic mRNA decay. Cell *147*, 1473–1483.

Bretes, H., Rouviere, J.O., Leger, T., Oeffinger, M., Devaux, F., Doye, V., and Palancade, B. (2014). Sumoylation of the THO complex regulates the biogenesis of a subset of mRNPs. Nucleic Acids Res. *42*, 5043–5058.

Brickner, J.H., and Walter, P. (2004). Gene recruitment of the activated INO1 locus to the nuclear membrane. PLoS Biol. *2*.

Brogaard, K., Xi, L., Wang, J.-P., and Widom, J. (2012). A map of nucleosome positions in yeast at base-pair resolution. Nature 2–7.

Brugiolo, M., Herzel, L., and Neugebauer, K.M. (2013). Counting on co-transcriptional splicing. F1000Prime Rep. *5*, 9.

Brune, C., Munchel, S.E., Fischer, N., Podtelejnikov, A. V, and Weis, K. (2005). Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export. RNA *11*, 517–531.

Buratowski, S. (2003). The CTD code. Nat. Struct. Biol. *10*, 679–680.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. a (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549–561.

Buske, F.A., Mattick, J.S., and Bailey, T.L. (2011). Potential in vivo roles of nucleic acid triple-helices. RNA Biol *8*, 427–439.

Buske, F.A., Bauer, D.C., Mattick, J.S., and Bailey, T.L. (2012). Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. Genome Res *22*, 1372–1381.

Cai, L., and Tu, B.P. (2012). Driving the Cell Cycle Through Metabolism. Annu. Rev. Cell Dev. Biol. *28*, 59–87.

Cai, L., Sutter, B.M., Li, B., and Tu, B.P. (2011). Acetyl-CoA Induces Cell Growth and Proliferation by Promoting the Acetylation of Histones at Growth Genes. Mol. Cell *42*, 426–437.

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193–198.

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007). Antisense RNA Stabilization Induces Transcriptional Gene Silencing via Histone Deacetylation in S. cerevisiae. Cell *131*, 706–717.

Camblong, J., Beyrouthy, N., Guffanti, E., Schlaepfer, G., Steinmetz, L.M., and Stutz, F. (2009). Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. Genes Dev *23*, 1534–1545.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global Analysis of Nascent RNA Reveals Transcriptional Pausing in Terminal Exons. Mol. Cell *40*, 571–581.

Casolari, J.M., Brown, C.R., Komili, S., West, J., Hieronymus, H., and Silver, P. a. (2004). Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. Cell *117*, 427–439.
Castelnuovo, M., and Stutz, F. (2015). Role of chromatin, environmental changes and single cell heterogeneity in non-coding transcription and gene regulation. Curr. Opin. Cell Biol. *34*, 16–22.
Castelnuovo, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F., and Zenklusen, D. (2013). Bimodal expression of PHO84 is modulated by early termination of antisense transcription. Nat. Struct. Mol. Biol. *20*, 851–858.
Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S., and Young, R. a (2001). Remodeling of yeast genome expression in response to environmental changes. Mol Biol Cell *12*, 323–337.
Chandrasekharan, M.B., Huang, F., and Sun, Z.-W. (2009). Ubiquitination of histone H2B regulates chromatin dynamics by enhancing nucleosome stability. Proc. Natl. Acad. Sci. U. S. A. *106*, 16686–16691.
Chang, M., French-Cornay, D., Fan, H.Y., Klein, H., Denis, C.L., and Jaehning, J. a (1999). A complex containing RNA polymerase II, Paf1p, Cdc73p, Hpr1p, and Ccr4p plays a role in protein kinase C signaling. Mol. Cell. Biol. *19*, 1056–1067.
Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science *318*, 1780–1782.
Chathoth, K.T., Barrass, J.D., Webb, S., and Beggs, J.D. (2014). A Splicing-Dependent Transcriptional Checkpoint Associated with Prespliceosome Formation. Mol. Cell *53*, 779–790.
Cheung, A.C.M., Sainsbury, S., and Cramer, P. (2011). Structural basis of initial RNA polymerase II transcription. EMBO J. *30*, 4755–4763.
Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: The exosome. Biochim. Biophys. Acta - Gene Regul. Mech. *1829*, 552–560.
Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. *11*, 3319–3326.
Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. Mol. Cell *44*, 667–678.
Chu, Y., Simic, R., Warner, M.H., Arndt, K.M., and Prelich, G. (2007). Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. EMBO J. *26*, 4646–4656.
Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.
Clark, D.J. (2010). Nucleosome positioning, nucleosome spacing and the nucleosome code. J. Biomol. Struct. Dyn. *27*, 781–793.
Cloutier, S.C., Wang, S., Ma, W.K., Petell, C.J., and Tran, E.J. (2013). Long Noncoding RNAs Promote Transcriptional Poising of Inducible Genes. PLoS Biol. *11*, 32–34.
Cohen, B. a, Mitra, R.D., Hughes, J.D., and Church, G.M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat. Genet. *26*, 183–186.
Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.
Core, L.J., Waterfall, J.J., Gilchrist, D. a., Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012). Defining the Status of RNA Polymerase at Promoters. Cell Rep. *2*, 1025–1035.
Cowling, V.H. (2010). Regulation of mRNA cap methylation. Biochem. J. *425*, 295–302.

Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet. *7*.
Crisucci, E.M., and Arndt, K.M. (2012). Paf1 Restricts Gcn4 Occupancy and Antisense Transcription at the ARG1 Promoter. Mol. Cell. Biol. *32*, 1150–1163.
Custódio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H.S., Grosveld, F., and Antoniou, M. (1999). Inefficient processing impairs release of RNA from the site of transcription. EMBO J. *18*, 2855–2866.
Cyert, M.S. (2001). Genetic analysis of calmodulin and its targets in Saccharomyces cerevisiae. Annu. Rev. Genet. *35*, 647–672.
David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. Proc. Natl. Acad. Sci. U. S. A. *103*, 5320–5325.
Dieppois, G., and Stutz, F. (2010). Connecting the transcription site to the nuclear pore: a multi-tether process that regulates gene expression. J. Cell Sci. *123*, 1989–1999.
Van Dijk, E.L., Chen, C.L., D'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Ne, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114–117.
Dohmen, R.J., Wu, P., and Varshavsky, A. (1994). Heat-inducible degron: a method for constructing temperature-sensitive mutants. Science *263*, 1273–1276.
Donis-Keller, H. (1979). Site specific enzymatic cleavage of RNA. Nucleic Acids Res. *7*, 179–192.
Douziech, M., Forget, D., Greenblatt, J., and Coulombe, B. (1999). Topological localization of the carboxyl-terminal domain of RNA polymerase II in the initiation complex. J Biol Chem *274*, 19868–19873.
Dower, K., Kuperwasser, N., Merrikh, H., and Rosbash, M. (2004). A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. RNA *10*, 1888–1899.
Dronamraju, R., and Strahl, B.D. (2014). A feed forward circuit comprising Spt6, Ctk1 and PAF regulates Pol II CTD phosphorylation and transcription elongation. Nucleic Acids Res. *42*, 870–881.
Duina, A. a (2011). Histone Chaperones Spt6 and FACT: Similarities and Differences in Modes of Action at Transcribed Genes. Genet. Res. Int. *2011*, 625210.
Dunn, E.F., Hammell, C.M., Hodge, C.A., and Cole, C.N. (2005). Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. Genes Dev. *19*, 90–103.
Eick, D., and Geyer, M. (2013). The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. Chem Rev *113*, 8456–8490.
Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.-K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., et al. (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. Mol. Cell. Biol. *24*, 3324–3336.
Eser, P., Demel, C., Maier, K.C., Schwalb, B., Pirkl, N., Martin, D.E., Cramer, P., and Tresch, A. (2014). Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. Mol Syst Biol *10*, 717.
Fan, X., Lamarre-Vincent, N., Wang, Q., and Struhl, K. (2008). Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments. Nucleic Acids Res. *36*, 1–7.
Farlik, M., Sheffield, N.C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J., and Bock, C. (2015). Single-Cell DNA Methylome Sequencing and Bioinformatic Inference of Epigenomic Cell-State Dynamics. Cell Rep. 1386–1397.
Fernandes, a. R., Mira, N.P., Vargas, R.C., Canelhas, I., and Sá-Correia, I. (2005). Saccharomyces cerevisiae adaptation to weak acids involves the transcription

factor Haa1p and Haa1p-regulated genes. Biochem. Biophys. Res. Commun. *337*, 95–103.
Ferrari, P., and Strubin, M. (2015). Uncoupling histone turnover from transcription-associated histone H3 modifications. Nucleic Acids Res. 1–14.
Fishburn, J., Tomko, E., Galburt, E., and Hahn, S. (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. Proc. Natl. Acad. Sci. *112*, 3961–3966.
Fleming, A.B., Kao, C.F., Hillyer, C., Pikaart, M., and Osley, M.A. (2008). H2B Ubiquitylation Plays a Role in Nucleosome Dynamics during Transcription Elongation. Mol. Cell *31*, 57–66.
Fox, M.J., Gao, H., Smith-Kinnaman, W.R., Liu, Y., and Mosley, A.L. (2015). The Exosome Component Rrp6 Is Required for RNA Polymerase II Termination at Specific Targets of the Nrd1-Nab3 Pathway. PLOS Genet. *10*, e1004999.
Frank-Kamenetskii, M.D., and Mirkin, S.M. (1995). Triplex DNA structures. Annu Rev Biochem *64*, 65–95.
Ganley, A.R., Ide, S., Saka, K., and Kobayashi, T. (2009). The effect of replication initiation on gene amplification in the rDNA and its relationship to aging. Mol Cell *35*, 683–693.
García-Martínez, J., Aranda, A., and Pérez-Ortín, J.E. (2004). Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. Mol. Cell *15*, 303–313.
Gasch, a P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Mol. Biol. Cell *11*, 4241–4257.
Ge, X.Q., and Lin, H. (2014). Noncoding RNAs in the regulation of DNA replication. Trends Biochem. Sci. *39*, 341–343.
Geisberg, J. V., Moqtaderi, Z., Fan, X., Ozsolak, F., and Struhl, K. (2014). Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell *156*, 812–824.
Geisler, S., Lojek, L., Khalil, A.M., Baker, K.E., and Coller, J. (2012). Decapping of Long Noncoding RNAs Regulates Inducible Genes. Mol. Cell *45*, 279–291.
Ghosh, A., and Lima, C.D. (2010). Enzymology of RNA cap synthesis. Wiley Interdiscip. Rev. RNA *1*, 152–172.
Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: A platform for interactive large-scale genome analysis. Genome Res. *15*, 1451–1455.
Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science *292*, 1876–1882.
Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.
Goel, S., Krishnamurthy, S., and Hampsey, M. (2012). Mechanism of start site selection by RNA polymerase II: Interplay between TFIIB and Ssl2/XPB helicase subunit of TFIIH. J. Biol. Chem. *287*, 557–567.
Gonatopoulos-Pournatzis, T., and Cowling, V.H. (2014). Cap-binding complex (CBC). Biochem. J. *457*, 231–242.
Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional Spliceosome Assembly Occurs in a Stepwise Fashion and Requires the Cap Binding Complex. Mol. Cell *19*, 53–63.
Green, E.M., Jiang, Y., Joyner, R., and Weis, K. (2012). A negative feedback loop at the nuclear periphery regulates GAL gene expression. Mol. Biol. Cell *23*, 1367–1375.
Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R.H., Michaelis, J., and Werner, F. (2011). The Initiation Factor TFE and the Elongation Factor

Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation. Mol. Cell *43*, 263–274.
Grzechnik, P., Gdula, M.R., and Proudfoot, N.J. (2015). Pcf11 orchestrates transcription termination pathways in yeast. Genes Dev *29*, 849–861.
Gu, M., Rajashankar, K.R., and Lima, C.D. (2010). Structure of the Saccharomyces cerevisiae Cet1-Ceg1 mRNA Capping Apparatus. Structure *18*, 216–227.
Guo, Z., and Sherman, F. (1996). 3'-end-forming signals of yeast mRNA. Trends Biochem. Sci. *21*, 477–481.
El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev *24*, 1546–1558.
Hahn, S., and Young, E.T. (2011). Transcriptional regulation in saccharomyces cerevisiae: Transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705–736.
Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millán-Zambrano, G., Barkai, O., Chávez, S., Pérez-Ortín, J.E., Darzacq, X., and Choder, M. (2013). Gene Expression Is Circular: Factors for mRNA Degradation Also Foster mRNA Synthesis. Cell *153*, 1000–1011.
Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J., et al. (2004). Scoring and Clustering and Conservation Testing Assignment of Single Motif to Each Regulator. Nature *431*, 1–5.
Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that Specify Promoter Nucleosome Location and Identity. Cell *137*, 445–458.
Hartzog, G. a., and Fu, J. (2013). The Spt4-Spt5 complex: A multi-faceted regulator of transcription elongation. Biochim. Biophys. Acta - Gene Regul. Mech. *1829*, 105–115.
Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes. Mol. Cell *31*, 925–932.
Heidemann, M., and Eick, D. (2012). Tyrosine-1 and threonine-4 phosphorylation marks complete the RNA polymerase II CTD phospho-code. RNA Biol. *9*, 1144–1146.
Heo, D.H., Yoo, I., Kong, J., Lidschreiber, M., Mayer, A., Choi, B.Y., Hahn, Y., Cramer, P., Buratowski, S., and Kim, M. (2013). The RNA polymerase II C-terminal domain-interacting domain of yeast Nrd1 contributes to the choice of termination pathway and couples to RNA processing by the nuclear exosome. J Biol Chem *288*, 36676–36690.
Hieronymus, H., and Silver, P.A. (2003). Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. Nat Genet *33*, 155–161.
Ho, J.W.K., Bishop, E., Karchenko, P. V, Nègre, N., White, K.P., and Park, P.J. (2011). ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. BMC Genomics *12*, 134.
Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA Polymerase II Collision Interrupts Convergent Transcription. Mol. Cell *48*, 365–374.
Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E.T., Nijmeijer, B., Kotthoff, C., Rybin, V., Amlacher, S., et al. (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. Nature *499*, 111–114.
Hongay, C.F., Grisafi, P.L., Galitski, T., and Fink, G.R. (2006). Antisense Transcription Controls Cell Fate in Saccharomyces cerevisiae. Cell *127*, 735–745.
Houseley, J., and Tollervey, D. (2009). The many pathways of RNA degradation. Cell *136*, 763–776.
Houseley, J., Kotovic, K., El Hage, A., and Tollervey, D. (2007). Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. EMBO J *26*, 4996–5006.

Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., and Vogelauer, M. (2008). A ncRNA Modulates Histone Modification and mRNA Induction in the Yeast GAL Gene Cluster. Mol. Cell *32*, 685–695.

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

Hurt, E., Luo, M.-J., Röther, S., Reed, R., and Strässer, K. (2004). Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. Proc. Natl. Acad. Sci. U. S. A. *101*, 1858–1862.

Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A.H., Dargemont, C., and Stutz, F. (2010). Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes Dev. *24*, 1927–1938.

Izban, M.G., and Luse, D.S. (1992). Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. J. Biol. Chem. *267*, 13647–13655.

Jain, A., Bacolla, A., Chakraborty, P., Grosse, F., and Vasquez, K.M. (2010). Human DHX9 helicase unwinds triple-helical DNA structures. Biochemistry *49*, 6992–6999.

Jamonnak, N., Creamer, T.J., Darby, M.M., Schaughency, P., Wheelan, S.J., and Corden, J.L. (2011). Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. RNA *17*, 2011–2025.

Johnson, S.A., Cubberley, G., and Bentley, D.L. (2009). Cotranscriptional Recruitment of the mRNA Export Factor Yra1 by Direct Interaction with the 3??? End Processing Factor Pcf11. Mol. Cell *33*, 215–226.

Johnston, M., Hillier, L., Riles, L., Albermann, K., Andre, B., Ansorge, W., Benes, V., Bruckner, M., Delius, H., Dubois, E., et al. (1997). The nucleotide sequence of Saccharomyces cerevisiae chromosome XII. Nature *387*, 87–90.

Jorgensen, P., Edgington, N.P., Schneider, B.L., Rupes, I., Tyers, M., and Futcher, B. (2007). The size of the nucleus increases as yeast cells grow. Mol. Biol. Cell *18*, 3523–3532.

El Kaderi, B., Medler, S., Raghunayakula, S., and Ansari, A. (2009). Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. J. Biol. Chem. *284*, 25015–25025.

Kao, C.F., Hillyer, C., Tsukuda, T., Henry, K., Berger, S., and Osley, M.A. (2004). Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. Genes Dev. *18*, 184–195.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096–1099.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362–366.

Kawashima, T., Douglass, S., Gabunilas, J., Pellegrini, M., and Chanfreau, G.F. (2014). Widespread use of non-productive alternative splice sites in Saccharomyces cerevisiae. PLoS Genet. *10*, e1004249.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013). Function of alternative splicing. Gene *514*, 1–30.

Keller, G., Ray, E., Brown, P.O., and Winge, D.R. (2001). Haa1, a Protein Homologous to the Copper-regulated Transcription Factor Ace1, is a Novel Transcriptional Activator. J. Biol. Chem. *276*, 38697–38702.

Keogh, M.C., Kurdistani, S.K., Morris, S. a., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., et al. (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593–605.

Kettenberger, H., Armache, K.J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol. Cell *16*, 955–965.
Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat Struct Mol Biol *17*, 1279–1286.
Kim, J., Kim, J.A., McGinty, R.K., Nguyen, U.T.T., Muir, T.W., Allis, C.D., and Roeder, R.G. (2013). The n-SET Domain of Set1 Regulates H2B Ubiquitylation-Dependent H3K4 Methylation. Mol. Cell *49*, 1121–1133.
Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature *365*, 520–527.
Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517–522.
Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct Pathways for snoRNA and mRNA Termination. Mol. Cell *24*, 723–734.
Kobayashi, T. (2003). The replication fork barrier site forms a unique structure with Fob1p and inhibits the replication fork. Mol. Cell. Biol. *23*, 9178–9188.
Kobayashi, T., and Ganley, A.R. (2005). Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. Science (80-.). *309*, 1581–1584.
Kobayashi, T., Heck, D.J., Nomura, M., and Horiuchi, T. (1998). Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. Genes Dev *12*, 3821–3830.
Kostrewa, D., Zeller, M.E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature *462*, 323–330.
Kowalik, K.M., Shimada, Y., Flury, V., Stadler, M.B., Batki, J., and Bühler, M. (2015). The Paf1 complex represses small-RNA-mediated epigenetic gene silencing. Nature.
Krishnamurthy, S., Ghazy, M. a, Moore, C., and Hampsey, M. (2009). Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. Mol. Cell. Biol. *29*, 2925–2934.
Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., et al. (2003a). Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Mol. Cell. Biol. *23*, 4207–4218.
Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., et al. (2003b). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: Linking transcriptional elongation to histone methylation. Mol. Cell *11*, 721–729.
Kuang, Z., Cai, L., Zhang, X., Ji, H., Tu, B.P., and Boeke, J.D. (2014). High-temporal-resolution view of transcription and chromatin states across distinct metabolic states in budding yeast. Nat. Struct. Mol. Biol. *21*, 854–863.
Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. Genes Dev. *26*, 1891–1896.
Kudlicki, A., Rowicka, M., and Otwinowski, Z. (2007). SCEPTRANS: An online tool for analyzing periodic transcription in yeast. Bioinformatics *23*, 1559–1561.
Kunkel, G.R., and Martinson, H.G. (1981). Nucleosomes will not form on double-stranded RNa or over poly(dA).poly(dT) tracts in recombinant DNA. Nucleic Acids Res. *9*, 6869–6888.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.
Lacadie, S. a., and Rosbash, M. (2005). Cotranscriptional Spliceosome Assembly Dynamics and the Role of U1 snRNA:5′ss Base Pairing in Yeast. Mol. Cell *19*, 65–75.
Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr. Protoc. Bioinforma.
Lardenois, A., Liu, Y., Walther, T., Chalmel, F., Evrard, B., Granovskaia, M., Chu, A., Davis, R.W., Steinmetz, L.M., and Primig, M. (2011). Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. Proc. Natl. Acad. Sci. U. S. A. *108*, 1058–1063.
Laxman, S., Sutter, B.M., and Tu, B.P. (2010). Behavior of a metabolic cycling population at the single cell level as visualized by fluorescent gene expression reporters. PLoS One *5*, 1–11.
Lebreton, A., Tomecki, R., Dziembowski, A., and Séraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature *456*, 993–996.
Lee, J.T. (2011). Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. Nat. Rev. Mol. Cell Biol. *12*, 815–826.
Lee, J.M., and Sonnhammer, E.L.L. (2003). Genomic gene clustering analysis of pathways in eukaryotes. Genome Res. *13*, 875–882.
Lee, J.S., Burkholder, G.D., Latimer, L.J., Haug, B.L., and Braun, R.P. (1987). A monoclonal antibody to triplex DNA binds to eucaryotic chromosomes. Nucleic Acids Res *15*, 1047–1061.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
Li, M., Phatnani, H.P., Guan, Z., Sage, H., Greenleaf, A.L., and Zhou, P. (2005). Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. Proc. Natl. Acad. Sci. U. S. A. *102*, 17636–17641.
Lidschreiber, M., Leike, K., and Cramer, P. (2013). Cap Completion and C-Terminal Repeat Domain Kinase Recruitment Underlie the Initiation-Elongation Transition of RNA Polymerase II. Mol Cell Biol *33*, 3805–3816.
Liu, Y., and Lu, X. (2012). Non-coding RNAs in DNA damage response. Am. J. Cancer Res. *2*, 658–675.
Liu, J., Zhang, J., Gong, Q., Xiong, P., Huang, H., Wu, B., Lu, G., Wu, J., and Shi, Y. (2011a). Solution structure of tandem SH2 domains from Spt6 protein and their binding to the phosphorylated rna polymerase II C-terminal domain. J. Biol. Chem. *286*, 29218–29226.
Liu, X., Bushnell, D.A., Wang, D., Calero, G., and Kornberg, R.D. (2010). Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. Science *327*, 206–209.
Liu, X., Bushnell, D.A., Silva, D.-A., Huang, X., and Kornberg, R.D. (2011b). Initiation complex structure and promoter proofreading. Science *333*, 633–637.
Longtine, M.S., McKenzie, A., Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953–961.
López-Maury, L., Marguerat, S., and Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat. Rev. Genet. *9*, 583–593.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. bioRxiv 1–21.

Lubliner, S., Keren, L., and Segal, E. (2013). Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. Nucleic Acids Res *41*, 5569–5581.
Luger, K., Mäder, a W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.
Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. *17*, 1195–1201.
Luo, W., Johnson, A.W., and Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: Implications for a unified allosteric-torpedo model. Genes Dev. *20*, 954–965.
Lykke-Andersen, S., and Jensen, T.H. (2007). Overlapping pathways dictate termination of RNA polymerase II transcription. Biochimie *89*, 1177–1182.
Machné, R., and Murray, D.B. (2012). The yin and yang of yeast transcription: Elements of a global feedback system between metabolism and chromatin. PLoS One *7*.
MacKellar, A.L., and Greenleaf, A.L. (2011). Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II. J Biol Chem *286*, 36385–36395.
Mahony, S., and Pugh, B.F. (2015). Protein–DNA binding in high-resolution. Crit. Rev. Biochem. Mol. Biol. 1–15.
Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mRNA decay. Elife *4*, 1–24.
Mao, X., Schwer, B., and Shuman, S. (1995). Yeast mRNA cap methyltransferase is a 50-kilodalton protein encoded by an essential gene. Mol. Cell. Biol. *15*, 4167–4174.
Marquardt, S., Hazelbaker, D.Z., and Buratowski, S. (2011). Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. Transcription *2*, 145–154.
Marquardt, S., Escalante-Chong, R., Pho, N., Wang, J., Churchman, L.S., Springer, M., and Buratowski, S. (2014). A chromatin-based mechanism for limiting divergent noncoding transcription. Cell *157*, 1712–1723.
Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature *429*, 571–574.
Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature *445*, 666–670.
Martinez-Rucobo, F.W., and Cramer, P. (2013). Structural basis of transcription elongation. Biochim. Biophys. Acta *1829*, 9–19.
Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C.M., and Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J. *30*, 1302–1310.
Martinez-Rucobo, F.W., Kohler, R., van de Waterbeemd, M., Heck, A.J.R., Hemann, M., Herzog, F., Stark, H., and Cramer, P. (2015). Molecular Basis of Transcription-Coupled Pre-mRNA Capping. Mol. Cell *58*, 1079–1089.
Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. *18*, 1073–1083.
Mayekar, M.K., Gardner, R.G., and Arndt, K.M. (2013). The recruitment of the Saccharomyces cerevisiae Paf1 complex to active genes requires a domain of Rtf1 that directly interacts with the Spt4-Spt5 complex. Mol. Cell. Biol. *33*, 3259–3273.

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

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science *336*, 1723–1725.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J. a, and Churchman, L.S. (2015). Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution. Cell *161*, 541–554.

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

Meinel, D.M., Burkert-Kautzsch, C., Kieser, A., O'Duibhir, E., Siebert, M., Mayer, A., Cramer, P., Soding, J., Holstege, F.C., and Strasser, K. (2013). Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II. PLoS Genet *9*, e1003914.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature *430*, 223–226.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. (2005). A structural perspective of CTD function. Genes Dev. *19*, 1401–1415.

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dumcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dolken, L., et al. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol *7*, 458.

Mischo, H.E., and Proudfoot, N.J. (2013). Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast. Biochim. Biophys. Acta - Gene Regul. Mech. *1829*, 174–185.

Morales, V., and Richard-Foy, H. (2000). Role of histone N-terminal tails and their acetylation in nucleosome dynamics. Mol. Cell. Biol. *20*, 7230–7237.

Morano, K. a., Grant, C.M., and Moye-Rowley, W.S. (2012). The response to heat shock and oxidative stress in saccharomyces cerevisiae. Genetics *190*, 1157–1195.

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

Mueller, C.L., Porter, S.E., Hoffman, M.G., and Jaehning, J. a. (2004). The Paf1 complex has functions independent of actively transcribing RNA polymerase II. Mol. Cell *14*, 447–456.

Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J., et al. (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. Science (80-.). *342*, 1238724.

Murray, S.C., Serra Barros, A., Brown, D.A., Dudek, P., Ayling, J., and Mellor, J. (2012). A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. Nucleic Acids Res *40*, 2432–2444.

Murray, S.C., Haenni, S., Howe, F.S., Fischl, H., Chocian, K., Nair, A., and Mellor, J. (2015). Sense and antisense transcription are associated with distinct chromatin architectures across genes. Nucleic Acids Res. *43*, 7823–7837.

Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C.L., and Greenblatt, J. (2003). Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. J. Biol. Chem. *278*, 33000–33010.

Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. (2008). The Glc7 Phosphatase Subunit of the Cleavage and Polyadenylation Factor Is Essential for Transcription Termination on snoRNA Genes. Mol. Cell *29*, 577–587.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038–1042.

Ng, H.H., Dole, S., and Struhl, K. (2003a). The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. J Biol Chem *278*, 33625–33628.

Ng, H.H., Robert, F., Young, R. a., and Struhl, K. (2003b). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell *11*, 709–719.

Nguyen, T., Fischl, H., Howe, F.S., Woloszczuk, R., Serra Barros, A., Xu, Z., Brown, D., Murray, S.C., Haenni, S., Halstead, J.M., et al. (2014). Transcription mediated insulation and interference direct gene cluster expression switches. Elife *3*, 1–21.

Noble, C.G., Hollingworth, D., Martin, S.R., Ennis-Adeniran, V., Smerdon, S.J., Kelly, G., Taylor, I. a, and Ramos, A. (2005). Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nat. Struct. Mol. Biol. *12*, 144–151.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526–540.

O'Duibhir, E., Lijnzaad, P., Benschop, J.J., Lenstra, T.L., van Leenen, D., Groot Koerkamp, M.J. a, Margaritis, T., Brok, M.O., Kemmeren, P., and Holstege, F.C.P. (2014). Cell cycle population effects in perturbation studies. Mol. Syst. Biol. *10*, 732.

O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N.J. (2004). Gene loops juxtapose promoters and terminators in yeast. Nat. Genet. *36*, 1014–1018.

Parker, R. (2012). RNA degradation in Saccharomyces cerevisae. Genetics *191*, 671–702.

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B Monoubiquitination Functions Cooperatively with FACT to Regulate Elongation by RNA Polymerase II. Cell *125*, 703–717.

Pearson, E., and Moore, C. (2014). The Evolutionarily Conserved Pol II Flap Loop Contributes to Proper Transcription Termination on Short Yeast Genes. Cell Rep. *9*, 821–828.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127–131.

Perales, R., and Bentley, D. (2009). "Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions. Mol. Cell *36*, 178–191.

Petty, E., and Pillus, L. (2013). Balancing chromatin remodeling and histone modifications in transcription. Trends Genet. *29*, 621–629.

Phatnani, H.P., and Greenleaf, a L. (2006). Phosphorylation and functions of the RNA polymerase IICTD. Genes Dev. *20*, 2922–2936.

Pinskaya, M., Gourvennec, S., and Morillon, A. (2009). H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. EMBO J. *28*, 1697–1707.

Piro, a. S., Mayekar, M.K., Warner, M.H., Davis, C.P., and Arndt, K.M. (2012). Small region of Rtf1 protein can substitute for complete Paf1 complex in facilitating global histone H2B ubiquitylation in yeast. Proc. Natl. Acad. Sci. *109*, 10837–10842.

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E. V., Borchers, C.H., Baumeister, W., Herzog, F., et al. (2015). Architecture of the RNA polymerase II–Mediator core initiation complex. Nature.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Tong, I.L., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517–527.
Porrua, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nat. Struct. Mol. Biol. *20*, 884–891.
Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nat Rev Mol Cell Biol *16*, 190–202.
Qi, L.S., Larson, M.H., Gilbert, L. a, Doudna, J. a, Weissman, J.S., Arkin, A.P., and Lim, W. a (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.
Qiu, H., Hu, C., and Hinnebusch, A.G. (2009). Phosphorylation of the Pol II CTD by KIN28 Enhances BUR1/BUR2 Recruitment and Ser2 CTD Phosphorylation Near Promoters. Mol. Cell *33*, 752–762.
Qiu, H., Hu, C., Gaur, N. a, and Hinnebusch, A.G. (2012). Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. EMBO J. *31*, 3494–3505.
Quinodoz, S., and Guttman, M. (2014). Long noncoding RNAs : an emerging link between gene regulation and nuclear organization. Trends Cell Biol. *24*, 651–663.
Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, 1707–1719.
Rando, O.J. (2007). Global patterns of histone modifications. Curr. Opin. Genet. Dev. *17*, 94–99.
Rao, A.R., and Pellegrini, M. (2011). Regulation of the yeast metabolic cycle by transcription factors with periodic activities. BMC Syst. Biol. *5*, 160.
Rao, B., Shibata, Y., Strahl, B.D., and Lieb, J.D. (2005). Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. Mol. Cell. Biol. *25*, 9447–9459.
Reiter, A., Hamperl, S., Seitz, H., Merkl, P., Perez-Fernandez, J., Williams, L., Gerber, J., Nemeth, A., Leger, I., Gadal, O., et al. (2012). The Reb1-homologue Ydr026c/Nsi1 is required for efficient RNA polymerase I termination in yeast. EMBO J *31*, 3480–3493.
Rhee, H.S., and Pugh, B.F. (2012a). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.
Rhee, H.S., and Pugh, B.F. (2012b). ChiP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. Curr. Protoc. Mol. Biol. 1–14.
Richmond, T.J., Richmond, T.J., Davey, C. a, and Davey, C. a (2003). The structure of DNA in the nucleosome core. Nature *423*, 145–150.
Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.
Rodriguez-Campos, A., and Azorin, F. (2007). RNA is an integral component of chromatin that contributes to its structural organization. PLoS One *2*, e1182.
Rondón, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-Safe Transcriptional Termination for Protein-Coding Genes in S. cerevisiae. Mol. Cell *36*, 88–98.
Ruiz-Roig, C., Noriega, N., Duch, a., Posas, F., and de Nadal, E. (2012). The Hog1 SAPK controls the Rtg1/Rtg3 transcriptional complex activity by multiple regulatory mechanisms. Mol. Biol. Cell *23*, 4286–4296.
Sambrook, J., Fritsch, E.F., Maniatis, T., and Fritschi, E.F. (1989). Molecular Cloning: A Laboratory Manual.
Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. Nat. Rev. Mol. Cell Biol. *7*, 557–567.
Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-wide mapping of yeast RNA polymerase II termination. PLoS Genet *10*, e1004632.

Schmid, M., Poulsen, M.B., Olszewski, P., Pelechano, V., Saguez, C., Gupta, I., Steinmetz, L.M., Moore, C., and Jensen, T.H. (2012). Rrp6p Controls mRNA Poly(A) Tail Length and Its Decoration with Poly(A) Binding Proteins. Mol. Cell *47*, 267–280.
Schmitz, K.M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev *24*, 2264–2269.
Schneider, M., Hellerschmied, D., Schubert, T., Amlacher, S., Vinayachandran, V., Reja, R., Pugh, B.F., Clausen, T., and Köhler, A. (2015). The Nuclear Pore-Associated TREX-2 Complex Employs Mediator to Regulate Gene Expression. Cell *162*, 1016–1028.
Schreieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L. a (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. Nat. Struct. Mol. Biol. *21*, 175–179.
Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell *155*, 1075–1087.
Schulze, J.M., Hentrich, T., Nakanishi, S., Gupta, A., Emberly, E., Shilatifard, A., and Kobor, M.S. (2011). Splitting the task: Ubp8 and ubp10 deubiquitinate different cellular pools of H2BK123. Genes Dev. *25*, 2242–2247.
Schwer, B., and Shuman, S. (2011). Deciphering the RNA Polymerase II CTD Code in Fission Yeast. Mol. Cell *43*, 311–318.
Segal, E., and Widom, J. (2009). From DNA sequence to transcriptional behaviour: a quantitative approach. Nat. Rev. Genet. *10*, 443–456.
Segref, A., Sharma, K., Doye, V., Hellwig, A., Huber, J., Lührmann, R., and Hurt, E. (1997). Mex67p, a novel factor for nuclear mRNA export. Binds to both poly(A)+ RNA and nuclear pores. EMBO J. *16*, 3256–3271.
Semerad, C.L., and Maher, L.J. (1994). Exclusion of RNA strands from a purine motif triple helix. Nucleic Acids Res. *22*, 5321–5325.
Shchyolkina, A.K., Timofeev, E.N., Borisova, O.F., Il'icheva, I.A., Minyat, E.E., Khomyakova, E.B., and Florentiev, V.L. (1994). The R-form of DNA does exist. FEBS Lett *339*, 113–118.
Shi, L., Sutter, B.M., Ye, X., and Tu, B.P. (2010). Trehalose is a key determinant of the quiescent metabolic state that fuels cell cycle progression upon return to growth. Mol. Biol. Cell *21*, 1982–1990.
Shibagaki, Y., Itoh, N., Yamada, H., Nagata, S., and Mizumoto, K. (1992). mRNA capping enzyme. Isolation and characterization of the gene encoding mRNA guanylytransferase subunit from Saccharomyces cerevisiae. J. Biol. Chem. *267*, 9521–9528.
Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu. Rev. Biochem. *75*, 243–269.
Silverman, S.J., Petti, A.A., Slavov, N., Parsons, L., Briehof, R., Thiberge, S.Y., Zenklusen, D., Gandhi, S.J., Larson, D.R., Singer, R.H., et al. (2010). Metabolic cycling in single yeast cells from unsynchronized steady-state populations limited on glucose or phosphate. Proc. Natl. Acad. Sci. U. S. A. *107*, 6946–6951.
Simon, M.D., Wang, C.I., Kharchenko, P. V., West, J. a., Chapman, B. a., Alekseyenko, a. a., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. Proc. Natl. Acad. Sci. *108*, 20497–20502.
Sinclair, D.A., and Guarente, L. (1997). Extrachromosomal rDNA circles--a cause of aging in yeast. Cell *91*, 1033–1042.
Singh, J., and Padgett, R. a (2009). Rates of in situ transcription and splicing in large human genes. Nat. Struct. & Mol. Biol. *16*, 1128–1133.
Singh, N., Ma, Z., Gemmill, T., Wu, X., DeFiglio, H., Rossettini, A., Rabeler, C., Beane, O., Morse, R.H., Palumbo, M.J., et al. (2009). The Ess1 Prolyl Isomerase Is Required

for Transcription Termination of Small Noncoding RNAs via the Nrd1 Pathway. Mol. Cell *36*, 255–266.
Sinturel, F., Navickas, A., Wery, M., Descrimes, M., Morillon, A., Torchet, C., and Benard, L. (2015). Cytoplasmic Control of Sense-Antisense mRNA Pairs. Cell Rep. 1–12.
Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. *28*, 1384–1396.
Slavov, N., and Botstein, D. (2011). Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast. Mol. Biol. Cell *22*, 1997–2009.
Slavov, N., Macinskas, J., Caudy, a., and Botstein, D. (2011). Metabolic cycling without cell division cycling in respiring yeast. Proc. Natl. Acad. Sci. *108*, 19090–19095.
Slavov, N., Airoldi, E.M., van Oudenaarden, a., and Botstein, D. (2012). A conserved cell growth cycle can account for the environmental stress responses of divergent eukaryotes. Mol. Biol. Cell *23*, 1986–1997.
Smale, S.T. (2009). Nuclear run-on assay. Cold Spring Harb. Protoc. *4*, 1–8.
Smith, J.E., Alvarez-Dominguez, J.R., Kline, N., Huynh, N.J., Geisler, S., Hu, W., Coller, J., and Baker, K.E. (2014). Translation of Small Open Reading Frames within Unannotated RNA Transcripts in Saccharomyces cerevisiae. Cell Rep. *7*, 1858–1866.
Smolle, M., and Workman, J.L. (2013). Transcription-associated histone modifications and cryptic transcription. Biochim. Biophys. Acta - Gene Regul. Mech. *1829*, 84–97.
Smolle, M., Venkatesh, S., Gogol, M.M., Li, H., Zhang, Y., Florens, L., Washburn, M.P., and Workman, J.L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. Nat. Struct. Mol. Biol. *19*, 884–892.
Sood, V., and Brickner, J.H. (2014). Nuclear pore interactions with the genome. Curr Opin Genet Dev *25*, 43–49.
Squazzo, S.L., Costa, P.J., Lindstrom, D.L., Kumer, K.E., Simic, R., Jennings, J.L., Link, A.J., Arndt, K.M., and Hartzog, G. a. (2002). The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. EMBO J. *21*, 1764–1774.
Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol Cell *24*, 735–746.
Strahl, B.D., Grant, P. a, Briggs, S.D., Sun, Z.-W., Bone, J.R., Caldwell, J. a, Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., et al. (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol. Cell. Biol. *22*, 1298–1306.
Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol *14*, 103–105.
Sugimoto, N., Wu, P., Hara, H., and Kawamoto, Y. (2001). pH and cation effects on the properties of parallel pyrimidine motif DNA triplexes. Biochemistry *40*, 9396–9405.
Suter, B., Schnappauf, G., and Thoma, F. (2000). Poly(dA.dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. Nucleic Acids Res. *28*, 4083–4089.
Szewczak, A.A., Ortoleva-Donnelly, L., Ryder, S.P., Moncoeur, E., and Strobel, S.A. (1998). A minor groove RNA triple helix within the catalytic core of a group I intron. Nat Struct Biol *5*, 1037–1042.
Taddei, A., Van Houwe, G., Hediger, F., Kalck, V., Cubizolles, F., Schober, H., and Gasser, S.M. (2006). Nuclear pore association confers optimal expression levels for an inducible yeast gene. Nature *441*, 774–778.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

Tardiff, D.F., Lacadie, S. a., and Rosbash, M. (2006). A Genome-Wide Analysis Indicates that Yeast Pre-mRNA Splicing Is Predominantly Posttranscriptional. Mol. Cell *24*, 917–929.

Terzi, N., Churchman, L.S., Vasiljeva, L., Weissman, J., and Buratowski, S. (2011). H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. Mol Cell Biol *31*, 3569–3583.

Thebault, P., Boutin, G., Bhat, W., Rufiange, A., Martens, J., and Nourani, A. (2011). Transcription regulation by the noncoding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. Mol. Cell. Biol. *31*, 1288–1300.

Thomas, M., White, R.L., and Davis, R.W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc. Natl. Acad. Sci. U. S. A. *73*, 2294–2298.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S., et al. (2010). Chemical-genomic dissection of the CTD code. Nat Struct Mol Biol *17*, 1154–1161.

Timson, D.J. (2007). Galactose Metabolism in Saccharomyces cerevisiae. Dyn. Biochem. Process Biotechnol. Mol. Biol. 63–73.

Topisirovic, I., Svitkin, Y. V., Sonenberg, N., and Shatkin, A.J. (2011). Cap and cap-binding proteins in the control of gene expression. Wiley Interdiscip. Rev. RNA *2*, 277–298.

Travers, A.A., and Klug, A. (1987). The bending of DNA in nucleosomes and its wider implications. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *317*, 537–561.

Trcek, T., Larson, D.R., Moldón, A., Query, C.C., and Singer, R.H. (2011). Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. Cell *147*, 1484–1497.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689–693.

Tsukamoto, T., Shibagaki, Y., Imajoh-Ohmi, S., Murakoshi, T., Suzuki, M., Nakamura, a, Gotoh, H., and Mizumoto, K. (1997). Isolation and characterization of the yeast mRNA capping enzyme beta subunit gene encoding RNA 5'-triphosphatase, which is essential for cell viability. Biochem. Biophys. Res. Commun. *239*, 116–122.

Tu, B.P., Kudlicki, A., Rowicka, M., and McKnight, S.L. (2005). Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. Science *310*, 1152–1158.

Tu, B.P., Mohler, R.E., Liu, J.C., Dombek, K.M., Young, E.T., Synovec, R.E., and McKnight, S.L. (2007). Cyclic changes in metabolic state during the life of a yeast cell. Proc. Natl. Acad. Sci. U. S. A. *104*, 16886–16891.

Tuck, A.C., and Tollervey, D. (2013). A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. Cell *154*, 996–1009.

Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., et al. (2014). Molecular basis for coordinating transcription termination with noncoding RNA degradation. Mol Cell *55*, 467–481.

Uhler, J.P., Hertel, C., and Svejstrup, J.Q. (2007). A role for noncoding transcription in activation of the yeast PHO5 gene. Proc. Natl. Acad. Sci. U. S. A. *104*, 8011–8016.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537–1550.
Vance, K.W., and Ponting, C.P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends Genet. *30*, 348–355.
Vasiljeva, L., and Buratowski, S. (2006). Nrd1 Interacts with the Nuclear Exosome for 3′ Processing of RNA Polymerase II Transcripts. Mol. Cell *21*, 239–248.
Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat Struct Mol Biol *15*, 795–804.
Venkatesh, S., Smolle, M., Li, H., Gogol, M.M., Saint, M., Kumar, S., Natarajan, K., and Workman, J.L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature *489*, 452–455.
Venters, B.J., and Pugh, B.F. (2009). How eukaryotic genes are transcribed. Crit Rev Biochem Mol Biol *44*, 117–141.
Vitaliano-Prunier, A., Menant, A., Hobeika, M., Géli, V., Gwizdek, C., and Dargemont, C. (2008). Ubiquitylation of the COMPASS component Swd2 links H2B ubiquitylation to H3K4 trimethylation. Nat. Cell Biol. *10*, 1365–1371.
Vitaliano-Prunier, A., Babour, A., Hérissant, L., Apponi, L., Margaritis, T., Holstege, F.C.P., Corbett, A.H., Gwizdek, C., and Dargemont, C. (2012). H2B ubiquitylation controls the formation of export-competent mRNP. Mol. Cell *45*, 132–139.
Vodala, S., Abruzzi, K.C., and Rosbash, M. (2008). The Nuclear Exosome and Adenylation Regulate Posttranscriptional Tethering of Yeast GAL Genes to the Nuclear Periphery. Mol. Cell *31*, 104–113.
Vojnic, E., Simon, B., Strahl, B.D., Sattler, M., and Cramer, P. (2006). Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. J. Biol. Chem. *281*, 13–15.
Wada, Y., Ohta, Y., Xu, M., Tsutsumi, S., Minami, T., Inoue, K., Komura, D., Kitakami, J., Oshida, N., Papantonis, A., et al. (2009). A wave of nascent transcription on activated human genes. Proc. Natl. Acad. Sci. U. S. A. *106*, 18357–18361.
Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.
Wang, L., Jiang, N., Wang, L., Fang, O., Leach, L.J., Hu, X., and Luo, Z. (2014). 3' Untranslated regions mediate transcriptional interference between convergent genes both locally and ectopically in Saccharomyces cerevisiae. PLoS Genet. *10*.
Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. Trends Biochem. Sci. *24*, 437–440.
Webb, S., Hector, R.D., Kudla, G., and Granneman, S. (2014). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. Genome Biol *15*, R8.
Werner-Allen, J.W., Lee, C.J., Liu, P., Nicely, N.I., Wang, S., Greenleaf, A.L., and Zhou, P. (2011). cis-proline-mediated ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. J. Biol. Chem. *286*, 5717–5726.
Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. Mol. Cell *54*, 601–612.
Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2003). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. J. Biol. Chem. *278*, 34739–34742.
Xiao, T., Kao, C., Krogan, N.J., Greenblatt, J.F., Osley, M.A., Strahl, B.D., and Sun, Z. (2005). Histone H2B Ubiquitylation Is Associated with Elongating RNA Polymerase II Histone H2B Ubiquitylation Is Associated with Elongating RNA Polymerase II. Society *25*, 637–651.

Xin, H., Takahata, S., Blanksma, M., McCullough, L., Stillman, D.J., and Formosa, T. (2009). yFACT Induces Global Accessibility of Nucleosomal DNA without H2A-H2B Displacement. Mol. Cell *35*, 365–376.
Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.
Xu, Z., Wei, W., Gagneur, J., Clauder-Munster, S., Smolik, M., Huber, W., and Steinmetz, L.M. (2011). Antisense expression increases gene expression variability and locus interdependency. Mol Syst Biol *7*, 468.
Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T., and Pugh, B.F. (2012). Genome-wide nucleosome specificity and directionality of chromatin remodelers. Cell *149*, 1461–1473.
Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. Genes Dev *21*, 160–174.
Youdell, M.L., Kizer, K.O., Kisseleva-Romanova, E., Fuchs, S.M., Duro, E., Strahl, B.D., and Mellor, J. (2008). Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. Mol Cell Biol *28*, 4915–4926.
Yudkovsky, N., Ranish, J. a, and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. Nature *408*, 225–229.
Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell Res. *21*, 564–578.
Zaugg, J.B., and Luscombe, N.M. (2012). A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast. Genome Res. *22*, 84–94.
Zenklusen, D., Vinciguerra, P., Strahm, Y., and Stutz, F. (2001). The yeast hnRNP-Like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p. Mol. Cell. Biol. *21*, 4219–4232.
Zenklusen, D., Vinciguerra, P., Wyss, J.-C., and Stutz, F. (2002). Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. Mol. Cell. Biol. *22*, 8241–8253.
Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodríguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. (2012). Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. J. Biol. Chem. *287*, 8541–8551.
Zhang, Z., Fu, J., and Gilmour, D.S. (2005). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor, Pcf11. Genes Dev. *19*, 1572–1580.
Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol. Mol. Biol. Rev. *63*, 405–445.
Zhurkin, V.B., Raghunathan, G., Ulyanov, N.B., Camerini-Otero, R.D., and Jernigan, R.L. (1994). A parallel DNA triplex as a model for the intermediate in homologous recombination. J Mol Biol *239*, 181–200.
Zid, B.M., and O'Shea, E.K. (2014). Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. Nature *514*, 117–121.