Gene regulation by mRNA isoforms and ncRNAs during budding yeast gametogenesis

Minghao Chia

Imperial College London

and

The Francis Crick Institute

PhD Supervisor: Folkert van Werven

A thesis submitted for the degree of

Doctor of Philosophy

Imperial College London

September 2018

Declaration

I, Minghao Chia, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Cell differentiation is fundamental to multi-cellular life because a single unspecialized zygote must give rise to all specialized cell types in the body. Cell fate specification requires the coordinated regulation of different genes across developmental time. How genes are regulated during cell differentiation is a longstanding question in biology. Gametogenesis in budding yeast is a cell differentiation program where a diploid cell gives rise to four haploid gametes and is a tractable model to understand gene regulation during development. Multiple studies in various vertebrate systems suggest that a significant fraction of genes are expressed as different transcript isoforms arising from alternative transcription start sites, or overlap with non-coding RNAs (ncRNAs) of unknown function. Here, I investigated how transcript isoforms or overlapping ncRNAs regulate gene expression during budding yeast gametogenesis. First, I showed that the kinetochore gene *NDC80* is repressed by transcription of a 5' extended mRNA isoform from an alternative promoter during early gametogenesis. This repressive mechanism required the co-transcriptional deposition of histone modifications and the stabilization of nucleosomes at the canonical *NDC80* promoter. *NDC80* repression could also be rapidly reversed to allow cells to re-enter mitosis if they were no longer starved. Next, I showed how a genome-wide approach called transcript end sequencing (TE-seq) could distinguish the ends of overlapping transcripts. TE-seq analysis showed over a thousand meiotic upstream transcripts or isoforms upregulated at specific stages of gametogenesis. Upstream overlapping transcription was associated with different regulatory outcomes on coding genes, and did not always repress them. TE-seq also identified examples of gene regulation which depended on different chromatin remodellers. Finally, most meiotic upstream alternative transcripts were rapidly repressed upon return to growth, similar to the *NDC80* example. These results provide a basis for understanding how mRNA isoforms and ncRNAs contribute to local gene regulation during development in eukaryotes.

299 words

Acknowledgements

I would like to thank my PhD supervisor, Folkert van Werven for his gentle mentorship and patience over the past four years in his lab. Your guidance has trained me to think critically and your kindness has made this PhD experience so much more enjoyable.

I would also like to acknowledge all past and present members of the lab for being wonderful colleagues and friends. I am not an expressive person by any measure, but all of you have treated me warmly. A big thank you to Fabien Moretto, Gianpiero Spedale, Radhika Warrier, Dora Sideri, Andrew Wu and Janis Tam for bouncing off ideas during lab meetings, for help with troubleshooting experiments and for your encouragements along the way. I am grateful to my thesis committee members: Jesper Svejstrup, Paola Scaffidi and Luis Aragon for their scientific advice and for pointing me in the right direction.

The work in this thesis would not have been possible without the collaboration of numerous scientists. I would like to thank the members of the Ünal lab at UC Berkeley: Elcin Ünal, Jingxun Chen and Amy Tresenrider for their input and work on the *NDC80* project. A shout-out to Harshil Patel and Mickael Escudero from the Francis Crick Bioinformatics & Biostatistics team who have helped me with processing RNA seq data. I would like to especially acknowledge Cai, Li from the Luscombe lab for his input in TE-seq data analysis and for teaching me how to interrogate genomic data using the command line. Thank you to members of the advanced sequencing facility (ASF) and the equipment park at the Francis Crick Institute for facilitating my work. I thank the Agency for Science, Technology and Research (A*STAR) of Singapore for generously funding my studies.

I am deeply appreciative of all the support received from my relatives (especially my mother and my wife) and my friends. To my dear wife Joy, your affection, forebearance and affirmation have been invaluable over the past 4 years. Last but not least, I thank the Lord for His manifold mercies and for this chance to study His work.

Table of Contents

Table of figures

List of tables

Abbreviations

AID	Auxin induced degron
AMP	Adenosine monophosphate
ANOVA	Analysis of Variance
ATP	Adenosine Triphosphate
BSA	Bovine Serum Albumin
BYTA	Buffered Yeast Extract Tryptone Acetate medium
CAGE	Cap Analysis of Gene Expression
cAMP	Cyclic adenosine monophosphate
cDNA	Complementary DNA
CDS	Coding sequence
COMPASS	Complex Proteins Associated with Set1p
CTD	Carboxy-terminal domain
$CuSO_4$	Copper(II) Sulphate
CUT	Cryptic unstable transcripts
DAPI	4′,6-diamidino-2-phenylindole
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
FACT	facilitates chromatin transcription
FDR	False Discovery Rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GDP	Guanosine diphosphate
GEF	Guanine nucleotide exchange factor
GFP	Green Fluorescent Protein
GTP	Guanosine Triphosphate
HAT	Histone acetyltransferase
HCl	Hydrochloric acid
HRP	Horseradish peroxidase
IAA	Indole-3-acetic acid
IGV	Integrative Genomics Viewer
IVT	In-vitro transcripts
KOH	Potassium hydroxide
LUTI/luti	Long undecodable transcript isoform
MSE	Middle sporulation element
MUTI	Meiotic upstream transcripts or isoforms
NAD	Nicotinamide adenine dinucleotide
ncRNAs	non-coding RNAs
NDR	Nucleosome depleted region
NET	Nascent elongating transcripts

NFR	Nucleosome free region
NMD	Nonsense mediated decay
NNS	Nrd1-Nab3-Sen1
NUT	Nrd1-unterminated transcript
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PBST	Phophate buffered saline with Tween-20
PCR	Polymerase chain reaction
PEST	Pro, Glu (and Asp), Ser, and Thr motif
PET	Paired-end tag
PIC	Pre-initiation complex
PKA	Protein kinase A
PVDF	Polyvinylidene difluoride
RBP	RNA binding protein
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
RT	Reverse transcription
RTG	Return to growth
SAGA	Spt-Ada-Gcn5-acetyltransferase complex
SAM	S-Adenosyl methionine
SDS	Sodium dodecyl sulphate
SEM	Standard error of the mean
SNP	Single nucleotide polymorphism
SPO	Sporulation medium
SUT	Stable unannotated transcript
TBE	Tris/Borate/EDTA
TCA	Trichloroacetic acid
TE-seq	Transcript end sequencing
TE	Tris/EDTA
TES	Transcript end site
TI	Transcriptional interference
TIF-seq	Transcript isoform sequencing
TORC1	Target of rapamycin complex 1
TPM	Tags per million
TSS	Transcription start site
TTS	Transcription terminator site
TU	Transcription unit
URS1	Upstream repressive sequence 1
UTP	Uridine triphosphate
UTR	Un-translated region
XUT	Xrn1-sensitive unstable transcripts
YPD	Yeast extract Peptone Dextrose
YPG	Yeast extract Peptone Glycerol

Chapter 1. Introduction

1.1 Introduction to cell differentiation

Cell differentiation as defined in the Gene Ontology database, is “the process in which relatively unspecialized cells acquire specialized structural and/or functional features that characterize the cells, tissues, or organs of the mature organism or some other relatively stable phase of the organism's life history” (Ashburner et al. 2000;The Gene Ontology 2017). Differentiation is fundamental to the development of multi-cellular organisms because a single unspecialized zygote must eventually give rise to all the various specialized cell types in the body. In each differentiating cell, networks of transcriptional regulators (e.g. transcription factors) activate or repress genes to specify distinct cell fates (Furlong 2010). Understanding how cell differentiation programs are controlled is a longstanding question in biology.

1.1.1 Gametogenesis as a key differentiation program

Gametogenesis is a conserved cell differentiation program common to all sexually reproducing organisms. Gametogenesis comprises of a single round of DNA replication, recombination between homologous chromosomes and two consecutive rounds of nuclear divisions called meiosis I and II (Marston and Amon 2004). In meiosis I, homologous chromosomes, each consisting of a pair of sister chromatids joined at the centromere, are pulled to opposite poles of the dividing cell. Meiosis I is also called reduction division because the number of chromosomes per cell is halved after chromosome segregation (Ohkura 2015). The products of meiosis I then undergo meiosis II, during which the sister chromatids are segregated to opposite poles of the cell. Similar to mitosis, meiosis II is an equational nuclear division because the number of chromosomes per cell remains the same after sister chromatid segregation (Ohkura 2015). As a result of gametogenesis, a single diploid parental cell generates four haploid daughter cells or gametes with half the number of chromosomes as the parent. Gametogenesis is a key differentiation program for sexually reproducing organisms because it maintains chromosome number between generations; the union of two haploid gametes during mating forms a zygote with the diploid number of chromosomes

once more. Another important function of gametogenesis is to generate novel combinations of alleles along chromosomes due to recombination between homologous chromosomes; this generates genetic diversity in sexually reproducing organisms. Given the importance of gametogenesis, it is pertinent to understand how this differentiation program is regulated. In the next section, I describe how gametogenesis in the budding yeast *Saccharomyces cerevisiae*, is an apt model to study the regulation of complex cell-fate decisions.

1.2 Introduction to budding yeast gametogenesis

1.2.1 Budding yeast gametogenesis is a model differentiation program

Gametogenesis or sporulation in budding yeast, follows the same cellular processes as described in the previous section, except that gametes are subsequently packaged to form spores (Marston and Amon 2004;van Werven and Amon 2011). Using budding yeast as a model organism has its advantages. The budding yeast genome is relatively small, at approximately 12 million base pairs and 6000 open reading frames (ORFs), facilitating genomic analyses (Goffeau et al. 1996). Large quantities of cellular material can be obtained due to its rapid doubling times (once every 90 minutes) and mutants can be easily made due to its experimental tractability (Duina et al. 2014). Several features also make budding yeast gametogenesis or sporulation, an apt developmental program to study how genes are regulated to achieve differential cell fates. First, cells must integrate multiple external and internal cues to initiate gametogenesis. Second, after the decision to enter gametogenesis has been made, cells must repress genes for the mitotic program (e.g. mitotic cyclins) and switch to expressing genes for meiosis. Furthermore, gametogenesis is conducted whilst the cells are starved for nutrients, which also necessitates switching from growth-related to stress-related genes (Lopez-Maury et al. 2008). Third, gametogenesis involves complex morphological changes that must occur in sequence. Therefore the meiotic transcriptional program must activate and repress different sets of genes in a robust and timely manner. Fourth, budding yeast cells which have not yet initiated meiotic nuclear divisions are able to quickly return back to the mitotic cell cycle if nutrients are re-introduced to the surroundings (Xu et al. 1995). This means that the early meiotic transcriptional program is dynamic to allow for rapid adaptation to sudden

environmental changes. These features are why I have adopted budding yeast gametogenesis as a model to study gene regulation in the context of cell differentiation.

In the coming sections of this introductory chapter, I will describe the signalling pathways and outputs responsible for initiating and maintaining budding yeast gametogenesis at different cell cycle stages. Given the relatively gene-dense budding yeast genome, I will then provide an overview of how gene expression can be regulated by pervasive and overlapping transcription of ncRNAs and transcript isoforms (i.e. mRNAs with alternative UTRs). Finally, I will end with a description of current genome-wide techniques to identify different classes of coding and non-coding transcripts, with a focus on mRNA isoforms.

1.2.2 Signalling during initiation of budding yeast gametogenesis

Budding yeast cells only enter gametogenesis in response to multiple cues **(Figure 1.1A)**. In order to make this cell-fate decision, cells must be deprived of fermentable carbon sources (e.g. glucose) and starved of nitrogen. This is to ensure that stress-resistant spores are only formed in response to severe nutrient stress; cells will otherwise reproduce asexually. In addition, cells must be diploid and formed of two opposite mating types (*MATa and MAT**α***). This ensures that a tetrad of two *MATa* and two *MAT**α*** gametes can be formed from a single parental diploid cell. More importanty, it ensures that haploid cells with a single mating type cannot enter sporulation and undergo a lethal meiosis (Mitchell and Bowdish 1992). Spores of opposite mating types can then mate again when the environmental conditions are favourable. The last cue is that cells must be respiratory competent. This is to meet the energetic demands of gametogenesis even when a fermentable carbon source is absent (Weidberg et al. 2016). Entry into gametogenesis is controlled by a master regulatory transcription factor called **I**nducer of **me**iosis **1** (Ime1) (Kassir et al. 1988). Ime1 is not found in most other fungi such as *Schizosaccharomyces pombe* or *Neurospora crassa*, although orthologs have been identified in other genera of the *Saccharomyces* family as well as in *Kluyveromyces lactis* and in *Ashbya gossypi* (Wapinski et al. 2007). In budding yeast, deletion of *IME1* is sufficient to abolish gametogenesis, while over-expression of *IME1*

promotes it (Kassir et al. 1988;Granot et al. 1989). Information about nutrient state, respiration and ploidy level are integrated onto the *IME1* promoter (**Figure 1.1 A**). Cells then enter gametogenesis once Ime1 levels accumulate past a certain threshold (Nachman et al. 2007). Therefore, control of *IME1* is critical for this cell-fate decision. In this subsection, I will briefly elaborate on how these signals are integrated onto the *IME1* promoter and how *IME1* expression triggers entry into gametogenesis.

Nutrient signalling pathways inhibit *IME1*. In the presence of glucose, the Ras/cyclic AMP (cAMP)/protein kinase A (PKA) pathway is activated. Ras-GEFs catalyse the switch from inactive Ras-GDP to active Ras-GTP, which then triggers a spike in intracellular second messenger cAMP. High cAMP levels activate PKA (orthologous to Tpk1, Tpk2 and Tpk3 in yeast) by relieving its repression from a regulatory subunit called Bcy1 (Matsuura et al. 1990;Santangelo 2006). On the other hand, the presence of nitrogen activates Target of Rapamycin Complex 1 (TORC1) (Weidberg et al. 2016). Experimental inhibition of both PKA and TORC1 signalling in budding yeast leads to high *IME1* and sporulation even in rich media (Weidberg et al. 2016). Conversely, constitutive activation of PKA by the *RAS2*^{val19} allele or overexpression of *TOR1* inhibits sporulation (Matsuura et al. 1990;Zheng and Schreiber 1997). It is now known that both the PKA and TORC1 pathways inhibit *IME1* expression through the recruitment of the Tup1-Cyc8 repressor (Mizuno et al. 1998;Weidberg et al. 2016). In addition, PKA and TORC1 signalling may repress *IME1* by promoting expression of G1 cyclins (Colomina et al. 1999;Mizunuma et al. 2013;Moreno-Torres et al. 2015). The result is that *IME1* expression and hence entry into gametogenesis is only initiated after glucose and nitrogen starvation. Of note, although this subsection focused on control of *IME1* expression, “back up” mechanisms exist for nutrient-mediated repression of gametogenesis. For example, nutrient signalling pathways also inhibit the Rim11 and Rim15 meiosis specific kinases which are required for Ime1 mediated transcriptional activation of downstream genes (Bowdish et al. 1994;Rubin-Bejerano et al. 1996;Vidan and Mitchell 1997;Pedruzzi et al. 2003;Sarkar et al. 2014).

IME1 expression is also subject to mating type control. Haploid *MATa and MATα* cells do not enter sporulation due to the binding of the Rme1 repressor in the *IME1* promoter. Diploid cells with both mating type loci express the a1-α2 repressor which inhibits *RME1* transcription and consequently facilitates *IME1* induction. Rme1 levels also explain why the some budding yeast strains sporulate very efficiently compared to other strains. For example, the SK1 strain has a single nucleotide polymorphism (SNP) which creates an additional binding site for the a1-α2 repressor in the *RME1* promoter (Deutschbauer and Davis 2005). Consequently SK1 cells have lower *RME1* and higher *IME1* levels and sporulate with high efficiency (Gerke et al. 2006).

In the absence of glucose, gametogenesis requires non-fermentable carbon sources (e.g. acetate) to serve as respiratory substrates for energy production. Cells which cannot respire due to mutations such as *pet100* or cells treated with respiratory inhibitors like sodium azide, antimycin or oligomycin do not express *IME1* under starvation conditions (Jambhekar and Amon 2008;Weidberg et al. 2016). The failure to induce *IME1* during nutrient starvation is due to a shortage of energy since *pet100* cells induce *IME1* in the presence of glucose, after inhibition of PKA and TORC1 pathways (Weidberg et al. 2016).

Regulation of *IME1* levels is critical because this master regulatory transcription factor controls the expression of a set of so called “early meiotic genes” which are essential for meiotic DNA replication and homologous recombination (Chu et al. 1998;Primig et al. 2000;Honigberg and Purnapatre 2003) **(Figure 1.1B)**. Examples of such genes are the meiosis specific cohesin Rec8, Spo11, which catalyses double stranded break formation, and the meiosis specific kinase Ime2, which activates other kinases to initiate meiotic S-phase (Smith et al. 1990;Williams et al. 2002;Schindler et al. 2003). Many of these early meiotic genes are not expressed prior to *IME1* induction because their promoters contain an upstream repressive sequence (URS1: 5’-TCGGCGGCT-3’) which is bound by a repressor called Ume6 (Williams et al. 2002). Indeed, more than 40 meiosis specific genes are de-repressed in *ume6* mutants during vegetative growth, resulting in a mild to moderate growth defect depending on the strain (Strich et al. 1994;Williams et al. 2002;Suzuki et al. 2003;Strich et al. 2011). Interestingly, Ume6 also plays a role in

coordinated activation of early meiotic genes by Ime1. This is necessary because the structure of Ime1 has no identifiable DNA-binding domain and no intrinsic DNA-binding activity (Rubin-Bejerano et al. 1996). Instead, Ime1 has an N-terminal transcriptional activation domain and a C-terminal interaction domain that can bind to Ume6, as demonstrated by studying different Ime1 domain-deletion mutants (Rubin-Bejerano et al. 1996). Upon initiation of gametogenesis, the starvation-responsive kinase Rim11 phosphorylates both Ime1 and Ume6; Subsequently, phosphorylated Ime1 and Ume6 form a transcription activating complex to initiate a wave of early meiotic gene expression (Bowdish et al. 1995;Rubin-Bejerano et al. 1996). After early meiotic gene expression, Ime1 signalling is repressed by a negative feedback loop. The early meiotic gene *IME2* encodes the Ime2 kinase which phosphorylates the C-terminal domain of Ime1; this phosphorylation targets Ime1 for degradation by the proteasome (Guttmann-Raviv et al. 2002). In summary, both transcriptional and post-transcriptional mechanisms ensure that Ime1 signalling is active at the appropriate time to drive expression of early meiotic genes.

1.2.3 Signalling during commitment and progression through gametogenesis

After initiation of gametogenesis, early meiotic genes are expressed in a coordinated fashion in order for cells to replicate their DNA and initiate homologous recombination. Another set of genes loosely termed as the “middle meiotic/sporulation genes” are subsequently required for prophase I exit and commitment to meiotic nuclear divisions (Chu and Herskowitz 1998;Primig et al. 2000). The transcription factor Ndt80, itself a middle sporulation gene, controls the expression of other middle genes and represents the commitment point to meiosis (Hepworth et al. 1998) **(Figure 1.1B)**.

How is *NDT80* expression temporally controlled after initiation of gametogenesis? The *NDT80* promoter contains two URS1 sites and two consensus sites called the middle sporulation element (MSE: 5’-YGNCACAAAA-3’) (Winter 2012). While URS1 is a binding site for Ume6, MSE is a binding site for another repressor called Sum1. During early meiotic gene expression, Ime1 itself

removes Ume6 repression but *NDT80* is still repressed by Sum1 (Pak and Segall 2002a;Pierce et al. 2003). An early meiosis kinase, Ime2 and active Cdk1 phosphorylate Sum1 to trigger its removal from MSE sites (Shin et al. 2010). After homologous recombination and repair of double stand breaks, removal of checkpoint inhibition leads to the binding of hyperphosphorylated Ndt80 to the MSE site in its own promoter, thus establishing a positive auto-regulatory loop (Pak and Segall 2002b;Sopko et al. 2002). Full Ndt80 signalling drives the transcription of a group of middle genes which also contain the MSE site in their promoters. Middle genes such as *CLB1* trigger exit from meiotic prophase I and initiate meiotic divisions (Acosta et al. 2011). The dependence of Ndt80 signalling on Ime2 and on passing the meiotic recombination checkpoint sets up timely expression of middle genes. Importantly, Ndt80 governs commitment to meiotic divisions. Cells which harbour the *ndt80* mutation are arrested at the pachytene stage of prophase I and can return to the mitotic cell cycle if nutrients are re-introduced to their surroundings (Xu et al. 1995). Conversely, addition of glucose after Ndt80 signalling does not prevent cells from completing meiosis (Friedlander et al. 2006). Taken together, Ndt80 signalling is key for commitment and progression through gametogenesis.

The last set of “late” genes to be induced in gametogenesis encode proteins responsible for spore wall formation and maturation (Primig et al. 2000). The temporal regulation of late genes is poorly understood, although MSE sites have been identified for some examples such as *DIT1* (Friesen et al. 1997). There is some evidence that the product of the middle gene *GIS1*, is necessary for expression of late genes such as *DIT1* and *SPS100*, thus explaining their delayed induction (Primig et al. 2000;Yu et al. 2010). More work needs to be done to understand how late gene expression is controlled.

To summarize, two regulatory transcription factors Ime1 and Ndt80, drive the sequential and coordinated transcription of early and middle-late meiotic genes respectively. Genome wide studies suggest that approximately 500 genes are upregulated and approximately 600 genes are repressed at specific stages throughout gametogenesis (Chu et al. 1998;Primig et al. 2000). How transcription

of meiotic genes is temporally coordinated with repression of other genes at different stages is still under investigation.

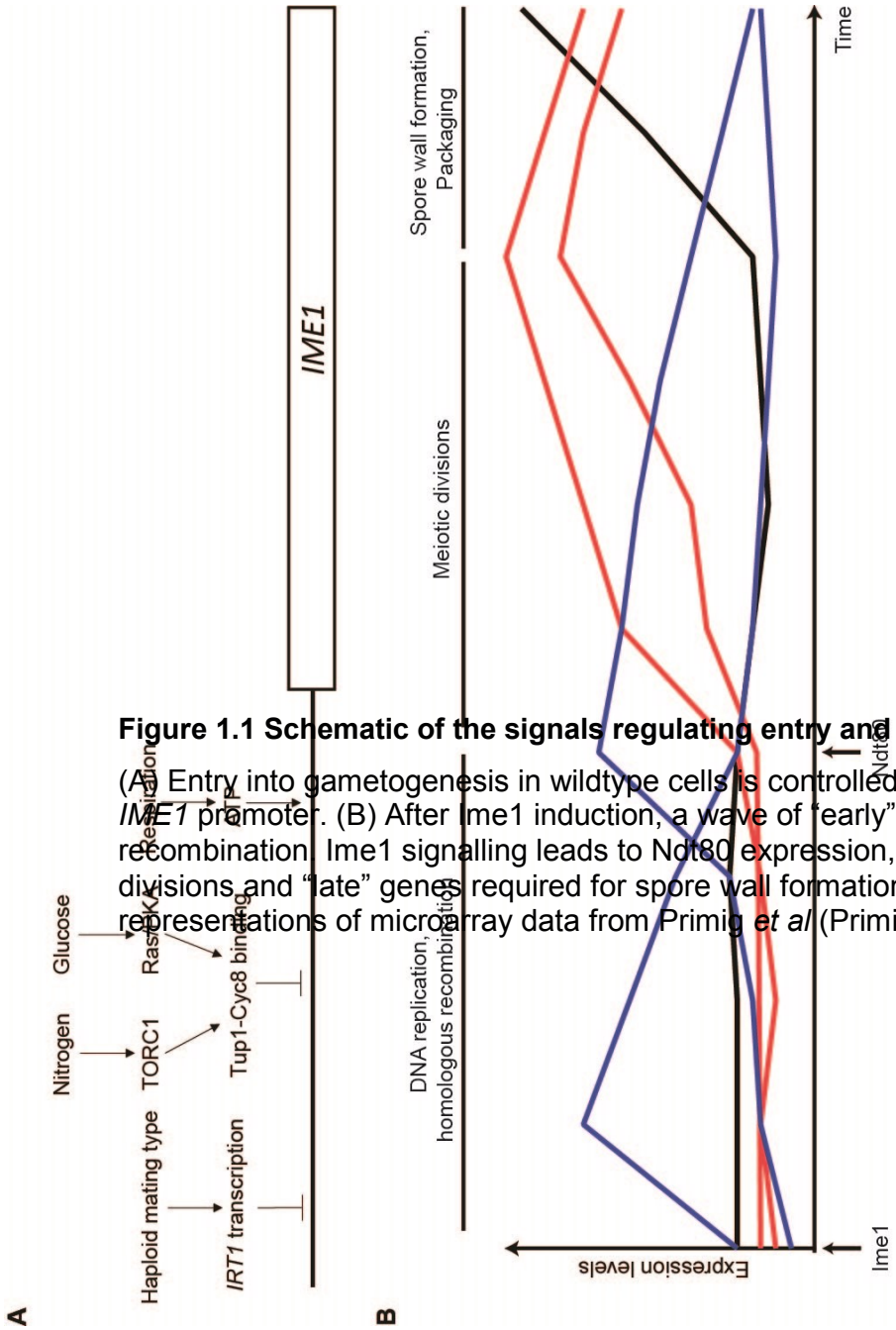


Figure 1.1 Schematic of the signals regulating entry and progression through buddi

(A) Entry into gametogenesis in wildtype cells is controlled by mating type, nutrient sign *IME1* promoter. (B) After Ime1 induction, a wave of "early" genes (blue) are expressed recombination. Ime1 signalling leads to Ndt80 expression, which triggers waves of "mi divisions and "late" genes required for spore wall formation and packaging of gametes. representations of microarray data from Primig *et al* (Primig et al. 2000).

1.3 Introduction to transcription and chromatin

1.3.1 The transcription cycle

A fundamental component of gene regulation involves the control of transcription, the mechanistic process by which DNA sequences are used as templates to synthesize complementary RNA molecules. Eukaryotes have three different RNA polymerases (Carter and Drouin 2009). RNA polymerase I transcribes rRNA genes except 5S rRNA, while RNA polymerase III transcribes 5S rRNA and transfer RNAs. RNA polymerase II (RNAPII) is responsible for transcribing mRNAs and non-coding RNAs. In this section, I explain some important features of transcription and chromatin that are relevant for describing how transcript isoforms and ncRNAs regulate gene expression.

The process of transcription by RNAPII is a complicated process involving multiple regulatory DNA sequences and numerous protein complexes. A gene is defined in the Sequence Ontology as “all of the sequence elements necessary to encode a functional transcript” (Eilbeck et al. 2005). A typical, simplified schematic of a gene comprises of a promoter, a transcription start site (TSS), an open reading frame (ORF) containing the start codon, exons, introns and stop codon, and a transcription terminator site (TTS). **(Figure 1.2)**. The sequence spanning a TSS and ending at a TTS is known as a transcription unit (TU) (Mellor et al. 2016). In order to initiate transcription, activating transcription factors (e.g. Ime1 or Ndt80) bind in the promoter region and recruit co-activator complexes such as SAGA and/or Mediator (Roberts and Winston 1997). Co-activator complexes in turn stimulate the recruitment and formation of the pre-initiation complex (PIC) at a region called the core promoter. The complete PIC comprises of RNAPII and 6 general transcription factors (GTFs): TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Lemon and Tjian 2000). RNAPII is a 12 subunit complex formed of Rpb1-12 (Bushnell and Kornberg 2003). Of special note is the Rpb1 subunit whose C-terminal domain (CTD) contains repeats of the consensus heptad amino acid sequence YSPTSPS, whose serine and threonine residues can be phosphorylated to regulate transcription (Stiller and Cook 2004). Following PIC formation, DNA

unwinding, and phosphorylation of Ser5 in the RNAPII CTD, transcription is initiated at the TSS. During early elongation, Ser5 phosphorylation enables modification of the 5' end of the nascent transcript with a 7-methyl-guanosine containing cap (m^7Gppp) that marks the TSS (Fabrega et al. 2003) **(Figure 1.2)**. After the first few hundred nucleotides have been transcribed, there is a gradual increase in Ser2 phosphorylation in the RNAPII CTD and a decline in phosphorylated Ser5 (i.e. during the switch from early to later elongation) (Komarnitsky et al. 2000). The elongation step proceeds until RNAPII encounters an A–rich and T-rich stretch and transcription is terminated downstream at the TTS. Concurrent with termination, the primary transcript is endonucleolytically cleaved at the 3' end and a poly(A) tail is added to the transcript end site (TES) (Kuehner et al. 2011) **(Figure 1.2)**. The pre-mRNA is spliced to remove the introns and to join the exons to form the coding sequence (CDS) beginning at the start codon and ending at the stop codon (Le Hir et al. 2003). The 5' and 3' untranslated regions (UTRs) flank the CDS of the mature mRNA transcript. RNAPII subunits are then recycled for a next round of transcription. Crucially in some cases, the very act of transcription itself (and not the RNA *per se*) can regulate gene expression at neighbouring loci (Mellor et al. 2016). This mode of gene regulation by the act of transcription itself has been observed in different eukaryotic models but is most well documented in budding yeast. These examples will be further described in section 1.5 of this chapter.

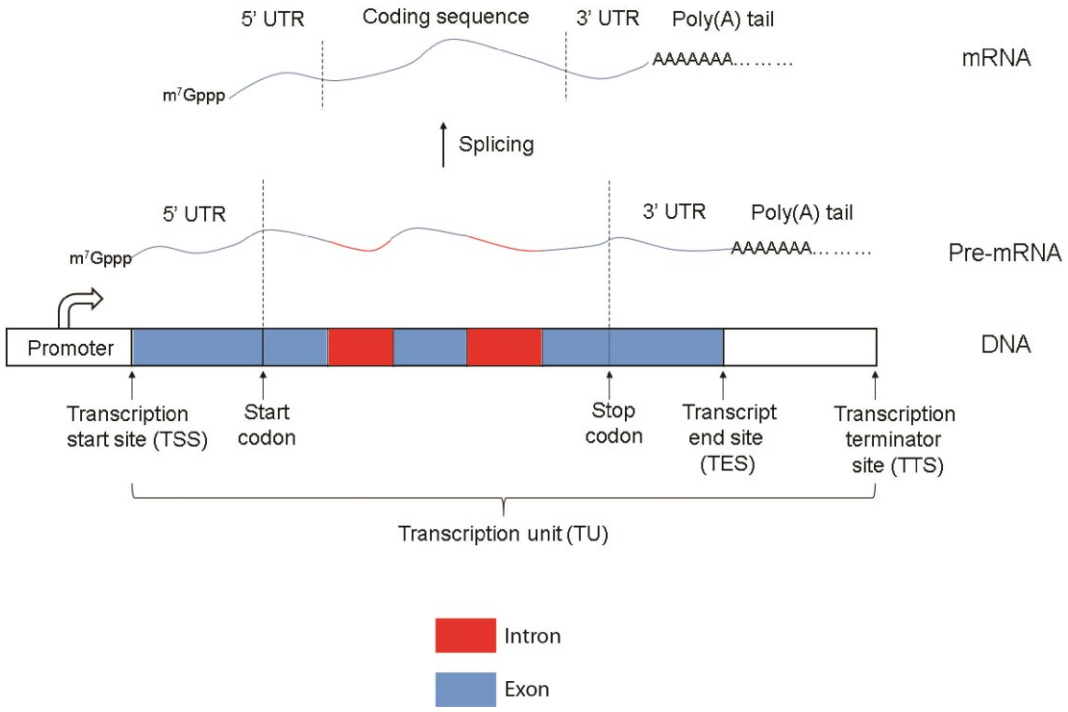


Figure 1.2 Schematic of a gene and its RNA product

A gene comprises of a promoter, a transcription start site (TSS), an open reading frame (ORF) containing the start codon, exons (blue), introns (red) and stop codon,and a transcription terminator site (TTS). The sequence spanning a TSS and ending at a TTS is known as a transcription unit (TU). In the pre-mRNA, the TSS is marked by the 7-methyl-guanosine containing cap (m^7Gppp). The transcript end site (TES) is the site where the poly(A) tail is added to the RNA. The pre-mRNA is spliced to remove the introns and to join the exons to form the coding sequence (CDS) beginning at the start codon and ending at the stop codon. The 5' and 3' untranslated regions (UTRs) flank the CDS of the mRNA.

1.3.2 Transcription and chromatin

In eukaryotes, the transcription machinery has to face an added complication: DNA does not exist as a naked polymer but is instead packaged as chromatin. Chromatin consists of repeated segments of DNA-protein complexes known as nucleosomes. Each nucleosome consists of 147 bp of DNA wrapped around an octameric complex of histone proteins; an octamer comprises of a pair each of the histones H2A, H2B, H3 and H4 (Luger et al. 1997). When packaged into nucleosomes, the amino-terminal (N-terminal) histone tails protrude from the

globular core (Luger et al. 1997). Amino acids in both the histone core, as well as the N-terminal tails can be chemically modified by the cell to regulate chromatin structure and gene expression (described in more detail in later paragraphs in this section) (Bannister and Kouzarides 2011).

Nucleosomes act as barriers to transcription because the histones block transcription factors and RNAPII from accessing the DNA template (Teves et al. 2014;Weber et al. 2014;Lai and Pugh 2017). As such, efficient transcription *in vivo* requires the alteration of chromatin structure and destabilization of nucleosomes mediated by proteins called chromatin remodellers and histone modifying enzymes (Schwabish and Struhl 2004). These proteins facilitate transcription in a myriad of ways. This includes the establishment and maintenance of either nucleosome-free or nucleosome-depleted regions (NFRs or NDRs) at the core promoter to allow PIC formation, as well as the destabilization and replacement of nucleosomes in the wake of elongating RNAPII (Orphanides et al. 1998;Cheung et al. 2008;Hartley and Madhani 2009;Krietenstein et al. 2016). Interestingly besides their function in the transcription process, some chromatin remodellers and histone modifying enzymes act in concert with RNAPII to activate or repress nearby TUs, thus constituting a form of local gene regulation. In the following subsections, I will provide an overview of these chromatin modifying proteins that are relevant to my work on gene regulation by transcript isoforms and ncRNAs.

1.3.3 The Set1/Set3C pathway

Set1 is a histone methyltransferase that catalyses the addition of methyl groups (Me) to the 4th lysine residue of the N-terminal tail of histone H3 (H3K4). Unlike mammalian cells which have at least six H3K4 methyltransferases, Set1 is solely responsible for all H3K4 methylation in budding yeast (Briggs et al. 2001;Shilatifard 2012). Set1 is a subunit of a larger complex known as COMPASS or Set1C (complex of proteins associated with Set1). Set1 localizes to the 5' ends of actively transcribed genes and was found to co-purify with Ser5 phosphorylated but not Ser2 phosphorylated RNAPII (Ng et al. 2003). In other words, Set1 is recruited during transcription initiation and early elongation to genes, where it catalyses H3K4 methylation.

Similar to Set1, genome wide ChIP data of H3K4 methylation in various organisms show that this conserved modification is associated with actively transcribed genes (Kirmizis et al. 2007;Guillemette et al. 2011). However, there is a graded distribution of different methylation states of H3K4 (tri-methylation to di-methylation to mono-methylation), spread from 5' to 3' of a gene **(Figure 1.3)** (Kirmizis et al. 2007;Soares et al. 2017). At an average gene, H3K4 tri-methylation (H3K4me3) is found within 200 base pairs downstream of the promoter, overlapping the +1 nucleosome. H3K4 di-methylation (H3K4me2) peaks at a few hundred nucleotides downstream of H3K4me3, while H3K4 mono-methylation (H3K4me1) is found even further downstream (Soares et al. 2017).

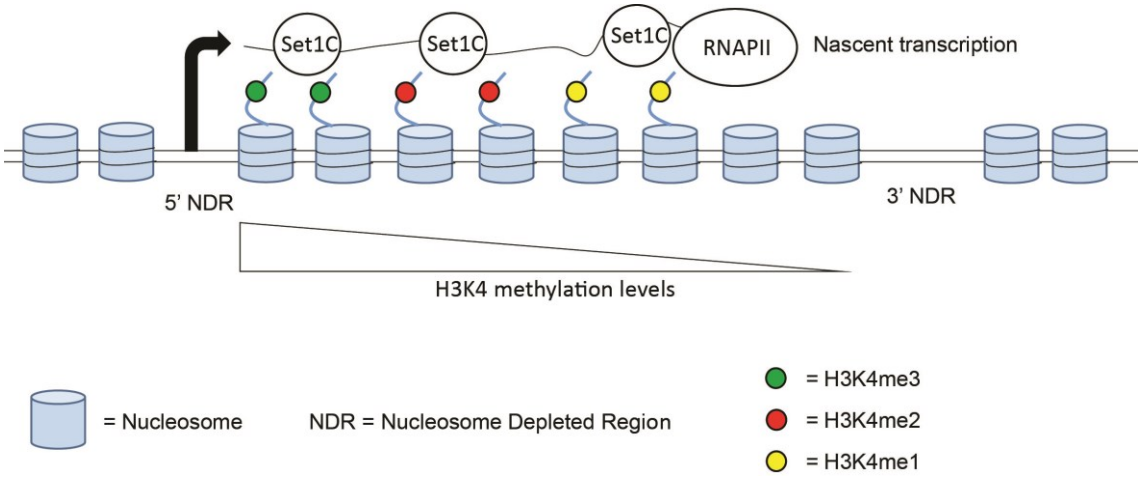


Figure 1.3 Co-transcriptional methylation of H3K4 by Set1C

Set1C is recruited to chromatin by the phosphorylated Ser5 CTD of RNAPII. After recruitment, Set1 catalyses the methylation of the H3K4 residue. There is a graded distribution of different methylation states of H3K4 (tri-methylation to di-methylation to mono-methylation), moving from the 5' to 3' of a gene.

The three methylation states of H3K4 may be correlated with different functions or regulatory outcomes. ChIP signals of H3K4me3 near the TSS increase in intensity with increasing levels of transcription (Santos-Rosa et al. 2002;Soares et al. 2017). It has been suggested that H3K4me3 activates transcription by acting as a recruiting mark for chromatin remodellers, but a causative activating role for H3K4me3 has been disputed because the loss of this mark has no effects on the transcription of most genes (Buratowski and Kim 2010;Howe et al. 2017). The role of H3K4me1 is less clear, although one recent study in budding yeast

demonstrated that this mark regulates the transcriptional response to osmotic stress (Nadal-Ribelles et al. 2015).

Although mostly indirect, evidence in the literature suggests a repressive role of H3K4me2 in transcription. Unlike H3K4me3, H3K4me2 is found at both active and inactive genes (Santos-Rosa et al. 2002). Cells expressing a mutant allele of Set1 (*SET1-ΔRRM*) that causes a complete loss of H3K4me3 but not H3K4me2, still showed near wildtype levels of histone de-acetylation at the 5' ends of genes (Kim and Buratowski 2009). In contrast, *set1Δ* mutants with no H3K4 methylation have markedly increased levels of acetylated histones at these regions (Kim and Buratowski 2009). Histone de-acetylation is thought to repress transcription because acetylated histones are associated with weaker affinity for DNA, higher nucleosome turnover rates and decreased nucleosome density, thereby facilitating access to DNA by RNAPII (Buratowski and Kim 2010;Zentner and Henikoff 2013). H3K4me2 is mechanistically linked to histone de-acetylation because this specific mark serves as a recruitment signal for the Set3 histone de-acetylase complex (Set3C) **(Figure 1.4)** (Kim and Buratowski 2009).

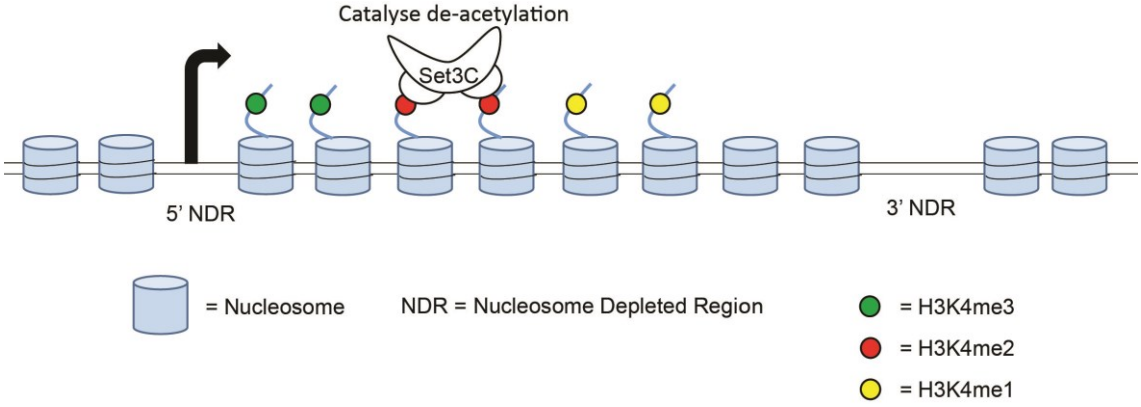


Figure 1.4 Histone de-acetylation at the 5' ends of transcribed regions by Set3C
Set3C is recruited to chromatin by H3K4me2, which is typically enriched in the 5' ends of transcribed regions. Set3C targets surrounding histones for de-acetylation.

Set3C comprises of the de-acetylases Hos2 and Hst1, as well as the Set3 subunit (Pijnappel et al. 2001). The PHD finger of Set3 specifically recognizes and binds to H3K4me2, which targets surrounding histones for de-acetylation **(Figure 1.4)** (Kim and Buratowski 2009). Loss of Set3C function leads to some meiotic

gene induction at earlier time points, suggesting that this complex plays a role in the dynamic repression of those genes (Pijnappel et al. 2001). More evidence for the dynamic role of Set3C in regulating gene expression came from comparing wildtype and *set3Δ* cells under either steady state conditions or under a series of carbon source shifts. Under steady state conditions, global transcription in *set3Δ* mutants was similar to wildtype (Lenstra et al. 2011). However, when cells were shifted from raffinose to galactose-rich media, Kim *et al.* found that *set3Δ*, or mutations in the Set3 PHD finger which abolished its binding to H3K4me2, led to more rapid induction of 113 genes (Kim et al. 2012). Interestingly, the promoters of most of these affected genes overlapped with another transcript. In at least 4 examined cases, the overlapping transcript was initiated from a distal TSS but *set3Δ* or *set1Δ* mutants exhibited increased histone acetylation and transcription only at the proximal TSS. Furthermore, *set3Δ* mutants also showed de-repression of cryptic transcripts initiating from within the 5' ends of some active genes. These results from Kim *et al.* suggest a model whereby transcription from a neighbouring, overlapping ncRNA deposits H3K4me2 by Set1 at the TSS of a coding gene. This recruits the Set3C de-acetylase which represses said promoter (Kim et al. 2012).

Till date, a similar repressive mechanism involving orthologs of Set1 and Set3C has not been described in mammalian cells, although some evidence suggests that H3K4me2 can recruit UpSET, the Drosophila homolog of Set3C, to de-acetylate histones at promoter regions for gene repression (Rincon-Arano et al. 2012;Ali et al. 2013). More work needs to be done to investigate if crosstalk between H3K4me2 and histone de-acetylation can regulate gene expression in mammalian cells.

1.3.4 The Set2/Rpd3S pathway

Set2 is a histone methyltransferase that catalyses the addition of methyl groups (Me) to the 36th lysine residue of the N-terminal tail of histone H3 (H3K36) **(Figure 1.5)**. Unlike mammalian cells which have at least eight H3K36 methyltransferases, Set2 is solely responsible for all H3K36 methylation in budding yeast (Strahl et al. 2002;Wagner and Carpenter 2012). Set2 co-purified with Ser2 phosphorylated RNAPII, which marks later transcription elongation (Krogan et al.

2003). Like Ser2 phosphorylated RNAPII, Set2 ChIP signals are strongest within the CDS of transcribed genes and very weak at the promoter or 3' untranslated regions (Komarnitsky et al. 2000;Krogan et al. 2003).

Similar to Set2, H3K36 methylation is enriched within gene bodies and towards the 3' ends of actively transcribed genes (Pokholok et al. 2005). There is also a graded distribution of different methylation states of H3K36 (mono-methylation to di-methylation to tri-methylation), moving from 5' to 3' of an average gene **(Figure 1.5)** (Pokholok et al. 2005;Venkatesh and Workman 2013).

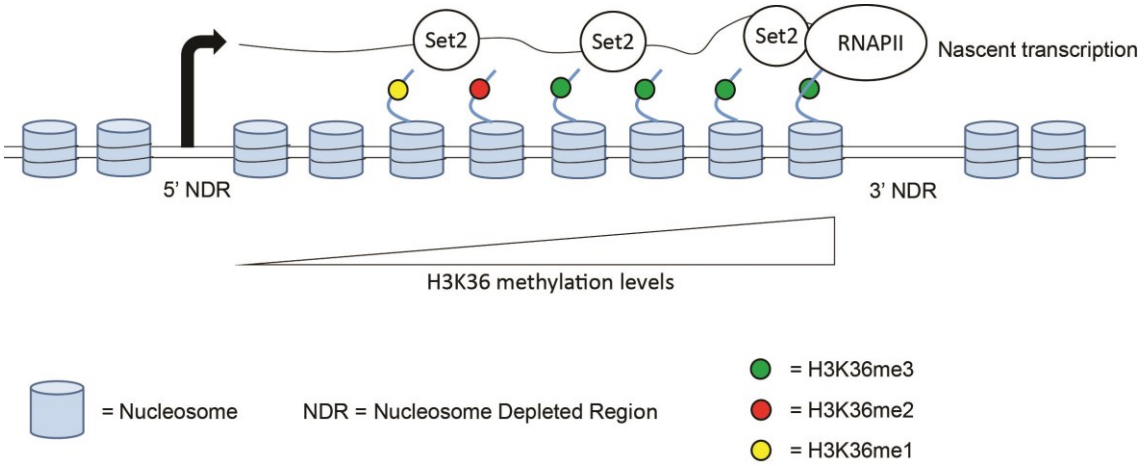


Figure 1.5 Co-transcriptional methylation of H3K36 by Set2

Set2 is recruited to chromatin by the phosphorylated Ser2 CTD of RNAPII. After recruitment, Set2 catalyses the methylation of the H3K36 residue. There is a graded distribution of different methylation states of H3K36 (mono-methylation to di-methylation to tri-methylation), moving from the 5' to 3' of a gene.

The repressive roles of H3K36 di-methylation (H3K36me2) and H3K36 tri-methylation (H3K36me3) are commonly associated with another histone de-acetylase called Rpd3S **(Figure 1.6)**. In this model, Set2 associates with elongating RNAPII and co-transcriptionally deposits H3K36me2/3 in the ORF and towards 3' ends of genes. H3K36me2/3 are recognized by the chromodomain of Eaf3 and the PHD domain of Rco1, which are subunits of the Rpd3S de-acetylase complex (Carrozza et al. 2005b;Joshi and Struhl 2005;Keogh et al. 2005). The activity of Rpd3S in turn generates a repressive chromatin state at the 3' ends of genes. Multiple lines of evidence support this model. The *rco1Δ*, *set2Δ*, *eaf3Δ,* Eaf3 chromodomain deletion or non-methylatable *H3K36A* mutants each showed

increased levels of acetylated histone H4 specifically at the 3' ends of genes (Carrozza et al. 2005b;Joshi and Struhl 2005;Keogh et al. 2005). Furthermore, different studies also showed that mutants with compromised Rpd3S or Set2 function have cryptic intragenic transcription phenotypes usually at the 3' ends of some genes. Interestingly, genes which are more susceptible to cryptic transcription in *set2Δ* tend to be longer and infrequently transcribed (Li et al. 2007). *In vitro* chromatin binding assays show that Rpd3S can bind to either H3K36me2 or H3K36me3 modified nucleosomes (Li et al. 2009). Furthermore, a *paf1Δ* mutant with no H3K36me3 but some H3K36me2 can still suppress cryptic transcription unlike *set2Δ* mutants (Li et al. 2009). Taken together, H3K36me2/3 recruits Rpd3S to maintain chromatin integrity and hence transcriptional fidelity at the 3' ends of genes.

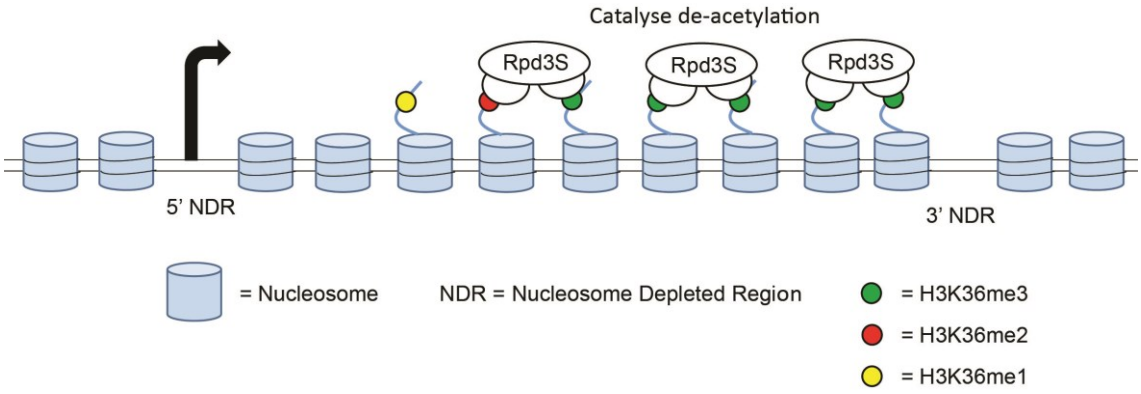


Figure 1.6 Histone de-acetylation at the 3' ends of transcribed regions by Rpd3S
Rpd3S is recruited to chromatin by H3K36me2 and H3K36me3, which is typically enriched in the 3' ends of transcribed regions. Rpd3S targets surrounding histones for de-acetylation.

More recent studies have provided additional mechanical elaborations of this repressive Set2/Rpd3S pathway in maintaining chromatin integrity. Apart from its role in recruiting Rpd3S, Set2 might also promote the retention of "old" H3K36me2/3 modified histones and inhibit the incorporation of acetylated "new" histones onto chromatin in the wake of RNAPII transcription (Smolle et al. 2012;Venkatesh et al. 2012). Asf1 is a histone chaperone which catalyses the loading of acetylated histones from the soluble histone pool onto DNA(Kim et al. 2007). The H3K36me2/3 modification inhibits the association of the histone chaperone Asf1 to nucleosomes and the *asf1Δ* mutation could partially suppress

the cryptic initiation phenotype of *set2Δ* (Venkatesh et al. 2012). Another study showed that H3K36me2/3 also recruits the chromatin remodeller Isw1b which suppresses histone exchange to maintain repressive chromatin (Smolle et al. 2012). Accordingly, the *isw1Δset2Δ* mutant phenocopied the *set2Δ* mutant in terms of increased H4 acetylation within gene bodies and increased cryptic transcription, supporting the model that Isw1b acts downstream of Set2 catalysed H3K36me2/3 (Smolle et al. 2012). Of note, this model has been recently challenged by another study showing that neither histone H3 methylation nor histone tail acetylation had an effect on histone turnover (Ferrari and Strubin 2015). Even so, H3K36me2/3 deposition and concomitant histone de-acetylation might still contribute to gene repression by the recruitment of remodellers and by increasing chromatin compaction at repressed promoters (Bannister and Kouzarides 2011).

The functional role of H3K36 mono-methylation (H3K36me) is unclear. In budding yeast, deletion of the H4 interaction motif of Set2 (*SET2–Δ31-39*) abolished H3K36me2 and H3K36me3 but not H3K36me (Du et al. 2008). *SET2–Δ31-39* mutants phenocopied *set2Δ* mutants in terms of increased levels of acetylated H4 at the 3' ends of genes, suggesting that H3K36me alone cannot establish repressive chromatin at those regions (Du et al. 2008).

In summary, the Set2/Rpd3S pathway and H3K36me2/3 can repress transcription. Importantly, budding yeast cells have also adapted this mechanism to suppress coding genes via transcription of overlapping ncRNAs or transcript isoforms. These examples can be found in section 1.5 of this introduction. It is unclear if a similar repressive mechanism exists in mammalian cells because knockdown of the mouse ortholog of Set2 (Setd2) did not lead to global changes in H3 or H4 acetylation (Edmunds et al. 2008). Instead, Setd2 catalysed deposition of H3K36me3 suppresses cryptic transcription in mouse ES cells by facilitating the recruitment of DNA methyltransferase Dnmt3b (Neri et al. 2017).

1.3.5 Elongation factors FACT and Spt6

The FACT (facilitates transcription) complex is a heterodimer of two highly conserved proteins: Spt16 and Pob3. This complex was initially identified as a factor necessary for productive transcription of chromatin templates *in vitro*; in the absence of FACT, the transcription reaction stalls at a location proximal to the promoter (Orphanides et al. 1998). Other *in vitro* assays demonstrated that the Spt16 subunit of FACT interacts with nucleosomes and destabilizes them by catalysing the loss of histone H2A-H2B subunits (Belotserkovskaya et al. 2003). Consistent with FACT's nucleosome destabilizing function, histone mutations that weaken nucleosome structure suppressed defects caused by the *spt16-11* allele (McCullough et al. 2011). Crucially, *in vitro* assays showed that FACT can also promote the deposition of histones onto DNA (Belotserkovskaya et al. 2003). Co-IP experiments show FACT co-purifying with RNAPII, while ChIP experiments showed that FACT prominently associates with active RNAPII gene bodies *in vivo*, and that Ser5 phosphorylation of RNAPII CTD (a mark of elongation) was required for this association (Krogan et al. 2002;Mason and Struhl 2003). Interestingly, conditional inactivation of Spt16 led to RNAPII enrichment at the 3' ends of some genes which resulted in the production of cryptic RNA transcripts originating from internal TSSs (Mason and Struhl 2003). Likewise, some mutations in *SPT16* suppress cryptic transcription in cells harbouring the L61W mutation in histone H3; these suppressor mutations are specific to the middle domain of *SPT16* and presumably restore FACT interaction with mutant H3 (Myers et al. 2011). Collectively, biochemical and genetic data support a model whereby FACT transiently destabilizes nucleosomes to facilitate the passage of elongating RNAPII **(Figure 1.7)**. In the wake of RNAPII transcription, FACT functions as a histone chaperone to restore nucleosome integrity. Mutations that inactivate FACT or that compromise its association with histones lead to defective elongation, as well as impaired re-assembly of nucleosomes after transcription. Consequently, aberrant transcription can initiate stochastically from cryptic promoters which are normally occluded by nucleosomes.

Spt6 is another elongation factor that associates with elongating RNAPII; Spt6 binds via its SH2 domain to the phosphorylated Ser2 CTD of RNAPII (Yoh et al. 2007;Sun et al. 2010). Like other elongation factors, Spt6 enhances the rate of

in vitro transcription elongation from chromatin (Endoh et al. 2004). *In vivo*, the *spt6-14* allele results in sensitivity to 6-azauracil and is synthetic lethal with a deletion in the elongation factor Ppr2 (TFIIS in humans) (Hartzog et al. 1998). Similar to FACT, Spt6 binds directly to histones and can function as a histone chaperone to assemble nucleosomes **(Figure 1.7)** (Bortvin and Winston 1996). Microarray analysis of mutant yeast harbouring the temperature sensitive *spt6-1004* allele show widespread aberrant transcription from intragenic cryptic promoters (Kaplan et al. 2003). As previously mentioned, this cryptic transcription phenotype is also due to the failure to restore chromatin integrity after RNAPII passage. Consistent with this hypothesis, the actively transcribed *FLO8* locus in the *spt6-1004* mutant is hyper-sensitive to micrococcal nuclease (MNase) digestion and shows reduced nucleosome occupancy (Kaplan et al. 2003). Conversely, MNase hyper-sensitivity and reduced nucleosome occupancy was not observed at the transcriptionally inactive *GAL1* locus. Taken together, the co-transcriptional recruitment and function of Spt6 is essential for maintaining transcriptional fidelity and suppressing the expression of cryptic ncRNAs.

Importantly, the chromatin remodellers mentioned in this chapter (Set3, Set2, Spt16 and Spt6) play roles in suppressing coding genes via transcription of overlapping ncRNAs or transcript isoforms. Some examples are described in section 1.5 of this introduction.

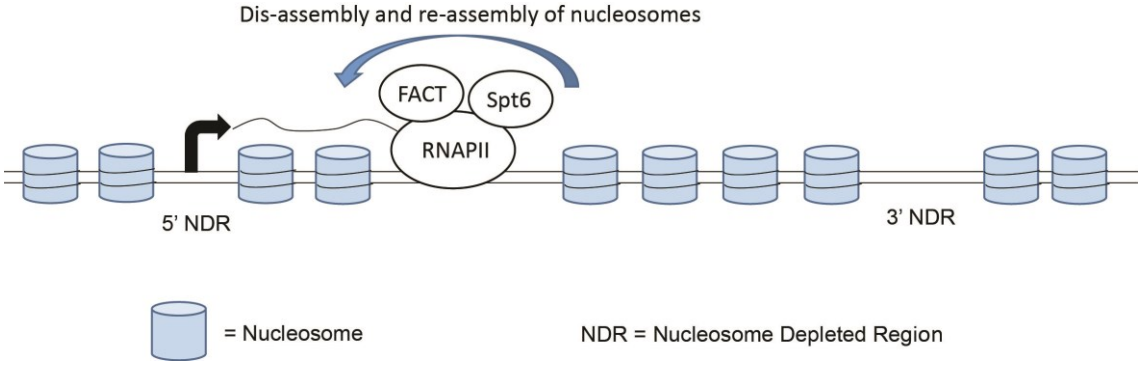


Figure 1.7 Chromatin remodelling by the FACT complex and Spt6 chaperones

The elongation factors FACT and Spt6 associate with elongating RNAPII. FACT and Spt6 remodel chromatin for productive transcription by transiently destabilizing nucleosomes. These factors also restore chromatin integrity by re-assembling nucleosomes in the wake of RNAPII.

1.4 ncRNAs and transcript isoforms

1.4.1 Complexity in the budding yeast transcriptome

The haploid budding yeast genome comprises of approximately 12 million base pairs and 6000 ORFs contained in 16 chromosomes (Goffeau et al. 1996). An estimated 75-85% of the budding yeast genome is transcribed during vegetative growth but only about 22% of these arise from protein coding transcripts. (David et al. 2006;Nagalakshmi et al. 2008). Thus, pervasive transcription arises from intergenic regions, intragenic TUs, anti-sense TUs as well as heterogeneity in TSSs and TESs **(Figure 1.8)** (David et al. 2006;Ito et al. 2008;Pelechano et al. 2013;Murray et al. 2015;Mellor et al. 2016). While many pervasive transcripts are non-coding RNAs (ncRNAs) that do not carry protein-coding information, alternative TSS and TES usage could lead to different mRNA isoforms with differing UTRs and translational capacities. Most of these ncRNAs and isoforms are transcribed from TUs which overlap protein coding genes, and the functional relevance for several examples have been documented in the literature (see section 1.5). In this section, I describe the different "classes" of ncRNAs and transcript isoforms that have been identified in wildtype and mutant budding yeast.

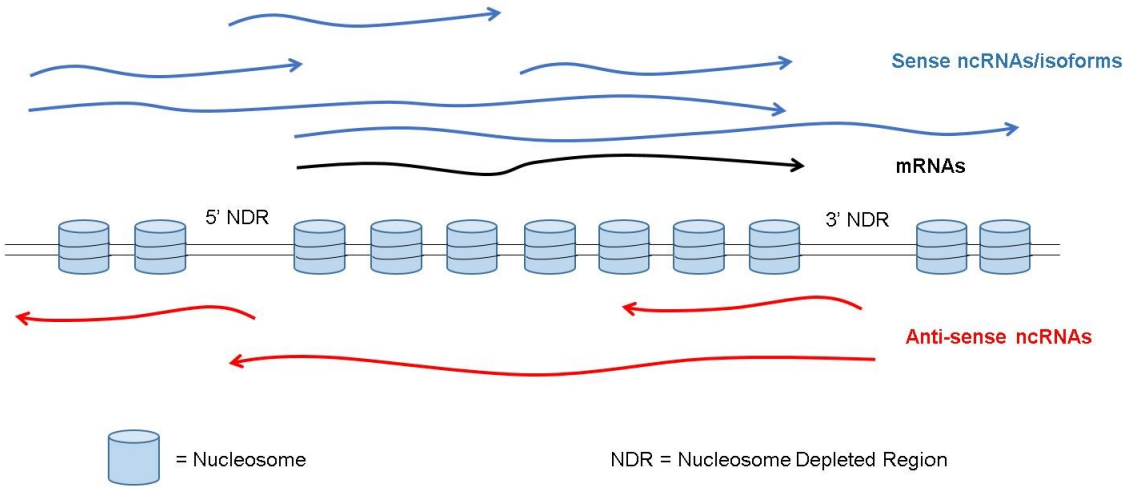


Figure 1.8 Pervasive transcription in the eukaryotic genome

The eukayotic genome is pervasively transcribed, leading to the production of different ncRNAs and/or mRNA isoforms which overlap with the transcription units of canonical mRNAs (black). Protein coding genes are flanked by nucleosome depleted regions (NDRs). Sense ncRNAs/isoforms (blue) can initiate from the

same 5' NDR as a protein coding gene, or from upstream intergenic regions, or from within gene bodies. Likewise, anti-sense ncRNAs (red) can initiate from the flanking 3' NDR, or from intergenic or intragenic regions.

1.4.2 Unstable transcripts

Although the budding yeast genome is pervasively transcribed, different classes of ncRNAs are rapidly degraded by various quality control pathways; these classes are not mutually exclusive and some examples fall into more than 1 category (Wyers et al. 2005;van Dijk et al. 2011;Schulz et al. 2013).

One such class of ncRNAs are called cryptic unstable transcripts (CUTs). CUTs were first identified in mutant yeast lacking a component of the nuclear exosome complex (*rrp6Δ*) (Wyers et al. 2005). These CUTs had short half-lives, were poly-adenylated at their 3' ends and were relatively short (250-600 nt) (Wyers et al. 2005). Subsequent studies found that CUTs initiate from NFRs shared with, and flanking protein coding ORFs; most of them were transcribed in the opposite orientation to the ORF due to the bi-directionality of promoters (Neil et al. 2009;Xu et al. 2009). Particularly due to the high gene density in budding yeast, transcription of many sense and anti-sense CUTs crosses into the promoters and/or ORFs of mRNA TUs (Neil et al. 2009;Xu et al. 2009). Despite their instability, transcription of some CUTs can still regulate gene expression due to the act of transcription itself (Neil et al. 2009). For example, transcription of *SRG1* runs into the promoter of the *SER3* gene and represses *SER3* expression (Martens et al. 2004).

Meiotic unannotated transcripts (MUTs) are a subclass of CUTs that are upregulated during gametogenesis. In mitotic cells, MUTs are rapidly degraded by Rrp6 (Lardenois et al. 2011). Upon entry into gametogenesis, MUTs are protected from degradation by an unknown mechanism and their levels increase as a consequence (Frenk et al. 2014). It is still unclear if MUTs play a role in regulating the meiotic transcriptional program.

Another class of unstable transcripts were identified as Xrn1-sensitive unstable transcripts (XUTs), so named because these ncRNAs were stabilized in *xrn1Δ* mutants (van Dijk et al. 2011). Xrn1 is a 5'-3' exonuclease which promotes

RNA turnover and translational repression at cytosolic P-bodies (Sheth and Parker 2003). XUTs are also poly-adenylated and most were transcribed anti-sense to a neighbouring gene (van Dijk et al. 2011). Interestingly van Dijk *et al.* identified 273 anti-sense XUTs that overlapped with the ORFs of protein coding genes; RNAPII occupancy in those genes was reduced in the *xrn1Δ* mutants. Transcription of these genes was partly rescued in *xrn1Δset1Δ* or in *xrn1Δ*H3K4A mutants, suggesting that some XUTs have a regulatory role dependant on the repressive H3K4me2 modification (van Dijk et al. 2011).

More recently, a third class of unstable transcripts called Nrd1-unterminated transcripts (NUTs) have been identified (Schulz et al. 2013). Nrd1 is a subunit of the Nrd1-Nab3-Sen1 (NNS) termination complex that recognizes short RNA sequence elements (NNS motifs) and facilitates transcription termination and degradation of many ncRNAs (Vasiljeva and Buratowski 2006). Upon Nrd1 depletion from the nucleus, this ncRNA termination pathway is compromised and read-through transcription produced more than 1500 NUTs (Schulz et al. 2013). NUTs originate from NDRs flanking ORFs and 942 of them are also transcribed in an anti-sense orientation to protein coding genes. Schulz *et al.* monitored changes in nascent expression of these coding genes in response to Nrd1 depletion and found 114 examples repressed by un-terminated NUT transcription. Interestingly, NUTs extended into the gene promoters in these 114 examples, suggesting that transcription of the anti-sense NUTs repressed transcription in the sense direction. Conversely, Schulz *et al.* also observed that NUT transcription could upregulate expression from a neighbouring ORF in some cases where the NUT TU had an upstream TSS in tandem with a downstream ORF TSS. Therefore, the NNS complex represents a quality control pathway to prevent overlapping readthrough ncRNA transcription from disrupting coding genes. Interestingly, the NNS pathway might also regulate ncRNA transcription in response to the cell cycle since the Sen1 subunit is specifically degraded during G1 (Mischo et al. 2018).

Taken together, CUTs, MUTs, XUTs and NUTs represent a diverse population of unstable ncRNAs whose transcription could have functional consequences for regulating neighbouring gene expression.

1.4.3 Transcripts arising in mutants with defective chromatin integrity

Defects in chromatin remodelling pathways can result in cryptic transcription phenotypes in part due to reduced nucleosome occupancy along transcribed loci (see previous section 1.2). At least 1000 cryptic transcripts initiate from within gene bodies when the elongation factors Spt6 or Spt16 are inactivated (Kaplan et al. 2003;Mason and Struhl 2003;Cheung et al. 2008;Feng et al. 2016). Similarly, loss of histone modifying enzymes such as Set3 or Set2 also results in spurious transcription from the 5' or 3' ends of genes respectively (Li et al. 2007;Kim et al. 2012). Aberrant expression of these transcripts could have implications for organismal fitness and responses to environmental stresses. For example, Venkatesh *et al.* found in a *set2Δ* mutant, upregulation of 853 anti-sense ncRNAs which overlap with protein coding genes (Venkatesh et al. 2016). Transcription of associated coding genes was not affected in the steady state (Venkatesh et al. 2016). However, within 120 minutes of nutrient depletion, *set2Δ* mutants showed decreased expression of hundreds of genes which were associated with increased transcription of intragenic cryptic anti-sense ncRNAs (McDaniel et al. 2017). Therefore, spurious transcription is suppressed by chromatin modifying enzymes to avoid interference with the dynamics of protein coding gene expression.

1.4.4 Stable unannotated transcripts

Stable unannotated transcripts (SUTs) were initially described as a set of transcripts whose levels were less sensitive to *rrp6Δ*, unlike CUTs (Xu et al. 2009). SUTs are stably expressed in wildtype cells and mostly initiate from shared 5' and 3' NFRs flanking protein coding genes. Some SUTs which overlap neighbouring gene promoters are transcribed in tandem from a shared 5' NFR (Xu et al. 2009). Other SUTs are transcribed anti-sense to an ORF from a 3' NFR and are termed natural anti-sense transcripts (NATs) (Mellor et al. 2016).

Regulation by SUTs (otherwise called ncRNAs once they are annotated) have been described in the context of stress responses or in developmental programs in budding yeast. Examples of the former case include SUTs which are involved in regulating the transcriptional responses to zinc starvation, phosphate

starvation and carbon source changes (Bird et al. 2006;Uhler et al. 2007;Kim et al. 2012;Huber et al. 2016). In terms of a developmental program, transcription of the ncRNA *IRT1* (formerly called *SUT643*) and the anti-sense transcript *IME4-AS* mediate mating type control of gametogenesis at the *IME1* and *IME4* loci respectively (van Werven et al. 2012). Although the functions of most SUTs and ncRNAs are unknown, some attempts have been made to profile functional SUTs and ncRNAs in a systematic, genome wide manner (Parker et al. 2017). By screening a library of ncRNA deletion strains, Parker *et al*. identified four SUTs whose expression was essential for vegetative growth; deletion of individual SUTs led to either up- or downregulation of neighbouring genes (Parker et al. 2018). The regulatory outcomes of overlapping SUT or ncRNA expression is dependent on genomic context and usually but not always, represses neighbouring gene transcription (mechanisms to be further elaborated in section 1.5).

1.4.5 Transcript isoforms

Transcript isoforms are defined as mRNAs with differing 5' and/or 3' UTRs due to alternative TSS and TES usage. The most detailed profiling of budding yeast transcript isoforms till date was conducted using a method called transcript isoform sequencing (TIF-seq) (Pelechano et al. 2013). Pelechano *et al.* identified more than 370,000 major transcript isoforms in a genome of only 6000 genes, pointing to a diverse and heterogeneous transcriptome (Pelechano et al. 2014). Interestingly, 743 genes were found to express overlapping isoforms. Although not formally tested in the TIF-seq study, it is possible that 5' extended isoforms could interfere with the transcription of their overlapping neighbours, similar to that reported for overlapping sense/anti-sense pairs (Xu et al. 2011).

Transcript isoforms could also be an important mode of post-transcriptional gene regulation through the RNA itself. In brief, isoforms could be translationally repressed by sequence elements in the 5' UTR e.g. upstream ORFs (uORFs), localized to different cellular compartments or even translated into a different polypeptide (Pelechano et al. 2014). Examples of these regulatory mechanisms are further described in section 1.5. Of note, the TIF-seq study was conducted in budding yeast cells subjected to carbon source shifts. Other studies have

subsequently demonstrated that transcript isoforms have potential functions in regulating gametogenesis and the unfolded protein response (Cheng et al. 2018;Van Dalfsen et al. 2018).

1.5 Consequences of overlapping transcription

1.5.1 Transcriptional interference represses genes

Transcriptional interference (TI) is defined as the repression of a transcription unit (TU) by transcriptional activity arising from another TU *in cis*. Overlapping TUs either arise naturally, by cryptic transcription or by defective termination and read-through transcription (Mellor et al. 2016). Numerous examples of TI have been documented at individual loci and the repressive mechanism depends on the orientation of the overlapping TUs as well as the relative strength of the interfering promoter. Importantly, ncRNAs generated by TI need not be stable since the regulatory mechanism involves the act of transcription itself rather than the RNA product *per se*.

TI can occur when two promoters are arranged in tandem and transcription from the upstream promoter represses the downstream promoter **(Figure 1.9A and B)**. In this scenario, transcription initiating upstream directs RNAPII passage across the downstream promoter; co-transcriptional chromatin remodelling then establishes a repressive state at the downstream promoter. TI represses genes in response to environmental conditions. During zinc starvation, the transcription factor Zap1 activates transcription of the *ZRR1* and *ZRR2* ncRNAs which repress the promoters of *ADH1* and *ADH3* respectively (Bird et al. 2006). When serine is readily available, budding yeast cells express the *SRG1* ncRNA which generates a repressive chromatin state at the downstream *SER3* promoter (Martens et al. 2004). Since Ser3 is part of the serine biosynthetic pathway, *SRG1* transcription prevents wastage of cellular resources when serine is already abundant. Interestingly, *SRG1* transcription represses the *SER3* promoter by increasing nucleosome occupancy (Hainer et al. 2011;Thebault et al. 2011). Here, *SRG1* transcription modifies chromatin into a repressive state because chromatin remodellers such as Spt6, Spt16 and Spt2 assemble nucleosomes in the wake of RNAPII (Hainer et al. 2011;Thebault et al. 2011). The *SRG1* ncRNA product itself

is unlikely to play a role in the repression mechanism since insertion of the *SRG1* promoter at an ectopic locus was sufficient to repress a downstream gene, without preserving the sequence of the ncRNA (Martens et al. 2004). Looking genome wide, at least 743 tandemly arranged gene pairs in budding yeast express overlapping transcript isoforms that could mediate TI in a similar manner (Pelechano et al. 2013).

TI in tandem promoters also regulates gene expression in the context of cell-fate programs. During mating type control of gametogenesis, transcription of the *IRT1* ncRNA represses the *IME1* promoter and prevents starved haploid cells from initiating gametogenesis (van Werven et al. 2012). In this mechanism, *IRT1* transcription results in increased nucleosome occupancy and reduced accessibility of transcription factors at the *IME1* promoter. Crucially, *IME1* repression is dependent on co-transcriptional chromatin remodelling in its promoter; in *set2Δ set3Δ* double mutants, *IME1* is de-repressed despite *IRT1* transcription (van Werven et al. 2012).

Although both Set2 and Set3 are required for *IME1* repression, not all TI examples follow this rule. It has been reported that Set2 and Set3 modulate the expression of different genes depending on the length of the adjacent ncRNAs which overlap with their promoters (Kim et al. 2012;Kim et al. 2016). This is due to the differential recruitment of Set3 and Set2 to the 5' and 3' ends of TUs respectively (Carrozza et al. 2005b;Kim and Buratowski 2009). Therefore, an adjacent ncRNA that is shorter in length should cause H3K4me2 enrichment and hence promoter repression by the Set1/Set3C pathway, whereas a longer ncRNA should result in H3K36me3 enrichment and promoter repression by Set2/Rpd3S instead. Indeed, Kim *et al.* demonstrate that during a series of carbon source shifts, genes whose promoters overlap with longer transcripts (~2.0 kb) are repressed by Set2/Rpd3S whereas those with shorter overlapping transcripts (~0.9 kb), are repressed by Set1/Set3C (Kim et al. 2016).

Alternatively, TI can occur between two convergent, overlapping TUs **(Figure 1.9C)**. *IME4*, another regulator for efficient entry into gametogenesis, is repressed in haploid cells by the transcription of an anti-sense ncRNA (*IME4-AS*)

which overlaps the *IME4* promoter (Hongay et al. 2006;van Werven et al. 2012). Glucose repression of the *GAL1-10* locus is also mediated by anti-sense ncRNA transcription (Houseley et al. 2008). The transcription factor Reb1 drives transcription of an overlapping anti-sense ncRNA (*GAL10-AS*) from the 3' end of the *GAL10* gene. Houseley *et al.* found that when the media contained low amounts of galactose (0.1 g/L) and glucose (0.1-0.2 g/L), transcription of *GAL10-AS* repressed both the levels and induction kinetics of the neighbouring *GAL10* and *GAL1* genes. By mutating the ncRNA promoter, they found that *GAL10-AS* transcription leads to the deposition of repressive H3K36me3 marks at the *GAL1-10* ORFs. Consistent with this observation, this TI mechanism was dependent on Eaf3, a subunit of the Rpd3S de-acetylase complex which recognizes H3K36me2/3 (Houseley et al. 2008). Thus, TI by the *GAL10-AS* ncRNA ensures that budding yeast preferentially metabolize glucose in a complex environment with low amounts of glucose mixed with galactose. More recently, Huber *et al.* conducted a systematic study to determine the function of anti-sense ncRNAs overlapping with 188 coding genes tagged with GFP. Premature termination of most (~75%) anti-sense ncRNAs had no effect on protein abundance (Huber et al. 2016). However, those anti-sense ncRNAs with TUs that overlap with the TSS of sense genes tended to be repressive; these could represent cases of TI due to the establishment of repressive chromatin at the sense TSS (Huber et al. 2016).

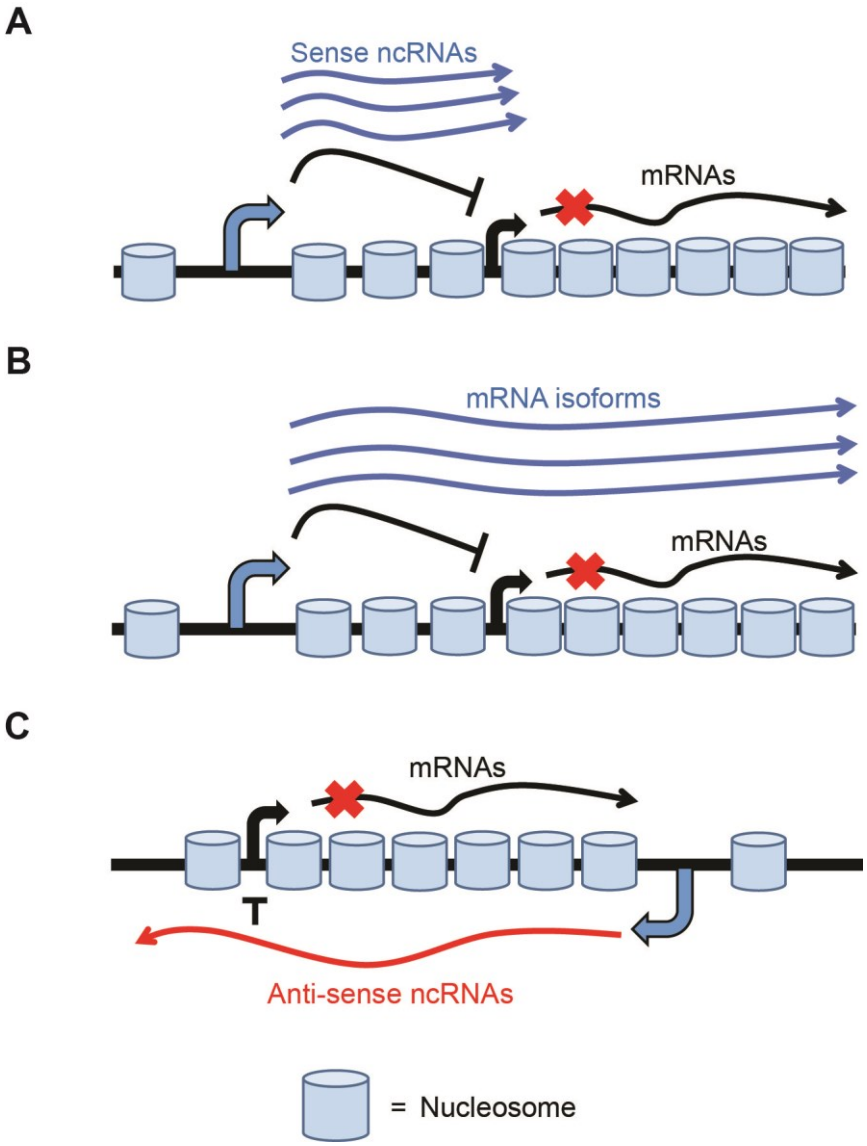


Figure 1.9 TI in different configurations resulting in repressive chromatin at the downstream promoter

(A) Tandem transcribed promoter configuration. Overlapping transcription is due to an intergenic transcript or a secondary transcript which initiates near the core promoter of another gene. **(B)** Tandem transcribed promoter configuration. Overlapping transcription is due to mRNA isoforms which share a CDS and a TES. **(C)** Convergent transcribed promoter configuration. Overlapping transcription is due to a sense/anti-sense pair. In each of these configurations, transcriptional interference occurs when overlapping transcription establishes repressive chromatin at the downstream promoter.

Besides the establishment of repressive chromatin, convergent sense/anti-sense TUs can interfere with each other due to head-to-head collisions with RNAPII transcribing opposite template strands (Hobson et al. 2012). Prescott and Proudfoot demonstrated this effect *in vivo* by rearranging the *GAL10* and *GAL7*

genes to overlap in a convergent orientation (Prescott and Proudfoot 2002). Transcription elongation of both nascent RNAs was inhibited when both promoters were induced simultaneously, likely due to RNAPII collision (Prescott and Proudfoot 2002;Hobson et al. 2012). In this model, productive transcription can only occur in one direction at a time **(Figure 1.10)** (Murray and Mellor 2016). This could explain why sense and anti-sense transcription from a convergent TU pair have been observed to be mutually exclusive in single cells (Castelnuovo et al. 2013). Perhaps individual cells can adopt one of three interconvertible states (transcriptionally silent, sense mRNA transcription or anti-sense ncRNA transcription) **(Figure 1.10)**. This could be a bet-hedging strategy for generating phenotypic variation in a genetically homogenous population, enabling robust responses to environmental changes (Snijder and Pelkmans 2011;Murray and Mellor 2016).

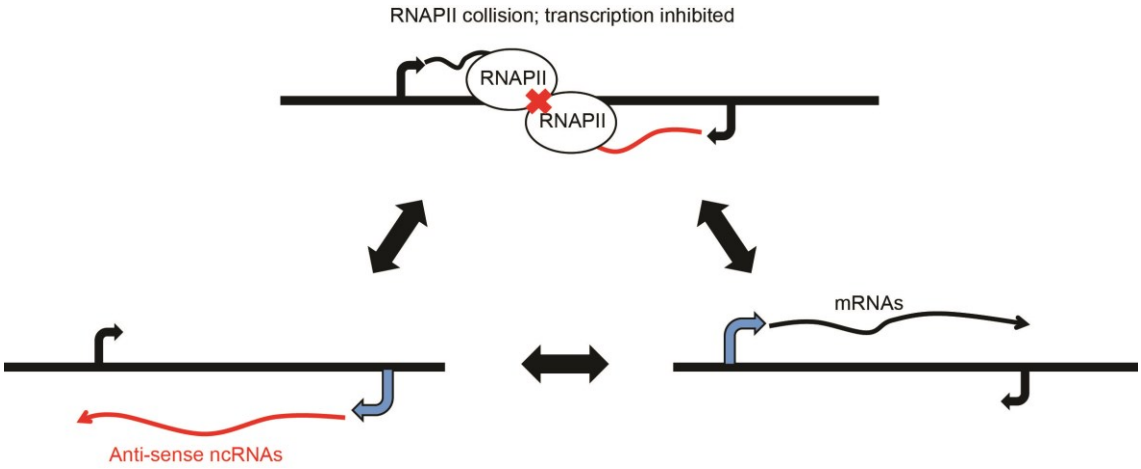


Figure 1.10 Model for dynamic variation in a convergent gene pair due to TI by colliding RNAPII

Nascent transcription from a sense and anti-sense gene pair is inhibited by collision between convergent RNAPII. A given cell can cycle between this transcriptionally silent "off" state, or transcribe productively in either the sense or the anti-sense direction.

Overlapping transcription can also interfere with sense transcription by insulating neighbouring TUs from productive transcription. This model is based off earlier studies suggesting that during transcription, the ends of a TU (TSS and TTS) are brought together by factors which associate with both sites such as Sua7 (TFIIB) and Ssu72 (O'Sullivan et al. 2004;Ansari and Hampsey 2005;Tan-Wong et al. 2012). Interference by TU insulation occurs when the transcription in the sense direction forms a loop which excludes the TSS and TTS of the neighbouring anti-sense TU and vice versa **(Figure 1.11 A)**. Gene regulation mediated by overlapping transcription and gene looping is illustrated by the *HMS2, SUT650* ncRNA and BAT2 neighbouring loci **(Figure 1.11 B)** (Nguyen et al. 2014). Most budding yeast cells metabolizing glucose transcribe *HMS2*, with a smaller fraction of them expressing a longer *HMS2* read-through transcript. It has been proposed that *HMS2* transcription forms gene loops which exclude the adjacent divergent *SUT650-BAT2* gene pair from productive transcription. However during galactose metabolism, transcription of the anti-sense *SUT650* ncRNA across the *HMS2* locus now excludes productive transcription from the sense *HMS2* TU but frees *BAT2* from transcriptional interference (Nguyen et al. 2014). Consistent with this model, mutants harbouring the *sua7-1* allele, which inhibits gene looping, demonstrate

increased levels of *HMS2* but decreased levels of *SUT650* during galactose metabolism (Nguyen et al. 2014). Thus, gene looping enhances transcriptional directionality and promotes reciprocal activation or repression of adjacent genes by transcriptional interference.

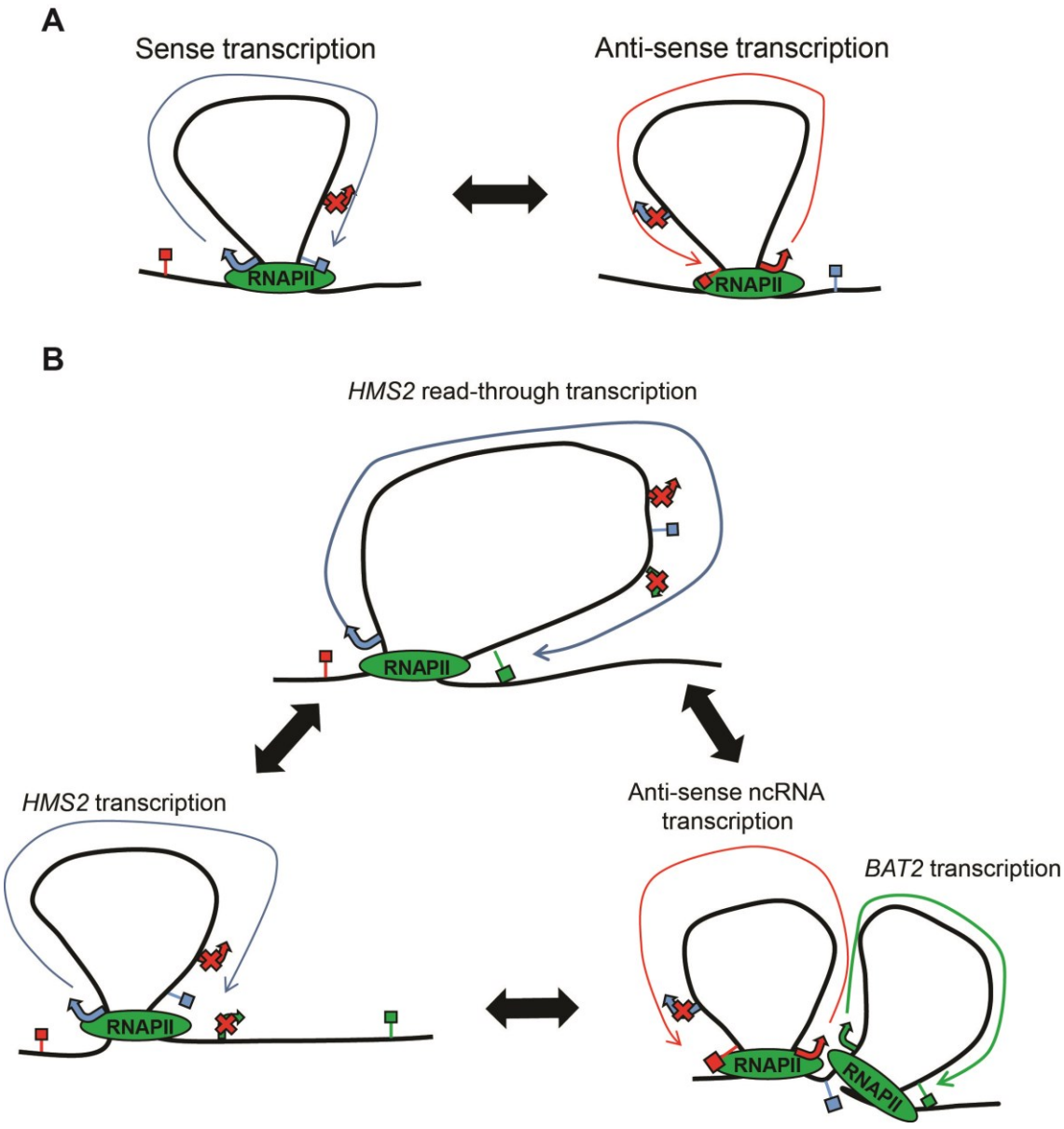


Figure 1.11 Transcriptional interference mediated by gene loops

(A) During transcription, the TSS (depicted as a directional arrow) and the TTS (depicted as a raised square) are brought together by factors (not depicted) which associate with RNAPII such as Sua7 (TFIIB) and Ssu72. Interference by TU insulation occurs when the transcription in the sense direction forms a loop which excludes the TSS and TTS of the neighbouring anti-sense TU and vice versa. **(B)** Complex regulation of the *HMS2-SUT650-BAT2* gene cluster during metabolic changes in budding yeast. During glucose metabolism, budding yeast cells express

HMS2 (blue) either as a regular transcript or an extended read-through transcript. In either these two states, the resultant gene loops prevent *SUT650* (red) and *BAT2* (green) from being productively transcribed. During galactose metabolism, cells express *SUT650* (red) and *BAT2* (green) while the *HMS2* TU (blue) is insulated from productive transcription. The double arrows indicate that cells can exist in one of three inter-convertible states, resulting in dynamic regulation of these loci in a population.

In addition, transcriptional interference mechanisms may also involve the formation of R-loops, which are a consequence of relatively stable hybridization between the nascent RNA and the template DNA (Chedin 2016;Belotserkovskii et al. 2017). During transcription, double-stranded DNA must be unwound to expose the template strand for initiation and elongation by RNAPII. This unwinding generates a positive supercoils (less relaxed) and negative supercoils (more relaxed) ahead and behind of RNAPII respectively **(Figure 1.12)**. These negative supercoils facilitate invasion of the template DNA strand by nascent RNA and subsequent hybridization as R-loops (Roy et al. 2010). The relative stability of R-loops block PICs from accessing the DNA template and could also lead to stalling and dissociation of elongating RNAPII (Belotserkovskii et al. 2017). In tandemly arranged promoters, transcription of an upstream overlapping RNA could form R-loops over the region containing the downstream promoter, resulting in TI **(Figure 1.12)**. R-loop formation has been demonstrated in the regulation of the human c-*MYC* and dihydrofolate reductase (*DHFR*) gene (Martianov et al. 2007;Yang et al. 2014). The human c-*MYC* gene is predominantly controlled by two tandem promoters and R-loops have been detected at these promoter regions (Marcu et al. 1997;Yang et al. 2014). R-loops in the c-*MYC* promoter region inhibit efficient transcription of this gene but accumulation of these promoter R-loops are inhibited by the topoisomerase TOP3B which is recruited by a scaffold protein TDRD3 (Yang et al. 2014). The repression of c-*MYC* transcription by R-loops could be physiologically important because overexpression of TDRD3 (and hence reduced R-loop accumulation) might contribute to c-*MYC* overexpression in some breast cancers (Groh and Gromak 2014). The human *DHFR* gene is also controlled by two tandem promoters (Masters and Attardi 1985). During serum starvation or quiescence, transcription from the upstream alternative *DHFR* promoter generates an RNA which represses the downstream promoter (Martianov et al. 2007). Importantly, a bandshift assay showed that R-loops were generated during *DHFR*

repression (Martianov et al. 2007). Taken together, R-loops constitute a possible mechanism for gene repression by TI.

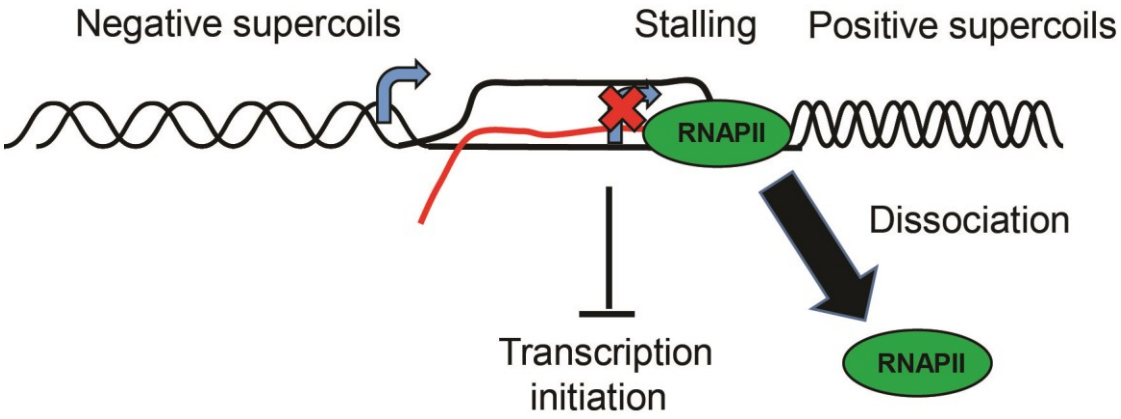


Figure 1.12 Transcriptional interference by R-loop formation

During transcription, double-stranded DNA must be unwound to expose the template strand for initiation and elongation by RNAPII. This unwinding generates a positive supercoils (less relaxed) and negative supercoils (more relaxed) ahead and behind of RNAPII respectively. R-loops are formed when an RNA product (red) invades negatively supercoiled DNA and hybridizes with the template DNA strand (black). R-loops can suppress transcription at the locus by either inhibiting transcription initiation at downstream tandem promoters or by stalling productive RNAPII elongation.

While different TI mechanisms have been described in this section, these models (interference by establishing repressive chromatin, gene looping or R-loop formation) are not necessarily mutually exclusive. Indeed, divergent promoters repressed by gene loops tend to be hypo-acetylated, a common characteristic of repressive chromatin (Tan-Wong et al. 2012). Both H3K36me3 and H3K4me2/3 are also enriched near the TSSs of genes with R-loop forming promoters (Sanz et al. 2016). Perhaps R-loop mediated stalling of RNAPII promotes the recruitment of histone methyltransferases to further reinforce gene repression (Chedin 2016). Several common themes emerge after considering different studies of repression by TI. Firstly, TI requires the promoter pair to be arranged in either a tandem or a convergent orientation. The extent of TI depends on the degree of overlap between the two TUs. Consistent with this statement, reducing overlapping transcription by inserting premature transcription terminator sequences can abolish TI (van Werven et al. 2012;Nguyen et al. 2014). Furthermore, the transcription of anti-sense ncRNA tends to be repressive if it passes through the TSS of a neighbouring gene (Huber

et al. 2016). Lastly, effective TI depends on the relative strengths of the two promoters. Relatively speaking, the stronger the interfering promoter, the more likely a neighbouring promoter will be sensitive to TI (Shearwin et al. 2005;Huber et al. 2016).

1.5.2 Other complex outcomes of overlapping transcription

Overlapping transcription can have other complex regulatory outcomes on neighbouring gene expression besides simple repression. In some cases, transcriptional overlap is necessary for full gene activation. In fission yeast, glucose starvation induces step-wise transcription of three upstream *fbp1*$^{+}$ ncRNAs which overlap with the canonical *fbp1+* promoter; promoter ncRNA transcription promotes an open chromatin state for transcription factors to bind and activate *fbp1+* (Hirota et al. 2008). Similarly, full activation of the human ε-globin gene cluster depends on overlapping ncRNA transcription initiating from the distal HS2 enhancer, into the ε-globin promoter (Ling et al. 2004;Ling et al. 2005). Similar to the *fbp1+* example, transcription of the intergenic ncRNA might maintain open chromatin at the ε-globin promoter for gene activation (Tuan et al. 1992;Gong et al. 1996;Gribnau et al. 2000).

Overlapping transcription can also modulate the dynamics of gene activation. When exposed to high phosphate, budding yeast cells repress *PHO5* and an anti-sense CUT/ncRNA (*PHO5-AS*) is initiated from 3' end of the *PHO5* ORF and transcribed across the *PHO5* promoter (Uhler et al. 2007). *PHO5-AS* transcription was not necessary for phosphate-mediated repression of *PHO5*. Instead during the switch to phosphate starvation, *PHO5-AS* transcription was associated with a more rapid establishment of a NDR at the *PHO5* gene, faster RNAPII recruitment and was thus necessary for proper kinetics of activation of *PHO5* (Uhler et al. 2007). How might overlapping ncRNA transcription lead to faster induction of the sense gene? In a genome-wide study, Murray *et al.* found that overlapping anti-sense transcription is associated with increased nucleosome occupancy in the promoters of their overlapping sense partners (Murray et al. 2015). However, these nucleosomes also tended to be enriched for acetylated

histones, reduced H3K36me3 and were turned over at higher rates with increasing anti-sense transcription. Therefore in this model, high anti-sense transcription establishes a dynamic chromatin environment and primes activation of sense gene promoters. It is still unclear how the effects of overlapping anti-sense transcription on chromatin can differ between loci (repressive vs. activating) (Houseley et al. 2008;Murray et al. 2015;Huber et al. 2016). Murray *et al.* suggest that these differences could be due to different compositions of elongating RNAPII complexes, especially for the examples where H3K36me3 is decreased following anti-sense transcription (Murray et al. 2015). Perhaps these differences in bulk histone modifications could also be a reflection of a highly heterogenous population of cells, with different sub-populations having dynamically inter-convertible chromatin states (Murray and Mellor 2016).

Overlapping transcription can also prime gene activation, in a model mediated by gene looping **(Figure 1.13)**. During osmotic stress, the Hog1 protein kinase is recruited to the 3' NDR of *CDC28* to activate transcription of an anti-sense ncRNA (*CDC28-AS*). This act of anti-sense transcription creates a gene loop which brings the *CDC28* sense promoter in close proximity to the *CDC28-AS* promoter, subsequently facilitating Hog1-dependent activation of *CDC28* sense transcription (Nadal-Ribelles et al. 2014) .

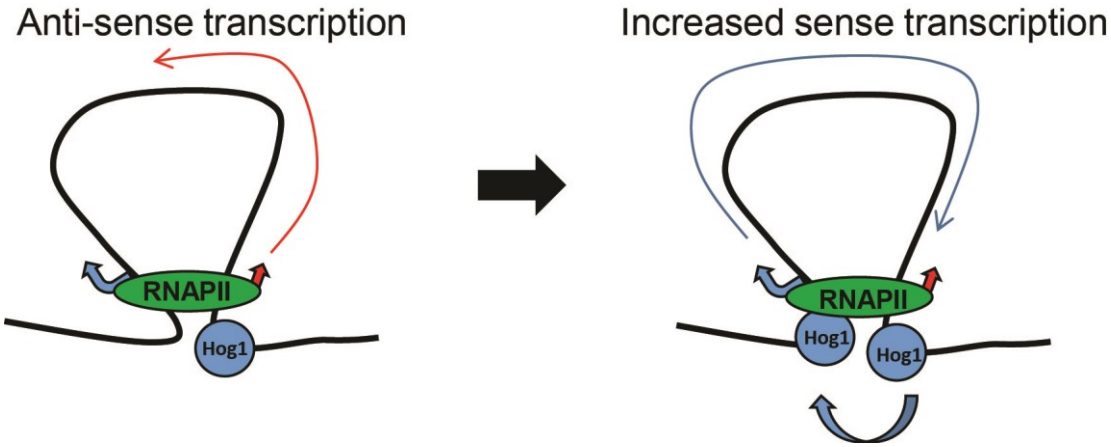


Figure 1.13 Overlapping transcription promotes gene activation by gene looping
During transcription, two different TSSs (depicted by blue and red directional arrows) are brought together by factors (not depicted) which associate with RNAPII such as Sua7 (TFIIB) and Ssu72. Gene activation occurs when anti-sense transcription (red) creates a gene loop which brings the sense promoter (blue) in

close proximity with the anti-sense promoter (red). This facilitates a “hand-off” of transcription factors such as Hog1 to the sense promoter and transcription in the sense direction is increased as a result.

In at least one reported case, a combination of both tandem and convergent TUs creates a transcriptional “toggle switch” to regulate the *FLO11* gene (Bumgarner et al. 2009). In this regulatory circuit, the ncRNA *ICR1* is transcribed in tandem with *FLO11*, and the ncRNA *PWR1* is convergent to *ICR1*. To silence *FLO11* expression, cells transcribe *ICR1* which interferes with the downstream *FLO11* promoter. On the other hand to activate *FLO11* expression, cells transcribe *PWR1* which interferes with *ICR1* transcription in the opposite strand. In another example, a combination of two tandem ncRNAs *IRT2* and *IRT1*, regulate *IME1* expression and form a regulatory circuit for mating type control of entry into budding yeast gametogenesis (Moretto et al. 2018). Additionally, overlapping transcription can be utilized by the cell to coordinate gene expression at multiple loci in response to environmental changes. As seen in the *HMS2-SUT650-BAT2* example, overlapping sense and anti-sense transcription mediates reciprocal activation and repression in some tandemly arranged genes during different phases of the yeast metabolic cycle **(Figure 1.11 B)** (Nguyen et al. 2014). These studies paint a more complex picture whereby mutually exclusive overlapping transcription events in a mixed cell population can generate multiple possible chromatin states at a given promoter and can also allow for swifter, coordinated responses to rapidly changing environmental conditions. Thus, cells can co-opt multiple overlapping TUs to achieve complex control of gene expression by TI.

Taken together, regulation by overlapping transcription cannot be generalized into simple binary outcomes of activation or repression. Instead, overlapping transcription can modulate coding gene transcription in different ways depending on genomic context and environmental signals. Regulatory flexibility involving overlapping transcription is reminiscent of how eukaryotic cells have evolved transcription factors which can either activate or repress gene expression depending on signalling cues. For example in budding yeast, the Ume6 DNA-binding protein can either act as a repressor when incorporated into the Rpd3L co-repressor complex during vegetative growth, or act as an activator when complexed with Ime1 (Rubin-Bejerano et al. 1996;Carrozza et al. 2005a). Similarly in mammalian cells, the estrogen receptor alpha (ERα) transcription factor

associates with various co-activators such as the p300 histone acetyltransferase (HAT) and the p160 family of proteins to activate gene expression (Shang et al. 2000). However, ERα can also repress certain genes directly and rapidly; this mechanism involves the initial recruitment of the p300 HAT, which in turn recruits the co-repressor CtBP1 (Stossi et al. 2009). It is not known how CtBP1 is preferentially associated with p300 in repressive, but not activating ERα complexes, although different post-translational modifications of p300 could be involved (Stossi et al. 2009). Therefore, both overlapping transcription and co-activator/co-repressor complexes contribute to diverse methods of regulating gene expression in eukayotes.

1.5.3 Examples from overlapping transcription from mammalian cells

Mammalian genomes are pervasively transcribed, with non-protein-coding gene transcripts accounting for most of it (Djebali et al. 2012). Interleaved TUs are an important feature of the human genome despite it being less gene dense compared to that of budding yeast. In fact, an estimated 25% of human genes have an overlapping anti-sense partner transcript (He et al. 2008;Mayer et al. 2015). In higher eukaryotes including human cells, a wide range of 5' mRNA isoforms are also expressed often in a cell type-specific manner (Wang et al. 2008;Aanes et al. 2013;Brown et al. 2014). For example, a recent study demonstrated thousands of TSS switching events during mouse cerebellar development (Zhang et al. 2017). Similar to budding yeast, transcription of some of these overlapping isoforms is also important for gene regulation in mammalian cells. During mouse embryonic stem (ES) cell differentiation, transcription of the overlapping anti-sense lncRNA *Airn* silences the *Igf2r* gene (Latos et al. 2012). In mouse cells, ncRNA transcription from the T early alpha promoter can interfere with expression of downstream genes in the T-cell receptor alpha chain (*Tcra*) locus (Abarrategui and Krangel 2007). In human primary arterial endothelial cells, exposure to cytokines like interferon-γ leads to repression of the immune gene *MICA* via the transcription of an overlapping 5'- extended *MICA* isoform (Lin et al. 2018).

Thus far, I have shown how TI is a conserved mechanism of gene regulation from the relatively simple budding yeast to more complex, higher eukaryotes. As expected, there are some similarities in TI mechanisms between budding yeast and mammals. One similarity is that TI involves the establishment of repressive nucleosomes in the downstream promoter, as seen in *SER3* repression in budding yeast, *tgp1+* repression in fission yeast and *MICA* repression in human cells (Hainer and Martens 2011;Ard et al. 2014;Lin et al. 2018). Genome-wide studies in both yeast and humans further suggest that overlapping anti-sense transcription is associated with increased nucleosome occupancy in the sense promoter, and it is possible that some of these genes are repressed by TI (Mayer et al. 2015;Murray et al. 2015). In contrast, TI mechanisms could also differ between yeast and higher eukaryotes due to added complexity in these organisms. For example, DNA methylation has been implicated in TI in mammals but not in budding yeast, which lack this genomic feature (Proffitt et al. 1984). In humans, a rare inherited form of α-thalassemia is caused by a deletion which positions a highly expressed, constitutively active antisense TU in close proximity to the α2-globin gene (*HBA2*). Overlapping read-through transcription from the ectopic antisense transcript induces DNA methylation in the CG-rich region of the *HBA2* promoter, establishes repressive chromatin and consequently silences *HBA2* expression (Tufarelli et al. 2003). A recent study in mouse ES cells showed that H3K36me3 recruits Dnmt3b to methylate DNA in gene bodies to suppress cryptic transcription (Neri et al. 2017). Speculatively, some mammalian coding mRNAs might be repressed by TI from an overlapping transcription event, if the deposition of H3K36me3 at their promoters also leads to Dnmt3b recruitment. Finally, R-loops might play a larger role in TI of mammalian promoters than in budding yeast. This is because unlike in budding yeast, about 60% of human protein coding genes have GC rich regions in their promoters (Illingworth and Bird 2009). These GC rich stretches are thermodynamically favourable for R-loop formation and widespread R-loop formation at these promoters was detected by DNA:RNA ImmunoPrecipitation sequencing (DRIP-seq) (Ginno et al. 2012). Despite these differences, budding yeast can serve as a good starting point to understand gene regulation by overlapping transcription in *cis*.

1.6 Gene regulation by ncRNA products *in trans*

Compared to *cis* regulation due to overlapping transcription, there are fewer reported cases of gene regulation in *trans* by the ncRNA product in budding yeast. In one example, ectopic expression of an anti-sense XUT (*TY1-RTL*) is sufficient to suppress transcription of the *TY1* retro-transposon (Berretta et al. 2008). In another example, the *PHO84* locus has two anti-sense CUTs (*PHO84-AS*) initiating from a region 20-80 bp downstream of the *PHO84* stop codon (Camblong et al. 2007). Ectopic expression of full length *PHO84-AS* transcripts can trigger silencing in *trans* at the endogenous *PHO84* locus, whereas post-transcriptional cleavage of *PHO84-AS* impairs silencing (Camblong et al. 2009). A recent study profiling functional ncRNAs in yeast identified *SUT075* as an essential ncRNA which could rescue lethality in a *SUT075Δ* mutant when expressed from a different locus, suggesting that this ncRNA functions *in trans* (Parker et al. 2018). However in these budding yeast examples, it is not clear how ncRNAs mechanistically regulate gene expression in *trans*.

How genes can be regulated *in trans* is better described in other model organisms. In the fission yeast *Schizosaccharomyces pombe*, entry into gametogenesis is regulated by a RNA binding-protein (RBP) Mei2 and a ncRNA called meiRNA (Watanabe and Yamamoto 1994). The meiRNA transcripts promote the association of Mei2 with another RBP called Mmi1 in the nucleus; this association inhibits exosome-mediated degradation of meiotic transcripts and promotes entry into gametogenesis (Watanabe and Yamamoto 1994;Yamashita et al. 1998;Harigaya et al. 2006). In human cells, a lncRNA called *Xist*, is tethered to extra copies of the X chromosome by a RBP called YY1; *Xist* association with the chromosome causes gene silencing by recruiting further chromatin remodellers such as histone deacetylases and the Polycomb Repressive Complex (Chu et al. 2015). In these cases, the RNA product in concert with RBPs, are part of the mechanism by which genes are regulated. Further work will uncover if more of such examples of regulation *in trans* exist in budding yeast.

1.7 Genome wide techniques to study ncRNAs and transcript isoforms

1.7.1 Microarrays and mRNA sequencing

Microarrays and short-read mRNA sequencing have been used to profile intergenic transcription in different cell types (Primig et al. 2000;He et al. 2008;Jacquier 2009;Brown et al. 2014). While useful, both these approaches involve converting RNA to DNA, and then fragmenting said libraries before hybridization on a microarray chip or before sequencing and alignment to a genome. These techniques have inherent difficulties when applied to the study of overlapping ncRNAs and transcript isoforms. It is difficult to determine the relative abundances of different isoforms at the same locus just from reads mapping to common regions. Additionally, microarrays or mRNA sequencing can neither identify the TSSs or TESs of transcripts in general, nor measure the levels of nascent transcription in wildtype cells. Robust measurements of isoform abundances or TSS/TES usage is critical for genome-wide studies TI or other regulatory outcomes due to overlapping transcription. These limitations have led to the development of other specialized protocols to study ncRNAs and transcript isoforms.

1.7.2 Profiling nascent transcripts

As previously discussed, transcription of overlapping mRNA isoforms or ncRNAs can affect gene expression locally. However, microarrays or mRNA sequencing can only measure steady state levels of relatively stable transcripts. Steady-state bulk RNA levels do not always reflect transcriptional activity or nascent transcription. More recently, specialized techniques have been developed to profile nascent transcription of all transcripts. One approach is called Global Run-On Sequencing (GRO-Seq), which involves isolating nuclei and incubating them with an rNTP mix where UTP is substituted with brominated-UTP (Br-UTP) (Gardini 2017). On-going transcription would utilize Br-UTP, labelled transcripts would be isolated with an anti-Br-UTP antibody and sequenced to provide a snapshot of nascent transcription at a resolution of 30-100 nt (Gardini 2017). An

improvement called PRO-seq involves similar steps except that biotin-rNTPs are used. Incorporation of biotin-rNTPs inhibits further RNAPII elongation, allowing for the analysis of nascent transcription at single nucleotide resolution(Mahat et al. 2016). The drawbacks of nuclear run-on assays like GRO-Seq and PRO-seq are that they require very large amounts of starting cellular material, are labour intensive and artifacts can be introduced during nuclei isolation (Mayer et al. 2015).

More recently, native elongating transcript sequencing or nascent elongating transcript sequencing (NET-seq) has been used as an alternative method to profile nascent transcription (Churchman and Weissman 2011;Mayer et al. 2015;Nojima et al. 2015). In NET-seq, stable elongating RNAPII complexes with their nascent RNAs are isolated either by immunoprecipitation or by cellular fractionation. RNAPII complexes are then denatured and formerly bound nascent RNAs are purified and sequenced. The advantages of NET-seq are that it measures nascent transcription *in vivo* at single nucleotide resolution, avoiding artifacts introduced during nuclei isolation (Mayer et al. 2015;Nojima et al. 2015).

1.7.3 Profiling the ends of transcripts

Although mRNA seq and NET-seq can be used to identify intergenic or promoter transcription events, the interpretation of these results is complicated by overlapping RNA seq reads from different transcripts. Without the ability to distinguish the ends of interleaved transcripts, it is not possible to accurately determine what isoforms or ncRNAs are transcribed at a locus. Thus a genome wide technique that profiles both TSS and TES usage quantitatively, is required to characterize overlapping ncRNAs or mRNA isoforms and to infer their roles in transcriptional regulation.

Specialized techniques have been used to determine the TSSs of transcripts, which hinge on identifying the 5' most nucleotide downstream of the m^7Gppp cap. One popular TSS mapping technique is called 5'-cap analysis of gene expression (5'-CAGE) (Takahashi et al. 2012). This method involves reverse transcription (RT) and chemical labelling of the m^7Gppp cap with biotin (Carninci et

al. 1996). Biotinylated RNA:cDNA hybrids are then isolated by magnetic streptavidin beads and first strand cDNA is used as a template for second strand cDNA synthesis. The use of customized RT and second strand primers results in double stranded DNA flanked on both ends with a restriction site for enzymes such as EcoP15I. EcoP15I cuts 27 bp downstream of the TSS to generate 27 bp CAGE tags which can then be amplified, sequenced and mapped to a reference genome; the first nucleotide following the EcoP15I site is the TSS. Another orthogonal technique called TSS-seq involves the de-phoshorylation of 5' un-capped transcripts (usually degradation intermediates), followed by treatment with an acid pyrophosphatase which removes the m^7Gppp cap to leave a 5' phosphate group (Arribere and Gilbert 2013). Hence, only those transcripts which had a cap (marking *bona fide* TSSs) would have a 5' end which is competent for ligation of a customized oligonucleotide. Processed RNA is then converted to double stranded cDNA, amplified, sequenced and mapped to a reference genome; the first nucleotide following the custom oligonucleotide sequence is the TSS.

Techniques used to determine the TES, or 3' most nucleotide adjacent to the poly(A) tail, must address the difficulties in sequencing the homopolymeric poly(A) tails of mRNAs (Quail et al. 2012). One approach called poly(A) site seq (PAS-seq) uses an anchored oligo d(T) primer with a stretch of 20Ts followed by A/G/C (V), followed by any nucleotide (N) for the RT reaction (Shepard et al. 2011). During RT, the VN di-nucleotide sequence hybridizes to the last 2 nucleotides preceding the poly(A) stretch. After second strand synthesis and amplification, the cDNA libraries have a stretch of only 20 As at one end. To avoid sequencing through the stretch of 20 As, a custom sequencing primer with 20 Ts is used in the Illumina flowcell. Therefore, sequencing would initiate at the junction between the 20As and the TES (Shepard et al. 2011). Wilkening *et al.* developed an orthogonal approach to map TESs called 3'T-fill (Wilkening et al. 2016). 3'T-fill also uses an anchored oligo d(T) primer for RT. However unlike PAS-seq, 3'T-fill modifies the Illumina clustering step, by filling in poly(A) stretches with unlabelled dTTPs. Sequencing would then initiate at the junction between the already filled 20 A/T stretch and the TES (Wilkening et al. 2016). A third strategy involves drastically shortening the poly(A) tails by enzymatic digestion. A technique called 3' Region Extraction And Deep Sequencing (3'-READS) involves the hybridization of mRNA fragments

containing the poly(A) tails to a custom oligonucleotide with 45 Ts at the 5' end and 5 Us at the 3' end (Hoque et al. 2013). After hybridization, treatment with RNAse H cleaves the duplex molecules at the junction between the Ts and Us, thereby shortening the Poly(A) stretch for subsequent sequencing. A separate approach called 3T-seq introduces a GsuI restriction site in the first strand cDNA, within the homopolymeric A/T region (Lai et al. 2015). GsuI is a type IIS restriction enzyme which makes a staggered cut 16 nucleotides downstream of its recognition site i.e. CTGGAG(16/14)^. This also produces double stranded cDNA with a truncated poly(A) tail for sequencing.

Individual TSS or TES mapping approaches have a limitation of failing to preserve information of both ends from a given transcript simultaneously. Both transcript isoform sequencing (TIF-seq) and Gene Identification Signature – Paired End Tagging (GIS-PET) address this limitation using different strategies. TIF-seq preserves 5' and 3' end information by circularizing cDNA and sequencing the fragments containing the paired-end junction (Pelechano et al. 2014). GIS-PET involves ligating cDNA into a plasmid vector (Ng et al. 2005). The cDNA insert is flanked on both ends by Mme1 restriction sites. Mme1 is a type IIS restriction enzyme which makes a staggered cut 20 nucleotides downstream of its recognition site i.e. TCCRAC(20/18)^. Thus, Mme1 digestion removes most of the CDS and preserves the sequences at the termini of the cDNA insert to form paired end tags (PETs). Multiple PETs, each representing both TSS and TES information from a transcript can then be concatenated, cloned into a vector and sequenced. While the preservation of both 5' and 3' information is useful for profiling transcript isoforms, both approaches require full length cDNA synthesis. The biases introduced due to differential ease of reverse transcription and other enzymatic manipulation of full-length RNAs of varying lengths complicates quantitative analysis of isoforms from either technique.

In summary, there are different approaches to profile nascent transcription, TSS or TES usage. Selection of a technique would depend on experimental and practical considerations such as cost, complexity of procedures, amount of required starting material as well as whether quantitative analysis is required.

1.8 Aims of this thesis

Cell differentiation programs require the coordinated regulation of different genes so that cellular changes can occur in a timely manner. The work in this thesis uses budding yeast gametogenesis as a model to understand how ncRNAs and transcript isoforms regulate gene expression during cell differentiation. Throughout gametogenesis, there is substantial transcription of long non-coding RNAs (ncRNAs) and long mRNA isoforms. However, it is unclear whether or how these transcripts contribute to gene regulation during gametogenesis. This thesis aims to address the following research questions. Do budding yeast cells regulate gene expression during gametogenesis by overlapping RNA isoform or ncRNA transcription, similar to how *IRT1* transcription represses *IME1*? How many such examples exist across its genome? What is the molecular mechanism by which mRNA isoform or ncRNA transcription regulate gene expression and do chromatin remodellers play a role here? Why is it advantageous for the cell to regulate gene expression via the transcription of overlapping mRNA isoforms or ncRNAs? Are there criteria common to genes repressed by transcriptional interference? What are the other effects of overlapping isoform or ncRNA transcription on downstream genes, apart from repression? To explore these questions, it is important to characterize the roles of ncRNAs and mRNA isoforms controlling gene expression during gametogenesis at both a single locus and at a genome-wide level.

In chapter 3, I describe an optimized approach of obtaining highly synchronous, sporulating cultures of budding yeast through inducible expression of *IME1*, the master regulator controlling entry into gametogenesis. Timely expression of *IME1*, coupled with nutrient starvation, enables synchronous progression through gametogenesis as assessed by flow cytometry and nuclear staining. Importantly, this approach does not require growth in acetate-containing media (BYTA), but can be performed in cells that were grown in rich medium (YPD) till saturation. This allows for higher densities of starting cultures to be obtained, ensuring that enough cellular material can be obtained in high resolution sporulation time courses. Furthermore, inducible expression of *IME1* and *NDT80* can be combined in the same cells, allowing for an even greater degree of synchrony from meiotic DNA replication to meiotic divisions. This system of

obtaining sporulating cultures facilitates downstream efforts to characterize the meiotic transcriptome.

In chapter 4, I describe and characterize how transcription of a 5' extended transcript isoform regulates the expression of the kinetochore protein *NDC80*. *NDC80* levels are repressed during meiotic prophase to enforce reductional chromosome segregation during meiosis I. I show how transcriptional control of *NDC80* during prophase I is brought about by a developmentally controlled switch to a transcript isoform with an extended 5' UTR. Transcription of this isoform, hereafter referred to as a **l**ong, **u**ndecodable **t**ranscript **i**soform (luti mRNA), represses the canonical *NDC80* promoter in a process partly dependent on the histone modifying enzymes Set2 and Set3. Thus, transcription of an extended mRNA transcript isoform represses a downstream promoter in a timely and dynamic manner.

The study of the *NDC80* locus raises the question of whether other genes are also regulated in a similar manner. While informative, standard RNA seq approaches neither satisfactorily reveal regulatory relationships between different transcript isoforms, nor allows for annotation of alternative 5' and 3' end usage during gametogenesis. This necessitates the use of specialized library preparation techniques to accurately map the ends of all capped and poly-adenylated transcripts during meiosis. In chapter 5, I review previously reported methods for transcript end mapping and describe an optimized protocol to study the diversity of mRNA isoforms and expression of ncRNAs.

In chapter 6, I use an optimized transcript end sequencing technique (TE-seq), to map both the 5' and 3' ends of transcripts during gametogenesis. Samples from time courses of wildtype cells were sequenced to identify new features of meiotic transcriptional regulation. I propose that transcription of non-coding mRNA isoforms can dynamically repress transcription of the coding mRNA isoform at distinct stages of gametogenesis. In other cases, overlapping transcription is instead associated with tuning or activation of gene expression. As a complementary approach, I have also applied this analysis to mutant cells to

explore if gene regulation could be compromised by the loss of chromatin remodellers (e.g. in *set2* or *set3* mutants).

This thesis details my efforts to understand how the transcription of overlapping ncRNA and mRNA isoforms contribute to the timely and dynamic regulation of genes during a developmental program. Using both locus specific and genome-wide approaches, hitherto unappreciated examples of transcriptional regulation are described, providing a rationale for extending such analyses to more complicated differentiation programs in higher eukaryotes.

Chapter 2. Materials & Methods

2.1 Construction of yeast strains

All yeast strains used in this thesis were derived from the sporulation proficient SK1 strain background except for the strains that harboured the LexA/*lexO* system for which the W303 strain background was used. The genotypes are listed in **Table 2.1**. The *CUP1* promoter fusion with *IME1* (*pCUP-IME1*) strain was made as follows. A region of the *pFA6a-KanMX6-pCUP1A-3HA* plasmid was amplified using the primers 5'-
GCATTGATATTTTCAAACTTATATAATTAATAATAATTAATAGCGCTTAGTTTAAA
GAAgaattcgagctcgtttaaac-3' and 5'-
GAAACCATCTTCTAAGGCAGCGTGAAGTTTTCCATG
CATATCCGCTTGCATgcactgagcagcgtaatctg-3'. Uppercase letters refer to *IME1* specific sequences while lowercase letters refer to the sequences for amplifying the *CUP1* promoter with N-terminal 3HA tags (Longtine et al. 1998). For the untagged version of *pCUP-IME1*, either the *pFA6a-KanMX6-pCUP1A-3HA* plasmid was amplified using the primers 5'-
GCATTGATATTTTCAAACTTATATAATTAATAATAATTAATAGCGCTTAGTTTAAA
GAAgaattcgagctcgtttaaac-3' and 5'-
CTAAGGCAGCGTGAAGTTTTCCATGCATATCCGCTTGCATtttatgtgatgattgattgatt
g-3'.

The *CUP1* promoter fusion with *IME4* (*pCUP-IME4*) strain was made as follows. A region of the *pFA6a-KanMX6-pCUP1A-3HA* plasmid was amplified using the primers 5'- CGACACACCTAAAACTGATTAGAATCGTTTCAAGATGCTT
gaattcgagctcgtttaaac-3' and 5'-
AATTCTGGATCAGAAAATGTACTAGTTTATCGTTAATCATgcactgagc
agcgtaatctg-3'. Uppercase letters refer to *IME4* specific sequences while lowercase letters refer to the sequences for amplifying the *CUP1* promoter with N-terminal 3HA tags (Longtine et al. 1998). For the untagged version of pCUP-IME4, the *pFA6a-KanMX6-pCUP1A-3HA* plasmid was amplified using the primers 5'-
CGACACACCTAAAACTGATTAGAATCGTTTCAAGATGCTT
gaattcgagctcgtttaaac-3' and 5'-

AATTCTGGATCAGAAAATGTACTAGTTTATCGTTAATCATtttatgtgatgattgattgattg-3'.

The strains were generated by a one-step promoter replacement protocol as described (Longtine et al. 1998). Subsequently, the haploid transformants were backcrossed, from which homozygous diploid cells were generated. In other strains, the *KanMX6* marker can be replaced with other selection markers by transformation with amplicons generated from different pringle plasmids.

2.2 Yeast strain genotypes

Strain	Genotype
FW1208	*MAT**a**, ho::LYS2, ura3, leu2::hisG, his3::hisG, trp1::hisG, UME6-3V5::His3MX* *MATα, ho::LYS2, ura3, leu2::hisG, his3::hisG, trp1::hisG, UME6-3V5::His3MX*
FW1472	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::KanMX, ime4::pCUP-3HA-IME4::KanMX, NDC80-3V5::KanMX, set2::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: KanMX, ime4::pCUP-3HA-IME4:: KanMX, NDC80-3V5::KanMX, set2::His3MX*
FW1509	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG*
FW1510	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG*
FW1511	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG*
FW1541	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3*
FW1810	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::KanMX, ime4::pCUP-3HA-IME4::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::KanMX, ime4::pCUP-3HA-IME4::KanMX*
FW1868	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::NatMX, ime4::pCUP-3HA-IME4::NatMX, NDC80-3V5::KanMX, NDC80::pndc80(600-300)-NDC80-3V5::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::NatMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX, NDC80::pndc80(600-300)-NDC80-3V5::His3MX*
FW1871	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::NatMX, ime4::pCUP-3HA-IME4::NatMX, NDC80-3V5::KanMX, NDC80::pndc80(600-500)-NDC80-3V5::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: NatMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX, NDC80::pndc80(600-500)-NDC80-3V5::His3MX*
FW1884	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-IME1::HphMX, ime4::pCUP-IME4::NatMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-IME1::HphMX, ime4::pCUP-IME4::NatMX*

FW1899	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: NatMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80::pndc80(600-300)::His3MX, NDC80-3V5::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: NatMX, ime4::pCUP-3HA-IME4:: NatMX*
FW1900	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: NatMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: NatMX, ime4::pCUP-3HA-IME4:: NatMX*
FW1902	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX, ime4::pCUP-3HA-IME4::NatMX, NDC80-3V5::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX*
FW1922	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX, ime4::pCUP-3HA-IME4::NatMX, NDC80-3V5::KanMX, set2::His3MX, set3::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX, set2::His3MX, set3::His3MX*
FW1923	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: NatMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: NatMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80::pndc80(600-300)-NDC80::His3MX*
FW2444	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::KanMX*
FW2480	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ime4::pCUP-3HA-IME4::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ime4::pCUP-3HA-IME4::KanMX*
FW2795	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3*
FW2912	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, set3::His3MX, set2::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, set3::His3MX, set2::His3MX*
FW2928	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX, ime4::pCUP-3HA-IME4::NatMX, NDC80-3V5::KanMX, set3::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX, set3::His3MX*
FW2929	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX, ime4::pCUP-3HA-IME4::NatMX, NDC80-3V5::KanMX, set2::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX, set2::His3MX*

FW2957	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX, ime4::pCUP-3HA-IME4::NatMX, SUA7-3V5::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, SUA7-3V5::KanMX*
FW3033	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX, ime4::pCUP-3HA-IME4::NatMX, set1::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, set1::KanMX*
FW3034	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::KanMX, ime4::His3MX, NDC80-3V5::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: KanMX, ime4::His3MX, NDC80-3V5::KanMX*
FW3058	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ime1::His6MX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ime1::His6MX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX*
FW3856	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX, ime4::pCUP-3HA-IME4::NatMX, NDC80-3V5::KanMX, ndt80::LEU2* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX, ndt80::LEU2*
FW4114	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX, ime4::pCUP-3HA-IME4::NatMX, NDC80-3V5::KanMX, upf1::NatMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80-3V5::KanMX, upf1::NatMX*
FW4644	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, NDC80-3V5::KanMX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, NDC80-3V5::KanMX*
FW4911	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ndt80::LEU2* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ndt80::LEU2*
FW5530	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, SUA7-3V5::KanMX, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80::pndc80(600-300)::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, SUA7-3V5::KanMX, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, NDC80::pndc80(600-300)::His3MX*
FW5767	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, set2::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, set2::His3MX*
FW5770	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, set3::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, set3::His3MX*
FW6083	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, SPT16::SPT16-3V5-AID::KanMX6, his3::pCUP-OsTIR::His3MX* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG,*

	irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, SPT16::SPT16-3V5-AID::KanMX6, his3::pCUP-OsTIR::His3MX
FW6109	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, SPT16::SPT16-3V5-AID::KanMX6* *MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1::HphMX , ndt80::pGAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, SPT16::SPT16-3V5-AID::KanMX6*
UB91	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ura3::pGPD1-GAL4(848).ER::URA3*
UB1217	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, HisMX:pGAL-Ndc80-3V5:KanMX, ura3::pGPD1-GAL4(848).ER::URA3* pGAL integrated 536 bp upstream of Ndc80 AUG
UB1218	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, His3MX::pGAL-Ndc80-3V5:KanMX, ura3::pGPD1-GAL4(848).ER::URA3* pGAL integrated 536 bp upstream of Ndc80 AUG
UB1235	*MATα, ho::LYS2, ura3, leu2::hisG, his3::hisG, trp1::hisG, set2::His3MX, set3::His3MX, HisMX::pGAL-Ndc80-3V5::KanMX, ura3::pGPD1-GAL4(848).ER::URA3* pGAL integrated 536 bp upstream of Ndc80 AUG
UB1236	*MATα, ho::LYS2, ura3, leu2::hisG, his3::hisG, trp1::hisG, set2::His3MX, His3MX::pGAL-Ndc80-3V5::KanMX, ura3::pGPD1-GAL4(848).ER::URA3* pGAL integrated 536 bp upstream of Ndc80 AUG
UB1237	*MATα, ho::LYS2, ura3, leu2::hisG, his3::hisG, trp1::hisG, set3::His3MX, His3MX::pGAL-Ndc80-3V5::KanMX, ura3::pGPD1-GAL4(848).ER::URA3* pGAL integrated 536 bp upstream of Ndc80 AUG
UB1240	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, Ndc80-3V5::KanMX, ura3::pGPD1-GAL4(848).ER::URA3*
UB1252	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, Ndc80-3V5::KanMX, ura3::pGPD1-GAL4(848).ER::URA3*
UB3338	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ura3::pGPD1-GAL4(848).ER::URA3, His3MX::pGAL-Ndc80* pGAL integrated 536 bp upstream of Ndc80 AUG
UB3351	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ura3::pGPD1-GAL4(848).ER::URA3, ndc80Δ:KanMX4, leu2::NDC80-3V5::LEU2*
UB3370	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, GAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, ndc80Δ:KanMX4, leu2::NDC80-3V5:LEU2*
UB3545	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, set2::His3MX, Ndc80-3V5:KanMX, ura3::pGPD1-GAL4(848).ER::URA3*
UB3547	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, set3::His3MX, Ndc80-3V5:KanMX, ura3::pGPD1-GAL4(848).ER::URA3*
UB3549	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, set2::His3MX, set3::His3MX, Ndc80-3V5:KanMX, ura3::pGPD1-GAL4(848).ER::URA3*
UB5154	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ura3::pGPD1-GAL4(848).ER::URA3, ndc80Δ::KanMX4, leu2::pGAL-NDC80-3V5::LEU2* pGAL integrated 536 bp upstream of Ndc80 AUG
UB6077	*MAT**a**, ho::LYS2, lys2, ura3, leu2::(-295::ADH1)-NDC80-3V5:LEU2, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, ndc80(-1000 and ORF):KanMX4* *MATα, ho::LYS2, lys2, ura3, leu2::(-295::ADH1)-NDC80-3V5:LEU2, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, ndc80(-1000 and ORF)::KanMX4*

UB6190	*MAT**a**, ho::LYS2, lys2, ura3, leu2::NDC80-3V5:LEU2, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, ndc80(-1000 and ORF)::KanMX4* *MATα, ho::LYS2, lys2, ura3, leu2::NDC80-3V5:LEU2, his3::hisG, trp1::hisG, irt1::pCUP-3HA-IME1:: HphMX, ime4::pCUP-3HA-IME4:: NatMX, ndc80(-1000 and ORF)::KanMX4*
UB8110	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ura3::pGPD1-GAL4(848).ER::URA3, Ndc80-3V5:KanMX, set2::HygB, set3::cNAT*
UB8114	*MAT**a**, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ura3::pGPD1-GAL4(848).ER::URA3, HISMX:pGAL-Ndc80-3V5:KanMX, set2::HygB, set3::cNAT* pGAL integrated 536 bp upstream of Ndc80 AUG
UB8358	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, phi+, KanMX:p1X-LexO-pCyc1-Ndc80luti, trp1::pGPD1-LexA-ER-HA-B112::TRP1* *W303* 1X-LexO-pCyc1 integrated 536 bp upstream of Ndc80 AUG, thus replacing the *NDC80^{luti}* promoter
UB8362	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, phi+, KanMX:p2X-LexO-pCyc1-Ndc80luti, trp1::pGPD1-LexA-ER-HA-B112::TRP1* *W303* 2X-LexO-pCyc1 integrated 536 bp upstream of Ndc80 AUG, thus replacing the *NDC80^{luti}* promoter
UB8366	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, phi+, KanMX:p3X-LexO-pCyc1-Ndc80luti, trp1::pGPD1-LexA-ER-HA-B112::TRP1* *W303* 3X-LexO-pCyc1 integrated 536 bp upstream of Ndc80 AUG, thus replacing the *NDC80^{luti}* promoter
UB8370	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, phi+, KanMX:p8X-LexO-pCyc1-Ndc80luti, trp1::pGPD1-LexA-ER-HA-B112::TRP1* *W303* 8X-LexO-pCyc1 integrated 536 bp upstream of Ndc80 AUG, thus replacing the *NDC80^{luti}* promoter
UB8374	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, phi+, trp1::pGPD1-LexA-ER-HA-B112::TRP1* *W303*
UB8686	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, psi+, KanMX:p3X-LexO-pCyc1-Ndc80luti, trp1::pGPD1-LexA-ER-HA-B112::TRP1, set2::HygB, set3::CNAT* *W303* 3X-LexO-pCyc1 integrated 536 bp upstream of Ndc80 AUG, thus replacing the *NDC80^{luti}* promoter
UB8691	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, psi+, trp1::pGPD1-LexA-ER-HA-B112::TRP1, set2::HygB, set3::CNAT* *W303*
UB8693	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, phi+, KanMX:p8X-LexO-pCyc1-Ndc80luti, trp1::pGPD1-LexA-ER-HA-B112::TRP1, set2::HygB, set3::CNAT* *W303* 8X-LexO-pCyc1 integrated 536 bp upstream of Ndc80 AUG, thus replacing the *NDC80^{luti}* promoter

UB9181	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ura3::pGPD1-GAL4(848).ER::URA3, ndc80Δ:KanMX4, leu2::pGAL-NDC80-3V5:LEU2, pGAL-NDT80::TRP1* pGAL integrated 536 bp upstream of Ndc80 AUG
UB9921	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, GAL-NDT80::TRP1, ura3::pGPD1-GAL4(848).ER::URA3, leu2::pGAL-mse-NDC80-3V5:LEU2, ndc80Δ:KanMX4* pGAL integrated 536 bp upstream of Ndc80 AUG
UB9923	*MATα, ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG, ura3::pGPD1-GAL4(848).ER::URA3, leu2::pGAL-mse-NDC80-3V5:LEU2, ndc80Δ:KanMX4* pGAL integrated 536 bp upstream of Ndc80 AUG
UB12945	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, phi+, NDC80-3V5:HisMX, trp1::pGPD1-LexA-ER-HA-B112::TRP1* *W303*
UB12947	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, phi+, NDC80-3V5:HisMX, trp1::pGPD1-LexA-ER-HA-B112::TRP1, set2::HygB, set3::CNAT* *W303*
UB12949	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, phi+, KanMX:p8X-LexO-pCyc1-Ndc80luti-NDC80-3V5:HisMX, trp1::pGPD1-LexA-ER-HA-B112::TRP1* *W303* 8X-LexO-pCyc1 integrated 536 bp upstream of Ndc80 AUG, thus replacing the *NDC80^{luti}* promoter
UB12951	*MAT**a**, ADE2, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, phi+, KanMX:p8X-LexO-pCyc1-Ndc80luti-NDC80-3V5:HisMX, trp1::pGPD1-LexA-ER-HA-B112::TRP1, set2::HygB, set3::CNAT* *W303* 8X-LexO-pCyc1 integrated 536 bp upstream of Ndc80 AUG, thus replacing the *NDC80^{luti}* promoter

Table 2.1 Genotypes of yeast strains used in this study

The strains FW1509, FW1510 and FW1511 were derived from the sporulation proficient SK1 strain background. All other FW strains were derived from genetic manipulation of FW1509 and/or FW1510. The UB strains were also from SK1 background, except for the strains that harboured the LexA/*lexO* system for which the W303 strain background was used.

2.3 Yeast culture conditions

2.3.1 Growth and Conditions for synchronous sporulation (YPD to SPO)

Cells were grown in YPD (1.0% (wt/vol) yeast extract, 2.0% (wt/vol) peptone, 2.0% (wt/vol) glucose, and supplemented with tryptophan (96 mg/l), uracil (24 mg/l) and adenine (12 mg/l) and grown to exponential phase (OD_{600} < 2.0) at 30°C and 300 rpm. While developing the protocol, I found that supplemented tryptophan can be left out from the YPD. For optimal aeration, the ratio of the total volume of the flask to the volume of media was at least 10:1. Approximately 0.05 OD of exponentially growing yeast were inoculated into new flasks containing reduced glucose YPD (1.0% (wt/vol) yeast extract, 2.0% (wt/vol) peptone, 1.0% (wt/vol) glucose, and supplemented with uracil (24 mg/l) and adenine (12 mg/l). Cultures reached OD_{600} ≥ 10.0 after 16-18 hours, and the majority of the cells (~90%) were single, unbudded cells as observed under a light microscope. The cells were then pelleted by centrifugation (2000*g*, 3 min, room temperature). The pellets were washed with sterile miliQ water, centrifuged again (2000*g*, 3 min, room temperature) and suspended in supplemented sporulation media (1.0% (wt/vol) potassium acetate, supplemented with adenine/uracil (40 mg/l each), histidine/leucine/tryptophan (20 mg/l each) and 0.02% (wt/vol) raffinose) at OD_{600} of 2.5. After two hours, $CuSO_4$ (50 μM) was added to induce expression from the *CUP1* promoter and initiate sporulation synchronously. For sporulation experiments involving *GAL4-ER*/*pGAL-NDT80* strains, 1 μM β-estradiol was added 6 hr after shifting to SPO to induce *NDT80* expression.

For the acetate supplementation experiments (Fig 2.5). Cells were first grown in YPD (1.0% (wt/vol) yeast extract, 2.0% (wt/vol) peptone, 2.0% (wt/vol) glucose, and supplemented with tryptophan (96 mg/l), uracil (24 mg/l) and adenine (12 mg/l) and grown to exponential phase (OD_{600} < 2.0) at 30°C and 300 rpm. Approximately 0.05 OD of exponentially growing yeast were inoculated into new flasks containing reduced glucose YPD with varying amounts of acetate (1.0% (wt/vol) yeast extract, 2.0% (wt/vol) peptone, 1.0% (wt/vol) glucose, 0% - 1.0% (wt/vol) potassium acetate and supplemented with uracil (24 mg/l) and adenine (12

mg/l)). After 16-18 hours, the cells were then pelleted by centrifugation (2000*g*, 3 min, room temperature). The pellets were washed with sterile miliQ water, centrifuged again (2000*g*, 3 min, room temperature) and suspended in regular sporulation media (0.3 % (wt/vol) potassium acetate and 0.02% (wt/vol) raffinose) at OD_{600} of 2.5. After two hours, $CuSO_4$ (50 µM) was added to induce expression from the *CUP1* promoter and initiate sporulation synchronously.

For the glutamate supplementation experiment (Fig 2.5). Cells were first grown in YPD (1.0% (wt/vol) yeast extract, 2.0% (wt/vol) peptone, 2.0% (wt/vol) glucose, and supplemented with tryptophan (96 mg/l), uracil (24 mg/l) and adenine (12 mg/l) and grown to exponential phase (OD_{600} < 2.0) at 30°C and 300 rpm. Approximately 0.05 OD of exponentially growing yeast were inoculated into new flasks containing reduced glucose YPD with glutamate (1.0% (wt/vol) yeast extract, 2.0% (wt/vol) peptone, 1.0% (wt/vol) glucose, 0.5% (vol/vol) L-glutamic acid and supplemented with uracil (24 mg/l) and adenine (12 mg/l). After 16-18 hours, the cells were then pelleted by centrifugation (2000*g*, 3 min, room temperature). The pellets were washed with sterile miliQ water, centrifuged again (2000*g*, 3 min, room temperature) and suspended in supplemented sporulation media (1.0% (wt/vol) potassium acetate, supplemented with adenine/uracil (40 mg/l each), histidine/leucine/tryptophan (20 mg/l each) and 0.02% (wt/vol) raffinose) at OD_{600} of 2.5. After two hours, $CuSO_4$ (50 µM) was added to induce expression from the *CUP1* promoter and initiate sporulation synchronously.

For strains expressing the *SPT16-AID* depletion allele and the *pCUP-TIR1* allele, cells were grown to saturation in YPD as described above. Two hours after shifting to SPO, $CuSO_4$ (50 µM) was added to induce *IME1* and *TIR1* expression from the *CUP1* promoter while 500 µM indole-3-acetic acid (IAA) was added to induce *SPT16-AID* depletion.

2.3.2 Growth and Conditions for synchronous sporulation (BYTA to SPO)

Cells were grown in YPD (1.0% (wt/vol) yeast extract, 2.0% (wt/vol) peptone, 2.0% (wt/vol) glucose, and supplemented with tryptophan (96 mg/l), uracil (24 mg/l) and adenine (12 mg/l) for 24 hr at 30°C and 300 rpm. Approximately 0.4 OD of saturated yeast cultures were inoculated into new flasks containing presporulation medium BYTA (1.0% (wt/vol) yeast extract, 2.0% (wt/vol) tryptone, 1.0% (wt/vol) potassium acetate, and 50 mM potassium phthalate). Cultures were grown 16-18 hours. The cells were then pelleted by centrifugation (2000*g*, 3 min, room temperature). The pellets were washed with sterile miliQ water, centrifuged again (2000*g*, 3 min, room temperature) and suspended in regular sporulation media (0.3% (wt/vol) potassium acetate and 0.02% (wt/vol) raffinose) at OD_{600} of 2.5. After two hours, $CuSO_4$ (50 µM) was added to induce expression from the *CUP1* promoter and initiate sporulation synchronously.

2.4 Assessment of sporulation synchrony

2.4.1 Budding index determination

Cells were grown in regular YPD (1% yeast extract, 2% peptone, 2% glucose, and supplemented with tryptophan (96 mg/l), uracil (24 mg/l) and adenine (12 mg/l) and grown to exponential phase (OD_{600} < 2) at 30°C and 300 rpm. Cells were transferred to new flasks (OD_{600} of 0.05) containing reduced glucose YPD (1% yeast extract, 2% peptone, 1% glucose, and supplemented with uracil (24 mg/l) and adenine (12 mg/l)) or regular YPD with 2% glucose. After 16-18 hours, budded and unbudded cells were counted under a light microscope.

2.4.2 Flow cytometry analysis

Pre-meiotic DNA replication was monitored by flow cytometry analysis (BD LSR Fortessa, BD Biosciences). Cells were pelleted by centrifugation (~2400*g*, 1 min, room temperature) and fixed in 80% (vol/vol) ethanol for at least 60 minutes before further processing. Fixed cells were pelleted by centrifugation (~2400*g*, 1 min) and re-suspended in 50 mM Tris-HCl pH 7.5. Cells were sonicated for a few seconds

before treatment with 0.2 mg/ml ribonuclease A in 50 mM Tris-HCl pH 7.5 at 37°C overnight. After ribonuclease A digestion, cells were stained with 50 µg/ml propidium iodide in FACS buffer (200 mM Tris-HCl pH 7.5, 211 mM NaCl and 78 mM $MgCl_2$) for one hour at room temperature before flow cytometry analysis. Propidium iodide stained cells were excited with a 561 nm yellow-green laser and signals were detected using a 610/20 yellow filter. Pulse shape analysis (pulse height against pulse area) was used to exclude clumps and doublets. DNA content from single cells was estimated with a histogram of counts against pulse area. At least 50,000 cells were used for the analysis of each sample.

2.4.3 Nuclei/DAPI counting

To monitor meiotic divisions by DAPI staining, cells were pelleted by centrifugation (~2400*g*, 1 min, room temperature) and fixed in 80% (vol/vol) ethanol for at least 60 minutes before further processing. Subsequently, samples were pelleted by centrifugation (~2400*g*, 1 min) and re-suspended in PBS with DAPI (1 µg/ml). Cells were sonicated for a few seconds and left in the dark at room temperature for at least 5 minutes. After DAPI staining, the proportion of cells containing one, two, three, or four DAPI masses were counted using a fluorescence microscope.

2.4.4 Calculating the synchrony of meiotic divisions

The synchrony of meiotic divisions for each time course experiment was approximated by fitting a linear trend line from the first time point when meiotic divisions were detected to the first time point when 75% or more of the cells completed meiotic divisions. From these analyses, I calculated the period, or time taken for 75% of the cells to complete meiotic divisions. A more synchronously dividing population would take a shorter time to complete meiotic divisions. The average values from three independent experiments and the standard error of the mean are included in the figures. To determine the statistical significance, I used one way ANOVA, post hoc multiple comparison testing and two tailed *t*-tests where appropriate (Prism 6, Graphpad). A *p*-value of 0.05 or less was considered statistically significant.

2.5 Western blotting

Levels of hemagglutinin (HA) epitope tagged Ime1 and Ime4 were determined by western blotting using the procedures as described previously (Berchowitz et al. 2013). In brief, cells were pelleted by centrifugation (~2400*g*, 1 min, room temperature) and re-suspended in cold 5.0% w/v trichloroacetic acid (TCA) for at least 10 minutes. The pellets were then washed with acetone, mixed with lysis buffer (50mM Tris pH 7.5, 1mM EDTA, 2.75mM dithiothreitol (DTT)) and cells were broken using a mini beadbeater (BioSpec). Lysates were mixed with SDS loading buffer (187.5 mM Tris pH 6.8, 6% v/v Beta-Mercaptoethanol, 30% v/v Glycerol, 9% w/v SDS, 0.05% w/v Bromophenol Blue) and boiled for 5 minutes for denaturation. Proteins were separated by PAGE and transferred onto PVDF membranes using the Mini Trans-Blot Cell (Bio-rad). The membranes were blocked for 60 minutes in blocking buffer (1.0% w/v BSA, 1.0% w/v milk) before incubation with mouse anti-HA (12CA5, Sigma-Aldrich) at a 1:1000 dilution overnight at 4°C. Membranes were washed in PBST (phosphate buffered saline with 0.01 % tween-20) and incubated with anti-mouse HRP secondary antibodies at a 1:5000 dilution (GE Healthcare). After addition of ECL substrate (GE Healthcare), membranes were imaged using Imagequant 600 RGB (GE Healthcare).

For Spt16 western blots, the membranes were blocked for 60 minutes in blocking buffer (1% w/v BSA, 1% w/v milk) before incubation with mouse anti-V5 (R96025, Sigma-Aldrich) at a 1:2000 dilution overnight at 4°C. Membranes were then washed in PBST (phosphate buffered saline with 0.01% (v/v) tween-20) and incubated with and incubated with anti-mouse HRP secondary antibodies at a 1:5000 dilution (GE Healthcare). After addition of ECL substrate (GE Healthcare), membranes were imaged using Imagequant 600 RGB (GE Healthcare).

For Ndc80 western blots, the membranes were blocked for 60 minutes in blocking buffer (1% w/v BSA, 1% w/v milk) before incubation with mouse anti-V5 (R96025, Sigma-Aldrich) at a 1:2000 dilution overnight at 4°C. Membranes were then washed in PBST (phosphate buffered saline with 0.01% (v/v) tween-20) and

incubated with IRDye 800CW goat anti-mouse secondary antibodies (LI-COR) at a 1:15000 dilution. Protein levels were detected using an Odyssey Imager (LI-COR).

For the Hxk1 loading controls, anti-hexokinase antibody (H2035, Stratech) was used at a 1:8000 dilution overnight at 4°C. The IRDye 680RD donkey anti-rabbit secondary antibody (LI-COR) was used at a 1:15000 dilution. Hxk1 levels were detected using an Odyssey Imager (LI-COR).

Where applicable, intensities of Ndc80 and Hxk1 bands on western blots were quantified using Image Studio Lite (LI-COR). Ndc80 levels were first normalized to Hxk1 levels and further normalized to that of the first time point on the same membrane.

To measure bulk histone H3 levels, membranes were blocked for 60 minutes in blocking buffer (5% w/v milk) before incubation with rabbit anti-H3, C terminus (07-690, Millipore) at a 1:3000 dilution overnight at 4°C. To measure bulk H3K36me3 levels, membranes were blocked for 60 minutes in blocking buffer (1% w/v BSA, 1% w/v milk) before incubation with rabbit anti-H3K36me3 (Ab9050, Abcam) at a 1:1000 dilution overnight at 4°C. Membranes were then washed in PBST and incubated with HRP conjugated ECL donkey anti-rabbit secondary antibodies (GE Healthcare) at a 1:8000 dilution. After addition of ECL substrate (GE Healthcare), protein levels were detected using an ImageQuant RGB 600 machine (GE Healthcare). At least two independent biological experiments were performed for each western blot experiment.

Uncropped western blots can be found in chapter 8, supplemental figures section **(Figure 8.3)**. The size markers used for all western blots was from the Precision Plus Protein™ Standards (Bio-Rad)

2.6 Northern blotting

The oligonucleotide sequences used for amplifying the *NDC80*, *CIT1* or *SCR1* templates are displayed in **Table 2.2**.

We adapted a northern blot protocol that was described previously (Koster et al. 2014). In short, RNA was extracted with TES buffer (10 mM Tris (pH7.5), 10mM EDTA and 0.5% v/v SDS), Acid Phenol:chloroform: Isoamyl alcohol (125:24:1) and precipitated in ethanol with 0.3 M sodium acetate. RNA samples were denatured in a glyoxal/DMSO mix (1 M deionized glyoxal, 50% v/v DMSO, 10 mM sodium phosphate (NaP_i) buffer pH 6.5-6.8) at 70°C for 10 minutes. Denatured samples were mixed with loading buffer (10% v/v glycerol, 2 mM NaPi buffer pH 6.5-6.8, 0.4% w/v bromophenol blue) and separated on an agarose gel (1.1% w/v agarose, 0.01M NaPi buffer) for at least 3 hours at 80 V. RNAs were then transferred onto nylon membranes overnight by capillary transfer. rRNA bands were visualized by methylene blue staining. The membranes were blocked for at least 3 hours at 42°C in ULTRAhyb® Ultrasensitive Hybridization Buffer (Ambion) before hybridization. All gene-specific radioactive probes were synthesized using a Prime-It II Random Primer Labeling Kit (Agilent), a *NDC80* DNA template and dATP [α-32P] (Perkin-Elmer). At least two independent biological experiments were performed for each northern blot experiment.

$NDC80^{ORF}$, $NDC80^{luti}$, CIT1 and *SCR1* levels were estimated from northern blots using ImageJ (Schneider et al. 2012). The net intensity of each band of interest was determined by subtracting the mean background intensity of the areas immediately above and below the band. Signals were first normalized to *SCR1* levels and further normalized to a specific band on the same membrane (usually the first time point when either *$NDC80^{ORF}$* or *$NDC80^{luti}$* appeared). One-tailed, unpaired *t*-tests were conducted to determine if the difference in *$NDC80^{ORF}$* levels between mutant and control strains were statistically significant.

Uncropped northern blots can be found in chapter 8, supplemental figures section **(Figure 8.2)**.

2.7 RT-PCR (qPCR)

The sequences of qPCR primers are displayed in **Table 2.2.**

Total RNA was treated with DNAse and purified (Macherey-Nagel). 750 ng of total RNA was used for the reverse transcription reaction using Superscript III (Life Technoloqies), and single stranded cDNA were quantified by real-time PCR (qPCR) using SYBR green mix (Life Technologies). To measure *IME1* or *NDC80^{luti}* RNA levels, random primers were used for the reverse transcription reaction. Since *IME4* has anti-sense transcription, an *IME4* sense-strand specific primer (5'-ATTCTGCTTGGCCTCAGCAT-3') and an *ACT1* sense-strand specific primer (5'-TTAGAAACACTTGTGGTGAA-3') was used during the reverse transcription reaction. The signals were normalized to *ACT1* transcript levels.

2.8 Spot growth assay

Spot assays presented in this thesis were done by collaborators at the University of California, Berkeley (Chia et al. 2017).

For strains harbouring *NDC80^{luti}* under control of the *GAL1-10* promoter, cells were first grown on YP plus 2% glycerol (YPG) plates overnight, and then re-suspended in milliQ H_2O to an OD_{600} of 0.2. Next, 5-fold serial dilutions were performed and diluted cells were spotted onto either YP-RG plates with no β-estradiol or YP-RG plates supplemented with1 μM β-estradiol. The cells were incubated at 30°C for 1–2 days. Note that the *GAL1-10* promoter in the SK1 strain background does not directly respond to galactose. At least two independent biological experiments were performed for each spot assay experiment.

For strains harbouring constructs in which *NDC80^{luti}* expression is driven by LexA/*lexO*, cells were grown on YPD plates, re-suspended in milliQ water to an OD_{600} of 0.2, serially diluted as above, and then spotted onto either YPD plates with no β-estradiol or YPD plates with different concentrations of β-estradiol (10, 15, 20, 25, or 30 nM). The cells were incubated at 30°C for 1 day before imaging. At least two independent biological experiments were performed for each spot assay experiment.

2.9 Chromatin immunoprecipiation

The oligonucleotide sequences used for ChIP-qPCR experiments are displayed in **Table 2.2**.

2.9.1 ChIP for V5 tagged proteins

Chromatin immunoprecipitation (ChIP) experiments were performed as described previously (van Werven et al. 2012). Cells were fixed in 1.0% w/v of formaldehyde for 15-20 min at room temperature and quenched with 100 mM glycine. After breaking cells using a mini beadbeater (BioSpec), crosslinked chromatin was sheared by sonication using Bioruptor (Diagenode, 6 cycles of 30 sec on/off). Extracts were incubated with anti V5 agarose beads (Sigma) for 2 hours, and beads were washed. Binding of proteins tagged with the V5 epitope was measured by qPCR using SYBR green mix (Life Technologies) and primers corresponding to the *IME4* promoter on a 7500 Fast Real-Time PCR system (Life Technologies). The hidden mating type locus *HMR* was used as a non-binding control (van Werven et al. 2012).

2.9.2 ChIP for histone marks

For the H3K4me2 and the H3K36me3 ChIP experiments described in chapter 4, cells were fixed in 1.0% w/v of formaldehyde for 15-20 min at room temperature and quenched with 100 mM glycine. Cells were broken using a mini beadbeater (BioSpec) and crosslinked chromatin was sheared by sonication using a Bioruptor (Diagenode, 7 cycles of 30 sec on/off). Extracts were incubated for 2 hours or overnight at 4 ˚C with magnetic Prot A beads (Sigma) coupled with a polyclonal antibody against Histone H3 tri methyl lysine 36 (Ab9050, Abcam), Histone H3 di methyl lysine 4 (Ab32356, Abcam) or Histone H3 (Ab1791, Abcam). Subsequently, reverse cross-linking was done in Tris-EDTA buffer (100 mM Tris pH 8.0, 10 mM EDTA, 1.0% v/v SDS) at 65°C overnight. After 2 hours of proteinase K treatment, samples were cleaned up and histone mark enrichment was measured by real-time PCR using SYBR green mix (Life Technologies) and primers corresponding to the *NDC80* promoter and the 5’ region of the *NDC80* open reading frame. Signals were either normalised to histone H3 levels or to the *HMR* locus.

2.9.3 ChIP on Micrococcal nuclease (MNase) treated chromatin extracts

To determine the chromatin structure at the *NDC80* locus, we extracted mononucleosomes using a MNase digestion protocol that was described previously followed by ChIP for histone H3 (Rando 2010;Rando 2011). Approximately 250 OD_{600} units of cells were crosslinked for 15 min with formaldehyde (1% v/v) and the reaction was quenched with glycine (125 mM). Subsequently, cells were resuspended in 20 ml of buffer Z (1 M sorbitol, 50 mM Tris-HCl pH 7.4**)** plus β-mercaptoethanol (10mM) and treated with 250 µg of T100 Zymolase (MP Biomedicals) for 60 min. Next, cells were resuspended in 2.5 ml NP buffer (0.5 mM spermidine, 1 mM β-mercaptoethanol (β -ME), 0.075% (w/v) Tergitol solution-type NP-40 detergent (NP-40), 50 mM NaCl, 10 mM Tris-HCl pH 7.4, 5 mM $MgCl_2$, 1 mM $CaCl_2$), and extract was treated with 5, 0.625, 0.2 or 0.04 µl of MNase (2 mg/ml, NEB) for 30 min at 37 °C, the reaction was quenched with EDTA (10 mM). The extract was adjusted to 0.1 M Hepes-KOH pH 7.5, 150mM NaCl, 0.1% w/v sodium deoxycholate, and 1% w/v Triton X-100. To check for the extent of MNase digestion, 60 µl of MNase treated and untreated extracts were reverse crosslinked overnight in SDS-TE (1% (w/v) SDS, 10 mM Tris pH 8, 1 mM EDTA), treated with RNase A, purified DNA fragments were separated by gel electrophoresis. The extracts of which the showed a mono-nucleosome pattern were used for ChIP with histone H3 antibodies. The ChIP was performed with 600 µl of extract as described in the chromatin immunoprecipitation section of the materials and methods. ChIP samples were quantified by qPCR on a 7500 FAST Real-Time PCR machine (Applied Biosystems). Scanning primer pairs covering the *NDC80* locus and upstream region were used for the analysis. Signals were quantified relative to untreated genomic DNA, and normalized over a primer pair directed against the *PHO5* core promoter (Chang and Vancura 2012).

2.10 mRNA sequencing

At least 2 µg of total RNA was treated with DNAse and purified on column (Macherey-Nagel). At least 400 ng of purified total RNA was used as input for the

KAPA mRNA Hyper Prep kit (KK8580, Roche). Libraries were prepared according to manufacturer's instructions. After bead based clean up, libraries were sequenced on an Illumina HiSeq 2500 to an equivalent of 75 bases single-end reads, at a depth of approximately 20 million reads per library.

2.11 Transcript end sequencing (TE-seq)

All custom oligonucleotide sequences used in this protocol are listed in **Table 2.3**.

2.11.1 Preparation of *in vitro* transcripts (IVTs)

A pool of *in vitro* transcripts (IVT) to use as spike-in controls for TE-seq libraries were prepared as previously described (Pelechano et al. 2014).

In brief, three different plasmids were used as templates for the *in vitro* transcription reaction. They were pGIBS-LYS (ATCC no. 87482), pGIBS-PHE (ATCC no. 87483) and pGIBS-THR (ATCC no. 87484). Each plasmid was linearized with NotI digestion for 1 hour at 37°C. Linearized templates were cleaned up with DNA columns (Macherey-Nagel). 200 ng of each template was used for the *in vitro* transcription using 20 units of T3 RNA polymerase (P2083, Promega) in a reaction buffer (1X transcription optimized buffer (P1181, Promega), 10 mM DTT, 0.5 mM rNTP and 0.5 µl RNasin Plus (N2611, Promega)), at 37°C for 2 hours. Subsequently, the DNA template was degraded using the TURBO DNA-free kit (AM1907, Ambion) at 37°C for 30 minutes. IVTs were then extracted with Acid Phenol:chloroform:Isoamyl alcohol (125:24:1) and precipitated in ethanol with 0.3 M sodium acetate. Quantification of IVTs was done using the Qubit RNA high sensitivity assay kit (Q32852, ThermoFisher Scientific) and IVT sizes were checked using a Bioanalyzer (Agilent Technologies). The expected lengths of the pGIBS-LYS, pGIBS-PHE and pGIBS-THR IVTs were 1106, 1407 and 2070 nucleotides respectively.

Purified IVTs were pooled at an approximate molecular ratio of 25 Lys:5 Phe:1 Thr. 1 µg of pooled IVTs were denatured at 65°C for 5 minutes, cooled on ice and then

subjected to a capping reaction (10 U Vaccinia capping enzyme (M2080, NEB), 1X capping buffer (M2080, NEB), 0.5 mM GTP, 0.1 mM SAM (B9003S, NEB) and 0.5 µl RNasin Plus). Pooled IVTs were then extracted with Acid Phenol:chloroform: Isoamyl alcohol (125:24:1) and precipitated in ethanol with 0.3 M sodium acetate.

2.11.2 5' end sequencing

This 5' end sequencing approach was adapted and modified from previously published protocols (Arribere and Gilbert 2013;Malabat et al. 2015;Adjalley et al. 2016).

50 ng of pooled IVTs were spiked into 1 mg of total RNA from each sample. At least 5 µg of mRNAs were purified from total RNA using the Poly(A)Purist MAG kit (AM1922, Ambion). mRNAs were fragmented for 3 minutes at 70°C using a Zinc-based alkaline fragmentation reagent (AM8740, Ambion). RNAs were cleaned up using RNeasy MinElute Cleanup Kits (74204, Qiagen) to enrich for 200-300 nt fragments. These fragments were dephosphorylated with 30 units of recombinant shrimp alkaline phosphatase (M0371, NEB) for 1 hour at 37°C with RNasin Plus, the phosphatase was heat inactivated and the RNA was extracted with Acid Phenol:chloroform: Isoamyl alcohol (125:24:1) and precipitated at -20°C overnight in ethanol with 0.3M sodium acetate and 1 µl linear acrylamide (AM9520, Ambion). RNA was then subjected to a decapping reaction with 2 units of Cap-Clip acid pyrophosphatase (C-CC15011H, Tebu-Bio) and with RNasin Plus. RNA was then extracted using acid Phenol:chloroform:isoamyl alcohol (125:24:1) and precipitated in ethanol. Some RNA from a 0 hour time point was set apart without the decapping reaction as a non-decapping control. Subsequently, the RNA was mixed with 10 µM of custom 5' adapter **(Table 2.3)** and the ligation reaction was done using T4 RNA ligase 1 (M0437M, NEB) and with RNasin Plus. The ligation reaction was cleaned up with the RNeasy MinElute Cleanup Kit and RNAs were mixed with 2.5 µM random hexamers (N8080127, ThermoFisher Scientific) and RNasin Plus, denatured at 65°C for 5 minutes and cooled on ice. Reverse transcription reactions were carried out using SuperScript IV reverse transcriptase (18090010, Invitrogen) at 23°C for 10 minutes, 50°C for 10 minutes, 80°C for 10 minutes and held at 4°C. The RNA templates were degraded by incubating reactions with 5 units of RNase H

(M0297, NEB) and 1.0 µl of RNase cocktail enzyme mix (AM2286, Ambion). DNA products were purified using 1.8x volume of HighPrep PCR beads (AC-60050, MagBio). Purified products were subjected to second strand synthesis using 0.3 µM of second strand biotinylated primer **(Table 2.3)** and the KAPA Hi-Fi hot start ready mix (KK2601, Roche). The second strand reaction was carried out at 95°C for 3 minutes, 98°C for 15 seconds, 50°C for 2 minutes, 65°C for 15 minutes and held at 4°C. Double stranded product (dsDNA) was purified with 1.8x volume HighPrep PCR beads and concentration was quantified using the Qubit dsDNA HS assay kit (Q32851, Invitrogen). At least 1 ng of dsDNA was then used as input for the KAPA Hyper Prep Kit (KK8504, Roche) and ligated to KAPA single indexed adapters Set A (KK8701, Roche) or Set B (KK8702, Roche). Samples were processed according to manufacturer's instructions with one exception: just prior to the library amplification step, samples were bound to MyOne Streptavidin C1 Dynabeads (65001, ThermoFisher Scientific) to capture biotinylated dsDNA. Library amplification was done on the biotinylated dsDNA fraction bound to the beads. Depending on the input amounts, 15-17 PCR cycles were used to generate libraries. Amplified libraries were quantified by Qubit, and adapter-dimers were removed by electrophoresing libraries on Novex 6% TBE gels (EC62655BOX, Invitrogen) at 120 V for 1 hour, and excising the smear above ~150 bp. Gel slices containing libraries were shredded by centrifugation at 13000 *g* for 3 minutes. Gel shreds were re-suspended in 500 µl crush and soak buffer (500 mM NaCl, 1.0 mM EDTA and 0.05% v/v SDS) and incubated at 65°C for 2 hours on a thermomixer (1400 rpm for 15 seconds, rest for 45 seconds). Subsequently, the buffer was transferred into a Costar SpinX column (8161, Corning Incorporrated) with two 1 cm glass pre-filters (1823010, Whatman). Columns were centrifuged at 13000 *g* for 1 minute. DNA libraries in the flowthrough were precipitated at -20°C overnight in ethanol with 0.3 M sodium acetate and 1 µl linear acrylamide (AM9520, Ambion). Purified libraries were further quantified and inspected on a Tapestation (Agilent Technologies) and sequenced on an Illumina HiSeq 2500 to an equivalent of 75 bases single-end reads, at a depth of approximately 20 million reads per library.

2.11.3 3' end sequencing

This 3' end sequencing approach was adapted and modified from previously published protocols ((Ng et al. 2005;Lai et al. 2015).

From the same pool of fragmented mRNAs for each sample, at least 1 µg was used for 3' end sequencing. RNA fragments were mixed with 2.5 µM GsuI20TVN primer **(Table 2.3)**, 0.5 mM 5-Methylcytosine-dNTPs (D1030, Zymo Research) and 0.5 µl RNasin Plus. Reaction mixtures were denatured at 65°C for 5 minutes and held at 50°C without allowing to cool. SuperScript IV, reaction buffer and 0.4 µg of Actinomycin D were added to the hot reaction mixtures and reverse transcription was performed at 50°C for 10 minutes, 80°C for 10 minutes and held at 4°C. Samples were cleaned with 1.8x volume HighPrep beads and biotinylated RNA:DNA hybrids were captured on MyOne Streptavidin C1 Dynabeads. After capture, streptavidin beads were washed once with 1x NEbuffer 2 (B7002S, NEB), re-suspended in water and subjected to second strand synthesis. The 50 µl second strand synthesis reaction consisted of 20 µl re-suspended streptavidin beads, 1X NEbuffer 2, 250 µM dNTPs, 26 µM NAD+ (B9007S), 2.5 units RNase H, 10 units *E.coli* DNA ligase (M0205S), and 15 units DNA polymerase I (M0209S). Second strand synthesis reactions were conducted at 16°C for 2.5 hours on a thermomixer (1400 rpm for 15 seconds, rest for 2 minutes). After reaction, beads were washed once with 1x binding and washing buffer (5.0 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1.0 M NaCl) and once with buffer B (10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 0.1 mg/ml BSA). Washed beads were re-suspended in 18 µl buffer B and digested with 10 units of GsuI (ER0461, ThermoFisher Scientific) at 30°C for 1 hour on a thermomixer (1400 rpm for 15 seconds, rest for 2 minutes). After digestion, the DNA fragments in the supernatant were extracted with Phenol/chloroform and precipitated at -20°C overnight in ethanol with 0.3 M sodium acetate and 1 µl linear acrylamide. The concentration was quantified using the Qubit dsDNA HS assay kit. At least 1 ng of dsDNA was then used as input for the KAPA Hyper Prep Kit (KK8504, Roche) and ligated to KAPA single indexed adapters Set A (KK8701, Roche) or Set B (KK8702, Roche). Samples were processed according to manufacturer's instructions. Amplified libraries were cleaned and purified by gel extraction using the procedures described in the previous section for 5' end

sequencing. Purified libraries were further quantified and inspected on a Tapestation (Agilent Technologies) and sequenced on an Illumina HiSeq 2500 to an equivalent of 75 bases single-end reads, at a depth of approximately 20 million reads per library.

2.12 Bioinformatics analyses

2.12.1 Adapter trimming and read alignment

For the RNA-seq data, adapter trimming was performed with cutadapt (version 1.9.1) (Martin 2011) with parameters "-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC --minimum-length=20". STAR (version 2.5.2) (Dobin et al. 2013) with parameters "--alignIntronMin 3 --alignIntronMax 5000" was used to perform the read mapping to the S. cerevisiae SK1 genome assembly from Keeney lab. Alignments with mapping quality of <10 or soft/hard-clipping were filtered. The tool "bedtools genomecov" (Quinlan and Hall 2010) was used to generate the RNA-seq coverage tracks across the genome. For the 5' end and 3' end sequencing data, adapter trimming of raw reads was performed with cutadapt (version 1.9.1) with parameters "-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC --minimum-length=20". In addition, for the 5' end data, the custom 5' adapter sequence specific to the protocol was removed by re-running cutadapt with the parameters "-g CACTCTGAGCAATACC -O 16 --minimum-length=20", and only the reads containing the adapter sequence were used for further analysis. STAR (version 2.5.2) with parameters "--alignIntronMin 2 --alignIntronMax 1" (i.e. not allowing introns) was used to align 5' end and 3' end sequencing reads to the SK1 genome assembly (plus three spike-in sequences). The alignments with mapping quality of >=10 were kept for further analysis. For 5' end alignments, the 5'-most nucleotide of reads were extracted and the genome-wide coverage tracks were generated. For 3' end alignments, we only kept the reads with soft-clipping at the 3' end (size of soft-clipping part <=10) and required at least two consecutive non-templated As in the soft-clipping part. Insertions/deletions were also not allowed for the 3' end alignments. The 3'-most nucleotide of aligned 3' end reads were extracted and genome-wide coverage tracks were generated.

2.12.2 Clustering of samples based on alignments

Using the genome-wide coverage tracks generated above, we did sample clustering for 5' end, 3' end and RNA-seq data respectively. First, for each sample we extracted the average coverage values for 100bp non-overlapping widows across the genome using the command "multiBigwigSummary" in deepTools(Ramirez et al. 2014), for 5' end, 3' end and RNA-seq data respectively. The resulting coverage matrixes (5'end, 3'end or RNA-seq) of 100bp windows were used to do hierarchical clustering of samples in R (version 3.4.1). The clustering was done by "hclust" with method="ward.D2" and distance function defined as (1-cor(log(data+1), method="pearson"))/2.

2.12.3 "Within sample" normalisation

RNA-seq read counts were normalised as fragments per kilobase of exon per million reads mapped (FPKM) (Mortazavi et al. 2008). This approach involves taking the read counts which map to an annotated gene, dividing this value by length of the transcript in kilobases and further dividing this per million reads sequenced. TSS or TES seq data are mapped as CAGEr "tags" (see section 2.12.6), and were normalised as tags per million reads mapped (TPM). Since the length of transcript isoforms or ncRNAs are unknown, the TPM formula simply takes the number of tags which map to a defined cluster and divides this per million reads sequenced.

Conversion of read counts to FPKM corrects for gene length and sequencing depth. Conversion of CAGEr tag counts to TPM corrects for sequencing depth. Both FPKM and TPM values are obtained from "within-sample" normalisation procedures and hence should not be used for direct comparison of gene expression between different time points or conditions.

2.12.4 "Between sample" normalisation and differential expression analysis

Differential expression analysis was performed with the DESeq2 package (version 1.18.1) (Love et al. 2014) in R (version 3.4.1). The raw read counts of called TSS/TES peaks, which are the input for running DESeq2, were extracted from aforementioned coverage tracks by bigWigAverageOverBed (Kent et al. 2010).

The DESeq2 package performs “between sample” normalisation for comparison of gene expression between different time points or conditions. In brief, DESeq2 creates a hypothetical “reference sample” by taking the geometric mean counts for all genes across all samples. Each sample is then normalised to this reference sample to control for library size for “between sample” comparisons. Next, DEseq2 estimates a dispersion parameter for each gene and models the variability of read counts within replicates of the same condition. This model is then used to identify differentially expressed genes, which will show a statistically significant difference in mean counts between a sporulation time point and a control reference time point (see next paragraph for comparisons). In this work, TSS/TES peaks with $\log_2$(fold change) >=1 and False Discovery Rate (FDR) < 0.05 in the DESeq2 results were considered as significantly differentially expressed peaks.

Each time point corresponding to early gametogenesis (Spo 3-6h) was compared to pre-meiotic cells immediately prior to *IME1* induction (Spo 2h). Each time point corresponding to mid-late gametogenesis (Spo 7-9h) was compared to cells in meiotic prophase, prior to *NDT80* induction (Spo 6h). The Spo 3M starvation controls were compared to pre-meiotic cells at Spo 2h. The Spo 7M starvation controls were compared to meiotic prophase cells at Spo 6h.

2.12.5 Determination of relative upstream promoter strength

To determine relative upstream promoter strength, normalized read counts from DESeq2 analysis were used. The number of reads associated with an upstream TSS was divided by the number of reads associated with the downstream coding TSS, measured from the same time point.

2.12.6 Peak calling with CAGEr and assignment to gene features

The 5’-most (for 5’ end data) and 3’-most (for 3’ end data) nucleotides were clustered into TSS or TES “peaks” using the CAGEr (Haberle et al. 2015). The key parameters were 1) clusterCTSS: method = "distclu", maxDist = 5, keepSingletonsAbove = 3; 2) aggregateTagClusters: tpmThreshold = 1, qLow = 0.05, qUp = 0.95, maxDist = 20. The assignment of the peaks to their nearby genes (genes within ±1kb of a peak) was done by “bedtools closest”(Quinlan and Hall

2010). Assignment of TSSs to the nearest TES was also done by “bedtools closest”. The output from “bedtools closest” was also used to determine the 5’ UTR length, i.e. the distance given in number of nucleotides from the apex of a TSS peak to the AUG of an annotated ORF.

2.12.7 Bi-directional promoter/divergent transcription analysis

If an up-regulated TSS peak had an up-regulated peak on the opposite strand within its upstream 500bp region using “bedtools intersect”(Quinlan and Hall 2010), we considered these two peaks as a pair of co-regulated peaks which might be generated by a bi-directional promoter. Similar criteria were applied to the down-regulated peaks.

2.13 Oligonucleotide sequences

Primer name	Oligonucleotide sequence (5’ to 3’)	Experiment
ACT1_FW	GTACCACCATGTTCCCAGGTATT	qPCR
ACT1_RV	AGATGGACCACTTTCGTCGT	qPCR
CIT1_probe_FW	CCGTGTTAGACCCCGAAGAAG	Northern blots
CIT1_probe_RV	GGGCAGAAACGTTACCACCTTC	Northern blots
HMR_FW	ACGATCCCCGTCCAAGTTATG	qPCR
HMR_RV	CTTCAAAGGAGTCTTAATTTCCCTG	qPCR
IME1_FW	CAACGCCTCCGATAATGTATATG	qPCR
IME1_RV	ACGTCGAAGGCAATTTCTAATG	qPCR
IME4_FW	CGTCTTTAGGCGGCTTTTGG	qPCR
IME4_RV	ACCGATCTTCCAGAATGCCG	qPCR
MCM5_probe_FW	GGGATCTTCTGCAGCTGGGTTAAC	Northern blots
MCM5_probe_RV	CTCTAGGGCATATAAGGCCTTATCCAG	Northern blots
NDC80_probe_FW	GGAGAGGTAGAATCGTCCCTG	Northern blots
NDC80_probe_RV	CTCCTCTTGAATAGCGCTTTGG	Northern blots
NDC80_1_FW	GCTCCTGTGTTCTCCATT	qPCR
NDC80_1_RV	GTGTGTTGATACTGCACTG	qPCR
NDC80_2_FW	ACCCGGATATCTGTTCAGCC	qPCR
NDC80_2_RV	TGTGGCGAATTGTTGCTCTT	qPCR
NDC80_3_FW	CGCCACAAGAAGGTCTC	qPCR

NDC80_3_RV	GCTTTTCGGACCTCCAAC	qPCR
NDC80_4_FW	GTTGGAGGTCCGAAAAGC	qPCR
NDC80_4_RV	GTTCAGTTATAACCATCTGGCAC	qPCR
NDC80_5_FW	GTGCCAGATGGTTATAACTGAAC	qPCR
NDC80_5_RV	CCGCTAATCGCAATAGACTG	qPCR
NDC80_6_FW	GGTTGAGAGCCCCGTTAAGT	qPCR
NDC80_6_RV	TTGGCACTTTCAGTATGGGT	qPCR
NDC80_7_FW	CCCATACTGAAAGTGCCAAAAGA	qPCR
NDC80_7_RV	GGGACGATTCTACCTCTCCTGTG	qPCR
NDC80_8_FW	GGAATACATTCACAGGAGAGG	qPCR
NDC80_8_RV	GGAATATATTATAGTACACCCTAACG	qPCR
NDC80_9_FW	TGCAAAGCTCAACAAGTACTGA	qPCR
NDC80_9_RV	TGCAGTTGGTATTTGGGACG	qPCR
NDC80_10_FW	CAAGGTCTAACCGACATGATC	qPCR
NDC80_10_RV	CATTTGTACCTCCTGCAAC	qPCR
ORC3_probe_FW	GATGTATCATACGATCTGTCACTTGTGG	Northern blots
ORC3_probe_RV	CCGGTGTTTCTGCTAGTTCTAAGAG	Northern blots
PHO5-TATA_FW	CCATTTGGGATAAGGGTAAACATC	qPCR
PHO5-TATA_RV	AGAGATGAAGCCATACTAACCTCG	qPCR
RAD16_probe_FW	TATCAGCCAGCCACTGACAG	Northern blots
RAD16_probe_RV	GTGATTTCACGAGCGTAGCA	Northern blots
SCR1_probe_FW	GAAGTGTCCCGGCTATAATAAA	Northern blots
SCR1_probe_RV	GACGCTGGATAAAACTCCCC	Northern blots
SWI4_probe_FW	GCGCCTCATTCAAACCATCC	Northern blots
SWI4_probe_RV	CCAGCCGAATCCAAGTCCTT	Northern blots

Table 2.2 List of oligonucleotide sequences used for qPCR and northern blots

Name	**Oligonucleotide sequence (5' to 3')**	**Use**
5' adapter	CACTCTrGrArGrCrArArUrArCrC	For ligation to the 5' ends of RNAs
Random hexamers	Random hexamers	For reverse transcription for 5' end sequencing
Second strand biotinylated primer	GCAC/iBiodT/GCACTCTGAGCAATACC	For second strand synthesis

GsuI20TVN	/5BiotinTEG/ GAGCTAGTTCTGGAGTTTTTTTTTTTTTTTTTTTTVN	For reverse transcription for 3' end sequencing

Table 2.3 List of custom oligonucleotide sequences used for transcript end sequencing

Chapter 3. Temporal expression of a master regulator drives synchronous sporulation in budding yeast

This research has been published in G3 (Bethesda) and has been modified for this results chapter (Chia and van Werven 2016). Chia M, van Werven FJ. Temporal Expression of a Master Regulator Drives Synchronous Sporulation in Budding Yeast. *G3: Genes|Genomes|Genetics*. 2016;6(11):3553-3560. doi:10.1534/g3.116.034983.

3.1 Abstract

Yeast cells enter and undergo gametogenesis relatively asynchronously, making it technically challenging to perform stage-specific genomic and biochemical analyses. Cell-to-cell variation in the expression of the master regulator of entry into sporulation, *IME1*, has been implicated to be the underlying cause of asynchronous sporulation. Here, I find that timing of *IME1* expression is of critical importance for inducing cells to undergo sporulation synchronously. When I force expression of *IME1* from an inducible promoter in cells incubated in sporulation medium for 2 hr, the vast majority of cells exhibit synchrony during meiotic S-phase replication and meiotic divisions. Inducing *IME1* expression too early or too late affects the synchrony of sporulation. Surprisingly, my approach for synchronous sporulation does not require growth in acetate-containing medium, but can be achieved in cells grown in rich medium until saturation. My system requires solely *IME1*, because the expression of the N6-methyladenosine methyltransferase *IME4*, another key regulator of early sporulation, is controlled by *IME1* itself. The approach described here can be combined easily with the *pGAL-NDT80* synchronization system, and thereby applied to study specific stages of sporulation, or the complete sporulation program.

3.2 Introduction

Gametogenesis, or sporulation is an important cell fate program for all sexually reproducing organisms. Sporulation in diploid budding yeast cells involves a single round of DNA replication followed by double strand break formation, homologous recombination, and two consecutive nuclear divisions called meiosis I and II to generate haploid progeny (Marston and Amon 2004). The products of meiotic divisions are subsequently packaged to form gametes or spores. These complex cellular changes during sporulation are driven by regulatory networks which activate, repress or tune gene expression in a dynamic manner. The high experimental tractability of this model organism and its well annotated genome make budding yeast gametogenesis an attractive model to investigate how genes are regulated during cell differentiation. Indeed, studies on this cell fate transition have yielded insights into the functional roles of some non-coding RNAs (Brar et al. 2012;van Werven et al. 2012). However, as mentioned in chapter 1 (introduction), much remains to be understood about the transcriptional program during sporulation.

Initiation of budding yeast gametogenesis is controlled by a single master regulatory transcription factor called inducer of meiosis 1 (*IME1*) (Kassir et al. 1988). Ime1 activates transcription of the “early genes” during sporulation (Honigberg and Purnapatre 2003). Signals conveying information about nutritional state and ploidy are integrated onto the *IME1* promoter such that only nutrient starved diploid cells enter gametogenesis (van Werven and Amon 2011). Another gene important for entry into gametogenesis is *IME4*, which encodes an enzyme that methylates the adenosine base of mRNAs to form N-6-Methyladenosine (m6A) (Shah and Clancy 1992;Clancy et al. 2002). Previous work indicated that Ime4 promotes the accumulation of *IME1* transcripts suggesting that there is positive regulation between the two genes during entry into sporulation (Shah and Clancy 1992). Another transcription factor encoded by the *NDT80* gene governs commitment to meiotic divisions and activates transcription of the “middle” and “late” sporulation genes (Xu et al. 1995;Winter 2012).

Although budding yeast is a highly tractable model organism, they enter and progress through sporulation relatively asynchronously. This makes it challenging to perform stage specific analyses using whole cell population based assays such as northern blots and RNA sequencing. To address this, some methods have been developed to synchronize sporulation. These methods utilize our understanding of the external and internal signals which drive entry and progression through sporulation. Whereas many of the laboratory yeast strains sporulate poorly, the sporulation proficient strain background SK1 can undergo pre-meiotic DNA replication and meiosis division with a certain degree of synchrony (Mai and Breeden 2000;Stuart 2008). The wildtype SK1 strain sporulates more efficiently than others because of three single nucleotide polymorphisms (SNPs). (Deutschbauer and Davis 2005). One particular SNP leads to reduced expression of *RME1*, a repressor of *IME1*; consequently SK1 cells have higher *IME1* levels during nutrient starvation (Gerke et al. 2006). As such, most sporulation protocols involve culturing SK1 cells in presporulation medium containing tryptone and potassium acetate (BYTA). By utilizing non-fermentable carbon sources, the cells are metabolically primed to initiate sporulation upon shifting to nutrient depleted medium (SPO) (Becker et al. 2015). In addition, controlled induction of either *IME1* or *NDT80* can synchronize the cells at different stages of the meiotic program (Carlile and Amon 2008;van Werven et al. 2012). Despite these methods, there is still a need to develop and optimize an approach to synchronize cells throughout sporulation, from DNA replication till the completion of meiotic divisions.

Here, I reveal new insights into *IME1* signalling, and describe a method to synchronize cells throughout meiotic DNA replication and meiotic divisions. I demonstrate that timed expression of *IME1* is sufficient to drive synchronous sporulation. In addition, I find that *IME1* expression leads to increased *IME4* expression, explaining my observation that cells readily enter sporulation highly synchronously when *IME1* is induced alone. I describe how the synchrony of sporulation can be assessed by flow cytometry and by counting nuclei. Surprisingly, my system does not require a pre-growth step in acetate-containing BYTA. Instead, synchronous sporulation can be induced directly in cells grown in reduced glucose YPD (rich medium) until saturation before shifting to SPO. The omission of the BYTA culturing step facilitates the use of larger cultures for detailed

sporulation time course experiments. My findings are in agreement with previous work by Nachman *et al*. showing that cell-to-cell variability in *IME1* levels causes variation in the timing of meiotic divisions, independent of nutritional history, cell size or cell cycle stage (Nachman et al. 2007). Finally, I show that inducible *IME1* and *NDT80* expression systems can be combined to improve sporulation synchrony. This optimized synchronization method is used in subsequent chapters which involve high resolution, stage-specific studies of the transcriptional program throughout sporulation.

3.3 Results

3.3.1 Timely induction of *IME1* and *IME4* leads to synchronous sporulation

The sporulation proficient strain background SK1 can undergo pre-meiotic DNA replication and meiosis division with a certain degree of synchrony (Mai and Breeden 2000;Stuart 2008). However, even more synchronous cell populations are necessary to study dynamic patterns of gene regulation during specific stages of sporulation or meiosis. This led to the development of different strategies to further improve the synchrony of sporulation (Wan et al. 2006;Carlile and Amon 2008;Berchowitz et al. 2013). Previous work showed that expressing *IME1* together with *IME4* from the inducible *CUP1* promoter (*CUP-IME1* and *CUP-IME4*) drives cells to undergo gametogenesis more synchronously compared to wildtype SK1 (Berchowitz et al. 2013). For this procedure, cells were grown in rich medium (YPD) till saturation, shifted to pre-sporulation medium (BYTA), and transferred to sporulation medium (SPO). Subsequently, cells were incubated in SPO for two hours before *IME1* and *IME4* were induced with $CuSO_4$.

I speculated that the timing of *IME1* and *IME4* induction in SPO could be an important factor in regulating sporulation kinetics, due to their known importance for meiotic entry and the fact that both transcripts are induced early on by the same nutritional cues (Shah JC; van Werven, 2011 #55;Chu, 1998 #13;Primig, 2000 #4}. To examine this, I expressed *IME1* and *IME4* at different times, and quantified the percentage of cells that completed meiotic divisions for a series of time points **(Figure 3.1 A)**. From these data, I estimated the synchrony of meiotic divisions by

computing the time or period taken for 75% of the cells to complete meiotic divisions (see *materials and methods* for details) **(Figure 3.1 A and B)**. The shorter the time or period, the more synchronous the meiotic divisions. I also conducted a one-way ANOVA and a post hoc Tukey's test to compare the effect of expressing *IME1* and *IME4* at different times on the mean period taken to complete meiotic divisions. The analysis of variance showed that the effect of expressing *IME1* and *IME4* at different times was statistically significant, $F(5,12) = 3.82$, $p = 0.0265$. I observed a statistically significant improvement in the synchrony of meiotic divisions when *IME1* and *IME4* were induced at two hours after shifting to SPO instead of at zero hours (4.19 h compared to 2.21 h, $p = 0.0112$) **(Figure 3.1 B).** Interestingly, inducing *IME1* and *IME4* either earlier or later did not improve the synchrony significantly ($p > 0.05$) suggesting that there is an optimal period to induce the two master regulators. The differences in kinetics cannot be explained by Ime1 and Ime4 protein levels since they were comparable between the different samples **(Figure 3.1 C)**. In conclusion, my result shows that the timing of *IME1* and *IME4* induction in sporulation medium contributes to synchronous meiotic divisions.

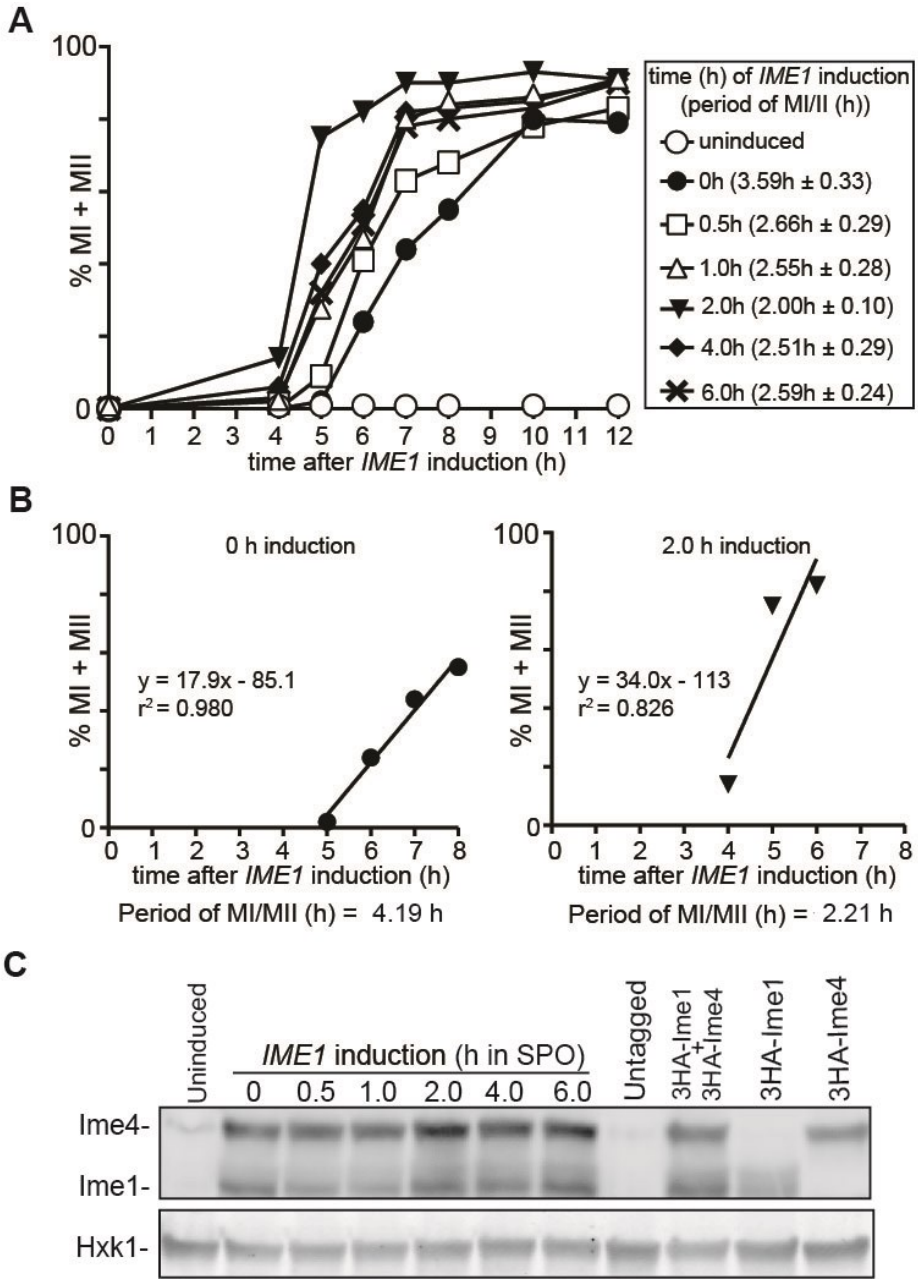


Figure 3.1 Synchronous sporulation requires specific timing of *IME1* and *IME4* induction

(A) Kinetics of meiotic divisions in diploid cells harbouring *CUP1* promoter fusions with *IME1* and *IME4* (*pCUP-IME1/pCUP-IME4*) (FW1810). Cells were grown overnight in rich medium (YPD), diluted to presporulation medium (BYTA), and grown for another 16 hr. Subsequently cells were transferred to sporulation medium (SPO), and *IME1* and *IME4* were induced at 0, 0.5, 1, 2, 4, and 6 hr in SPO. Samples were collected at 4 hr after induction up to 12 hr with a 1-hr interval, fixed in ethanol, nuclei were stained with DAPI, and DAPI masses were counted. Cells that harboured two, three, or four DAPI masses were classified as cells undergoing meiosis I or meiosis II (% MI + MII, *y*-axis). For each time point, at least

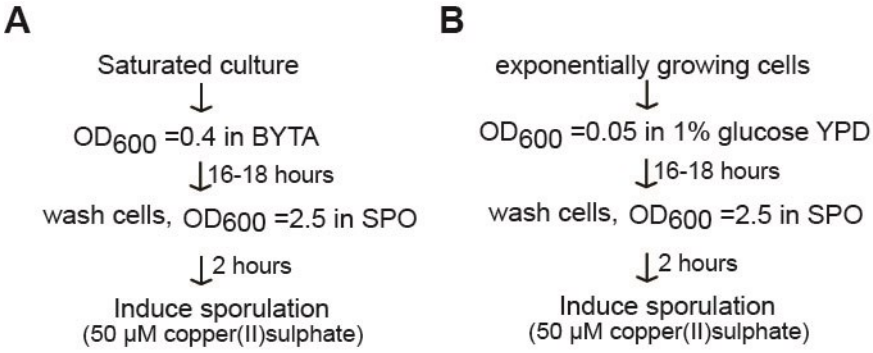
200 cells were counted. The time after *IME1/IME4* induction is plotted on the *x*-axis. This panel is an example of one repeat out of three independent experiments, n = 3. From each time course experiment, I also computed the time or period taken for 75% of the cells to complete meiotic divisions (see *Materials and Methods* for details). Values in brackets next to the legend are the mean number of hours and the standard error of the mean (SEM) from three independent experiments, n = 3. One-way ANOVA and a post hoc Tukey's test were used to find statistically significant differences between the group means (see main text for details). **(B)** Graph to illustrate how I determined the time or period taken to complete meiotic divisions when *IME1* and *IME4* were induced at 0 or 2 hr after shifting cells to SPO medium as described in (A). A linear trend line was fitted from the first time point where meiotic divisions were detected, to the time point where 75% or more of the cells completed meiotic divisions. From the function, I calculated the period or time taken for 75% of the cells to complete meiotic divisions (MI/II). This panel is an example of one repeat out of three independent experiments. **(C)** Western blot showing Ime1 and Ime4 protein levels in cells described in (A). Samples were taken at 2 hr after inducing *IME1* and *IME4*. Ime1 and Ime4 levels were detected by anti-hemagglutinin (HA) antibodies. I also measured Ime1 and Ime4 in an untagged control (FW1511), and in cells that contain HA-tagged *IME1* (FW2444), or *IME4* (FW2480) alone. To control for loading, Hxk1 levels were also determined. This blot is representative of three independent experiments, n = 3.

3.3.2 Prior growth in BYTA pre-sporulation medium is not necessary for synchronous sporulation

Efficient *IME1* transcription requires glucose and nitrogen starvation, and the presence of a non-fermentable carbon source in the growth medium (Kassir et al. 1988). To metabolically prime cells to express high levels of *IME1* upon shifting to sporulation medium (SPO), cells are usually pre-grown in acetate containing medium (BYTA) which lacks glucose **(Figure 3.2 A)**. In contrast, induction of transcription from the *CUP1* promoter solely requires the presence of copper ions in the medium. If variability in the onset of meiotic divisions is largely dependent on *IME1* and *IME4* levels, then pre-growth in BYTA should be dispensable when entry into sporulation is induced from the *CUP1* promoter.

I tested if the *pCUP-IME1*/*pCUP-IME4* system can induce gametogenesis synchronously when cells were pre-grown in glucose containing medium (YPD) and shifted to SPO directly **(Figure 3.2 B)**. As established in the previous section, optimal synchrony of meiosis was achieved by inducing *IME1* and *IME4*, 2 hrs after shifting to SPO. Synchronous sporulation experiments should also start with a population of non-budding, starved cells in the G1 phase because the mitotic and

meiotic programs are incompatible with each other (Colomina et al. 1999). To ensure that most cells were arrested as unbudded G1 cells, I grew cells in YPD with reduced glucose (1.0 % wt/vol instead of 2.0 % wt/vol) for 16-18 hours. Whereas the majority of cells were budding when grown in medium with standard glucose levels, the reduced glucose condition enriched for unbudded cells (about 90 percent) after an overnight culture **(Figure 3.2 C)**. Cells pre-grown in either YPD or BYTA also gave rise to viable spores **(Figure 3.2 D)**. Thus, I used the reduced glucose YPD medium for subsequent YPD to SPO experiments in my thesis.



C

% glucose	% budding cells
2% w/v	48.25 ± 5.29
1% w/v	9.25 ± 0.52

D

condition	% viable spores
BYTA to SPO	96.13 ± 0.13
YPD to SPO	95.63 ± 1.87

Figure 3.2 Two protocols for synchronous sporulation

(A) Flowchart for inducing synchronous sporulation by pre-growing the cells in acetate-containing presporulation medium (BYTA). Diploid *pCUP-IME1/pCUP-IME4* (FW1810) cells were grown to saturation for 24 hr in YPD. Cells were subsequently diluted to an OD_{600} of 0.4 in BYTA and grown for 16–18 hr. Cells were then pelleted by centrifugation, washed with sterile water and resuspended to a final OD_{600} of 2.5 in SPO; 50 µM $CuSO_4$ was added 2 hr after the cells were transferred to SPO to induce *IME1* and *IME4*. **(B)** Flowchart for inducing synchronous sporulation without pre-growing the cells in acetate containing medium (YPD to SPO). Diploid *pCUP-IME1/pCUP-IME4* (FW1810) cells were grown to exponential phase for 6–7 hr in YPD. Cells were subsequently diluted to YPD medium with 1% glucose and grown for 16–18 hr to an OD_{600} of 10–12 to obtain mostly unbudded G_1 cells. Cells were then pelleted by centrifugation, washed with sterile water and resuspended to a final OD_{600} of 2.5 in SPO; 50 µM $CuSO_4$ was added 2 hr after the cells were transferred to SPO to induce *IME1* and *IME4*. **(C)** Budding index of diploid *pCUP-IME1/pCUP-IME4* (FW1810) cells cultured for 16 hr in YPD with different glucose

concentrations. The mean and SEM from three independent experiments, n = 3, is shown, and 400 cells were counted for each repeat. **(D)** Spore viability of the *pCUP-IME1*/*pCUP-IME4* strain (FW1810). Cells were grown overnight in YPD, and induced to sporulate in SPO after transfer from YPD or presporulation media (BYTA). Sporulation was induced using standard protocols (BYTA to SPO), or by using the method described in (B). $CuSO_4$ was added 2 hr after the cells were transferred to SPO. Tetrads were collected 24 hr after *IME1* induction, dissected, and assayed for viability. The mean value of three independent experiments, n = 3 plus the SEM is shown. 160 spores were dissected per repeat.

Next, I examined whether cells prepared under the two different protocols (BYTA to SPO vs. YPD to SPO) showed any differences in sporulation synchrony. This was done by measuring the kinetics of meiotic S-phase and meiotic divisions in repeated time course experiments. By flow cytometry, I observed that cells pre-grown in BYTA or YPD both completed meiotic DNA replication in approximately 60 minutes **(Figure 3.3 A)**. Remarkably, in the YPD to SPO condition the DNA profile showed intermediate peaks for several time points (between 2C and 4C) indicating that the population of cells underwent DNA replication with a high degree of synchrony **(Figure 3.3 A,** right**)**. When the cells were pre-grown in BYTA, these intermediate peaks were less pronounced **(Figure 3.3 A,** left**)**.

In addition, I measured the rate of meiotic divisions and found the majority of cells competed meiotic divisions within comparable periods in both conditions **(Figure 3.3 B)**. A two-tailed t-test showed that the time taken to complete meiotic divisions under both conditions were not statistically significantly different ($p > 0.05$). Overall, my results show that synchronous DNA replication and meiotic divisions can be induced from cells pre-cultured till saturation in nutrient rich medium containing glucose.

The composition of SPO medium could influence the synchrony of meiotic divisions of cells. To test this, cells pre-cultured in reduced glucose YPD were shifted to either regular SPO (0.3% wt/vol potassium acetate) or to supplemented SPO (1.0% wt/vol potassium acetate with other amino acids and adenine; see *materials and methods* for details). Cells started meiotic divisions slightly earlier when using the supplemented SPO instead of the regular SPO **(Figure 3.3 C)**. Hence, unless otherwise stated, supplemented SPO was used for most other experiments in this thesis.

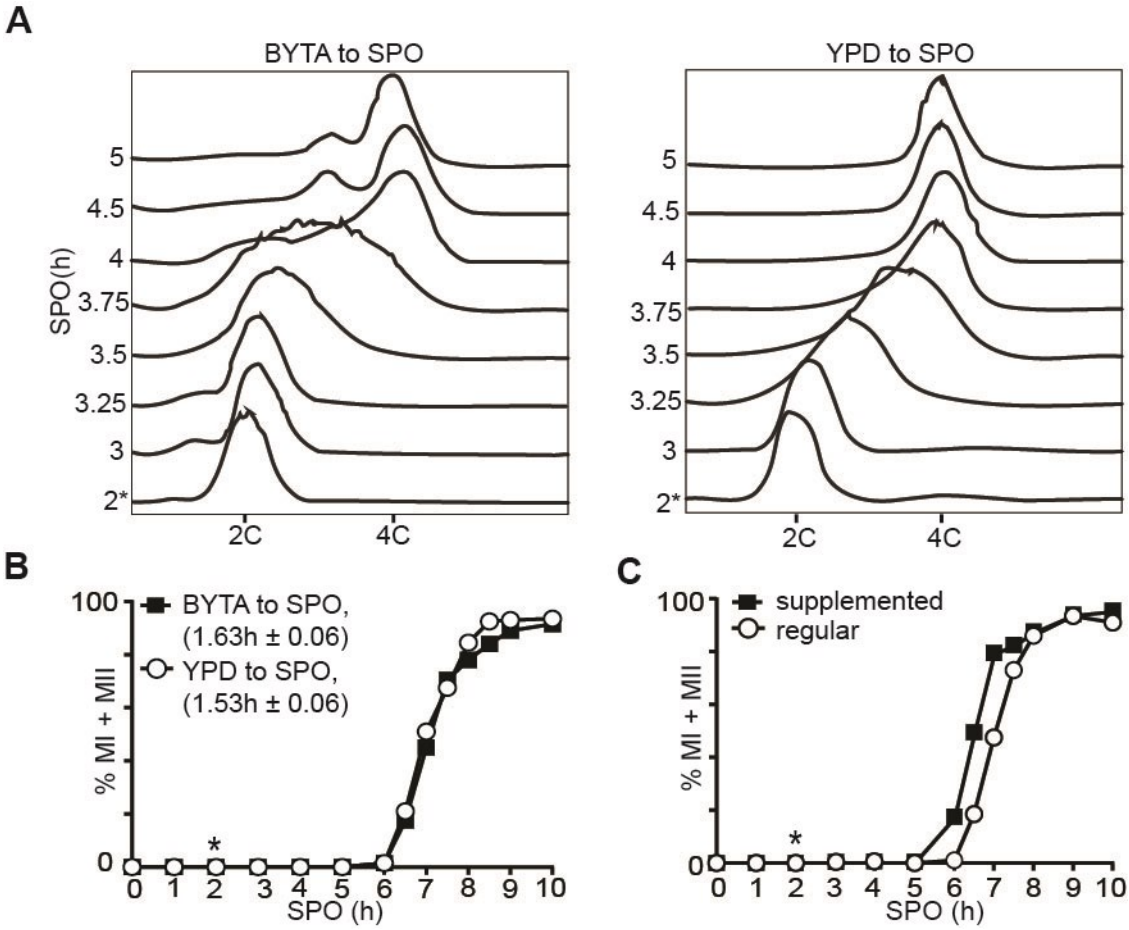


Figure 3.3 Pre-growing cells in BYTA is not necessary for synchronous sporulation

(A) Flow cytometry analysis of DNA content of diploid *pCUP-IME1/pCUP-IME4* (FW1810) cells cultured in either reduced glucose YPD or BYTA before shifting to SPO. Samples were taken at indicated time points, fixed, and DNA content was measured by propidium iodide staining. At least 50,000 cells were analyzed at each time point. Result is representative of three independent repeats, n = 3. **(B)** Kinetics of meiotic divisions in the *pCUP-IME1*/*pCUP-IME4* strain (FW1810) as described in (A). For determining the kinetics of meiotic divisions, samples were taken at the indicated time point, fixed, and DAPI masses were counted. Cells that harboured two, three, or four DAPI masses were classified as cells undergoing meiosis I or meiosis II (% MI + MII). For each time point, at least 200 cells were counted. This panel is an example of one repeat out of three independent experiments, n = 3. I also computed the time or period taken for 75% of the cells to complete meiotic divisions (see *Materials and Methods* for details). This number is displayed in brackets next to the legend, and represents the mean number of hours followed by the SEM of three independent experiments, n = 3. **(C)** Kinetics of meiotic divisions of the *pCUP-IME1*/*pCUP-IME4* strain (FW1810) except that sporulation was induced in either regular SPO or supplemented SPO

(see *Materials and Methods*). *, time of induction of *IME1*, 2 hr after the cells were transferred to SPO. Result is representative of three independent repeats, n = 3.

3.3.3 The HA tag on *IME1* and *IME4* does not compromise synchrony

The *pCUP-IME1*/*pCUP-IME4* strains used in previous sections have a 3x hemagglutinin (3HA) tag fused to the N-termini of Ime1 and Ime4. While the 3HA tags facilitate monitoring of Ime1 and Ime4 protein levels, it is unclear if they have any effect on protein function and sporulation. To test this, I compared the kinetics of meiotic S-phase in cells with the 3HA tag to those without. Both strains were pre-grown in reduced glucose YPD and were shifted to SPO. $CuSO_4$ was added 2 hr after the cells were transferred to SPO to induce either tagged or untagged *IME1* and *IME4*. FACS analysis showed that both strains underwent meiotic S-phase synchronously in either regular or supplemented SPO **(Figure 3.4 A and B)**. As a further check, I measured the kinetics of meiotic division in both strains and found them to be comparable **(Figure 3.4 C)**. Thus, both the tagged and untagged pCUP-IME1/pCUP-IME4 strains were suitable for synchronous sporulation experiments and the 3HA tag did not appear to inhibit Ime1 and Ime4 activity.

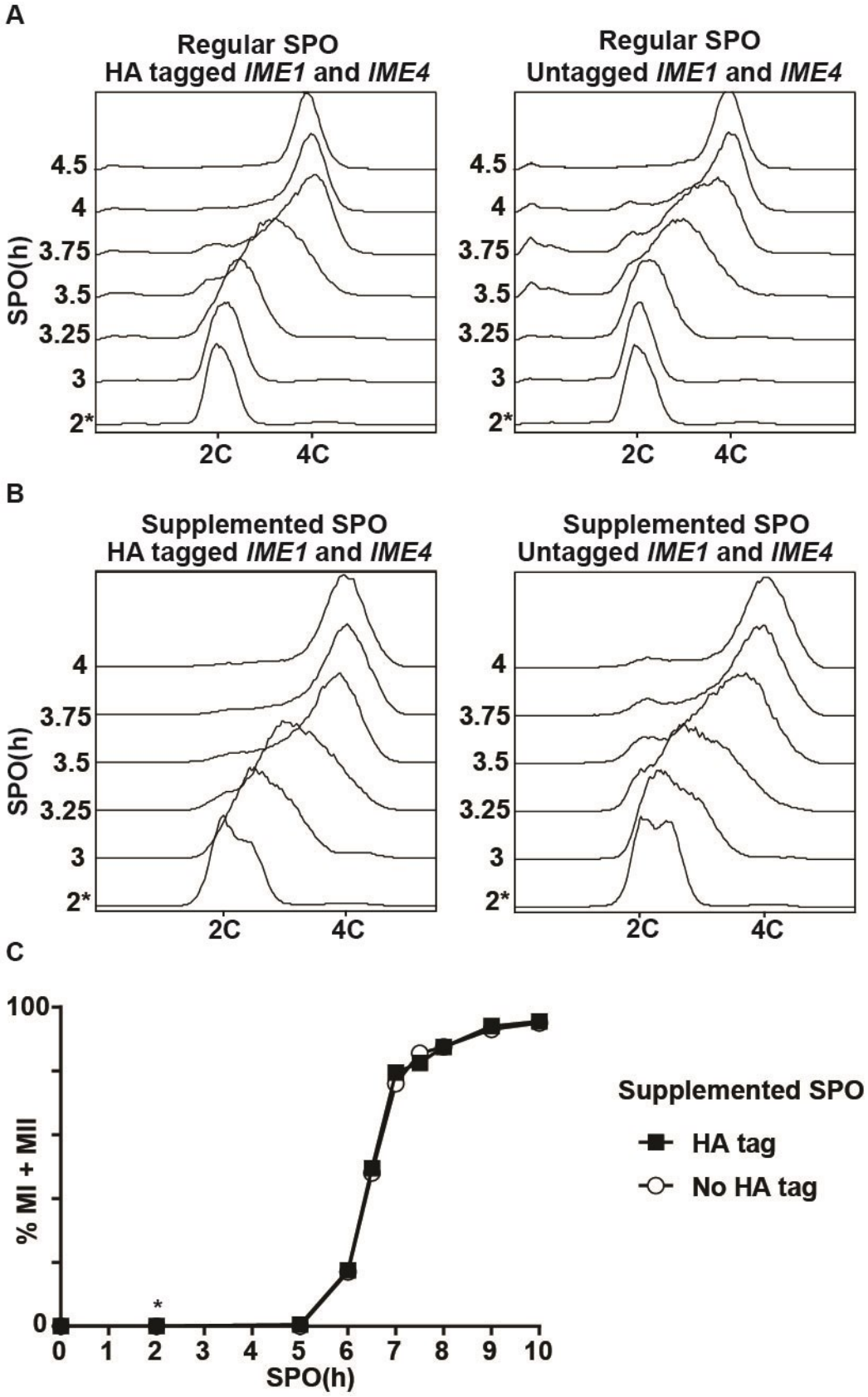


Figure 3.4 Tagged and untagged strains sporulate synchronously

(A) Flow cytometry analysis of DNA content of cells with N-terminal HA tagged *IME1* and *IME4* (FW1810, left) or untagged *IME1* and *IME4* (FW1884, right). Both strains were cultured in 1% glucose YPD before shifting to regular SPO. $CuSO_4$ was added 2 hr after the cells were transferred to SPO to induce *IME1* and *IME4*. Samples were taken at indicated time points, fixed, and DNA content was measured by propidium iodide staining. At least 50,000 cells were analyzed at each time point. Data is representative of two independent repeats, n = 2. **(B)** Same as in (A), except that supplemented SPO (see *Materials and Methods*) was used. Data is representative of two independent repeats, n = 2. **(C)** Kinetics of meiotic divisions in cells with N-terminal HA tagged *IME1* and *IME4* (FW1810) or untagged *IME1* and *IME4* (FW1884), in supplemented SPO. For determining the kinetics of meiotic divisions, samples were taken at the indicated time point, fixed, and DAPI masses were counted. Cells that harboured two, three, or four DAPI masses were classified as cells undergoing meiosis I or meiosis II (% MI + MII). For each time point, at least 200 cells were counted. *, time of induction of *IME1* and *IME4*, 2 hr after the cells were transferred to SPO. Data is representative of two independent repeats, n = 2.

3.3.4 Supplementing YPD with acetate or glutamate does not improve synchrony

Budding yeast cells must coordinate their metabolism to the demands of sporulation. Upon glucose starvation, budding yeast cells metabolize acetate, a non-fermentable carbon source, through the tricarboxylic acid cycle (TCA). This switch from glucose to acetate utilization primes cells to transit from vegetative growth to meiosis (Becker et al. 2015). Acetate metabolism through the TCA generates intermediates for glutamate synthesis (Dickinson et al. 1983). Indeed, different studies have shown that cells maximize glutamate synthesis during early meiosis (Dickinson et al. 1983;Ray and Ye 2013). I tested if the addition of acetate or glutamate in the growth medium improved the kinetics of meiosis in the *pCUP-IME1*/*pCUP-IME4* strains. Adding different concentrations of acetate to the YPD media did not improve sporulation synchrony measured by the rate of meiotic divisions **(Figure 3.5 A)**. Acetate in the growth media is dispensable for synchronous progression through meiotic S-phase **(Figure 3.3 A and Figure 3.5 B)**. FACS analysis also showed that additional glutamate in the growth medium was unnecessary for synchronous sporulation **(Figure 3.5 C)**. Taken together, these results suggest that the YPD to SPO protocol reported in this chapter **(Figure 3.2 B)** is sufficient to generate highly synchronous cell populations for stage specific analysis of the sporulation program.

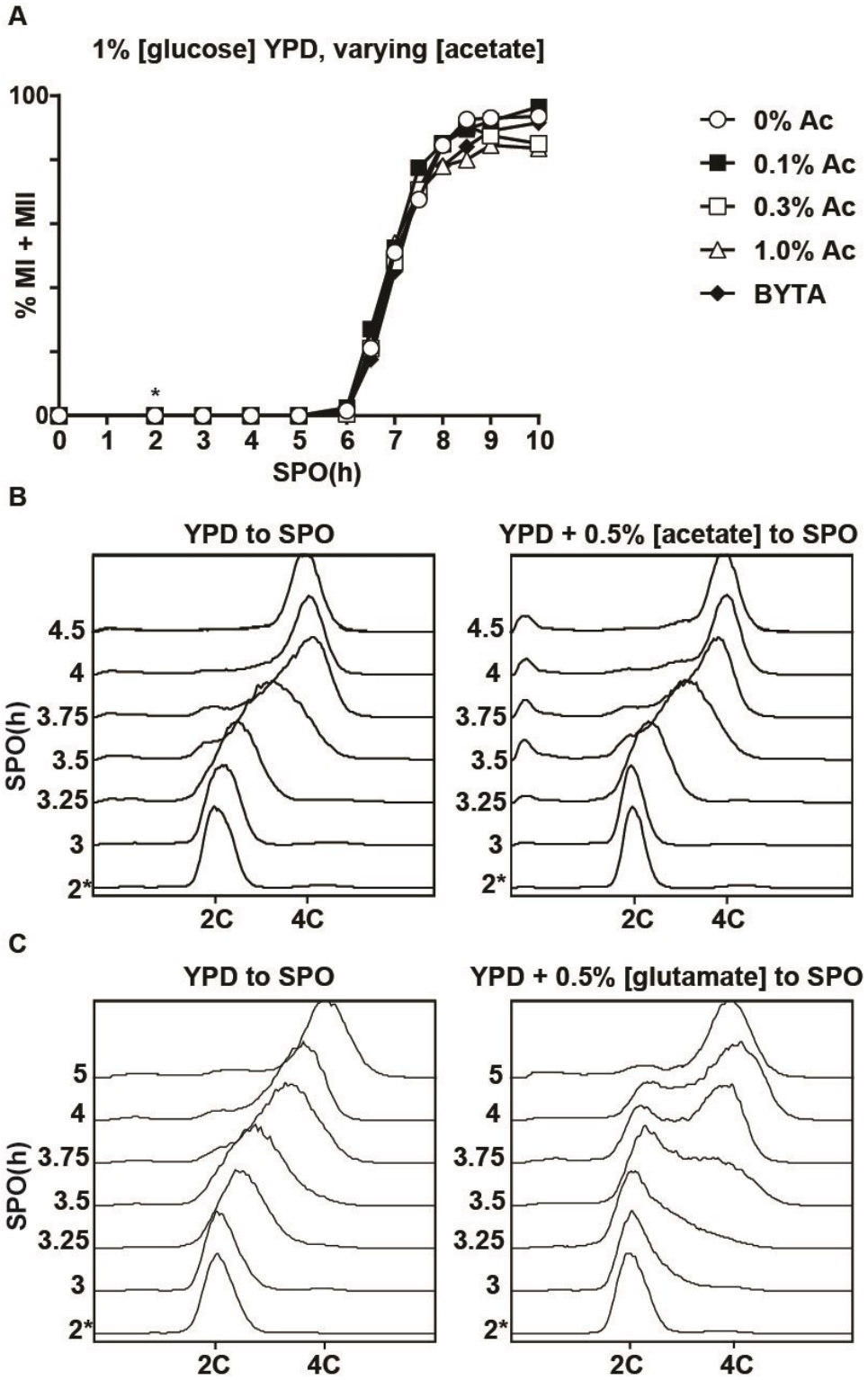


Figure 3.5 Supplementing YPD with acetate or glutamate does not improve synchrony

(A) Kinetics of meiotic divisions of the *pCUP-IME1*/*pCUP-IME4* strain (FW1810). Cells were grown for 16-18 hr in either 1% YPD medium supplemented with

varying concentrations of potassium acetate (Ac) or in regular presporulation medium (BYTA). Subsequently cells were transferred to regular SPO and *IME1* and *IME4* were induced at 2 hr. Samples were collected at the indicated time points, fixed in ethanol, nuclei were stained with DAPI, and DAPI masses were counted. Cells that harboured two, three, or four DAPI masses were classified as cells undergoing meiosis I or meiosis II (% MI + MII, *y*-axis). For each time point, at least 200 cells were counted. Data is representative of two independent repeats, n = 2. **(B)** Flow cytometry analysis of DNA content of the *pCUP-IME1*/*pCUP-IME4* strain (FW1810). Strains were either cultured in 1% glucose YPD (left) or 1% glucose YPD with 0.5% potassium acetate, before shifting to regular SPO. $CuSO_4$ was added 2 hr after the cells were transferred to SPO to induce *IME1* and *IME4*. Samples were taken at indicated time points, fixed, and DNA content was measured by propidium iodide staining. At least 50,000 cells were analyzed at each time point. Note that the left panel is the same flow cytometry experiment as displayed in Figure 3.4(A), left. Data is representative of two independent repeats, n = 2. **(C)** Flow cytometry analysis of DNA content of the *pCUP-IME1*/*pCUP-IME4* strain (FW1810). Strains were either cultured in 1% glucose YPD (left) or 1% glucose YPD with 0.5% glutamate, before shifting to supplemented SPO. $CuSO_4$ was added 2 hr after the cells were transferred to SPO to induce *IME1* and *IME4*. Samples were taken at indicated time points, fixed, and DNA content was measured by propidium iodide staining. At least 50,000 cells were analyzed at each time point. *, time of induction of *IME1* and *IME4*, 2 hr after the cells were transferred to SPO. Data is representative of two independent repeats, n = 2.

3.3.5 *IME1* induction alone is sufficient to induce synchronous sporulation

Both *IME1* and *IME4* have been shown to promote entry into sporulation (van Werven and Amon 2011). Whereas *ime4* deletion mutants do not sporulate in certain strain backgrounds but do in others, *IME1* is essential for sporulation in *S. cerevisae* (Kassir et al. 1988;Shah and Clancy 1992;Hongay et al. 2006). *IME4* has been implicated to positively regulate *IME1* expression (Shah and Clancy 1992). I hypothesized that if *IME1* and *IME4* regulate each other, then synchronous sporulation should require controlled expression of either *IME1* or *IME4* alone. To test this, I measured the period taken to complete meiotic divisions when *IME1*, *IME4*, or both were induced from the *CUP1* promoter. One-way ANOVA showed that there was a statistically significant difference between the group means, $F(3,8) = 6.97$, $p = 0.0127$. I found that the kinetics of meiotic divisions of cells that express *pCUP-IME4* only were comparable to wildtype control cells **(Figure 3.6 A)**. In contrast, the kinetics of meiotic division in cells harbouring *pCUP-IME1* alone or both *pCUP-IME1* and *pCUP-IME4* significantly improved when compared to the wildtype control, which was confirmed by a post hoc Dunnett's test (compare 1.63 h

and 1.79 h to 4.30 h, p = 0.0166 and p = 0.0223, respectively). The results were similar when I examined the kinetics of meiosis I and meiosis II separately **(Figure 3.6 B)**. To investigate more closely whether induction of *pCUP-IME1* alone is sufficient for synchronous sporulation, I also monitored the kinetics of pre-meiotic DNA replication **(Figure 3.6 C)**. I found that cells harbouring either *pCUP-IME1* or *pCUP-IME1/pCUP-IME4*, underwent pre-meiotic DNA replication synchronously within approximately 45 minutes, and gave rise to viable spores **(Figure 3.6 D)**. In conclusion, temporal expression of *IME1* alone is sufficient to induce synchronous sporulation.

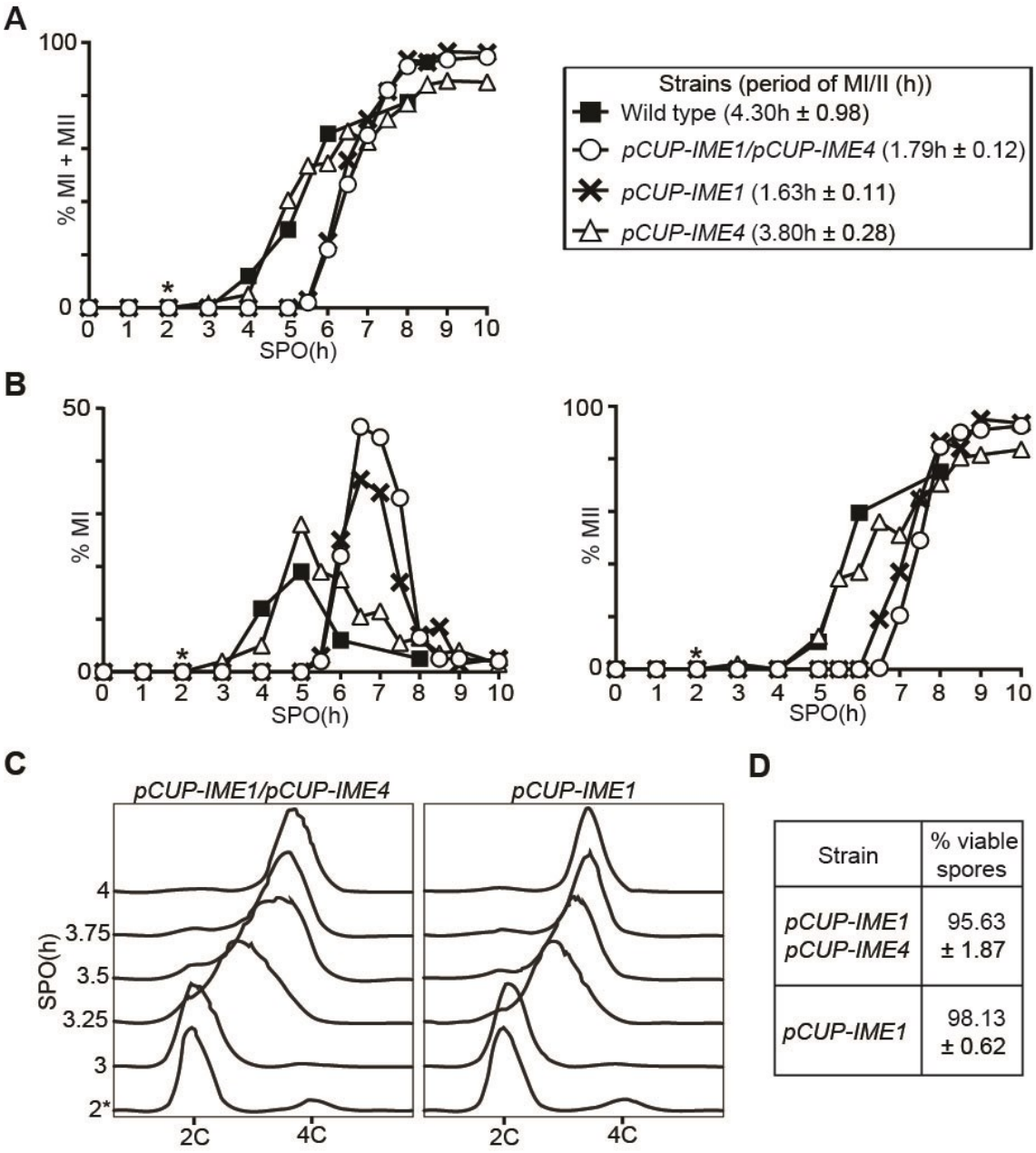


Figure 3.6 *IME1* induction is sufficient for synchronous DNA replication and meiotic divisions

(A) Kinetics of meiotic divisions in wildtype cells (FW1511), cells harbouring *IME1* and *IME4* fused to the *CUP1* promoter (*pCUP-IME1/pCUP-IME4*) (FW1810), *pCUP-IME1* (FW2444), or *pCUP-IME4* (FW2480). Cells were grown overnight in YPD, and shifted to supplemented SPO (see *Materials and Methods*); 50 μM $CuSO_4$ was added 2 hr after the cells were transferred to SPO. Samples were taken at the indicated time point, fixed in ethanol, nuclei were stained with DAPI, and DAPI masses were counted. Cells that harboured two, three, or four DAPI masses were classified as cells undergoing meiosis I or meiosis II (% MI + MII). This panel is an example of one repeat out of three independent experiments, n = 3. For each time point, at least 200 cells were counted. I also computed the time or period taken for 75% of the cells to complete meiotic

divisions (see *Materials and Methods* for details). The values in brackets next to the legend are the mean number of hours followed by the SEM of three independent experiments, n = 3. One-way ANOVA and a post hoc Dunnett's test were used to find statistically significant differences between the group means (see main text for details). **(B)** Similar to A except that the percentages of bi- (left panel), tri-, and tetra-nucleate (right panel) cells are shown. This plot is representative of three independent experiments. **(C)** Flow cytometry analysis of DNA content of *pCUP-IME1*/*pCUP-IME4* (FW1810) and the *pCUP-IME1* (FW2444) cells that were induced to sporulate as described in (A). Samples were taken at indicated time points, fixed, and DNA content was measured by propidium iodide staining; 50 µM $CuSO_4$ was added 2 hr after the cells were transferred to SPO. At least 50,000 cells were analyzed at each time point. This result is representative of three independent experiments. **(D)** Spore viability of the *pCUP-IME1*/*pCUP-IME4* (FW1810) and the *pCUP-IME1* (FW2444) cells that were induced to sporulate as described in (A). Tetrads were collected 24 hr after induction, dissected, and assayed for viability (*n* = 160 spores). *, time of induction of *IME1*, 2 hr after the cells were transferred to SPO. Values are the mean and SEM from three independent experiments, n = 3.

The observation that temporal expression of *IME1* but not of *IME4*, generates a high degree of synchrony during DNA replication and meiotic divisions, prompted me to revisit how the two genes regulate each other. I hypothesized that *IME1* directly or indirectly regulates *IME4* expression. To examine this possibility, I measured *IME4* transcript levels in cells harbouring *pCUP-IME1* in the presence or absence of $CuSO_4$. As expected, *IME1* transcript levels increased when copper ions were added to the SPO medium **(Figure 3.7 A, left)**. Since *IME4* is also regulated by an anti-sense transcript, I specifically quantified *IME4* sense mRNA using a transcript specific primer in the reverse transcription reaction (Hongay et al. 2006;Gelfand et al. 2011). I found that *IME4* transcript levels significantly increased when *IME1* was induced suggesting that Ime1 stimulates *IME4* transcription **(Figure 3.7 A, right)**. Data from a genome-wide study indicated that *IME4* is directly regulated by the repressor Ume6 (Williams et al. 2002). During early sporulation Ime1 interacts with Ume6 to form a transcription activating complex for the expression of early meiotic genes (Bowdish et al. 1995;Rubin-Bejerano et al. 1996). To test whether Ume6 indeed binds the *IME4* promoter, I identified the canonical URS1 motif (TAGGCGGC) sequence at -234 base pairs (bp) upstream in the *IME4* promoter. More importantly, I found that Ume6 was directly bound to the *IME4* promoter as shown by ChIP **(Figure 3.7 B)**. In conclusion, my results showed that *IME1* directly regulates the expression of *IME4* explaining why *IME1* can single-handedly induce synchronous sporulation. These results also suggest that

IME1 and *IME4* act in a positive feedback loop to stimulate the expression of each other.

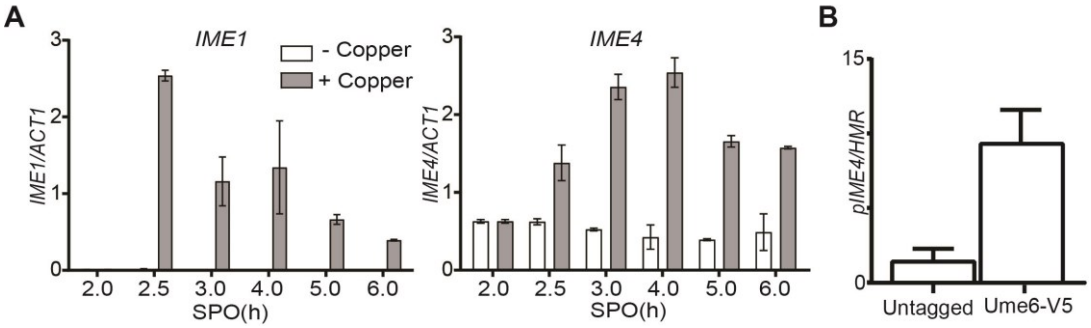


Figure 3.7 *IME1* directly regulates *IME4* expression

(A) Induction of *IME1* promotes *IME4* expression. Diploid cells harbouring *pCUP-IME1* (FW2444) were transferred to SPO. Cells were either untreated (control) or treated with $CuSO_4$ (induced), and samples were collected at different time points. Total RNA was isolated, reverse transcribed, and *IME1* (left panel) or *IME4* (right panel) mRNA levels were measured by quantitative PCR. To quantify *IME4* levels, primers specific for the *IME4* and *ACT1* sense strand were used in the reverse transcription reaction. Signals were normalized to *ACT1* levels. Values displayed correspond to the mean and the bars correspond to the range. Two independent repeats were done, n = 2. **(B)** Ume6 binds to the promoter of *IME4*. Diploid cells harbouring Ume6 tagged with the V5 epitope (FW1208) and a wildtype control strain (FW1511) were grown in YPD to saturation. Cells were fixed with formaldehyde, and cells were processed for ChIP analyses (see *Materials and Methods* for details). DNA fragments specific to the *IME4* promoter (*pIME4*) were amplified and quantified by qPCR. Signals were normalized to the *HMR* locus. . *, time of induction of *IME1,* 2 hr after the cells were transferred to SPO. The mean is displayed with the SEM. Three independent repeats were done, n = 3.

3.3.6 Controlled induction of *IME1* and *NDT80* is a suitable tool to study the sporulation program

Different genetic approaches have been used to synchronize cells at different stages of sporulation (Benjamin et al. 2003;Wan et al. 2006;Carlile and Amon 2008). One approach makes use of controlled expression of the transcription factor *NDT80* and as a result, cells undergo meiotic divisions synchronously (Benjamin et al. 2003;Carlile and Amon 2008). The Ndt80 transcription factor promotes the expression of numerous genes that regulate meiotic divisions, also known as middle genes (Xu et al. 1995;Chu et al. 1998). Effective induction of

NDT80 is achieved by controlling its expression from the *GAL1* promoter (*pGAL-NDT80)* and by the transcription factor *GAL4-ER* consisting of the Gal4 DNA binding domain fused to the estrogen receptor binding domain. In the presence of β-estradiol, *pGAL-NDT80* cells induce *NDT80* and exit from pachytene arrest to undergo meiotic divisions (Benjamin et al. 2003;Carlile and Amon 2008). The *pGAL-NDT80* system specifically synchronizes meiotic divisions during gametogenesis, but unlike the *pCUP-IME1* system, this method does not synchronize the events prior to meiotic chromosome segregation. To examine whether it is possible to combine the *pCUP-IME1* system with the *pGAL-NDT80* system, I generated a diploid strain with both synchronization systems. The early and middle stages of sporulation were initiated by *IME1* at 2 hours and *NDT80* at 6 hours after shifting cells to SPO, respectively **(Figure 3.8 A)**. I found that the *pCUP-IME1* and *pGAL-NDT80* cells had a similar degree of synchrony of meiotic divisions **(Figure 3.8 B)**. The *pCUP-IME1/pGAL-NDT80* strain showed a minor improvement in synchrony, which was not statistically significant ($p > 0.05$), when compared to cells expressing *pCUP-IME1* or *pGAL-NDT80* alone (1.18 h compared to 1.63 h and 1.63 h) **(Figure 3.8 B)**. I observed a similar trend when I examined meiosis I and meiosis II divisions separately, showing that the *pCUP-IME1* and *pGAL-NDT80* systems can be combined **(Figure 3.8 C)**. With the *pCUP-IME1/ pGAL-NDT80* system, I could synchronize cells at the level of pre-meiotic DNA replication till completion of meiotic divisions, with the added advantage of being able to control entry into both the early and middle stages of gametogenesis.

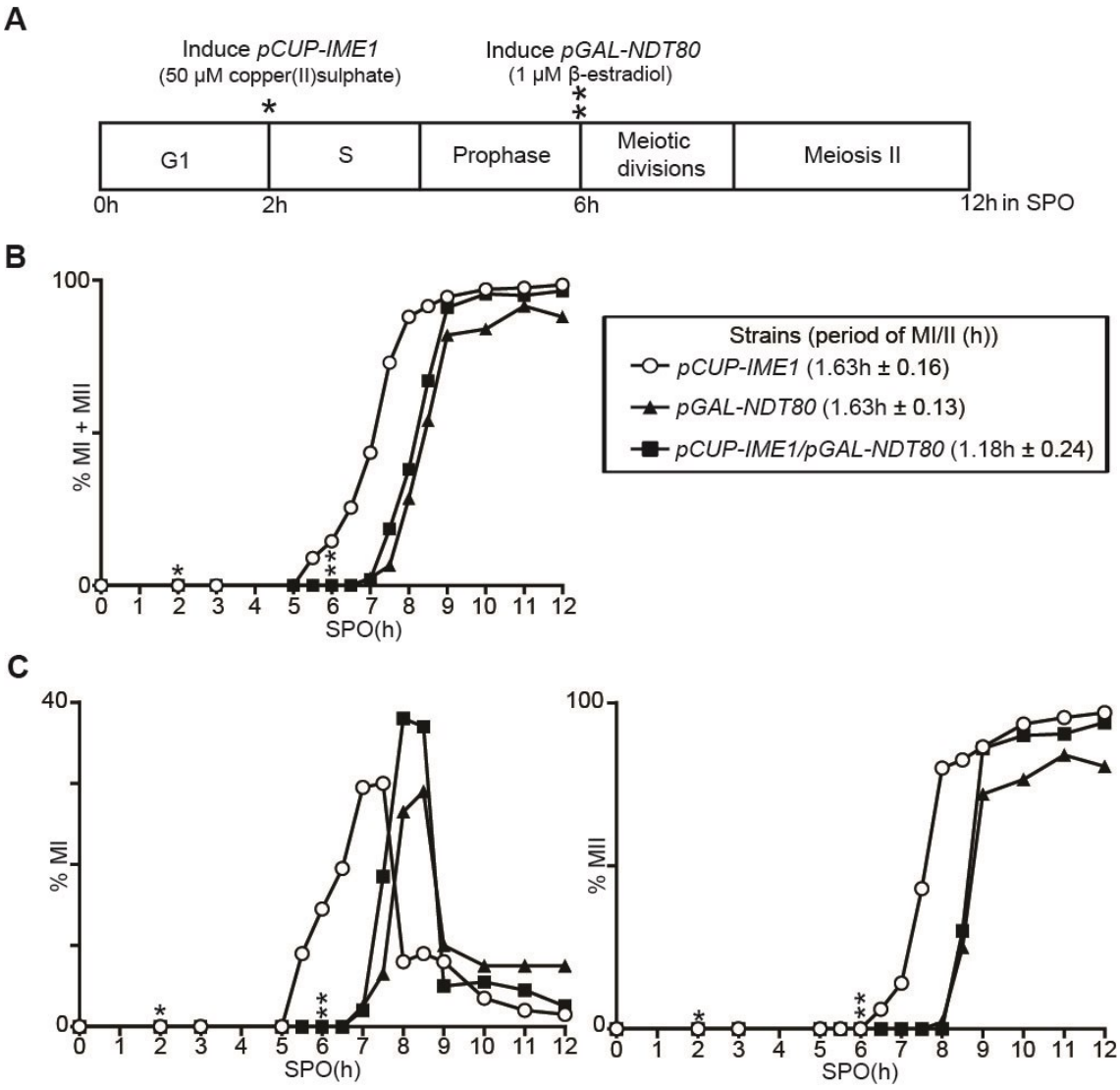


Figure 3.8 Combining the *pCUP1-IME1* and the *pGAL-NDT80* systems improves the synchrony of sporulation

(A) Scheme of experimental setup. The diploid cells harbouring *IME1* fused to the *CUP* promoter (FW2444), *NDT80* expressed from the *GAL* promoter together with Gal4 fused to the estrogen receptor (*GAL4.ER pGAL-NDT80*) (FW1541) or a strain expressing both (*pCUP-IME1* and *GAL4.ER pGAL-NDT80*) (FW2795) were grown in YPD overnight. Cells harbouring *GAL4.ER pGAL-NDT80* (FW1541) were transferred to pre-sporulation medium (BYTA). Subsequently, cells were pelleted by centrifugation, washed with sterile water and resuspended to a final OD_{600} of 2.5 in SPO; 50 µM $CuSO_4$ was added 2 hr after the cells were transferred to SPO, and 1 µM β-estradiol was added 6 hr after transfer to SPO. **(B)** Kinetics of meiotic divisions in strains, using procedures described in (A). Samples were taken at the indicated time points, fixed in ethanol, nuclei were stained with DAPI, and DAPI masses were counted. Cells that harboured two, three, or four DAPI masses were classified as cells undergoing meiosis I or meiosis II (% MI + MII). For each time point, at least 200 cells were counted. This panel is an example of one repeat out of three independent experiments, n = 3. I also computed the time or period taken

for 75% of the cells to complete meiotic divisions (see *Materials and Methods* for details). Values in brackets next to the legend, are the mean number of hours followed by the SEM of three independent experiments, n = 3. One-way ANOVA was used to find statistically significant differences between the group means (see main text for details). **(C)** Similar to (B) except that the percentages of bi- (left panel), or tri- and tetra-nucleate (right panel) cells (*n* = 200 cells) of strains described in (A) were determined. *, time of *IME1* induction; **, time of *NDT80* induction. This panel is an example of one repeat out of three independent experiments, n = 3.

3.4 Discussion

Here I demonstrate that temporal expression of a single gene, *IME1*, induced sporulation synchronously in budding yeast. My approach required neither pre-culturing in acetate containing BYTA medium nor expressing *IME4* from a heterologous promoter. The system described here can be easily combined with the *pGAL-NDT80* synchronization method and will be of use for studying specific stages of sporulation or the complete sporulation program.

My data indicate that synchronous sporulation requires a specified timing of *IME1* induction in *pCUP-IME1/pCUP-IME4* strains. Optimal synchrony was achieved when *IME1* was induced at 2 hrs in SPO. This is consistent with other studies reporting that the SK1 wildtype *IME1* promoter is switched on within 2 hrs of shifting to SPO (Inai et al. 2007;Nachman et al. 2007). The specific timing for inducing *IME1* in *pCUP-IME1/pCUP-IME4* strains suggests that other factors, such as the nutritional state of the cell, influence the kinetics of sporulation too. Although Nachman *et al*. report that nutrient signalling has little independent contribution to the variability of early gene expression, their study used wildtype strains in which the expression of *IME1* is already high in saturated cultures, prior to starvation in SPO. As such, the decision to initiate meiosis is coordinated with the metabolic state of the cells. In contrast, *IME1* expression is de-coupled from metabolic state in the *pCUP-IME1/pCUP-IME4* strains, which could inhibit sporulation kinetics if *IME1* is induced too early or too late in SPO. How might the metabolic state of the cell also control the timing of sporulation, after ectopic overexpression of *IME1*? Inducing *IME1* too late could still affect sporulation because cells which have been starved for prolonged times could have reduced rates of transcription and translation (Jona et al. 2000;Jambhekar and Amon 2008). On the other hand,

inducing *IME1* too early does not result in optimal sporulation because cells are not ready. In line with this hypothesis, in a recent report showed that a certain level of nutrient sensing, target of rapamycin complex (TORC1) activity is needed for sporulation (Weidberg et al. 2016). Too much or too little TORC1 activity affects sporulation negatively. Perhaps, TORC1 activity is most optimal for synchronous sporulation at two hours in SPO. Another explanation is that *IME1* or its early gene targets are not properly activated because the meiosis promoting kinases Rim11 and Rim15 are inhibited by residual PKA activity (Bowdish et al. 1994;Rubin-Bejerano et al. 1996;Vidan and Mitchell 1997;Pedruzzi et al. 2003;Sarkar et al. 2014). More work is needed to pinpoint why timing of *IME1* expression is critical for synchronous sporulation.

Existing sporulation protocols typically involve culturing the cells in pre-sporulation media containing nitrogen and acetate without glucose (BYTA), before shifting to SPO (Carlile and Amon 2008;Falk et al. 2010;van Werven et al. 2012). This is based on previous work showing that budding yeast sporulate optimally when allowed to adapt to conditions that promote respiration through the tricarboxylic acid cycle (Codon et al. 1995). Cells in early meiosis also increase intracellular glutamate for later protein synthesis and spore wall formation (Ray and Ye 2013;Walther et al. 2014). To test the efficacy of the copper inducible system in differing pre-sporulation culture conditions, I first cultured budding yeast cells overnight in either reduced glucose YPD or in BYTA. After inducing sporulation, I determined that spore viability and sporulation efficiency were similar in both tested conditions. Surprisingly, meiotic DNA replication was more synchronous in cells shifted from YPD to SPO compared to that from BYTA to SPO. The relative asynchrony of DNA replication in cells that are shifted from BYTA to SPO is consistent with previous work using similar strains (Berchowitz et al. 2013). Separately, I found that supplementing YPD with acetate or glutamate was not necessary for synchronous sporulation. Taken together, my protocol (reduced glucose YPD to SPO) allows for more synchronous sporulation than existing published methods. By dispensing with the BYTA step, my protocol is also shortened by one day. Furthermore, a greater density of cells can be harvested from my protocol for sporulation experiments. This is especially useful for later

experiments described in my thesis which require relatively large quantities of cellular material at different time points.

My work sheds light on how *IME1* and *IME4* regulate each other in wildtype cells. It has been shown that *ime4* mutant cells have diminished levels of *IME1* (Shah and Clancy 1992). In addition, one report showed that *IME1* transcripts contain the m6A modification suggesting that Ime4 controls *IME1* directly (Bodi et al. 2010). However, genome-wide sequencing of m6A did not identify the modification in *IME1* (Schwartz et al. 2013). In this study I demonstrate that *IME1* can also directly regulate *IME4* expression. My data showed that *IME4* levels increased when *IME1* was induced. I also found that the Ume6 repressor was bound to the promoter of *IME4*. Others have shown that *IME4* transcripts accumulate later in sporulation than *IME1,* which also supports the idea that *IME4* can be downstream of *IME1* (Primig et al. 2000;Nachman et al. 2007). I propose that *IME1* and *IME4* can positively regulate each other. The advantage of this positive feedback loop is that it allows for rapid accumulation of both transcripts when cells are ready to undergo sporulation. Using the *pCUP1-IME1* as opposed to the *pCUP1-IME1/pCUP1-IME4* strain also has the advantage of reducing the number of selection markers required for synchronous meioses. This facilitates genetic crosses of other mutants into the *pCUP1-IME1* background.

Several other approaches have been used to synchronize cells throughout or at specific stages of sporulation. First, the sporulation proficient SK1 strain background can undergo sporulation efficiently and with some degree of synchrony when specific growth conditions are adopted (Kane and Roth 1974;Falk et al. 2010;Borner and Cha 2015). I show that the *pCUP-IME1* strain reached a much better synchrony in comparison to wildtype SK1. Second, mutations that cause cells to arrest in specific stages of gametogenesis are also used to synchronize cell populations. For example, *ime2* mutants arrest prior to DNA replication, whereas *ndt80* mutants arrest in meiotic prophase I (Xu et al. 1995;Dirick et al. 1998). Although these approaches are useful for studying specific stages, they have several limitations. For example, only one stage per mutant can be studied, and not all stages can be arrested. Third, there are other “block and release” genetic approaches that reversibly arrest and then synchronize cell populations in certain

stages of sporulation. For example, the *pGAL-NDT80* system synchronizes meiotic divisions (see previous section) (Carlile and Amon 2008). Another example of stage specific synchronization is via the analogue sensitive allele of *CDC7* (*cdc7-as3*), which is used to arrest cells following meiotic S-phase and synchronizes cells through homologous recombination and meiosis I (Wan et al. 2006;Lo et al. 2008;Wan et al. 2008). However, these approaches only synchronize cells for a selective part of gametogenesis. My data shows that the *pCUP-IME1* system can achieve a high degree of synchrony during meiotic DNA replication and meiotic divisions. Fourth, some protocols utilize centrifugal elutriation to isolate starved G1 cells (Stuart 2017). However, these methods are time consuming and require specialized equipment for cell isolation. In contrast, I could easily obtain a population of mostly G1 cells by culturing them in reduced glucose YPD.

The *pCUP1-IME1* system can be used alone or combined with the *pGAL-NDT80* system as a tool to profile gene expression or protein production patterns throughout gametogenesis. The high degree of synchrony achieved by my method will be useful in dissecting the different stages in finer detail or to study temporal coordination and regulation of events during gametogenesis. In chapter 4, I use the *pCUP1-IME1* system to characterize the function of a long *NDC80* mRNA isoform, specifically in early meiosis. In chapters 5 and 6, I use the *pCUP1-IME1*/ *pGAL-NDT80* system to investigate the regulatory roles of ncRNAs at different stages of meiosis. By controlling the expression of these two transcription factors, I can identify and dissect the function of *IME1* or *NDT80* specific ncRNAs.

Chapter 4. Transcription of a 5' extended mRNA isoform directs dynamic chromatin changes and interference of a downstream promoter

This research has been published in two separate papers in eLife and has been modified for this results chapter (Chen et al. 2017;Chia et al. 2017). Chen J, Tresenrider A, Chia M, McSwiggen DT, Spedale G, Jorgensen V, Liao H, van Werven FJ, Unal E. 2017. Kinetochore inactivation by expression of a repressive mRNA. Elife 6:e27417; Chia M, Tresenrider A, Chen J, Spedale G, Jorgensen V, Unal E, van Werven FJ. 2017. Transcription of a 5' extended mRNA isoform directs dynamic chromatin changes and interference of a downstream promoter. Elife 6:e27420

4.1 Abstract

Cell differentiation programs require dynamic regulation of gene expression. During meiotic prophase in *Saccharomyces cerevisiae*, expression of the kinetochore complex subunit Ndc80 is downregulated but how this is achieved is not known. Here I demonstrate a transcriptional interference mechanism that is responsible for inhibiting expression of the coding *NDC80* mRNA isoform. Transcription of a 5' extended *NDC80* long undecoded transcript isoform (*NDC80*^{luti}) from a distal promoter is responsible for repressing the coding-competent *NDC80* isoform (*NDC80*^{ORF}) during meiotic S-phase and prophase. *NDC80*^{luti} transcription directs Set1-dependent histone H3K4 dimethylation and Set2-dependent H3K36 trimethylation to establish a repressive chromatin state in the downstream canonical *NDC80* promoter. As a consequence, *NDC80* expression is repressed during meiotic prophase. The transcriptional mechanism described here is rapidly reversible, adaptable to fine-tune gene expression, and relies on the Set2 histone methyltransferase and the Set3 histone deacetylase complex. Thus, expression of a 5' extended mRNA isoform causes transcriptional interference at the downstream promoter. I demonstrate that this is an effective mechanism to promote dynamic changes in gene expression during cell differentiation.

4.2 Introduction

The previous results chapter described optimized methods to synchronize sporulation in budding yeast. In this chapter, I employ these approaches to study how an essential gene for chromosome segregation *NDC80*, is dynamically regulated at specific stages of the meiotic program. Ndc80 is a component of the kinetochore, a structure which mediates attachment of spindle microtubules to the centromere(Ciferri et al. 2007). As such, Ndc80 is essential for chromosome segregation during nuclear divisions(Ciferri et al. 2007). Ndc80 is downregulated during meiotic prophase and is rapidly induced during meiotic chromosome segregation, at the onset of meiotic divisions (Miller et al. 2012;Meyer et al. 2015). In the presence of a spindle, mis-expression of Ndc80 during meiotic prophase causes aberrant separation of sister chromatids during meiosis I (Miller et al. 2012). Thus, the dynamic control of Ndc80 expression is critical for meiotic divisions. Surprisingly, RNA seq data from our lab and others show reads mapping upstream of the *NDC80* core promoter and its coding sequence during meiotic prophase, at a time when Ndc80 levels are supposed to be repressed. How Ndc80 is repressed during the early stages of meiosis remained elusive. Here, I present my findings about this repressive mechanism.

It is not known how *NDC80* transcription is regulated during early gametogenesis. During S phase and meiotic prophase, microarray, mRNA seq and ribosome profiling data suggest that different mRNA isoforms are expressed from numerous loci including the *NDC80* locus (Brar et al. 2012;Kim Guisbert et al. 2012;Lardenois et al. 2015). However, only a limited number of these ncRNAs have been assigned a biological function. For example, in cells with a single mating type locus (*MAT***a** or *MAT*α) transcription of ncRNAs represses *IME1* and *IME4*, two regulators of entry in meiosis (Hongay et al. 2006;van Werven et al. 2012). Throughout meiosis a subset of genes show stage specific expression of mRNA isoforms with often reduced translational capabilities (Brar et al. 2012). However, it is not well understood how ncRNAs and mRNA isoforms contribute to dynamic control of gene expression during yeast meiosis.

In this chapter, I show that the transcription of a 5' extended *NDC80* long undecoded transcript isoform (*NDC80*^{luti}) is responsible for repressing the coding-competent *NDC80* isoform (*NDC80*^{ORF}) during meiotic S-phase and prophase. Furthermore, *NDC80*^{luti} cannot be translated into Ndc80 protein due to translation of the upstream open reading frames in this mRNA isoform (Chen et al. 2017). Altogether, these results demonstrate that *NDC80*^{luti} functions in a regulatory manner, in which its transcription is both necessary and sufficient to downregulate *NDC80*^{ORF} levels during meiotic prophase.

In addition, I describe the mechanism by which the 5' extended *NDC80*^{luti} mRNA isoform represses *NDC80*^{ORF}. *NDC80*^{luti} transcription interferes with the downstream *NDC80*^{ORF} promoter by establishing a repressive chromatin state. This repression requires both Set3 and Set2. *NDC80*^{luti}-mediated repression can be rapidly reversed to suit the physiological needs of the cell. Furthermore, the repression mechanism described here can be adapted to fine-tune gene expression. Thus, transcription of a 5' extended mRNA isoform mediates transcriptional interference of the downstream promoter, allowing dynamic control of gene expression.

4.3 Results

4.3.1 A *NDC80* mRNA isoform with an extended 5' leader is expressed specifically during early gametogenesis

Genome wide transcriptome data from other studies indicated that there are at least two different mRNA isoforms expressed from the *NDC80* locus during early meiosis (Brar et al. 2012;Kim Guisbert et al. 2012;Lardenois et al. 2015). To verify this, I collected RNA from different time points in synchronously sporulating cells which were then sent for RNA sequencing **(Figure 4.1 A)**. In pre-meiotic cells when *IME1* is not induced or during meiotic divisions, an *NDC80* mRNA is produced from its core promoter. However, in early meiosis (S phase and prophase), mRNA reads map approximately 500 base pairs (bp) upstream of the *NDC80* start codon **(Figure 4.1 A)**. This could be due to either a short *NDC80*

promoter transcript or a longer *NDC80* mRNA isoform with an extended 5' untranslated region (5' UTR).

To investigate the expression pattern of the *NDC80* mRNA isoforms more closely in cells undergoing meiosis, I collected RNA at specific intervals throughout meiosis and performed northern blot analyses using a probe that recognizes both the ORF and upstream region of *NDC80*. In cells with the wildtype *IME1* and *IME4* alleles, both longer and shorter *NDC80* mRNA isoforms were expressed **(Figure 4.1 B**, left panel (control)**)**. After 3 hours in SPO, the expression of the short isoform (*) decreased, whereas the levels of the longer isoform increased (**). At 5 hours the short *NDC80* form was induced **(Figure 4.1 B**, left panel (control)**)**, which corresponds to cells entering meiotic divisions (Chen et al. 2017).

I also measured *NDC80* expression in a strain in which the *IME1* and *IME4* genes were fused to a copper inducible promoter (*pCUP-IME1* and *pCUP-IME4*). The induction of *pCUP-IME1* and *pCUP-IME4* ensures that cells can enter meiosis synchronously, as described in chapter 3 (Berchowitz et al. 2013;Chia and van Werven 2016). Only the short mitotic isoform was clearly detected in cells arrested before entry into meiosis **(Figure 4.1 B**, right panel (induced)**)**. Strikingly, soon after *IME1* and *IME4* were induced at two hours in SPO, expression of the long *NDC80* isoform increased and levels of the mitotic *NDC80* isoform decreased **(Figure 4.1 B**, right panel (induced)**)**. The mitotic *NDC80* isoform was repressed throughout meiotic S-and prophase. Mirroring changes in the levels of the mitotic *NDC80* transcript, Ndc80 protein levels also decreased during entry into meiosis, meiotic S-phase and prophase **(Figure 4.1 B)**. As expected, expression of the mitotic *NDC80* isoform and Ndc80 protein remained relatively constant when meiosis was not induced **(Figure 4.1 B**, middle panel (uninduced)**)**. I conclude that during meiotic entry, meiotic S phase and prophase, the expression of a longer *NDC80* form is induced and the mitotic form of *NDC80* is repressed. The distinct pattern of isoforms seen in *pCUP-IME1*/*pCUP-IME4* cells compared to the wildtype control underscores the importance of studying this regulatory mechanism in a highly synchronous population.

The observation that the expression of the mitotic and longer *NDC80* mRNA isoforms are inversely correlated during early meiosis, suggests that there may be a direct effect of the longer *NDC80* isoform on mitotic *NDC80* repression. Expression of the longer *NDC80* mRNA isoform is responsible for the decline in mitotic *NDC80* levels during early meiosis (Chen et al. 2017). Furthermore, *Chen et al.* showed that nine short upstream open reading frames in the extended 5' region of the long isoform inhibit translation of Ndc80 protein from this mRNA isoform (Chen et al. 2017). Thus, the long *NDC80* mRNA isoform is translationally inert. Hence, this transcript is defined as the *NDC80* long undecoded transcript isoform (*NDC80*^{luti}). The short *NDC80* protein coding mRNA isoform is called *NDC80*^{ORF}. This nomenclature is used thereafter.

Both *IME1* and *IME4* are necessary for efficient entry into sporulation (van Werven and Amon 2011). To examine if *NDC80*^{luti} transcription is dependent on *IME1* or *IME4*, I conducted meiotic time course experiments in cells with both the *pCUP-IME1* and *pCUP-IME4* alleles (control), cells with the *pCUP-IME1* allele in an *ime4Δ* background (*ime4Δ*) and cells with the *pCUP-IME4* allele in an *ime1Δ* background (*ime1Δ*) (**Figure 4.1 C)**. *IME1* and/or *IME4* expression was induced after two hours in SPO. In control cells, *NDC80*^{luti} was expressed within one hour of inducing *IME1* and *IME4* **(Figure 4.1 C,** left**)**. During meiotic divisions, *NDC80*^{ORF} mRNA and Ndc80 protein levels increased. In *ime4Δ* cells, induction of *NDC80*^{ORF} mRNA and Ndc80 protein was delayed, indicative of inefficient progression through meiosis **(Figure 4.1 C,** middle**)**. However, *NDC80*^{luti} was still expressed in a timely manner correlating with *IME1* induction. As expected, *ime1Δ* cells failed to enter meiosis and did not express *NDC80*^{luti} or downregulate *NDC80*^{ORF} **(Figure 4.1 C,** right**)**. This shows that *NDC80*^{luti} expression and correlated *NDC80*^{ORF} repression is controlled by Ime1 and is not merely a starvation induced signal.

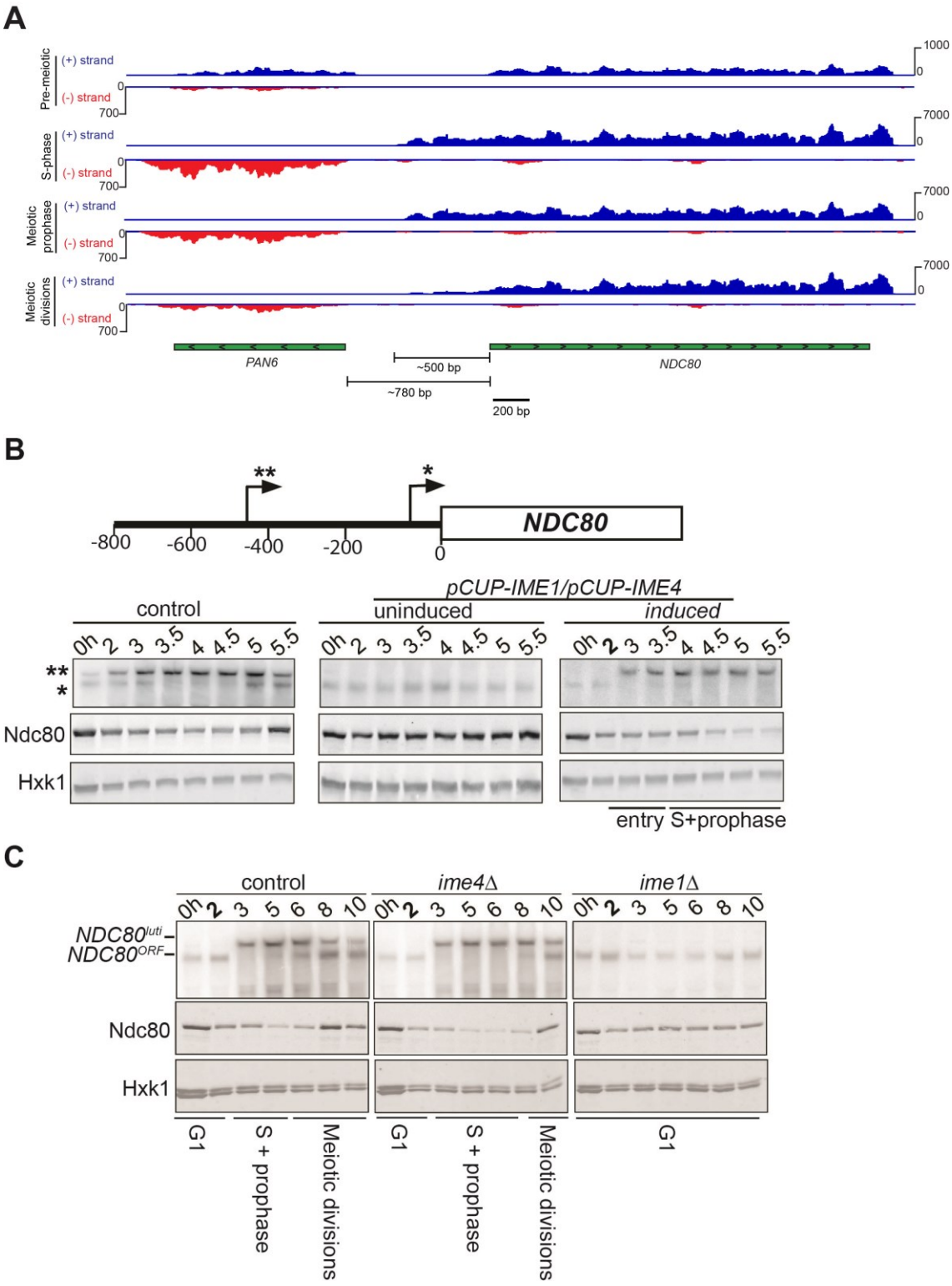


Figure 4.1 A longer *NDC80* mRNA transcript isoform is expressed during gametogenesis in an *IME1* dependent manner

(A) mRNA read counts (FPKM) mapping to the *NDC80* locus at different time points, as visualized on the Integrative Genomics Viewer (IGV). Diploid cells

harbouring the *pCUP-IME1* and *GAL4.ER pGAL-NDT80* alleles (FW2795) were grown overnight to saturation and subsequently transferred to sporulation medium (SPO). After two hours in SPO, *IME1* was induced with $CuSO_4$ (50 μM) to set in motion synchronous meiosis. RNA at different time points were extracted, purified and were used as inputs for mRNA library preparation. "Pre-meiotic" refers to cells in saturated culture, before *IME1* induction. S-phase is during Spo 4h. Meiotic prophase is during Spo 5h. Meiotic divisions is during Spo 8h. Transcription in the anti-sense (-) strand is also shown. For context, the neighbouring *PAN6* locus is also shown. Scales for values and distances (bp) are also given. Visualized reads are representative of three independent repeats, n = 3. (**B**) Expression pattern of two *NDC80* mRNA isoforms during starvation and early meiosis. Diploid control cells (FW4644) or cells harbouring the *CUP1* promoter fused with *IME1* and *IME4* (*pCUP-IME1*/*pCUP-IME4*) (FW1902) were grown overnight in rich medium, shifted to pre-sporulation medium, and subsequently transferred to sporulation medium (SPO). These cells also harboured *NDC80* tagged at the carboxy-terminus with three copies of the V5 epitope. After two hours in SPO (bolded time point), *IME1* and *IME4* were induced with $CuSO_4$ (50 μM) to set in motion synchronous meiosis in *pCUP-IME1*/*pCUP-IME4* cells. Samples were taken at the indicated time points for northern and western blot analyses. To detect the two different *NDC80* mRNA isoforms, RNA was extracted, separated by gel electrophoresis, blotted, and hybridized with a probe that spans the *NDC80* promoter and the 5' end of the coding region. Ndc80 protein levels were determined by western blot using anti V5 antibodies. As a loading control, I also detected Hxk1 levels with anti-Hxk1 antibodies. A schematic of the *NDC80* locus is shown over the northern blot. The single asterisk denotes the transcription start site of the short mitotic *NDC80* mRNA isoform. The double asterisk denotes the transcription start site of the distal 5' extended long *NDC80* mRNA isoform. The distance in base pairs from the *NDC80* start codon is also displayed. This experiment is representative of two independent repeats, n = 2. **(C)** Expression pattern of two *NDC80* mRNA isoforms in diploid control cells harbouring the *CUP1* promoter fused with *IME1* and *IME4* (*pCUP-IME1*/*pCUP-IME4*) (FW1902, left), in diploid cells with the *pCUP-IME1/ime4* genotype (FW3034, middle) or in diploid cells with the *ime1/ pCUP-IME4* genotype (FW3058). Cells were grown as described in B. After two hours in SPO (bolded time point), *IME1* and/or *IME4* were induced with $CuSO_4$ (50 μM) to set in motion synchronous meiosis. Samples were taken at the indicated time points for northern and western blot analyses. To detect the two different *NDC80* mRNA isoforms, RNA was extracted, separated by gel electrophoresis, blotted, and hybridized with a probe that spans the *NDC80* promoter and the 5' end of the coding region. Ndc80 protein levels were determined by western blot using anti-V5 antibodies. As a loading control, I also detected Hxk1 levels with anti-Hxk1 antibodies. The timing of the pre-meiotic (G1), S and prophase, and meiotic divisions are indicated. This experiment is representative of three independent repeats, n = 3.

4.3.2 *NDC80^{luti}* is degraded by nonsense mediated decay

The anti-correlation between the levels of the *NDC80^{luti}* transcript and the Ndc80 protein suggests that this isoform cannot be translated efficiently. Translational control of *NDC80^{luti}* is caused by the presence of 9 upstream open

reading frames (uORFs) in the 5' UTR of *NDC80*^{luti} (Chen et al. 2017). These uORFs are out of frame with the coding ORF of *NDC80* and thus contain premature stop codons (Chen et al. 2017). Transcripts containing uORFs like *NDC80*^{luti} are expected to be good substrates for nonsense mediated decay (NMD) in both yeast and mammalian cells (He et al. 2003;Hurt et al. 2013). To test this, I deleted *UPF1,* a helicase necessary for the NMD pathway. Diploid *pCUP-IME1/pCUP-IME4* cells with the *upf1Δ* mutation showed increased levels of *NDC80*^{luti} by both northern blots and qPCR, compared to control cells **(Figure 4.2 A and B)**. Therefore, *NDC80*^{luti} is degraded by the NMD pathway, consistent with the requirement for Upf1, and with this transcript being translationally inert.

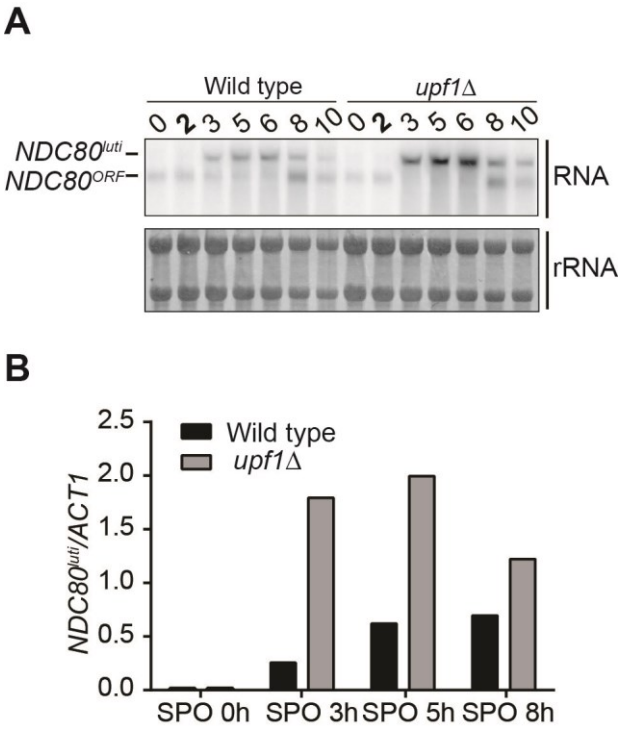


Figure 4.2 *NDC80*^{luti} levels are elevated in an *upf1Δ* mutant

(A) Control (FW1902) and *upf1Δ* (FW4114) cells harbouring *pCUP-IME1*/*pCUP-IME4* were grown in rich medium, transferred to pre-sporulation medium (BYTA), and then shifted to sporulation medium. After 2 hr (bolded time point), *IME1* and *IME4* expression were induced, and samples were taken at the indicated time points. Northern blot membranes were hybridized with a probe that spans the *NDC80* promoter and the 5' end of the coding region. For reference, rRNA bands after methylene blue staining are displayed. This result is representative of two independent repeats, n = 2. **(B)** *NDC80*^{luti} levels are elevated in the *upf1Δ* mutant, as measured by qPCR. Total RNA was isolated, reverse transcribed, and *NDC80*^{luti} mRNA levels were measured by qPCR. *NDC80*^{luti} levels were normalized to *ACT1* levels. This result is only from one repeat; samples from this experiment came from the same experiment in panel A.

4.3.3 Transcription of *NDC80^{luti}* represses transcription of the canonical *NDC80^{ORF}* mRNA

The mechanism by which *NDC80^{luti}* represses the downstream *NDC80^{ORF}* promoter might be related to a transcriptional interference mechanism during which intergenic transcription or transcription over promoter regions establishes a repressive chromatin state and prevents transcription factors from binding (Martens et al. 2004;Hainer et al. 2011;van Werven et al. 2012). To further investigate whether the mechanism of *NDC80^{luti}* -mediated gene repression also shares other features of transcriptional interference, I tested whether *NDC80^{luti}* transcription alters the association of transcription factors with the *NDC80^{ORF}* promoter. The binding of the basal transcription factor Sua7, which is homologous to human TFIIB, changed during meiosis across the *NDC80* locus **(Figure 4.3 A)**. Before entry into meiosis, Sua7 was bound to the core promoter of *NDC80^{ORF}*. However, after *IME1* and *IME4* induction (four hours in SPO) when *NDC80^{luti}* transcription occurred and cells underwent meiotic S phase, Sua7 binding to the *NDC80^{ORF}* core promoter (around −100 bp from AUG) was reduced, while binding to the *NDC80^{luti}* promoter (around −600 bp from AUG) increased **(Figure 4.3 A)**. Of note, the signal for Sua7 binding also showed a peak at −800 bp, which might be due to increased expression of the adjacent *PAN6* gene in the divergent direction **(Figure 4.1 A)**. I also examined Sua7 binding at the *NDC80^{ORF}* promoter in a mutant that does not transcribe *NDC80^{luti}* by deleting the region -300 to -600 bp upstream of *NDC80^{ORF}* (*ndc80$^{300\text{-}600}$*) (Chen et al. 2017). In the *ndc80$^{-300\text{-}600}$* mutant, no change in Sua7 binding around the *NDC80^{ORF}* promoter was observed even after timely induction of *IME1* and *IME4* **(Figure 4.3 B)**. Thus, *NDC80^{luti}* prevents TFIIB recruitment at the *NDC80^{ORF}* promoter during early meiosis. Importantly, this result suggests a functional role for *NDC80^{luti}* transcription in downregulating *NDC80^{ORF}*.

The reduction in TFIIB recruitment to the *NDC80^{ORF}* promoter could be due to the establishment of a repressive chromatin state. For example, transcription of an intergenic ncRNA across the *SER3* promoter directs nucleosome assembly in the promoter, which is essential for *SER3* repression in budding yeast (Hainer et al. 2011). Therefore, I examined how the chromatin structure in the *NDC80^{ORF}*

promoter is modified by *NDC80*^{luti} transcription. To identify where the nucleosomes stably associate with the *NDC80* locus, I performed chromatin immunoprecipitation (ChIP) of histone H3 on micrococcal nuclease (MNase) treated chromatin extracts **(Figure 4.3 C and D)**. In pre-meiotic cells (labeled 2 h), there was a relatively low signal around the core promoter of *NDC80*^{ORF}, which is indicative of a nucleosome free region (NFR) and consistent with active *NDC80*^{ORF} transcription. During meiotic prophase (labeled 4 hr), when *NDC80*^{luti} was transcribed, the signal around the core promoter increased, indicating that nucleosome occupancy was increased. These findings are consistent with the notion that transcription of *NDC80*^{luti} inhibits TFIIB recruitment and establishes a repressive chromatin state at the *NDC80*^{ORF} promoter.

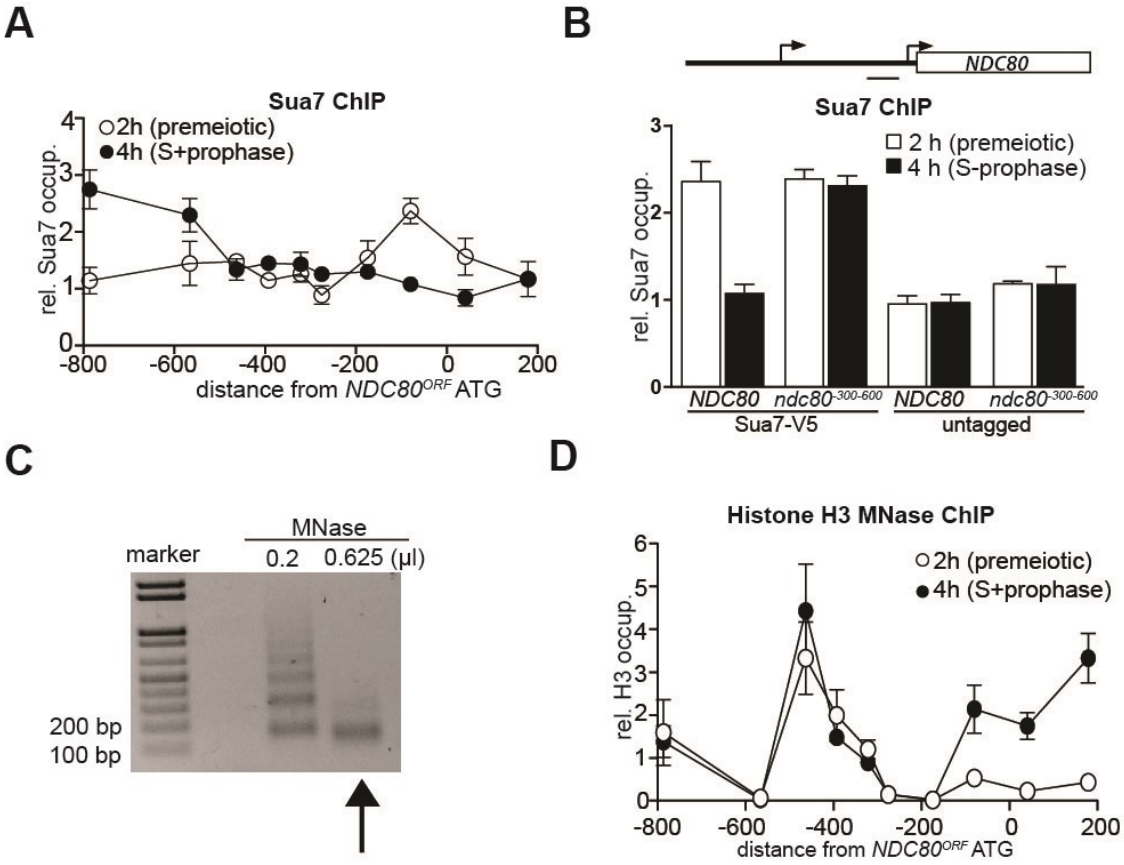


Figure 4.3 NDC80^{luti} **transcription represses the canonical *NDC80*^{ORF} promoter**

(A) *NDC80*^{luti} transcription correlates with reduced TFIIB (Sua7) binding at the *NDC80*^{ORF} promoter. Cells harbouring Sua7 tagged with three copies of V5 (Sua7-V5) and the *pCUP-IME1*/*pCUP-IME4* alleles (FW2957) were induced to undergo meiosis synchronously. Samples for chromatin immunoprecipitation were taken at two hours (2 hr (pre-meiotic), no *NDC80*^{luti} transcription) and four hours

after transfer to sporulation medium (SPO) (4 hr (S + prophase), *NDC80*^{luti} transcription). Cells were fixed with formaldehyde, chromatin extracts were prepared and Sua7-V5 bound DNA fragments were immunoprecipitated using agarose beads coupled with anti-V5 antibodies. The recovered DNA fragments were quantified by qPCR using ten different primer pairs scanning the *NDC80* locus, and were normalized over a primer pair directed against the *HMR* locus. The midpoint position of each primer pair is indicated in the x-axis. The y-axis shows relative histone H3 occupancy. The mean normalised signal for each primer pair and the SEM from three independent experiments is displayed, n = 3. **(B)** *NDC80*^{luti} transcription is required for inhibiting Sua7 binding in the *NDC80*^{ORF} promoter during meiotic prophase. A mutant strain harbouring a deletion upstream in the *NDC80*^{luti} promoter region (*ndc80*$^{-300-600}$, FW5530) and untagged strains (FW1902 and FW1868) were included in the analysis. A primer pair directed against the *NDC80*^{ORF}core promoter was used for the quantification of Sua7 binding at the *NDC80*^{ORF} promoter. The mean normalised signal and the SEM from three independent experiments is displayed, n = 3. **(C)** Example of extract with mononucleosomes prepared from cells (FW1902, 2 hr (pre-meiotic)). In short, cells were fixed with formaldehyde, treated with zymolase, and subsequently treated with different concentrations of micrococcal nuclease (MNase). To check for the extent of MNase digestion, part of the sample from each extract was reverse crosslinked, purified, and separated by gel electrophoresis. The arrow indicates the extract that was used for subsequent ChIP analysis. **(D)** *NDC80*^{luti} transcription correlates with the establishment of repressive chromatin in the promoter of *NDC80*^{ORF}.Chromatin structure at the *NDC80* locus was determined by ChIP of histone H3 on micrococcal nuclease (MNase) treated extracts in cells that also harboured *pCUP-IME1/pCUP-IME4* (FW1902). Samples were taken at two hours (2 hr (pre-meiotic), no *NDC80*^{luti} transcription) and four hours after transfer to SPO (4 hr (S + prophase), *NDC80*^{luti} transcription), fixed with formaldehyde, and chromatin extracts were treated with micrococcal nuclease. Extracts that predominantly contained mononucleosomes were used for ChIP assays with histone H3 antibodies (see *Materials and methods* for details). The recovered DNA fragments were quantified by qPCR using ten different primer pairs directed against the *NDC80* locus relative to a no MNase input. The signals from each primer pair were then normalized over a primer pair directed against the *PHO5* core promoter. The midpoint position of each primer pair is indicated in the x-axis. The mean normalised signal for each primer pair and the SEM from three independent experiments is displayed, n = 3.

4.3.4 Deletion or truncation of *NDC80*^{luti} leads to de-repression of *NDC80*^{ORF} during early meiosis

As shown in the previous section, the *ndc80*$^{-300-600}$ mutant showed no change in Sua7 binding around the *NDC80*^{ORF} promoter in meiotic prophase. This result means that deletion of *NDC80*^{luti} should lead to a de-repression of *NDC80*^{ORF} during early meiosis. To investigate this, I performed time course experiments of control cells with the wildtype *NDC80* allele and *ndc80*$^{-300-600}$ mutant cells, both in the *pCUP-IME1/pCUP-IME4* background **(Figure 4.4 A)**. As expected, the *ndc80*$^{-}$

$^{300\text{-}600}$ mutant cells failed to repress *NDC80*^{ORF} mRNA and Ndc80 protein during meiotic S and prophase. Similar to the control, mutant cells still upregulated expression of *NDC80*^{ORF} at the 6h time point during meiotic divisions. This is consistent with the finding that *NDC80*^{ORF} is induced by the mid-meiotic transcription factor Ndt80 (Chen et al. 2017). To exclude any possibility of artifacts caused by deleting a relatively large 300 bp region upstream of *NDC80*^{ORF}, I performed the same analysis on *ndc80*$^{-500\text{-}600}$ mutants, which carried a smaller 100 bp deletion in the *NDC80*^{luti} promoter **(Figure 4.4 B)**. *NDC80*^{ORF} is also de-repressed in the *ndc80*$^{-500\text{-}600}$ mutant during meiotic S and prophase.

As an alternative to promoter deletion, transcription of full length *NDC80*^{luti} can also be interrupted by introducing an *ADH1* terminator sequence in the luti transcript (Chen et al. 2017). Collaborators from the University of California, Berkeley generated strains with *ndc80Δ* at the endogenous locus, covered by ectopic integration of either a wildtype *NDC80* sequence or an allele with a terminator sequence inserted 220 bp downstream of the *NDC80*^{luti} transcriptional start site **(Figure 4.4 C)**. After inducing synchronous sporulation, Chen *et al*. showed that the *NDC80*$^{luti\text{-}Ter}$ strains de-repress *NDC80*^{ORF} and Ndc80 during early prophase. A shorter *NDC80*$^{luti\text{-}Ter}$ transcript was made, but unlike a full length luti transcript, was unable to repress *NDC80*^{ORF}. Altogether, data from both the deletion and the truncation mutants indicate that transcription of full length *NDC80*^{luti} is necessary for repressing *NDC80*^{ORF} transcription.

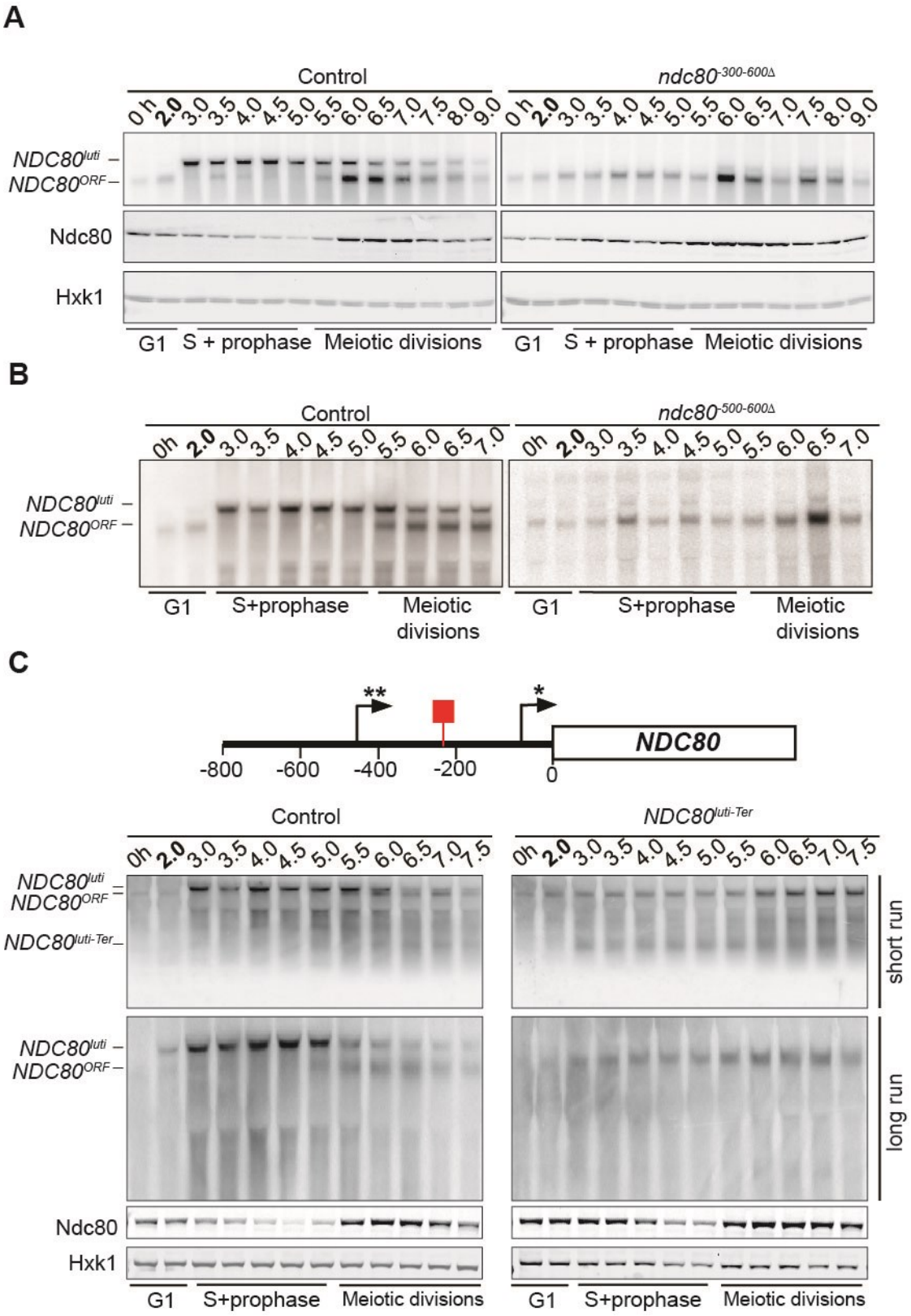


Figure 4.4 Loss of full length *NDC80^{luti}* transcription leads to upregulation of *NDC80^{ORF}* during prophase

(**A**) Diploid control cells (FW1902) and cells harbouring a deletion upstream in the *NDC80* promoter region (*ndc80*$^{-300-600}$, FW1861) were grown overnight in rich medium, shifted to pre-sporulation medium, and subsequently transferred to sporulation medium (SPO). These cells also harboured *NDC80* tagged at the carboxy-terminus with three copies of the V5 epitope. After two hours in SPO (bolded time point), *IME1* and *IME4* were induced with $CuSO_4$ (50 μM) to set in motion synchronous meiosis. Samples were taken at the indicated time points. RNA was extracted, separated by gel electrophoresis, blotted, and hybridized with a probe that spans the *NDC80* promoter and the 5' end of the coding region. Ndc80 protein levels were determined by western blot using anti V5 antibodies. As a loading control, I also detected Hxk1 levels with anti-Hxk1 antibodies. The timing of the pre-meiotic (G1), S and prophase, and meiotic divisions are indicated. This experiment is representative of two independent repeats, n = 2. While this experiment was conducted in the van Werven lab, this panel is reproduced from Chen *et al*. (Chen et al. 2017). (**B**) Diploid control cells (FW1902) and cells harbouring a deletion upstream in the *NDC80* promoter region (*ndc80*$^{-500-600}$, FW1871) were grown and induced to enter synchronous gametogenesis as described in A. Samples were taken at the indicated time points. RNA was extracted, separated by gel electrophoresis, blotted, and hybridized with a probe that spans the *NDC80* promoter and the 5' end of the coding region. The timing of the pre-meiotic (G1), S and prophase, and meiotic divisions are indicated. This experiment is representative of two independent repeats, n = 2. (**C**) Diploid control cells (UB6190) and *NDC80*$^{luti\text{-}Ter}$ cells which harbour a terminator sequence inserted 220 bp downstream of the *NDC80*^{luti} transcriptional start site (UB6077) were grown and induced to enter synchronous gametogenesis as described in A. Samples were taken at the indicated time points. RNA in blots labelled "short run" were electrophoresed for only 1.5 hr. RNA in blots labelled "long run" were electrophoresed for 3.0 hr. The *NDC80*^{luti} and *NDC80*^{ORF} isoforms could be sufficiently resolved only in the long run conditions, while the truncated *NDC80*^{luti} transcript (*NDC80*$^{luti\text{-}Ter}$) could only be detected in the short run conditions. Ndc80 protein levels were determined by western blot using anti V5 antibodies. As a loading control, they also detected Hxk1 levels with anti-Hxk1 antibodies. The timing of the pre-meiotic (G1), S and prophase, and meiotic divisions are indicated. This experiment is representative of two independent repeats, n = 2. The data presented in this panel is the work of collaborators from the University of California, Berkeley (Chen et al. 2017).

4.3.5 Transcriptional repression by *NDC80*^{luti} occurs in *cis* and not in *trans*

Gene regulation by transcriptional interference is expected to occur in *cis*, since it is largely the act of upstream transcription itself that represses a downstream promoter (Martens et al. 2004;van Werven et al. 2012). To test this, three heterozygous diploid strains were generated and subjected to synchronous spoulation. The control strain has one *NDC80-V5* allele and one wildtype *NDC80* allele **(Figure 4.5,** left**)**. The second strain has the *ndc80*$^{-300-600}$

deletion in *trans* with the *NDC80-3V5* allele **(Figure 4.5,** middle**)** while the third strain has the same deletion in *cis* with the *NDC80-3V5* allele **(Figure 4.5,** right**)**. Consequently, V5 epitope tagged Ndc80 protein levels during early meiosis were a proxy for locus specific *NDC80* regulation. $NDC80^{ORF}$ and Ndc80 were repressed during meiotic S and prophase in the control strain. Elevated levels of $NDC80^{ORF}$ were observed in both mutant strains due to the heterozygous $ndc80^{-300-600}$ deletion. Crucially, V5 tagged Ndc80 levels were only de-repressed in the mutant which has the deletion in *cis* with the *NDC80-3V5* allele **(Figure 4.5,** compare middle and right panels). This shows that transcriptional repression by $NDC80^{luti}$ occurs in *cis* and not in *trans*, typical of a transcriptional interference mechanism.

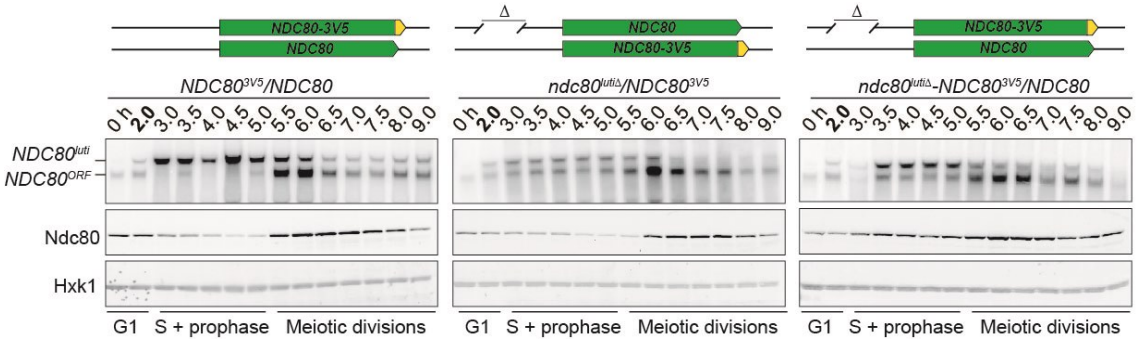


Figure 4.5 Transcriptional repression by $NDC80^{luti}$ occurs in *cis* and not in *trans*

Diploid control cells with one *NDC80-V5* allele and one wildtype *NDC80* allele (FW1900, left), cells heterozygous for the $ndc80^{-300-600}$ deletion in *trans* with a *NDC80-3V5* allele ((FW1923, middle) and cells heterozygous for the $ndc80^{-300-600}$ deletion in *cis* with a *NDC80-3V5* allele (FW1899, right) were grown overnight in rich medium, shifted to pre-sporulation medium, and subsequently transferred to sporulation medium (SPO). After two hours in SPO (bolded time point), *IME1* and *IME4* were induced with $CuSO_4$ (50 µM) to set in motion synchronous meiosis. Samples were taken at the indicated time points. RNA was extracted, separated by gel electrophoresis, blotted, and hybridized with a probe that spans the *NDC80* promoter and the 5' end of the coding region. Ndc80 protein levels were determined by western blot using anti V5 antibodies. As a loading control, I also detected Hxk1 levels with anti-Hxk1 antibodies. The timing of the pre-meiotic (G1), S and prophase, and meiotic divisions are indicated. This experiment is representative of two independent repeats, n = 2. While this experiment was conducted in the van Werven lab, this panel is adapted from Chen *et al*. (Chen et al. 2017).

4.3.6 Transcription of *NDC80^{luti}* promotes H3K4me2 and H3K36me3 deposition in the promoter and 5' region of *NDC80^{ORF}*

Co-transcriptional recruitment of chromatin modifying enzymes regulates the chromatin state of genes in the wake of elongating RNA polymerase II. For example, repressive chromatin marks, such as histone 3 lysine 4 dimethylation (H3K4me2) and lysine 36 trimethylation (H3K36me3) are deposited co-transcriptionally within gene bodies by the Set1 and Set2 methyltransferases (Hampsey and Reinberg 2003) (Kim and Buratowski 2009). The histone deacetylase complexes Set3C and Rpd3S recognize H3K4me2 and H3K36me3, respectively, and repress cryptic transcription from chromatin carrying these modifications (Carrozza et al. 2005b;Keogh et al. 2005;Kim and Buratowski 2009;Govind et al. 2010). Set1/Set3C and Set2/Rpd3S have also been implicated in transcription-coupled repression of gene promoters (Houseley et al. 2008;Kim et al. 2012;van Werven et al. 2012;Ard and Allshire 2016).

To investigate whether *NDC80^{luti}* mediated repression of *NDC80^{ORF}* also requires Set1/Set3C and Set2/Rpd3S, I measured the distribution of H3K4me2 and H3K36me3 marks at the *NDC80* locus **(Figure 4.6 A and B)**. I observed almost no enrichment in the *NDC80^{ORF}* promoter of either marks in pre-meiotic cells (labelled 2h), but H3K36me3 and to a lesser extent, H3K4me2 increased at the *NDC80^{ORF}* promoter in meiotic prophase cells (labeled 4h). As expected, the enrichment of H3K4me2 and H3K36me3 depended on Set1 and Set2, respectively **(Figure 4.6**, *set1*Δ and *set2*Δ**)**. I also normalised the ChIP data to the transcriptionally silent hidden mating type locus *HMR* and observed enrichment of H3K4me2 and H3K36me3 at the *NDC80^{ORF}* promoter during prophase. **(Figure 8.1 A and B)**. This suggests that H3K4me2 and H3K36me3 were enriched at the *NDC80^{ORF}* promoter during meiotic prophase.

In *ndc80$^{-300-600}$* cells that do not express *NDC80^{luti}*, the deposition of the H3K36me3 and H3K4me2 marks in the *NDC80^{ORF}* promoter was reduced **(Figure 4.6 C and D)**. Thus, *NDC80^{luti}* transcription promotes the deposition of repressive H3K4me2 and H3K36me3 marks within the *NDC80^{ORF}* promoter. Taken together,

deposition of repressive chromatin marks in the *NDC80*^{ORF} promoter requires *NDC80*^{luti} transcription.

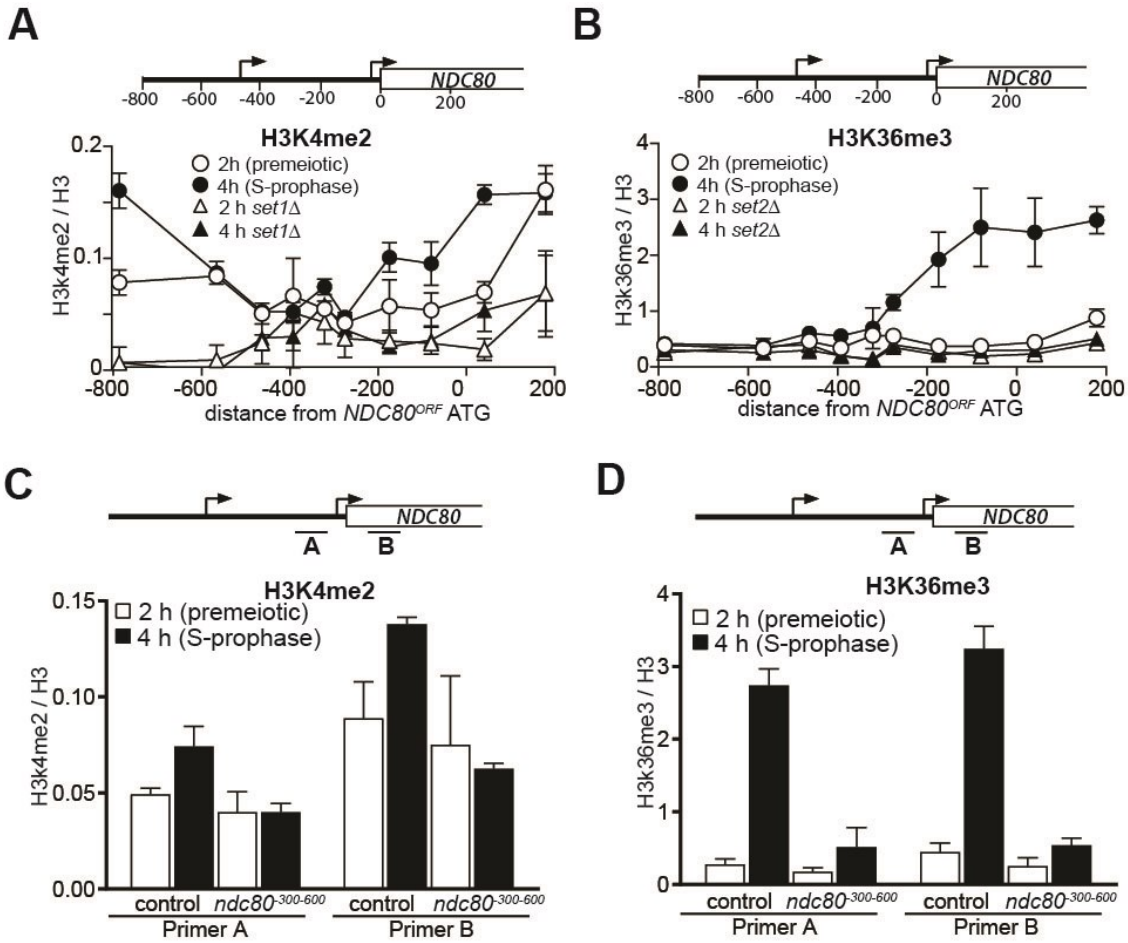


Figure 4.6 Transcription of *NDC80*^{luti} promotes H3K4me2 and H3K36me3 in the promoter and 5' region of *NDC80*^{ORF}

(A) Wildtype (FW1902) and *set1Δ* (FW3033) cells harbouring the *pCUP-IME1*/*pCUP-IME4* alleles were induced to undergo meiosis synchronously. Samples for chromatin immunoprecipitation were taken at two hours (2 hr (pre-meiotic), no *NDC80*^{luti} transcription) and four hours in sporulation medium (SPO) (4 hr (S + prophase), *NDC80*^{luti} transcription). Cells were fixed with formaldehyde, chromatin extracts were prepared and H3K4me2 or H3 enriched fragments were immunoprecipitated using magnetic Prot A beads coupled with anti-H3K4me2 or anti-H3 antibodies, respectively. The recovered DNA fragments were quantified by qPCR using ten different primer pairs scanning the *NDC80* locus. The midpoint position of each primer pair is indicated in the x-axis. The H3K4me2 signal was normalized over histone H3. The mean normalised signal and the SEM from three independent experiments is displayed, n = 3. **(B)** Similar to panel A, except that histone H3 lysine 36 trimethylation (H3K36me3) abundance was determined by ChIP. Wildtype (FW1902) and *set2Δ* (FW1472) cells harbouring the *pCUP-IME1*/*pCUP-IME4* alleles were used for the analysis. The mean normalised signal and the SEM from three independent experiments is displayed, n = 3. **(C)** Similar to panel A, except that the ChIP for H3K4me2 was performed in control cells

(FW1902) and cells harbouring the *ndc80*$^{-300\text{-}600}$ deletion which removes *NDC80*^{luti} transcription (FW1868). For the analyses we used primer pairs directed against the *NDC80*^{ORF} promoter "region A", and the 5' region of the *NDC80* gene "region B". The signals were normalized to the levels of H3. The mean normalised signal and the SEM from three independent experiments is displayed, n = 3. **(D)** Similar to panel C except that H3K36me3 levels were determined by ChIP. The mean normalised signal and the SEM from three independent experiments is displayed, n = 3.

4.3.7 Set2 and Set3 mediate *NDC80*^{luti} induced gene repression of *NDC80*^{ORF}

Having established that H3K36me3 and H3K4me2 marks localize to the *NDC80*^{ORF} promoter when *NDC80*^{luti} is transcribed, I next examined if the Set1/Set3C and Set2/Rpd3S pathways contributed to *NDC80*^{ORF} repression. Since it was not possible to measure *NDC80*^{ORF} levels by qPCR due to the presence of the overlapping *NDC80*^{luti} mRNA, I used northern blot data to determine the relative levels of both transcripts at different time points.

I first tested this quantification approach in both wildtype cells and *pCUP-IME1/pCUP-IME4* cells **(Figure 4.7)**. Measuring the levels of transcripts in northern blots required identifying a loading control whose expression remained constant throughout gametogenesis. Through RNA seq data, Chen *et al.* only found 4 candidate RNAPII genes that were abundantly expressed and whose levels differed by less than 4 fold across meiotic time courses (Chen et al. 2017). One promising candidate for loading controls was *CIT1*. However, we found that *CIT1* levels fluctuated considerably in early meiosis, making it unsuitable for normalizing *NDC80*^{ORF} levels **(Figure 4.7 A and B)**. As an alternative, we blotted for the RNAPIII gene *SCR1*, which has been used by other groups as a loading control for northern blots (Marquardt et al. 2014;Kim et al. 2016). *SCR1* levels remained relatively constant throughout sporulation and was used as a loading control for northern blot quantification in subsequent figures in this chapter **(Figure 4.7 B)**. After normalizing to *SCR1*, there was a clear increase in *NDC80*^{luti} levels and a decrease in *NDC80*^{ORF} levels during early meiosis in control or induced cells relative to uninduced cells **(Figure 4.7 C and D**, see 3.0 h to 5.5 h time points). Ndc80 protein levels clearly decreased in induced cells, compared to control or uninduced cells, mirroring the fall in *NDC80*^{ORF} levels **(Figure 4.7 C and E)**. This

quantification method was then used to estimate relative *NDC80*^{ORF} levels in *set3* and *set2* mutants.

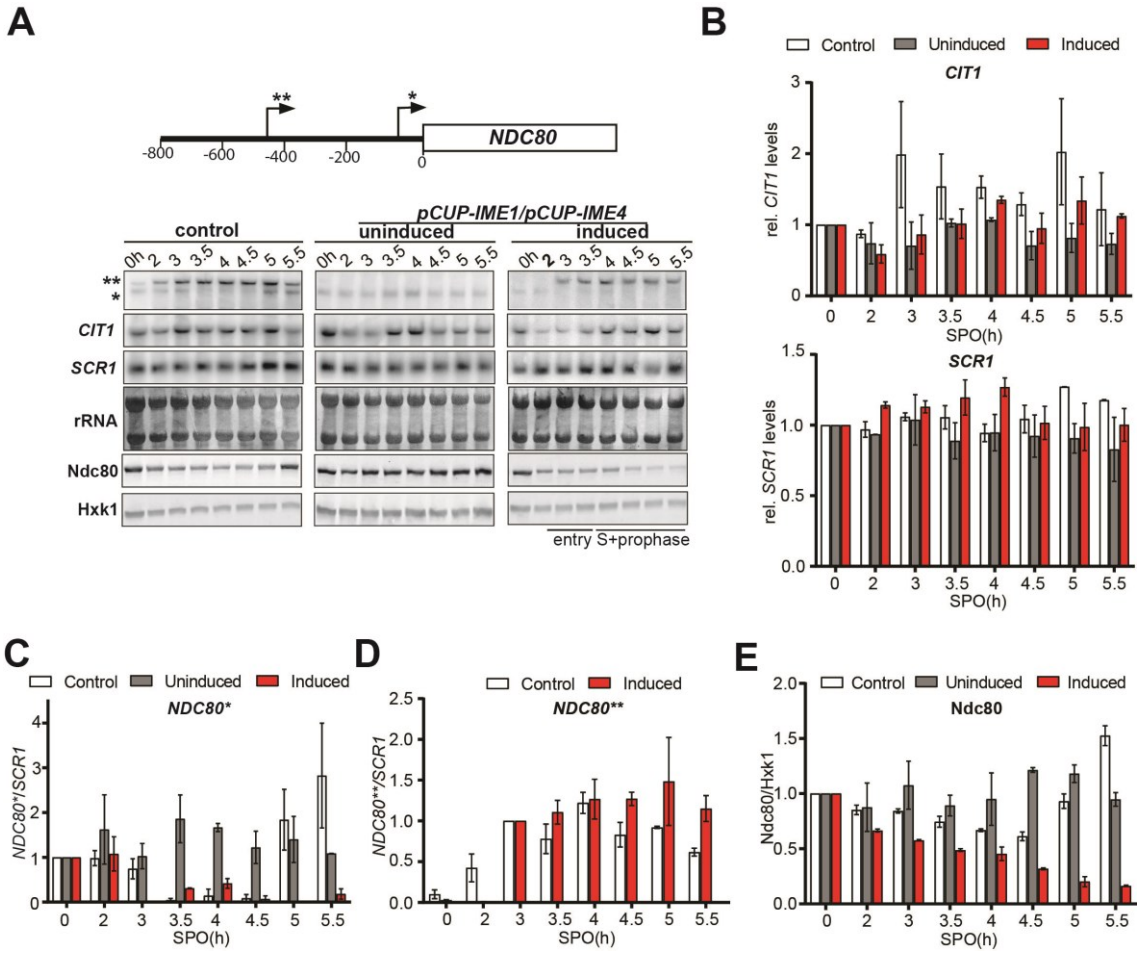


Figure 4.7 Example quantification of NDC80 mRNA isoforms and Ndc80 protein levels, normalized to *SCR1* and Hxk1 respectively

(A) The same northern blot and western blot of *NDC80* expression during a meiotic time course as described in **Figure 4.1 B**, except that *CIT1, SCR1* and ribosomal RNA expression levels are also displayed. Blot is representative of two independent repeats, n =2 **(B)** Quantification of *CIT1* and *SCR1* northern blot signals from A up till the 4.5 hr time point (see *materials and methods* for details). To control for technical variation between experiments and blots, the 0 hr time point was set to one. The mean is displayed and the bars represent the range. This panel represents two independent repeats, n = 2. **(C)** Quantification of expression of *NDC80*^{ORF} in panel A (labelled with one asterisk *). The signal was normalized over *SCR1*. The 0 hr time point was set to one to control for technical variation between experiments and blots. The mean is displayed and the bars represent the range. This panel represents two independent repeats, n = 2. **(D)** Quantification of expression of *NDC80*^{luti} in panel A (labelled with two asterisks **). The signal was normalized over *SCR1*. Since the long isoform is not expressed in pre-meiotic conditions, the 3 hr time point (as opposed to 0 hr) was set to one to control for technical variation between experiments and blots. The mean is displayed and the bars represent the range. This panel represents two independent repeats, n = 2.

(E) Quantification of Ndc80 protein levels in panel A. The signal was normalized over Hxk1. To control for technical variation between experiments and blots, the 0 hr time point was set to one. The mean is displayed and the bars represent the range. This panel represents two independent repeats, n = 2.

Next, I examined whether Set1/Set3C and Set2/Rpd3S contribute to *NDC80*^{ORF} repression. Since Set1 also plays an important role in meiotic recombination, *SET3* was deleted instead to test how the Set1/Set3C pathway regulates the *NDC80* locus (Borde et al. 2009;Acquaviva et al. 2013;Sommermeyer et al. 2013). All the strains used in this experiment also had the *pCUP-IME1/pCUP-IME4* alleles. In the *set2*Δ*set3*Δ double mutant, but not the single mutants, both *NDC80*^{luti} and *NDC80*^{ORF} transcripts were detected throughout multiple time points in early meiosis, and the steady-state level of Ndc80 protein remained high **(Figure 4.8 A and B**, compare the time points from two to five hours between control and mutant cells**)**. Importantly, the de-repression of *NDC80*^{ORF} in the *set2*Δ*set3*Δ double mutant was not due to reduced expression of *NDC80*^{luti} (**Figure 4.8 A and C)**.

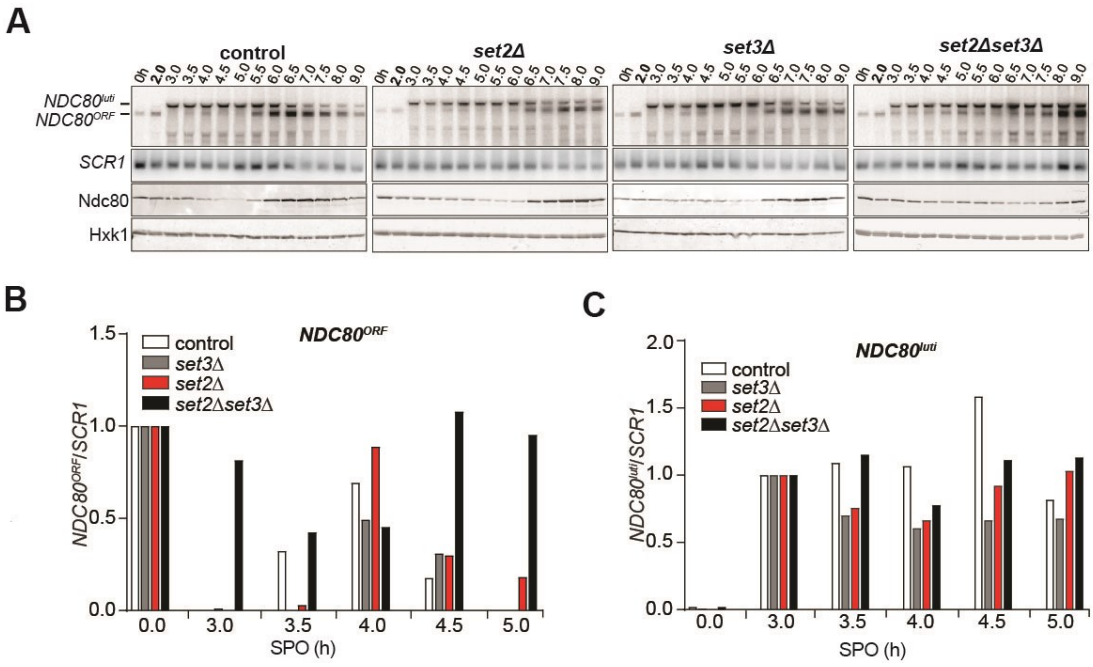


Figure 4.8 *NDC80*^{ORF} levels are de-repressed in the *set2Δset3Δ* mutant during meiotic prophase

(A) Control (FW1902), *set2Δ* (FW2929), *set3Δ* (FW2928) and *set2Δset3Δ* (FW1922) cells harbouring *pCUP-IME1*/*pCUP-IME4* and *NDC80-V5* were grown in rich medium, transferred to pre-sporulation medium, and then shifted to SPO medium. After 2 hr (bolded time point), *IME1* and *IME4* expression were induced, and samples for northern and western blot analyses were taken at the indicated time points. Northern blot membranes were hybridized with a probe

that spans the *NDC80* promoter and the 5' end of the coding region. As a loading control, membranes were also hybridized with a probe targeting *SCR1*. Ndc80 protein was detected with anti-V5 antibodies and Hxk1 levels were determined with anti-hexokinase antibodies. This result is representative of two independent repeats, n = 2. **(B)** Quantification of *NDC80^{ORF}* levels in the experiment shown in panel A. Signals were normalized to *SCR1.* To control for variation in overall signal between different northern blots, the *NDC80^{ORF}* signal at the 0 hr time point was set to one. **(C)** Quantification of *NDC80^{luti}* levels in the experiment shown in panel A. Signals were normalized to *SCR1.* To control for variation in overall signal between different northern blots, the *NDC80^{luti}* signal at the 3 hr time point was set to one.

An argument could be made that the de-repression of *NDC80^{ORF}* in the *set2*Δ*set3*Δ double mutant was due to poor sporulation synchrony. In other words, the co-existence of *NDC80^{luti}* and *NDC80^{ORF}* transcripts in early meiosis could be due to a sub-population of cells that never entered meiosis and continued to express the mitotic *NDC80^{ORF}* mRNA isoform. Indeed, the kinetics of meiotic divisions in the *set2*Δ*set3*Δ double mutant were much more delayed compared to the single mutants or to control cells, using BYTA to SPO culture conditions **(Figure 4.9 A)**. I improved the kinetics of meiosis by adopting the YPD to SPO synchronization protocol **(Figure 4.9 B**, also see chapter 3**)**. Instead of growing cells in BYTA, I shifted them directly to SPO after they reached saturation in nutrient rich conditions. I then induced *IME1* and *IME4*. This synchronization procedure reduced the delay in meiotic divisions in the double mutant **(Figure 4.9 A and B)**. In addition, meiotic S phase was completed in more than 75 percent of cells after 6 hours, indicating that the majority of cells had entered meiosis **(Figure 4.9 C)**. Importantly, *NDC80^{luti}* mediated repression was still compromised in *set2*Δ *set3*Δ double mutant cells despite improved sporulation synchrony **(Figure 4.9 D-F**, compare the time points from three-to five hours for the control with three-to six hours for the mutant cells**)**.

To minimize the effects of blot-to-blot technical variation on my estimates for *NDC80^{ORF}* levels, RNA from selective time-points in early meiosis (3.5 and 4.5 hours) were ran on the same gel and transferred onto the same membrane **(Figure 4.9 G)**. Subsequent quantification confirmed that there were significant differences in *NDC80^{ORF}* levels between the control and the *set2*Δ *set3*Δ double mutant, but not the single mutants **(Figure 4.9 H)**. As usual, *NDC80^{luti}* levels were comparable

between the control and all mutants **(Figure 4.9 I)**. Thus, transcriptional interference of *NDC80*^{ORF} requires both Set2 and Set3.

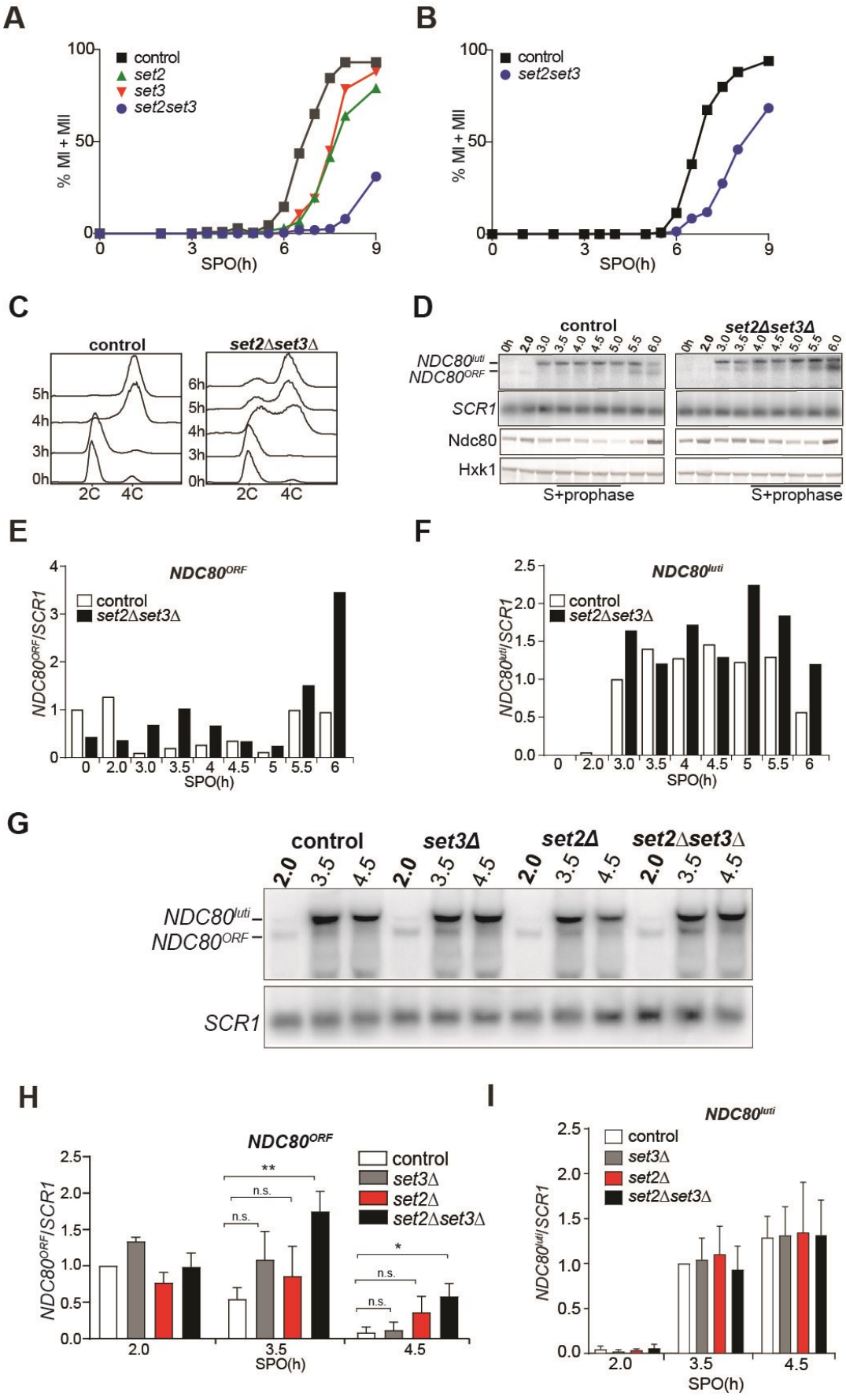

A
control
set2
set3
set2set3
% MI + MII
SPO(h)
B
control
set2set3
% MI + MII
SPO(h)
C
control
set2Δset3Δ
5h
4h
3h
0h
2C
4C
D
control
set2Δset3Δ
NDC80luti
NDC80ORF
SCR1
Ndc80
Hxk1
S+prophase
E
NDC80ORF
control
set2Δset3Δ
NDC80ORF/SCR1
SPO(h)
F
NDC80luti
control
set2Δset3Δ
NDC80luti/SCR1
SPO(h)
G
control
set3Δ
set2Δ
set2Δset3Δ
NDC80luti
NDC80ORF
SCR1
H
NDC80ORF
control
set3Δ
set2Δ
set2Δset3Δ
NDC80ORF/SCR1
n.s.
SPO(h)
I
NDC80luti
control
set3Δ
set2Δ
set2Δset3Δ
NDC80luti/SCR1
SPO(h)

Figure 4.9 *NDC80^{ORF}* levels are still de-repressed in the *set2Δset3Δ* mutant even in a more synchronous population of sporulating cells

(A) *set2Δ set3Δ* mutants undergo meiosis with delayed kinetics. Kinetics of meiotic divisions (MI + MII) in control (FW1902), *set2Δ* (FW2929)*, set3Δ (*FW2928) and *set2Δset3Δ* (FW1922) cells harbouring *pCUP-IME1*/*pCUP-IME4* and *NDC80-V5.* Cells were grown in rich medium, transferred to pre-sporulation medium (BYTA), and then shifted to sporulation medium. After 2 hr, *IME1* and *IME4* expression were induced, and samples were taken at the indicated time points, fixed, and stained with DAPI. The percentage of cells with one, two or more DAPI masses was determined for at least 200 cells per time point. Result is representative of three independent repeats, n = 3. **(B)** Similar to A, except that cells were grown in rich medium (YPD) to OD_{600} of 1–2, shifted to reduced glucose medium (YPD, 1% glucose), grown overnight to saturation, and then transferred to SPO. After 2 hr, *IME1* and *IME4* were induced. DAPI counting result is representative of two independent repeats, n = 2. **(C)** Flow cytometry analysis of DNA content in control (FW1902) and *set2Δset3Δ* (FW1922) strains. Synchronous meiosis was induced as described in panel B. Samples were taken at the indicated time points after transfer to SPO and were stained with propidium iodide. At least 50,000 cells were analysed per time point. Result is representative of three independent repeats, n = 3 (**D**) Control (FW1902) and *set2Δset3Δ* (FW1922) cells harbouring *pCUP-IME1*/*pCUP-IME4* and *NDC80-V5* were grown in conditions as described in panel B. After 2 hr (bolded time point), *IME1* and *IME4* expression were induced, and samples for northern and western blot analyses were taken at the indicated time points. Northern blot membranes were hybridized with a probe that spans the *NDC80* promoter and the 5' end of the coding region. As a loading control, membranes were also hybridized with a probe targetting *SCR1*. Ndc80 protein was detected with anti-V5 antibodies and Hxk1 levels were determined with anti-hexokinase antibodies. Result is representative of three independent repeats, n = 3 **(E and F)** Quantification of *NDC80^{ORF}* and *NDC80^{luti}* levels in the experiment shown in panel D. Signals are normalized to *SCR1.* The relative expression (*NDC80^{ORF}/SCR1*) with respect to the 0 hr time point is displayed in panel E. The relative expression (*NDC80^{luti}/SCR1*) with respect to the 3 hr time point is displayed in panel F. **(G)** Control (FW1902) and *set2Δset3Δ* (FW1922) strains were grown to undergo a synchronous meiosis as described in panel B (YPD to SPO), and selective time points were taken for northern blot analysis of *NDC80^{luti}* and *NDC80^{ORF}* transcripts on the same membrane. As a loading control, the northern membranes were hybridized with a probe targetting *SCR1*. **(H)** The *NDC80^{ORF}* levels in panel G were quantified and the mean and SEM from three independent experiments, n = 3 are displayed. One-tailed, unpaired t-tests were conducted to test if the differences in *NDC80^{ORF}* levels were statistically significant. A single asterisk * denotes *p*-value<0.05. A double asterisk ** denotes *p*-value<0.01. 'n.s.' means 'not significant'. To control for technical variation between different northern blots, the *NDC80^{ORF}* signal from the 2 hr time point from the control strain of each blot was set to one. **(I)** Same experiment as described in panels G and H, except that the *NDC80^{luti}* levels on northern blots were quantified, first normalized to *SCR1* and then normalized to the intensity of *NDC80^{luti}* at 3.5 hr in control cells. The mean and the SEM from three independent experiments are displayed, n = 3.

Previous work showed that *set2*Δ mutant exhibit increased nucleosome dynamics leading to de-repression of cryptic promoters (Venkatesh et al. 2012). In addition, *set3*Δ mutants display reduced histone H3 density in the 5' region of transcribed genes (Kim and Buratowski 2009). Set2 and Set3 are also required for transcription coupled chromatin changes in the *IME1* promoter by the long non-coding RNA *IRT1* (van Werven et al. 2012). These findings prompted us to examine whether Set2 and Set3 are necessary for *NDC80*^{luti} mediated nucleosome assembly in the *NDC80*^{ORF} promoter. Even though *NDC80*^{luti} was efficiently transcribed in *set2*Δ*set3*Δ double mutant cells during early meiosis **(Figure 4.8 C)**, repressive chromatin was not established at the *NDC80*^{ORF} promoter in the mutants **(Figure 4.10 A and B)**. Therefore, both Set2 and Set3 are necessary for transcriptional interference of *NDC80*^{ORF} by increasing nucleosome occupancy in its promoter.

Taken together, the data in this sub-section provide strong evidence that *NDC80*^{luti} transcription deposits repressive H3K4me2 and H3K36me3 marks in the *NDC80*^{ORF} core promoter. These two marks are localized to the same region and cooperate to repress *NDC80*^{ORF} transcription. *NDC80*^{ORF} is de-repressed in early meiosis only when both Set1/Set3C and Set2/Rpd3S are knocked out together. This means that *NDC80*^{luti} transcription alone is insufficient to establish repressive chromatin in the *NDC80*^{ORF} core promoter. Gene repression in this case, requires both luti transcription across the targeted promoter and the formation of repressive chromatin mediated by co-transcriptional histone modifications.

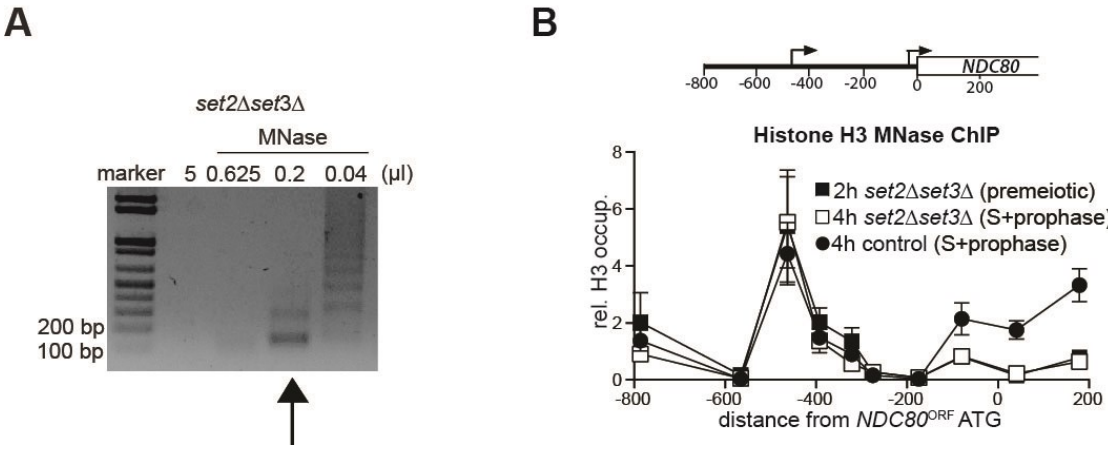


Figure 4.10 Set2 and Set3 mediate *NDC80^{luti}* induced repression of *NDC80^{ORF}* by establishing repressive chromatin

(A) Example of extract with mononucleosomes from *set2Δset3Δ* cells (FW1922, 2 hr, pre-meiotic) prepared as described in *Materials and methods*. In short, cells were fixed with formaldehyde, treated with zymolase, and subsequently treated with different concentrations of micrococcal nuclease (MNase). Part of the sample from each extract was reverse crosslinked, purified, and separated by gel electrophoresis. The arrow indicates the extract that was used for subsequent ChIP analysis. **(B)** *NDC80^{luti}* transcription requires Set2 and Set3 to establish a repressive chromatin state at the promoter of *NDC80^{ORF}.* Chromatin structure at the *NDC80* locus was determined by ChIP of histone H3 on micrococcal nuclease (MNase) treated extracts in control (FW1902) and *set2Δset3Δ* (FW1922) cells. Samples were taken prior to *IME1*/*IME4* induction at 2 hr in SPO (2 hr, pre-meiotic) and after induction at 4 hr in SPO (4 hr, S + prophase), fixed with formaldehyde, and mononucleosome fragments were isolated. The recovered DNA fragments were quantified by qPCR using ten different primer pairs directed against the *NDC80* locus relative to a no MNase input. The signals from each primer pair were then normalized over a primer pair directed against the *PHO5* core promoter. The midpoint position of each primer pair is indicated in the x-axis. The mean normalised signal for each primer pair and the SEM from three independent experiments is displayed, n = 3.

4.3.8 *NDC80^{luti}* mediated repression is reversible

Ndc80 is an essential kinetochore protein required for chromosome segregation in both mitosis and meiosis (Ciferri et al. 2007). Thus, Ndc80 levels must be restored when cells either transit from prophase to meiotic divisions or when cells re-enter the mitotic cycle before meiotic commitment. This transition necessitates that the repression by *NDC80^{luti}* transcription be rapidly reversible so that cells can resume cell divisions in a timely manner. Indeed, *NDC80^{ORF}* levels swiftly increase prior to meiotic divisions to facilitate chromosome segregation (Chen et al. 2017). To examine how gene repression by *NDC80^{luti}* transcription can adapt to changes in cell fate, we measured *NDC80^{luti}* and *NDC80^{ORF}* expression in meiotic prophase cells that re-entered the mitotic cell cycle **(Figure 4.11 B)**. Cells carrying the *ndt80*Δ mutation were arrested in meiotic prophase, and returned to growth by providing them with rich medium **(Figure 4.11 A)**. Strikingly, almost no *NDC80^{luti}* mRNA could be detected 15 min after return to growth (RTG) **(Figure 4.11 B and C)**. Concomitantly, *NDC80^{ORF}* was almost fully re-expressed at the same time and Ndc80 protein levels increased after 30 min **(Figure 4.11 C and D)**. Therefore, *NDC80^{luti}* mediated repression of *NDC80^{ORF}* is rapidly reversible,

allowing for adaptation to physiological needs and re-entry into the mitotic cell cycle.

We hypothesized that the dynamic changes of *NDC80*^{luti} mediated repression during RTG would be reflected in the chromatin state of the *NDC80*^{ORF} promoter. Previously, we had shown that the H3K36me3 ChIP signal was strongly enriched in the *NDC80*^{ORF} promoter when it was repressed by transcriptional interference **(Figure 4.6 B)**. This repressive histone modification is reported to be stable and it is unclear if this mark would persist after the upregulation of *NDC80*^{ORF} during RTG (Sein et al. 2015). We tested if the expression of *NDC80*^{ORF} during RTG is associated with the loss of the repressive H3K36me3 in its promoter. Indeed, we found that H3K36me3 at the 5' end of *NDC80*^{ORF} was strongly reduced within 15 min and almost completely lost within 30 min after cells returned to a nutrient rich environment **(Figure 4.11 E)**. The loss was specific to the *NDC80* locus because the levels at the 3' end of the *ACT1* gene increased slightly while bulk H3K36me3 levels did not change **(Figure 4.11 E and F)**. Thus *NDC80*^{luti}-mediated gene repression is reversible, allowing for rapid and dynamic changes in gene expression and chromatin state.

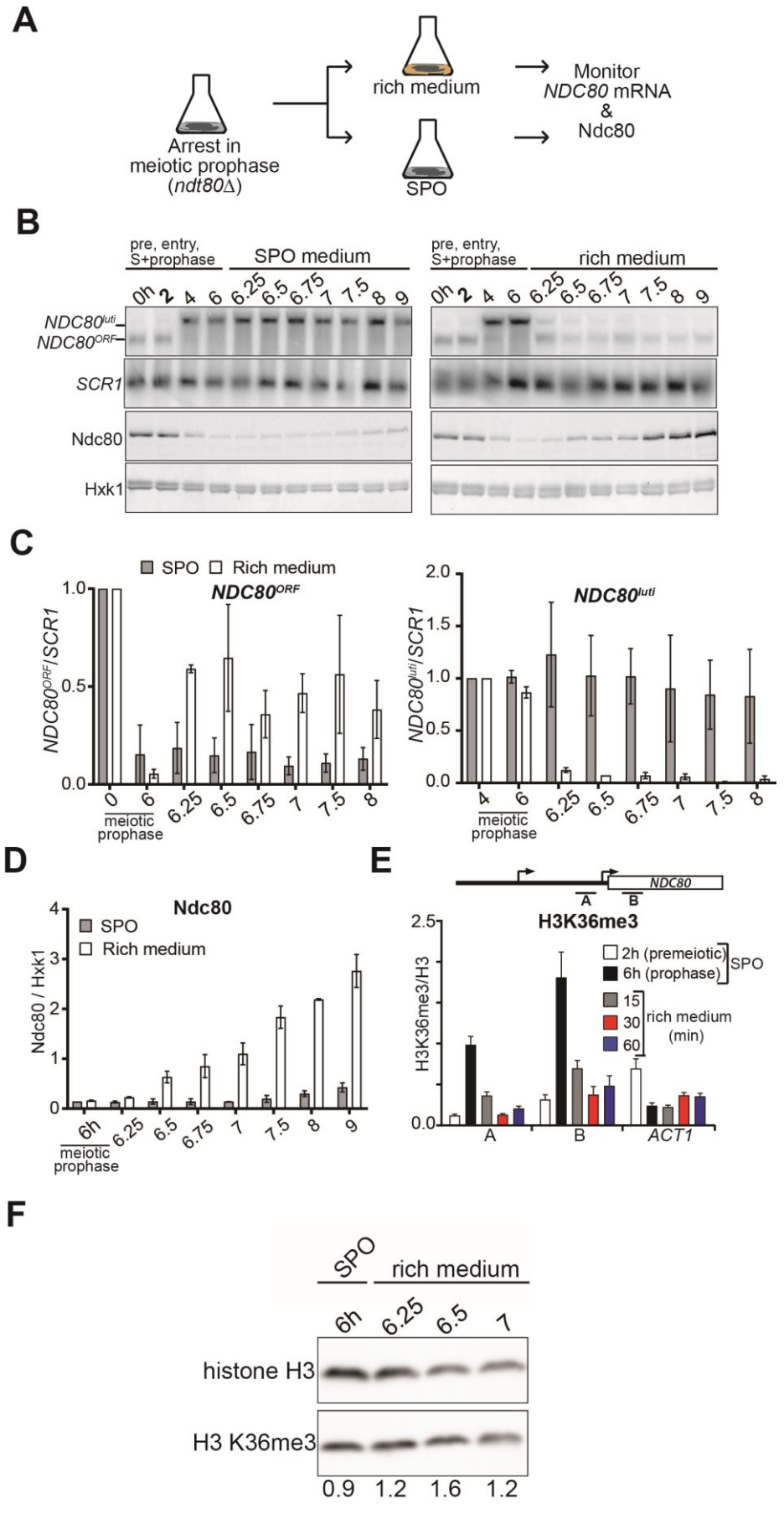

A
Arrest in
meiotic prophase
(ndt80Δ)
rich medium
SPO
Monitor
NDC80 mRNA
&
Ndc80
B
pre, entry,
S+prophase
SPO medium
rich medium
0h 2 4 6 6.25 6.5 6.75 7 7.5 8 9
NDC80luti
NDC80ORF
SCR1
Ndc80
Hxk1
C
SPO
Rich medium
NDC80ORF
NDC80luti
NDC80ORF/SCR1
NDC80luti/SCR1
meiotic prophase
D
Ndc80
Ndc80 / Hxk1
E
NDC80
A
B
H3K36me3
H3K36me3/H3
2h (premeiotic)
6h (prophase)
SPO
15
30
60
rich medium
(min)
ACT1
F
SPO
rich medium
6h 6.25 6.5 7
histone H3
H3 K36me3
0.9 1.2 1.6 1.2

Figure 4.11 *NDC80^{luti}* mediated repression is rapidly reversible

(A) Schematic overview of the experimental set-up for the return to growth experiments. **(B)** Cells repress *NDC80^{luti}* and induce *NDC80^{ORF}* when returned to a nutrient rich environment (YPD). Cells harbouring *ndt80Δ* and *pCUP-IME1/pCUP-IME4* (FW3856) were grown in rich medium, shifted and grown in pre-sporulation medium, and subsequently transferred to SPO. Samples were taken prior to induction of *IME1* and *IME4* at 0 hr and 2 hr (pre-meiotic [pre]), and after induction at 4 hr (S-phase) and 6 hr (prophase) in SPO. After 6 hr, cells were either transferred to rich medium or kept in SPO medium. Samples for RNA and protein were taken at the indicated time points. To detect *NDC80^{luti}* and *NDC80^{ORF}* expression, RNA was extracted, separated by gel electrophoresis, blotted, and hybridized with a probe that spans the *NDC80* promoter and coding region. As a loading control for northern blots, we also probed membranes for *SCR1*. Ndc80 protein levels were determined by western blot using anti-V5 antibodies. As a loading control we also detected Hxk1 levels with anti-Hxk1 antibodies. Blots are representative of two independent repeats, n = 2. **(C)** Quantification of *NDC80^{ORF}* and *NDC80^{luti}* levels from the northern blot in panel B. The signals were normalized over *SCR1*. The error bars represent the standard error of the mean from two independent experiments. To control for technical variation between experiments the 0 hr and 4 hr time points were set to one for *NDC80^{ORF}* and *NDC80^{luti}*, respectively. The mean is displayed and the bars represent the range. This panel represents two independent repeats, n = 2. **(D)** Quantification of Ndc80 protein levels during return to rich medium from the western blot in panel B. The Ndc80 protein levels were normalized to Hxk1 protein abundance. The relative levels with respect to the 6 hr time point are displayed. The mean is displayed and the bars represent the range. This panel represents two independent repeats, n = 2. **(E)** H3K36me3 is rapidly lost from the *NDC80^{ORF}* promoter and 5' region after return to growth. Growth conditions were similar to panel B and histone H3 lysine 36 trimethylation (H3K36me3) levels were quantified at the *NDC80^{ORF}* promoter during return to growth in a nutrient rich environment. Samples for chromatin immunoprecipitation were taken at the indicated time points. Cells were fixed with formaldehyde, chromatin extracts were prepared and H3K36me3 and histone H3 enriched fragments were immunoprecipitated with anti-H3K36me3 or anti-H3 antibodies, respectively. The recovered DNA fragments were quantified by qPCR using a primer directed against the *NDC80^{ORF}* promoter (region A) and a primer directed against the 5' region of the *NDC80* gene (region B). We also analyzed the signal at the 3' end of the *ACT1* open reading frame. The H3K36me3 signals were normalized to the histone H3 signal. The mean normalised signal for each primer pair and the SEM from three independent experiments is displayed, n = 3. **(F)** Bulk histone H3K36me3 levels remain constant during return to a nutrient rich environment. Cells harbouring *ndt80Δ* (FW4911) were grown in rich medium, shifted and grown in pre-sporulation medium, and subsequently transferred to SPO. After 6 hr, cells were transferred to rich medium. Samples were taken at the indicated time points. H3K36me3 and histone H3 levels were determined by western blot using anti-H3K36me3 and anti-histone H3 antibodies, respectively. The relative abundance of H3K36me3 normalized to histone H3 is indicated. This panel represents two independent repeats, n = 2.

4.3.9 Gene repression by *NDC80^{luti}* is tunable, and can bypass the requirement for Set2 and Set3 in luti mediated repression

Work from *Escherichia coli* showed that gene regulation by transcriptional interference is not binary with an on or off state, but can be utilized to fine-tune gene expression levels (Bordoy et al. 2016;Hao et al. 2016). Collaborators from UC Berkeley investigated whether transcriptional interference by *NDC80^{luti}* could also be tunable, allowing for incremental changes in *NDC80^{ORF}* levels (Chia et al. 2017). To scale the level of *NDC80^{luti}* expression, they used a tightly controlled, inducible system. The system utilizes a heterologous, chimeric transcriptional activator (LexA-ER-AD) whose activity is induced in a concentration-dependent manner by β-estradiol (Ottoz et al. 2014). Varying the number of LexA-binding sites (*lexO*) in the *NDC80^{luti}* promoter and titrating the concentration of β-estradiol, enabled scalable transcriptional induction of *NDC80^{luti}* **(Figure 4.12 A and B)**. Since *NDC80^{ORF}* is an essential gene, ectopic induction of *NDC80^{luti}* should cause lethality in actively dividing cells; this can be observed in a spot assay. The growth defect caused by ectopic *NDC80^{luti}* expression in mitosis was more severe with elevated concentrations of β-estradiol and higher number of *lexO* sites in the *NDC80^{luti}* promoter **(Figure 4.12 A)**. The higher the *luti* transcription, the greater the inhibition of *NDC80^{ORF}* expression. Thus, modulating *NDC80^{luti}* transcription levels enables tunable transcriptional repression of *NDC80^{ORF}* and Ndc80 in a population of cells.

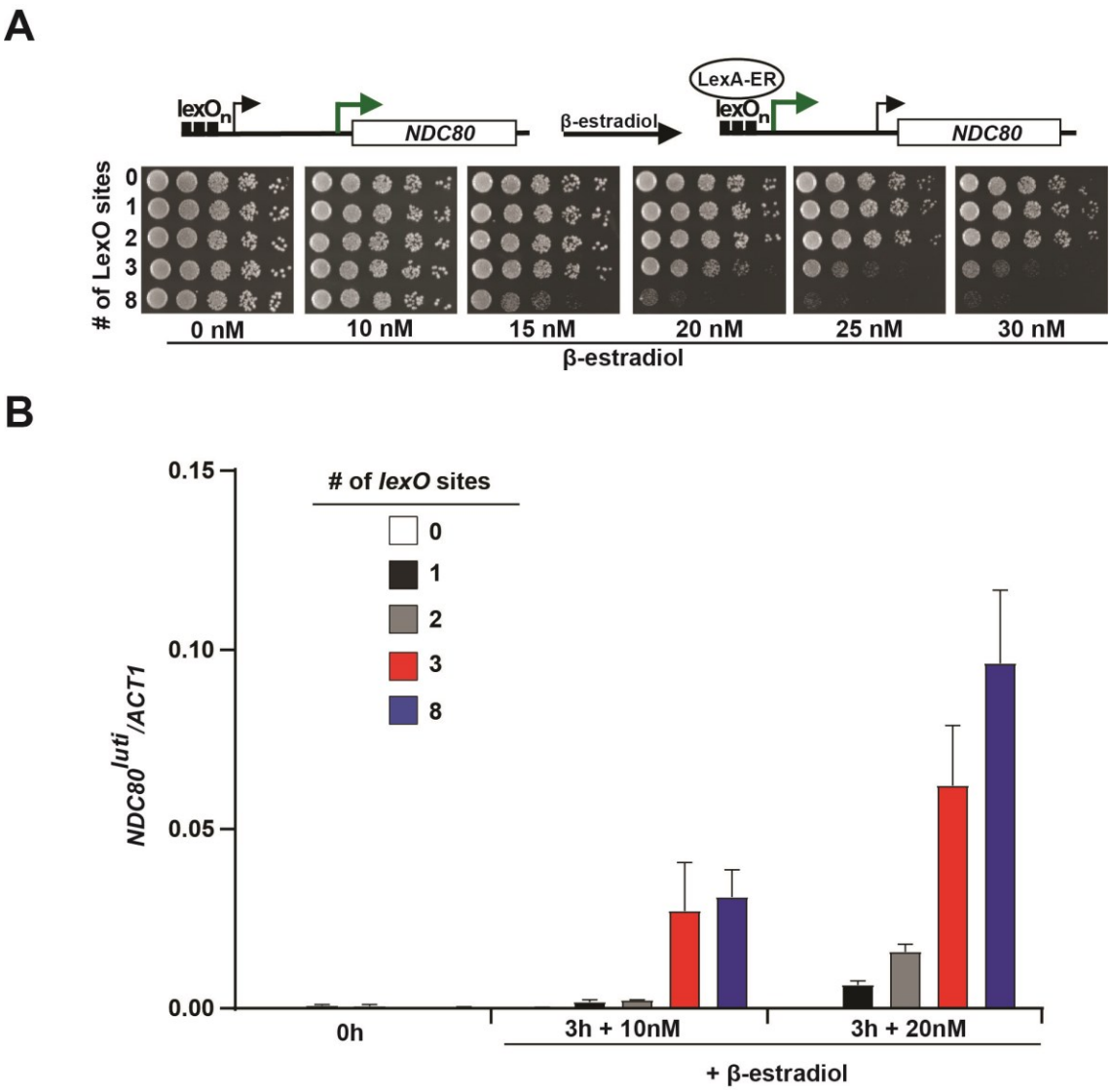


Figure 4.12 Scalable *NDC80*^{luti} induction using the LexA-ER-AD system

The experiments shown in this figure were conducted by collaborators at the University of California, Berkeley (Chia et al. 2017).
(A) Adjustable expression of *NDC80*^{luti} using the LexA-*lexO* system. Spot assay of cells harbouring 0, 1, 2, 3, or 8 *lexO* binding sites in the *NDC80*^{luti} promoter (UB8374, UB8358, UB8362, UB8366, UB8370) in the presence of different concentrations of β-estradiol. These cells also expressed LexA fused to an activation domain (AD) and the human estrogen receptor (ER) (LexA-ER-AD). Cells were grown overnight, diluted in sterile water, and spotted on YPD plates in the absence or presence of different concentrations of β-estradiol. Result is representative of three independent experiments, n = 3. **(B)** *NDC80*^{luti} levels in the presence of variable number of *lexO* sites in the *NDC80*^{luti} promoter. Cells harbouring 0, 1, 2, 3, or 8 *lexO* and LexA-ER-AD (UB8374, UB8358, UB8362, UB8366, and UB8370) were grown in YPD overnight. Subsequently, cells were diluted and exponentially growing cells were treated with 10 or 20 nM β-estradiol for 3 hr. RNA was extracted, reverse transcribed, and *NDC80*^{luti} mRNA levels were determined by quantitative PCR. Signals were normalized to *ACT1*. The mean and the SEM from three independent experiments are displayed, n = 3.

During transcription, nucleosomes are disassembled and reassembled by histone chaperones that associate with RNA polymerase (Venkatesh and Workman 2015). Therefore, higher levels of *NDC80*^{luti} transcription could lead to an increased rate of nucleosome deposition in the *NDC80*^{ORF} promoter and thus scalable *NDC80*^{ORF} repression. If so, then sufficiently high levels of *NDC80*^{luti} transcription should be sufficient for repressing *NDC80*^{ORF} without requiring Set1/Set3C and Set2/Rpd3S to maintain repressive chromatin. Collaborators from UC Berkeley tested this via spot assays again **(Figure 4.13 A)** (Chia et al. 2017). Cells with both pathways compromised (*set2*Δ*set3*Δ) and harbouring three or eight *lexO* sites did not show a growth defect when exposed to intermediate levels of β-estradiol (15 nM), whereas control cells did **(Figure 4.13 A)**. This result was expected because in the *set2*Δ*set3*Δ mutant background, *NDC80*^{luti} mediated repression is impaired **(Figure 4.9 G)**. Surprisingly at higher concentrations of β-estradiol (25 nM), *set2*Δ*set3*Δ mutant cells harbouring three *lexO* sites exhibited a moderate growth defect while cells with eight *lexO* sites exhibited a severe growth defect. To explain this, Ndc80 protein levels were measured in control and *set2*Δ*set3*Δ mutant cells harbouring 0 or 8 copies of *lexO* sites (Chia et al. 2017). The growth defects and rescue observed in Figure 4.13 A were reflected in the Ndc80 protein levels **(Figure 4.13 B and C)**. *set2*Δ*set3*Δ mutant cells with 0 *lexO* sites had wildtype levels of Ndc80 throughout. When exposed to intermediate levels of β-estradiol (15 nM), *set2*Δ*set3*Δ mutant cells with 8 *lexO* sites initially showed reduced Ndc80 at 2h and subsequent recovery at 4h **(Figure 4.13 B and C)**. In contrast, at high levels of β-estradiol (25 nM), the same cells had reduced Ndc80 at 2h and 4h **(Figure 4.13 B and C)**. These data suggest that high levels of *NDC80*^{luti} transcription could bypass the requirement for Set2 and Set3 in *NDC80*^{ORF} repression.

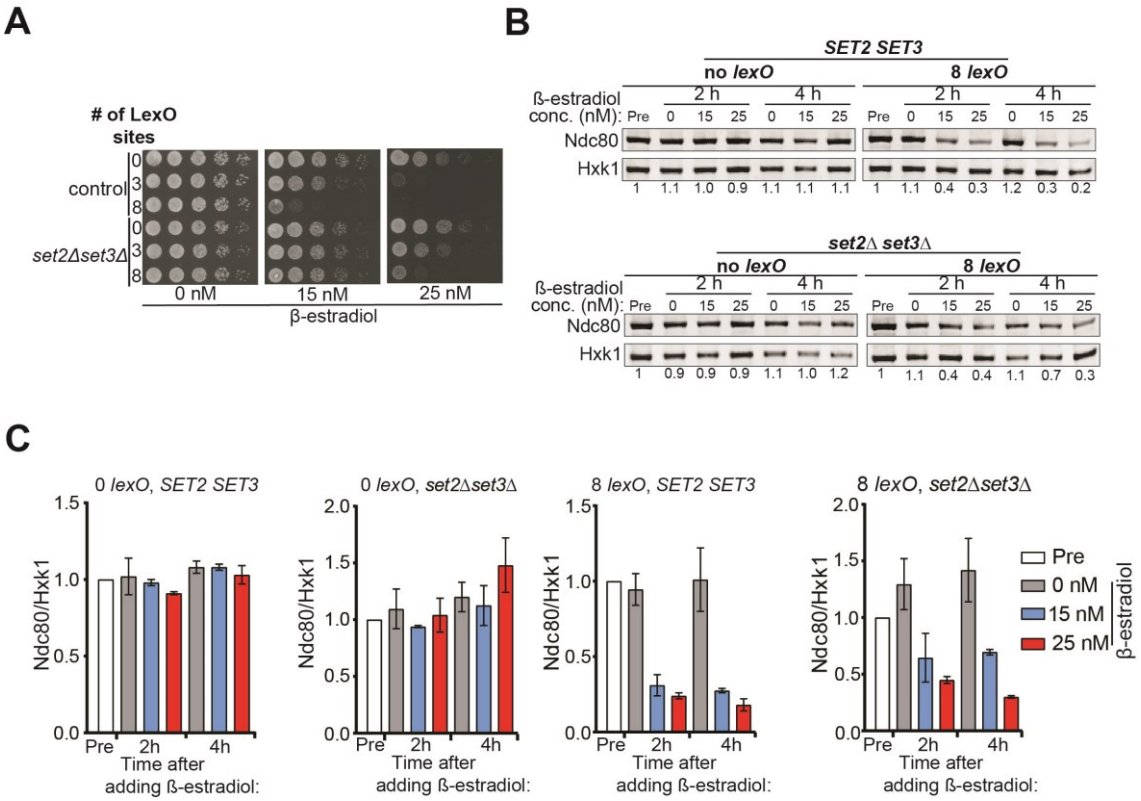


Figure 4.13 High levels of *NDC80^{luti}* transcription bypasses the requirement for Set2 and Set3 in *NDC80^{ORF}* repression.

The experiments shown in this figure were conducted by collaborators at the University of California, Berkeley (Chia et al. 2017).
(A) Spot assay of cells harbouring 0, 3 or 8 *lexO* binding sites in the *NDC80^{luti}* promoter in control cells (UB8374, UB8366, UB8370) or in *set2Δset3Δ* mutant cells (UB8691, UB8686 and UB8693). These cells also expressed LexA fused to an activation domain (AD) and the human estrogen receptor (ER) (LexA-ER-AD). Cells were grown overnight, diluted in sterile water, and spotted on YPD plates in the absence or presence of different concentrations of β-estradiol. Data is representative of two independent experiments, n =2. **(B)** Ndc80-V5 protein levels in *SET2 SET3* cells harbouring none (UB12945) or 8 *lexO* sites (UB12949), or *set2*Δ *set3*Δ cells with none (UB12947) or 8 *lexO* sites (UB12951). All four strains carry LexA-ER-AD. Ndc80 protein was detected by anti-V5 immunoblot. Hxk1 levels were used as a loading control. 'Pre' denotes pre-induction. Exponentially growing cells were treated with ethanol (0 nM), 15 nM, or 20 nM β-estradiol. Samples were taken at 2 hr or 4 hr after β-estradiol induction. Ndc80 levels were normalized to Hxk1 level, and the number under each lane shows the Ndc80/Hxk1 ratio normalized to that in the pre-induction condition. Data is representative of two independent experiments, n =2. **(C)** Same as panel B except that the mean is displayed and the bars represent the range. This panel represents two independent repeats, n = 2.

Ndt80 induces the expression of *NDC80^{ORF}* via the middle sporulation element (MSE) in the *NDC80^{ORF}* promoter, at the onset of meiotic divisions (Chen

et al. 2017). Ndt80 dependent induction of *NDC80*^{ORF} during mid gametogenesis occurs even under relatively high *NDC80*^{luti} expression **(Figure 4.4 A)**. This means that transcriptional interference of *NDC80*^{ORF} could be bypassed by a strong inducing signal. To examine this, collaborators from UC Berkeley increased the levels of *NDC80*^{ORF} by ectopically expressing the meiotic transcription factor Ndt80 in mitotic cells **(Figure 4.14 A)** (Chia et al. 2017). When the MSE site is present in the *NDC80*^{ORF} promoter, Ndt80 induction leads to increased *NDC80*^{ORF} expression, which suppressed the growth defect caused by *NDC80*^{luti} transcription. **(Figure 4.14 A**, compare MSE positive versus negative in the presence of *pGAL-NDT80* and *pGAL-NDC80*^{luti}**)**. Ndc80 protein levels were also measured in the strains described above (Chia et al. 2017). Both wildtype and *pGAL-NDT80* control cells showed moderate increases in Ndc80 over time because there was no ectopic *NDC80*^{luti} expression **(Figure 4.14 B and C**, see control and *pGAL-NDT80* strains**)**. When *NDC80*^{luti} was expressed, cells lacking either *NDT80* expression or the MSE site in *NDC80* had reduced Ndc80 levels due to uninterrupted transcriptional interference **(Figure 4.14 B and C**, see *pGAL-NDC80*^{luti}, and both *pGAL-mse-NDC80*^{luti} strains**)**. Crucially, when both *NDC80*^{luti} and *NDC80*^{ORF} were co-expressed, Ndc80 levels remained relatively constant, meaning that transcriptional interference was bypassed **(Figure 4.14 B and C**, see *pGAL-NDC80*^{luti}*/pGAL-NDT80* strain**)**. These results were consistent with the spot assay shown in **Figure 4.14 A**. Thus, increased Ndt80 driven transcription from the *NDC80*^{ORF} promoter can bypass *NDC80*^{luti} mediated repression. These results also show that the strength of the *NDC80*^{ORF} promoter influences the effectiveness of *NDC80*^{luti} mediated repression.

Taken together, the spot assays and western blots presented in this section demonstrate that transcriptional interference by expression of a 5' extended transcript can be tuned by adjusting the relative strengths of the distal and proximal promoters. Hence, this mechanism can be adapted as a regulatory module to generate a range of gene expression outputs.

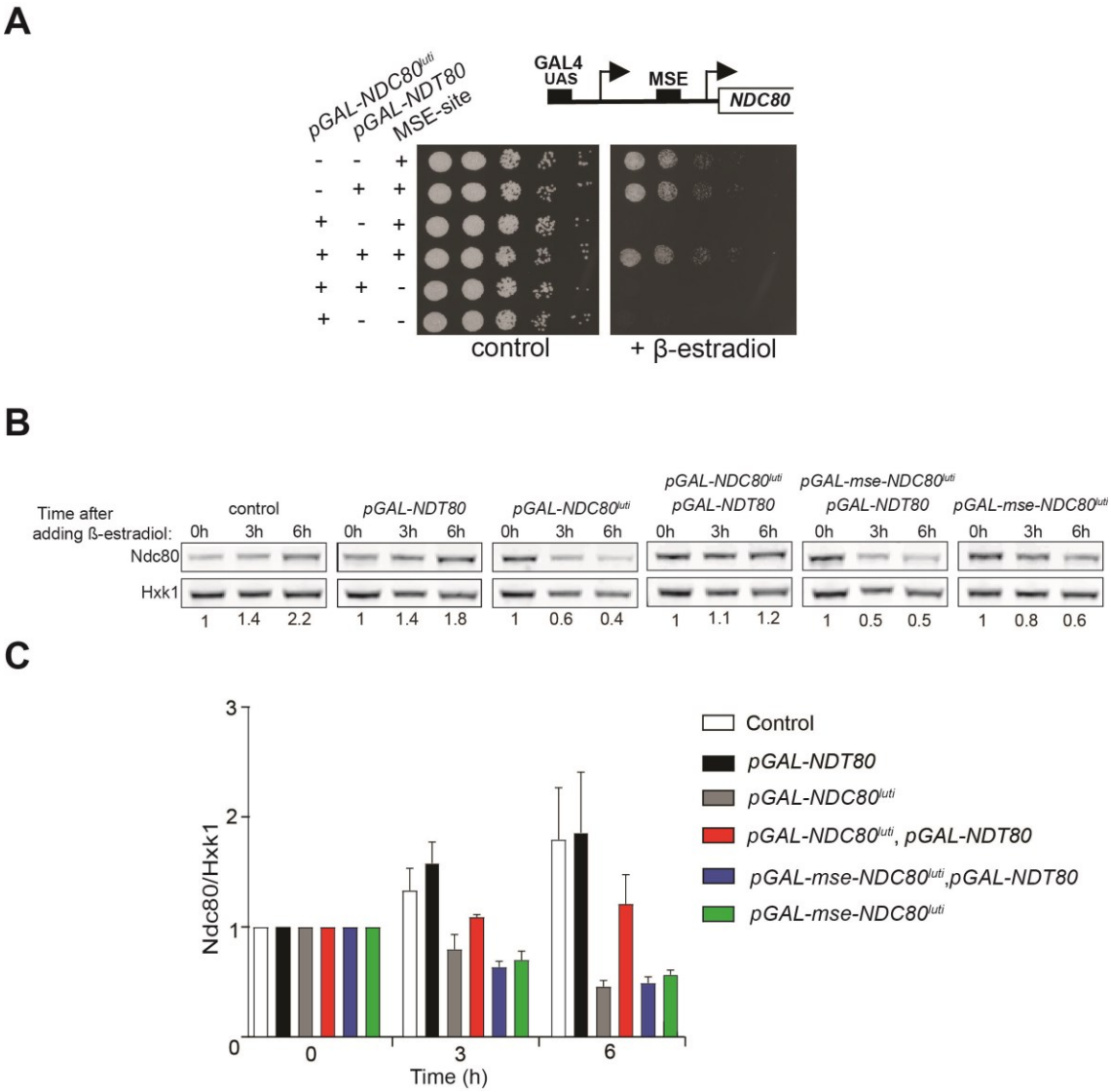


Figure 4.14 Increased *NDC80^{ORF}* promoter activity bypasses *NDC80^{luti}* mediated repression

The experiments shown in this figure were conducted by collaborators at the University of California, Berkeley (Chia et al. 2017).
(A) For this spot assay, three sets of strains were used: (1) Cells with a wildtype *NDC80* and with a functional MSE site (+MSE), but with either a wildtype *NDT80* (UB3351) or a *pGAL1-10* driven *NDT80* (*pGAL-NDT80*, UB3370); (2) cells with *pGAL-NDC80^{luti}* and with a functional MSE site, along with either a wildtype *NDT80* (UB5154) or *pGAL-NDT80* (UB9181); (3) cells with *pGAL-NDC80^{luti}* and a non-functional MSE site, along with either *pGAL-NDT80* (UB9921) or wildtype *NDT80* (UB9923). These cells also expressed Gal4-ER to activate *pGAL* driven expression. Cells were grown overnight, diluted in sterile water, and spotted on YP +raffinose +galactose plates in the absence or presence of β-estradiol (1 μM). This panel is representative of three independent experiments, n = 3. **(B)** For this western blot, the same strains described in panel A were used. Exponentially growing cells were treated with ethanol or 1 μM β-estradiol. Samples were taken at 3 hr or 6 hr after β-estradiol induction. The amount of samples loaded corresponded to similar OD units of cells collected from each culture. Ndc80 levels

were normalized to Hxk1 levels, and the number under each lane shows the Ndc80/Hxk1 ratio normalized to that in the pre-induction condition. This panel is representative of three independent experiments, n = 3. **(C)** Quantification of Ndc80 protein levels from panel B. The Ndc80 protein levels were normalized to Hxk1 protein abundance. The relative levels with respect to the 0 hr time point are displayed. The mean and the SEM from three independent experiments are displayed, n = 3.

4.4 Discussion

Eukaryotic cells have evolved diverse ways of regulating gene expression to drive changes in cell fate. In this chapter, I have done a detailed investigation of how transcription of the kinetochore gene *NDC80*, is regulated during budding yeast gametogenesis. During mitosis, cells express the shorter *NDC80*^{ORF} mRNA, which is essential for nuclear divisions. However during meiotic prophase, repression of Ndc80 is critical for setting up faithful chromosome segregation during meiosis I (Miller et al. 2012). Upon entry into gametogenesis, high levels of Ime1 signalling lead to transcription of the longer *NDC80*^{luti} mRNA isoform from an upstream promoter, which represses transcription of the shorter *NDC80*^{ORF} coding mRNA from the downstream promoter.

Levels of *NDC80*^{luti} are negatively correlated with Ndc80 protein levels. Why is *NDC80*^{luti} not translated into Ndc80, despite the long isoform containing the protein coding sequence? Chen *et al.* found 9 uORFs in the 5' leader sequence of *NDC80*^{luti} which are necessary and sufficient for translational repression of this isoform (Chen et al. 2017). Data from ribosome profiling suggest that these uORFs titrate ribosomes away from the canonical ORF, rendering *NDC80*^{luti} incompetent for productive translation (Brar et al. 2012;Cheng et al. 2018). Translational control of *NDC80*^{luti} ensures that protein levels are reduced when this isoform is expressed during early gametogenesis.

How does the cell toggle between transcription of the coding *NDC80*^{ORF} or the repressive *NDC80*^{luti} mRNA isoforms? The data is consistent with a model whereby *NDC80*^{luti} transcription interferes with the downstream *NDC80*^{ORF} promoter. First, during synchronous sporulation the expression of *NDC80*^{luti} and *NDC80*^{ORF} are anti-correlated by northern blots. In support of this data, Chen *et al* used fluorescent *in situ* hybridization (FISH) with one set of probes complementary

to the CDS in both isoforms, and another set of probes complementary to the 5' extended region of *NDC80*^{luti}, to concurrently measure levels of both isoforms in individual cells (Chen et al. 2017). These FISH experiments confirmed that levels of *NDC80*^{ORF} decline and levels of *NDC80*^{luti} increase in the same cell, during meiotic prophase (Chen et al. 2017). This also indicates that the negative correlation between these two isoforms is not due to a heterogenous population of cells expressing either *NDC80*^{ORF} or *NDC80*^{luti}. Second, mutants of *NDC80*^{luti} cannot repress *NDC80*^{ORF} (Chen et al. 2017). Transcription of truncated *NDC80*^{luti} (*NDC80*$^{luti\text{-}TER}$) fails to repress *NDC80*^{ORF}, thus excluding promoter competition as a mechanism for repression. Third, transcription of *NDC80*^{luti} prevents pre-initiation complex formation as measured by Sua7 ChIP. Fourth, *NDC80*^{luti} transcription is correlated with increased nucleosome occupancy at the *NDC80*^{ORF} core promoter and loss of the nucleosome free region (NFR). This indicates a repressive chromatin state since nucleosomes are barriers to transcription. Finally, *NDC80*^{luti} transcription deposits H3K34me2 and H3K36me3 marks at the *NDC80*^{ORF} core promoter; full repression of this promoter depends on Set1/Set3C and Set2/Rpd3S activity. This model is summarized in a schematic **(Figure 4.15)**.

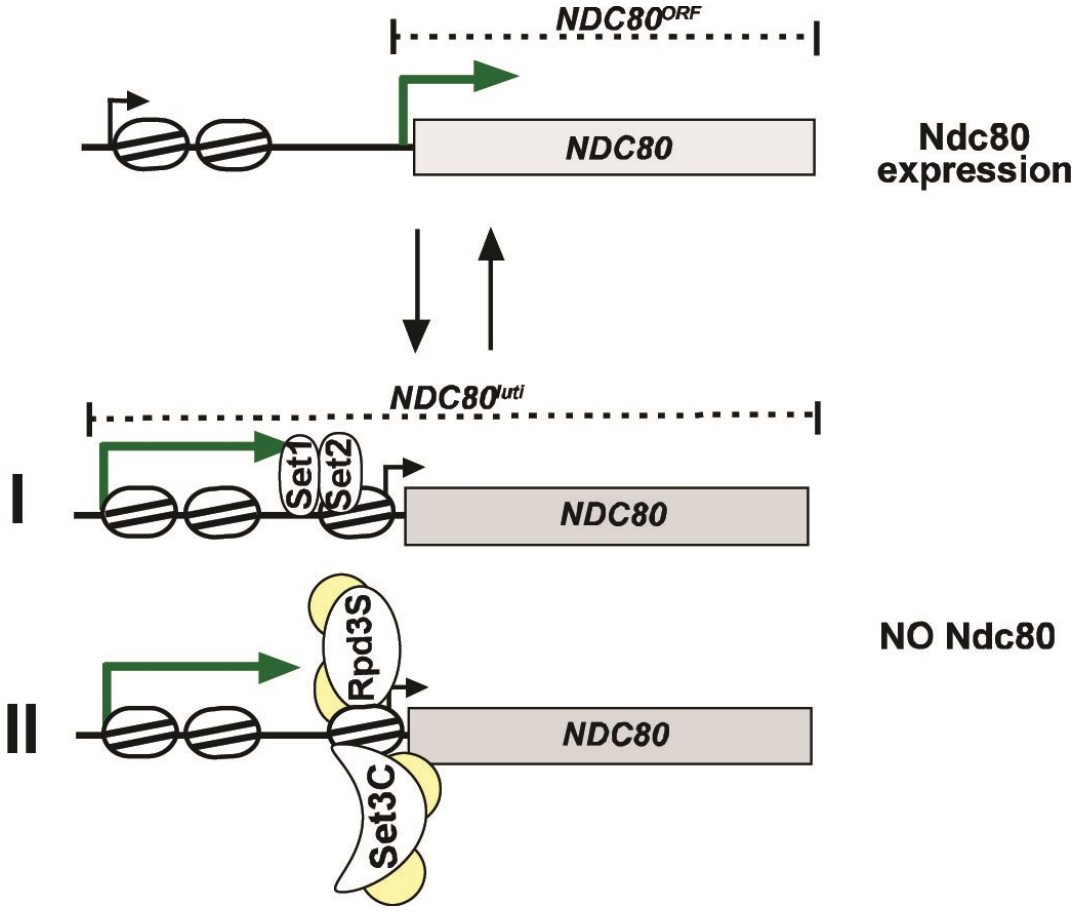


Figure 4.15 Model for *NDC80*^{luti} mediated repression of *NDC80*^{ORF}

During mitosis the *NDC80*^{ORF} promoter has an open chromatin structure and *NDC80*^{ORF} is transcribed. During meiotic S-phase and meiotic prophase, transcription of *NDC80*^{luti} deposits Set1-dependent H3K4me2 and Set2-dependent H3K36me3 in the same area of the *NDC80*^{ORF} promoter. Co-transcriptional deposition of these marks facilitates the recruitment of the histone deacetylase complexes Set3C and Rpd3S. This leads to increased nucleosome occupancy in the *NDC80*^{ORF} promoter and *NDC80*^{ORF} transcription is repressed. The set of reversing arrows indicates that *NDC80*^{luti} mediated repression of *NDC80*^{ORF} is reversible.

There are some possible critiques to my data and my proposed model of repression by changes in nucleosome occupancy in the tandem promoter region of the two *NDC80* isoforms. For instance, nucleosome occupancy at the upstream luti promoter (-500 region) did not change between pre-meiotic cells (no *NDC80*^{luti} transcription) and meiotic prophase (*NDC80*^{luti} transcription) **(Figure 4.3)**. Why are nucleosomes not lost upon *NDC80*^{luti} transcription? One possible explanation is that in pre-meiotic cells with no *NDC80*^{luti} transcription, the Ume6 repressor is bound just upstream of the *NDC80*^{luti} promoter. Indeed, analysis of the URS1

binding motif and ChIP data showed that Ume6 binds around the -550 region, about 50 bp upstream of the luti promoter (Chen et al. 2017). It is known that Ume6 represses promoters as part of a HDAC complex called Rpd3L, which stabilizes nucleosomes *in vitro* and *in vivo* (Chen et al. 2012). Thus in pre-meiotic cells, Ume6 and the Rpd3L HDAC could stabilize the nucleosome at the *NDC80*^{luti} promoter to repress it. However when *NDC80*^{luti} is transcribed during prophase, the RNAPII PIC together with chromatin remodellers like RSC could stabilize the positioning of the +1 nucleosome (Struhl and Segal 2013). The MNase H3 ChIP data in this work is not of sufficient resolution to distinguish between a repressive nucleosome (pre-meiotic cells) from the stable +1 nucleosome (prophase cells) at the *NDC80*^{luti} promoter. Another critique of my model is based on the observation that luti transcription does not correlate with increased nucleosome occupancy just further upstream (-200 to -300) of the *NDC80*^{ORF} promoter **(Figure 4.3)**. One likely explanation for this nucleosome depleted region is that it is bound by the Sum1 repressor in both pre-meiotic and prophase cells. Indeed, analysis of the MSE binding motif suggests that Sum1 binds around the -200 region (Chen et al. 2017). A follow-up experiment to test this would be to perform Sum1 ChIP at this region and monitor its association in pre-meiotic and prophase cells.

One limitation of my current data and model is that I have not provided direct evidence for the recruitment of HDACs by H3K4me2 or H3K36me3, nor have I measured levels of acetylated histones at the *NDC80* locus. Thus, important follow up experiments could include ChIP for acetylated histones, Set3 and Rpd3S subunits (e.g. Rco1) in control, *set1Δ* and *set2Δ* cells during gametogenesis. If my model is correct, then Set3 and Rco1 recruitment and a reduction in histone acetylation should be observed at the *NDC80*^{ORF} promoter during prophase. Conversely, the loss of H3K4 methylation in *set1Δ* mutants and the loss of H3K36 methylation in *set2Δ* mutants should reduce the recruitment of HDACs and interfere with de-acetylation at the *NDC80*^{ORF} promoter during prophase. These experiments are required to provide stronger evidence for my proposed model.

Another unresolved question is how regulation of histone acetylation or de-acetyation can modulate chromatin structure at the *NDC80*^{ORF} promoter. Histone de-acetylation is commonly but not always associated with gene repression.

Conversely, histone acetylation is usually but not always associated with gene activation. Acetylation of lysines on histone tails (e.g. histone H3 lysine 14 (H3K14ac) or histone H4 lysine 16 (H4K16ac) might recruit bromodomain containing chromatin remodelers such as the SWI/SNF complex or basal transcription factor TFIID to stimulate transcription (Hassan et al. 2001;Bannister and Kouzarides 2011;van Nuland et al. 2013). Histone tail lysine acetylation might also promote a more permissive chromatin structure for transcription by weakening inter-nucleosome contacts (Robinson et al. 2008). Likewise, acetylation of lysine 56 in the globular domain of histone H3 (H3K56ac) de-stabilizes chromatin and promotes transcription (Williams et al. 2008). In addition, H3K56ac might also increase histone turnover by promoting histone eviction and re-assembly, thus contributing to dynamic chromatin permissive for transcription (Kaplan et al. 2008). As a corollary, histone de-acetylation might repress gene expression and maintain repressive chromatin by reversing the abovementioned effects. However, multiple studies have also reported opposite effects of histone acetylation/de-acetylation on nucleosome dynamics, nucleosome occupancy and transcription. For instance, some HDACs such as Hos2 (part of Set3C) are required for activating *GAL* genes in budding yeast, and overlapping anti-sense transcription is associated with increased nucleosome occupancy together with increased histone acetylation in the sense promoter (Wang et al. 2002;Murray et al. 2015). Furthermore, another recent study found no link between different histone modifications and histone turnover (Ferrari and Strubin 2015). How does regulation of histone acetylation or de-acetylation modulate chromatin structure at the $NDC80^{ORF}$ promoter? If my model is correct, a *rco1Δ set3Δ* mutant should show increased acetylated histone ChIP signals in the $NDC80^{ORF}$ promoter, even when $NDC80^{luti}$ is transcribed. The *rco1Δ set3Δ* mutant should phenocopy the *set2Δ set3Δ* mutant by de-repressing $NDC80^{ORF}$ in meiotic prophase, as measured by northern and western blotting. Likewise, my model predicts reduced nucleosome occupancy at the $NDC80^{ORF}$ promoter during prophase in the *rco1Δ set3Δ* mutant; this can be tested by a H3 MNase ChIP. In addition, it would be informative to test if histone turnover is reduced at the $NDC80^{ORF}$ promoter in meiotic prophase in both the *set2Δ set3Δ* mutant and the *rco1Δ set3Δ* mutant. Measuring changes in H3K56ac incorporation at the $NDC80^{ORF}$ promoter during meiosis in control and mutant cells could be one way of estimating histone replacement rates, since this mark is characteristic of

freshly incorporated histones onto chromatin (Rufiange et al. 2007;Kaplan et al. 2008). Alternatively, other histone exchange assays can be used to measure this directly (Ferrari and Strubin 2015). As such, more work is needed to understand the exact role of histone de-acetylation in my model of *NDC80*^{ORF} regulation.

Importantly, nucleosomes could be deposited at the *NDC80*^{ORF} promoter *NDC80*^{luti} transcription itself, independent of mechanisms involving histone modifications. During transcription, nucleosomes in gene bodies are disassembled and reassembled by histone chaperones such as Spt6 and FACT that associate with RNA polymerase (Venkatesh and Workman 2015). Likewise, the passage of elongating RNAPII through the downstream *NDC80*^{ORF} promoter could direct nucleosome assembly and reduce the size of the NDR, thus inhibiting transcription initiation (Hainer et al. 2011).

The repressive mechanism described in my model for *NDC80*^{luti} regulation has similarities to other loci regulated by transcriptional interference associated with intergenic or promoter transcripts. For example when serine is available, budding yeast cells express the *SRG1* ncRNA from an upstream transcription start site (TSS). *SRG1* ncRNA transcription increases nucleosome occupancy and reduces transcription factor binding at the downstream *SER3* promoter(Martens et al. 2004). In another example, mating type control of budding yeast gametogenesis is mediated by transcription of a non-coding *IRT1* transcript in the *IME1* promoter (van Werven et al. 2012). Although *IRT1* terminates within the *IME1* promoter and is not a luti transcript, *IRT1* transcription also establishes a repressive chromatin state in the *IME1* promoter. Despite these similarities, there are also some differences in these transcriptional interference mechanisms. Unlike the *NDC80*^{luti} and *IRT1* examples, interference by *SRG1* does not require Set1, Set2 or Set3; interference is instead dependent on the histone chaperones and elongation factors Spt6, Spt16 and Spt2 (Hainer et al. 2011;Thebault et al. 2011). Both the *NDC80*^{luti} and the *IRT1* interference mechanisms require Set1/Set3C and Set2/Rpd3S. However, the H3K4me2 and H3K36me3 marks are deposited at different parts of the *IME1* promoter, while these marks overlap in the *NDC80*^{ORF} promoter (van Werven et al. 2012). Despite these mechanistic differences, it is clear that transcriptional interference is an important strategy for regulating cell fate decisions and stress responses in budding yeast.

In contrast to the act of luti transcription, the *NDC80*^{luti} transcript per se is unlikely to have a major regulatory function such as those reported for some other lncRNAs like HOTAIR (Rinn et al. 2007). First, *NDC80*^{luti} mediated repression occurs in *cis* and not in *trans*. Second, *NDC80*^{luti} does not localize to a discrete region in the nucleus by FISH analysis (Chen et al. 2017). Third, there is poor conservation of the 5' nucleotide sequence of *NDC80*^{luti} in *Saccharomyces* species, making a regulatory role for luti RNA secondary structure unlikely (Chen et al. 2017). That being said, it is still possible that nascent *NDC80*^{luti} RNA contributes to repression of *NDC80*^{ORF}. Work done by two different groups showed that both Set1 and Set2 bind to nascent RNA, and this interaction facilitates recruitment and stabilization of these histone methyltransferases during transcription elongation (Battaglia et al. 2017;Sayou et al. 2017).

Set2 and Set3 is essential for establishing a repressive chromatin state and for inhibiting *NDC80*^{ORF} transcription. How does *NDC80*^{luti} transcription cooperate with Set2 and Set3 to repress *NDC80*^{ORF}? Our suggested model involves a two-step mechanism **(Figure 4.15)**. First, transcription of *NDC80*^{luti} deposits Set1 mediated H3K4me2 and Set2 mediated H3K36me3 in the *NDC80*^{ORF} promoter. Second, co-transcriptional deposition of these marks facilitates the recruitment of the histone deacetylase complexes Set3C and Rpd3S (Carrozza et al. 2005b;Keogh et al. 2005;Buratowski and Kim 2010). Notably, the two marks localize to the same area of the *NDC80*^{ORF} promoter, perhaps indicating that there is redundancy between the two pathways. Previous work established a role for Set2 in suppressing histone exchange and promoting nucleosome stability through chromatin remodelers (Smolle et al. 2012;Venkatesh et al. 2012). Although the role of histone modifications in regulating histone turnover is currently disputed, hypo-acetylated histones are associated with more compact chromatin that could inhibit transcription (Bannister and Kouzarides 2011;Ferrari and Strubin 2015;Venkatesh and Workman 2015). As indirect evidence, my data show that mutant cells lacking both Set2 and Set3 have reduced nucleosome occupancy in the *NDC80*^{ORF} promoter, and *NDC80*^{ORF} transcription is unimpeded despite active *NDC80*^{luti} transcription.

The Set1/Set3C and Set2/Rpd3S pathways have well characterized roles in preventing cryptic transcription and regulating gene expression via long non-coding RNA transcription (Carrozza et al. 2005b;Keogh et al. 2005;Kim and Buratowski 2009;Kim et al. 2012;van Werven et al. 2012;Ard and Allshire 2016;Venkatesh et al. 2016). It has been reported that Set2 and Set3 modulate the expression of different genes depending on the length of the adjacent ncRNAs which overlap with their promoters (Kim et al. 2012;Kim et al. 2016). The reason for this is because H3K4me2 is enriched at the 5' ends of transcribed regions whereas H3K36me3 appears within gene bodies. Therefore, an adjacent ncRNA that is shorter in length should cause H3K4me2 enrichment over the repressed promoter, whereas a longer ncRNA should result in H3K36me3 enrichment instead. Kim *et al.* demonstrate that during a series of carbon source shifts, genes whose promoters overlap with longer transcripts (~2.0 kb) are repressed by Set2/Rpd3S whereas those with shorter overlapping transcripts (~0.9 kb), are repressed by Set1/Set3C (Kim et al. 2016). According to their classification, *NDC80*^{luti} is a short overlapping transcript. However, *NDC80*^{luti} mediated repression of *NDC80*^{ORF} was compromised in the *set2*Δ*set3*Δ double mutant cells, but not in the single mutants. In this instance, Set1/Set3C and Set2/Rpd3S act redundantly during *NDC80*^{luti} mediated repression of *NDC80*^{ORF}. The H3K4me2 and H3K36me3 marks overlap in the *NDC80*^{ORF} promoter suggesting that both modifications control the same promoter region. Perhaps, H3K4me2 and H3K36me3 occur on the same nucleosome as part of a repressive combinatorial histone code.

Set2 and Set3 are necessary for *NDC80*^{luti} mediated regulation during meiotic S and prophase. However, data from spot assays and western blots suggest that Set2 and Set3 are dispensable when the luti RNA is overexpressed from heterologous promoters. In fact, the higher the levels of *NDC80*^{luti} transcription, the better the repression of *NDC80*^{ORF} becomes. One possible explanation is that the rate of nucleosome deposition at the *NDC80*^{ORF} promoter is increased during higher levels of *NDC80*^{luti} transcription. In this situation, the requirement for histone deacetylase complexes to stabilize nucleosomes becomes obsolete. Alternatively, elongating RNA polymerase might physically interfere with the *NDC80*^{ORF} promoter when *NDC80*^{luti} is highly expressed. While increased nucleosome occupancy at the *NDC80*^{ORF} promoter plays an important role in gene repression described here, this

model does not exclude alternative mechanisms of regulating *NDC80* transcription. It is possible that *NDC80*^{luti} transcription could result in R-loop formation over the *NDC80*^{ORF} promoter. R-loop formation could then inhibit transcription initiation by interfering with PIC formation or by establishing repressive chromatin through Set2 recruitment; this possibility can be tested using DRIP-seq (Ginno et al. 2012;Chedin 2016;Sanz et al. 2016;Belotserkovskii et al. 2017). Another possibility is that *NDC80*^{luti} transcription could form gene loops which exclude the *NDC80*^{ORF} TU; this hypothesis can be tested by investigating *NDC80*^{ORF} repression during prophase, in *sua7-1* mutants which are defective in gene loop formation (Nguyen et al. 2014).

Another important feature of regulation by *NDC80*^{luti} transcription is that it is dynamic and rapidly reversible. This ensures that different meiotic events occur in a timely manner. During meiotic prophase Ndc80 levels decline, but as soon as cells enter meiotic divisions, Ndc80 levels increase. How is this achieved despite *NDC80*^{luti} transcription? First, levels of *NDC80*^{luti} decrease in the middle phase of gametogenesis, likely due to negative regulation of Ime1 activity by kinases such as Ime2 (Guttmann-Raviv et al. 2002). Second, Chen *et al.* showed that the transcription factor Ndt80 activates the meiotic wave of *NDC80*^{ORF} expression (Chen et al. 2017). Expression of Ndt80 can bypass *NDC80*^{luti} induced repression suggesting that repression is not refractory to increasing levels of *NDC80*^{ORF} transcription. This ensures that *NDC80*^{ORF} can be rapidly produced after meiotic prophase, when Ndc80 becomes essential for proper meiotic divisions. Interestingly, the MSE site (approximately 200 bp upstream from the AUG) in the *NDC80*^{ORF} promoter is not protected by nucleosomes even during *NDC80*^{luti} transcription, which may explain the ability of Ndt80 to activate *NDC80*^{ORF} even while *NDC80*^{luti} is expressed.

Another example of dynamic regulation by *NDC80*^{luti} can be seen in the return to growth time courses. Prior to meiotic divisions, cells that are exposed to a nutrient rich environment again can rapidly re-enter the mitotic cell cycle instead of doing meiotic divisions (Winter 2012). When cells arrested in meiotic prophase were returned to nutrient rich YPD, *NDC80*^{luti} expression was lost and *NDC80*^{ORF} was rapidly induced. The rapid loss of *NDC80*^{luti} expression is likely due to active degradation or depletion of nuclear Ime1. Under nutrient rich conditions, *IME1*

transcription shuts down and Ime1 protein translocates to the cytoplasm (Colomina et al. 2003;van Werven and Amon 2011). In addition, Ime1 has a half-life of only a few minutes, facilitating rapid clearance of Ime1 protein during return to growth (Chia et al. 2017). The mechanism for the reversibility of *NDC80*^{luti} mediated repression is not fully understood. Perhaps transcriptional activators and chromatin remodelers stimulate nucleosome eviction in the *NDC80*^{ORF} promoter during its activation. In line with this idea, levels of the repressive H3K36me3 in the *NDC80*^{ORF} promoter are rapidly lost when the repression is reversed. This is despite the purported stability of the H3K36me3 mark due to its role in reducing histone turnover (Smolle et al. 2012;Venkatesh et al. 2012;Sein et al. 2015). Nonetheless, the reversibility of repression allows cells to quickly adjust cell fate decisions (mitosis or meiosis) to sudden changes in the environment.

NDC80^{luti} transcription is an evolutionarily simple mechanism to regulate gene expression. This system co-opts existing transcription factors such as Ime1 to achieve either gene activation or gene repression in different contexts. This system only requires the evolution of transcription factor binding sites and promoter sequences upstream of another target promoter (Tresenrider and Unal 2018). Developmentally coordinated activation and repression of different sets of genes can thus be achieved by the activity of a stage specific transcription factor (Tresenrider and Unal 2018). Excitingly, *NDC80*^{luti} is an example of an mRNA that does not function in protein production but rather represses gene expression. Luti-mediated transcriptional regulation could in part, account for the poor correlation between mRNA and protein levels observed during gametogenesis (Liu et al. 2016;Cheng et al. 2018).

Transcriptional interference by 5' extended isoforms might be wide-spread in yeast and across species. Transcript isoform sequencing in yeast revealed that the 5' and 3' ends of mRNAs are extremely heterogeneous (Pelechano et al. 2013). During budding yeast meiosis, hundreds of genes have ribosome footprints upstream of the canonical AUG (Brar et al. 2012;Cheng et al. 2018). Some examples have been further confirmed and show a clear inhibitory effect on expression from the downstream promoter (Liu et al. 2015;Xie et al. 2016). However, how the expression of different mRNA isoforms regulates transcription

remains unexplored at the genome-wide level. In higher eukaryotes including human cells, a wide range of 5' extended mRNA isoforms are also expressed often in a cell type-specific manner (Wang et al. 2008;Aanes et al. 2013;Brown et al. 2014). Understanding the principles underlying gene regulation by 5' extended mRNA isoforms during yeast meiosis will deepen our understanding of how complex differentiation programs in higher eukaryotes are regulated. Although Cheng *et al.* have reported transcript isoform toggling during meiosis, their claims are only indirect inferences drawn from regular mRNA seq and ribosome profiling data (Cheng et al. 2018). To my knowledge, there is no existing study that has accurately profiled TSS usage in a highly synchronized population of sporulating cells. Thus in the next chapter, I describe an approach to profile the 5' and 3' ends of all mRNAs for quantitative comparisons of isoform usage throughout gametogenesis.

Chapter 5. Transcript end sequencing (TE-seq) profiles the 5' and 3' ends of mRNA isoforms.

5.1 Abstract

The transcription of overlapping mRNA isoforms with alternative TSSs and TESs is of regulatory importance during budding yeast gametogenesis. There is a shortage of detailed studies profiling the exact 5' and 3' ends of transcripts across developmental time. Here, I present an optimized protocol called transcript end sequencing (TE-seq) to characterize the termini of all mRNAs. By filtering out artifacts and by analysing internal controls, I demonstrate that TE-seq can identify TSSs/TESs at nucleotide resolution and can reliably quantify the relative abundances of reads associated at the start and end of mRNAs. This method enables quantitative comparisons of TSS/TES usage which will be useful for inferring regulatory relationships between mRNA isoforms or ncRNAs.

5.2 Introduction

In chapter 4, I described how transcription of the long undecodable transcript isoform (luti) *NDC80*^{luti} repressed the canonical *NDC80*^{ORF} promoter during meiotic S and prophase. A clear follow up question is to ask how many genes are regulated by luti mRNA transcription during gametogenesis. Previous studies have found more than 190 genes which have ribosome footprints upstream of the canonical AUG (Brar et al. 2012;Cheng et al. 2018). However without profiling actual transcription start site (TSS) usage across gametogenesis, it is unclear which and how many of these examples are *bona fide* TSS toggling events like the *NDC80* example. Although mRNA seq data can be used to identify intergenic or promoter transcription events, the interpretation of these results is complicated by overlapping RNA seq reads from different transcripts. Bulk mRNA seq data cannot accurately determine what isoforms are transcribed from a locus nor distinguish the transcript end site (TES) of interleaved, alternative transcripts. Thus a genome wide technique that profiles both TSS and TES usage quantitatively, is required to

characterize lncRNAs or mRNA isoforms and to infer their roles in transcriptional regulation.

Furthermore, data from the previous chapter suggests that repression by luti transcription is not binary, but can vary depending on the relative strengths of the luti and canonical promoters. It is possible that luti transcription at other loci functions to tune gene expression like a molecular rheostat, rather than as a toggle switch between on/off states. It is unclear how many genes are repressed by upstream transcription during gametogenesis, similar to the *NDC80* example. Other ncRNA or isoform transcription events could have minimal or no effect on highly active promoters (Lenstra et al. 2015). Intergenic or promoter transcription can also be associated with gene activation. In the fission yeast *Schizosaccharomyces pombe*, glucose starvation induces step-wise transcription of 4 distinct *fbp1*$^{+}$ transcripts (including the canonical mRNA); promoter ncRNA transcription removes repressors and promotes an open chromatin state for gene activation (Hirota et al. 2008). In a different study, Uhler *et al*. show in budding yeast that an anti-sense ncRNA also stimulates *PHO5* gene activation through transcription-coupled histone eviction in the *PHO5* promoter (Uhler et al. 2007). Till date, there is no study examining the different effects of upstream overlapping transcription on downstream gene regulation during a developmental program like budding yeast gametogenesis.

A recent study by Pelechano *et al.* showed extensive heterogeneity in the 5' and 3' ends of yeast mRNAs using a method called transcript isoform sequencing (TIF seq) (Pelechano et al. 2013). However, TIF seq cannot accurately quantify the relative amounts of different isoforms due to its reliance on reverse transcribing full length mRNAs and inefficiencies associated with the circularization of long mRNAs; shorter isoforms will be over-represented in TIF-seq libraries. Therefore, a different approach must be used for quantitative analysis of overlapping mRNA isoforms or ncRNAs, in order to infer regulatory relationships between transcripts e.g. identifying examples similar to how *NDC80*^{luti} transcription represses *NDC80*^{ORF}.

These considerations led me to optimize an approach to profile TSS and TES usage of mRNAs across a high resolution meiotic time course called transcript end

sequencing (TE-seq). In this chapter, I give an overview of TE-seq and show why this method is suitable for quantitative comparisons of TSS or TES usage. This method enables a robust and detailed study of transcriptional regulation by ncRNAs and mRNA isoforms during budding yeast gametogenesis.

5.3 Results

5.3.1 An optimized protocol to sequence the 5' ends of mRNAs

To profile the 5' ends of poly(A)+ RNA/mRNAs, I used an approach adapted from previously published techniques (Arribere and Gilbert 2013;Malabat et al. 2015;Adjalley et al. 2016) (see *materials and methods* for details). As a quality control for all 5' seq libraries, three different capped and poly-adenylated *in vitro* transcripts (IVTs) of varying lengths were synthesized: pGIBS-LYS, pGIBS-PHE and pGIBS-THR **(Figure 5.1)**. These IVTs were pooled in an approximate molar ratio of 25 pGIBS-LYS:5 pGIBS-PHE:1 pGIBS-THR and were spiked into all total RNA samples before 5' seq library preparation. Quality control analysis using these IVTs will be discussed in a later section of this results chapter.

A flowchart summarizing 5' seq library preparation is shown in **Figure 5.2.** Total RNA with IVT spike-ins were fragmented into a mode size of ~200 nucleotides. Incubating RNA in a hot alkaline zinc solution initially generates fragments with 5' hydroxyl groups and 2',3' cyclic-phosphate ends (Dallas et al. 2004). Subsequent treatment with alkaline phosphatase further removed any remaining 5' phosphate groups; the resultant 5' hydroxyl groups are incompetent for ligation. The 2',3' cyclic-phosphate ends were also dephosphorylated to give ligation-competent 3' hydroxyl ends. Throughout the fragmentation and dephosphorylation steps, the 5' ends of mRNAs are protected by the 7-methyl-guanosine cap (m^7Gppp). Subsequent removal of the m^7Gppp cap exposes 5' phosphate groups which are ligation competent. Therefore, the majority of fragments which can be ligated at the 5' ends to a custom oligonucleotide, represent the 5' ends of mRNAs. These short fragments were reverse transcribed and converted to biotinylated double stranded cDNA which were used as inputs for DNA library preparation. The biotin group on the cDNA enabled selective capture and purification of *bona fide* 5' fragments and a gel extraction step removed any

excess un-ligated Illumina adapters that would contaminate the sequencer. Purified libraries were sequenced to an equivalent of 75 bases single-end reads, at a depth of approximately 20 million reads per library. The high sequencing depth relative to genome size (~12 Mb) and the use of 3 biological repeats for each sample increases the reliability of identifying most of the TSSs throughout gametogenesis.

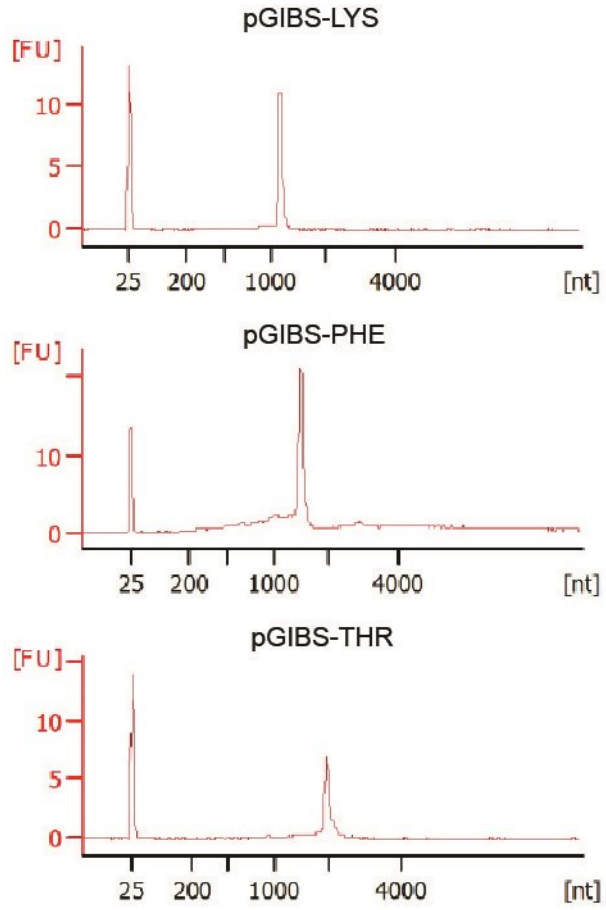


Figure 5.1 Bioanalyzer profiles of three *in vitro* transcripts (IVTs) used as internal spike-in controls for TE-seq

IVTs were synthesized from three plasmid templates: pGIBS-LYS, pGIBS-PHE and pGIBS-THR. Purified IVTs were run on a bioanalyzer and each trace shows a sharp peak corresponding to the size of the IVTs.

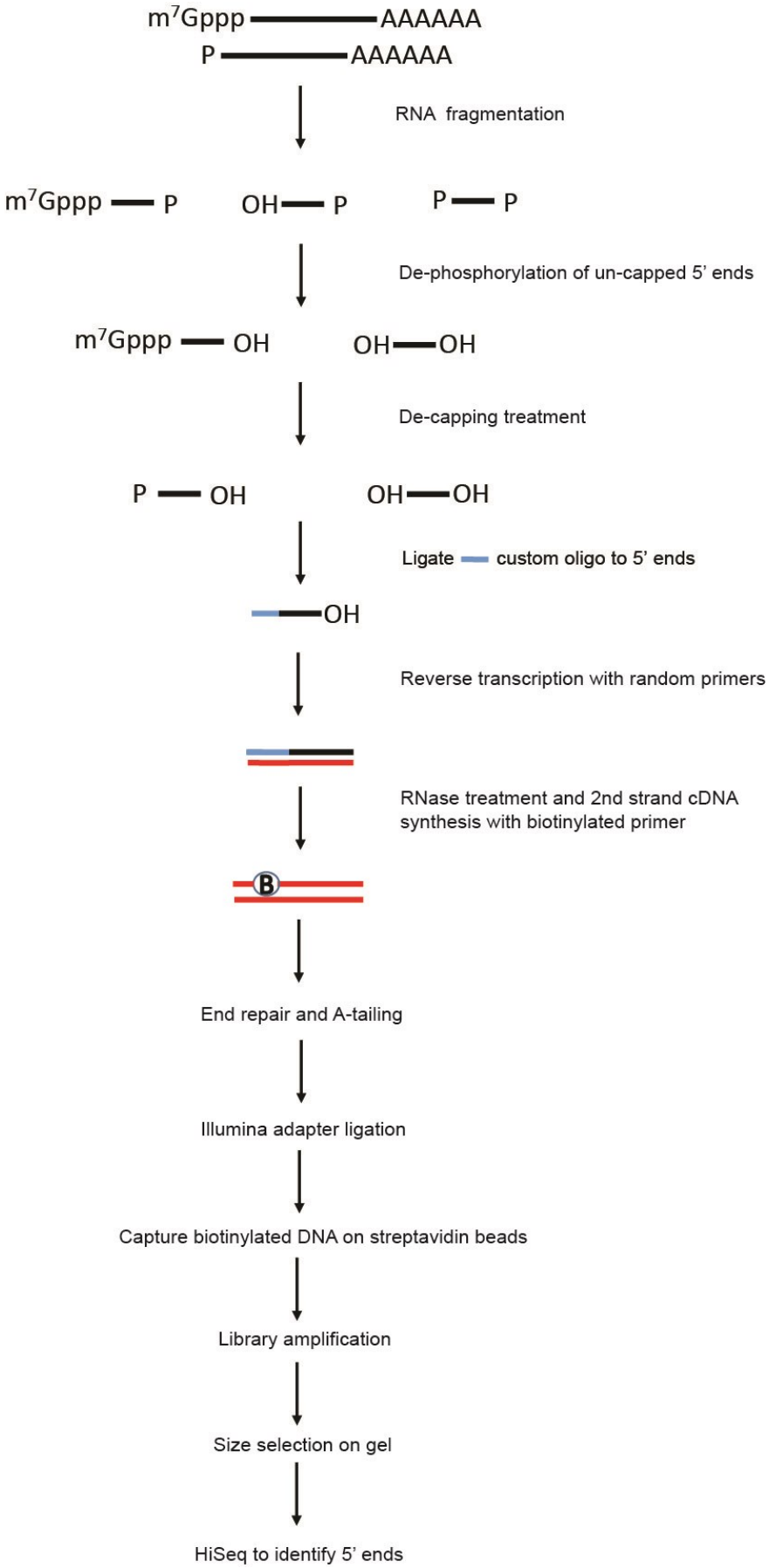

m7Gppp AAAAAA
P AAAAAA
RNA fragmentation
m7Gppp P
OH P
P P
De-phosphorylation of un-capped 5' ends
m7Gppp OH
OH OH
De-capping treatment
P OH
OH OH
Ligate custom oligo to 5' ends
OH
Reverse transcription with random primers
RNase treatment and 2nd strand cDNA synthesis with biotinylated primer
B
End repair and A-tailing
Illumina adapter ligation
Capture biotinylated DNA on streptavidin beads
Library amplification
Size selection on gel
HiSeq to identify 5' ends

Figure 5.2 A schematic view of 5' seq library preparation

Fragmented mRNAs (black) were dephosphorylated and then treated with Cap-Clip acid pyrophosphatase to remove the m^7Gppp cap. Some RNA from a 0 hour time point was set apart without the decapping reaction as a non-decapping control. The 5' ends of RNA were ligated to a custom 5' adapter (blue). Fragments were reverse transcribed with random primers to generate first strand cDNA (red). The RNA templates were degraded with RNase and purified products were used as a template for second strand synthesis using a biotinylated primer. Double stranded DNA (red) was purified and then used as input for the library preparation. Prior to the library amplification step, samples were captured on MyOne Streptavidin C1 Dynabeads. Library amplification was done on the biotinylated dsDNA fraction bound to the beads. Amplified libraries above ~150 bp were further purified by gel extraction to remove un-ligated adapters. Purified libraries were sequenced on an Illumina HiSeq 2500 to an equivalent of 75 bases single-end reads, at a depth of approximately 20 million reads per library.

5.3.2 An optimized protocol to sequence the 3' ends of mRNAs

To profile the 3' ends of poly(A)+ RNA/mRNAs, I used an approach adapted from previously published techniques (Ng et al. 2005;Lai et al. 2015) (see materials and methods for more details). A key challenge of profiling TESs is due to the difficulties in sequencing the homopolymeric poly(A) tails of mRNAs (Quail et al. 2012). My approach to this problem involves enzymatically shortening the poly(A) stretch during 3' library preparation, as seen in **Figure 5.3**.

In this protocol, a separate matching aliquot of mRNA fragments from each sample was reverse transcribed with a custom biotinylated and anchored oligo$(dT)_{20}$ primer in the presence of dNTPs and 5-methyl-dCTP. The relatively high annealing temperature (50°C) maximises the chances of the VN anchor in the RT primer hybridizing to the last two nucleotides preceding the poly(A) tail and minimizes internal priming. 1st strand cDNA:RNA hybrid molecules were purified with streptavidin beads and subjected to a second strand synthesis step known as "nick translation"(Gubler and Hoffman 1983). During nick translation, RNase H cleaves the RNA template in the cDNA:RNA hybrid at numerous places. This converts each RNA molecule into a series of shorter fragments, which prime *E. coli* DNA polymerase I (Pol I) for second strand cDNA synthesis using the existing DNA template. The 5' to 3' exonuclease activity of Pol I removes the RNA primers whilst adding dNTPs to the 3' ends of the newly synthesized daughter strand. Finally, *E.*

coli DNA ligase seals any residual gaps left in the second strand product, resulting in double stranded cDNA. Although this approach causes the loss of some information at the 5' ends of the second strand, the sequences immediately preceding the poly(A) tail are still reliably preserved in the final product. After second strand synthesis, the resultant hemi-methylated cDNA molecules have a GsuI restriction site next to the poly dA:dT stretch. GsuI is a type IIS restriction enzyme which makes a staggered cut 16 nucleotides downstream of its recognition site i.e. CTGGAG(16/14)^. Importantly, GsuI is Dcm methylation sensitive, which prevents it from acting on internal restriction sites in the hemi-methylated cDNA. Therefore after GsuI digestion, most of the poly(A) tails are shortened to 4 nts in length. These molecules were used as inputs for DNA library preparation. The gel extraction step removed any excess un-ligated Illumina adapters that would contaminate the sequencer. Purified libraries were sequenced to an equivalent of 75 bases single-end reads, at a depth of approximately 20 million reads per library. The high sequencing depth relative to genome size (~12 Mb) and the use of 3 biological repeats for each sample increases the reliability of identifying most of the TESs throughout gametogenesis.

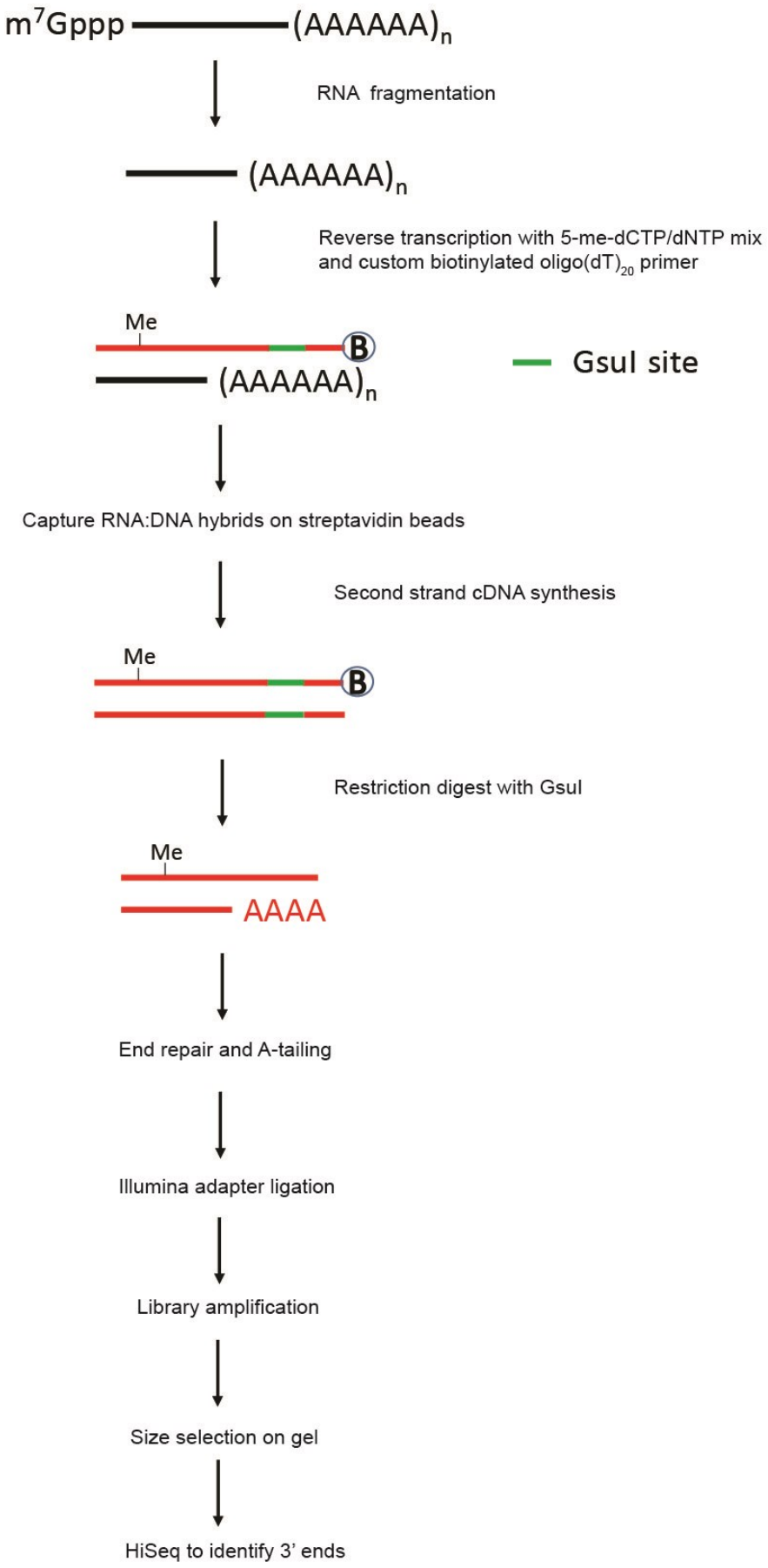

m7Gppp
(AAAAAA)n
RNA fragmentation
(AAAAAA)n
Reverse transcription with 5-me-dCTP/dNTP mix
and custom biotinylated oligo(dT)20 primer
Me
B
(AAAAAA)n
GsuI site
Capture RNA:DNA hybrids on streptavidin beads
Second strand cDNA synthesis
Me
B
Restriction digest with GsuI
Me
AAAA
End repair and A-tailing
Illumina adapter ligation
Library amplification
Size selection on gel
HiSeq to identify 3' ends

Figure 5.3 A schematic view of 3' seq library preparation

Fragmented mRNAs (black) were were reverse transcribed with 5-Methylcytosine-dNTPs and an anchored oligo d(T) primer with a GsuI restriction site (green). The resulting first strand cDNA is methylated (red). RNA:DNA hybrids were captured on MyOne Streptavidin C1 Dynabeads and subjected to second strand synthesis to generate hemi-methylated double stranded DNA (red). Bound cDNA was digested with GsuI and released products were used as input for the library preparation. Amplified libraries above ~150 bp were further purified by gel extraction to remove un-ligated adapters. Purified libraries were sequenced on an Illumina HiSeq 2500 to an equivalent of 75 bases single-end reads, at a depth of approximately 20 million reads per library.

5.3.3 Processing TE-seq reads

I conducted the bioinformatics analysis with the help of Cai, Li from the Luscombe lab. All the RNA used for experiments reported in this chapter were obtained from three independent biological repeats. They consisted of saturated phase cells (Spo 0h), cells in early gametogenesis (Spo 4h) and cells in mid-late gametogenesis (Spo 7h).

The 5' end reads were trimmed to remove the custom adapter sequence 5'-CACTCTGAGCAATACC-3' and between 20-45% of reads could be mapped to the budding yeast SK1 genome **(Table 5.1)**. The 3' ends were identified as those sequences with non-templated As or Ts depending on which end the molecule was sequenced from. As expected, most of the 3' reads had a series of 4 terminal As due to GsuI's digestion of the Poly(A) tail **(Figure 5.4)**. A small fraction of reads had fewer than 4 terminal As, possibly due to some inaccuracies in GsuI cutting. Subsequently, reads with 2 or more As were selected for mapping to the genome. To reduce artifacts due to non-specific hybridization by the oligo d(T) primer, we also excluded any reads adjacent to a stretch of 20 A/Ts in the genome (allowing for up to 8 mis-matches). Approximately 20% of 3' reads were usable after these steps **(Table 5.1)**. 5' and 3' reads were clustered into TSS or TES "peaks" using the CAGEr software and assigned to neighbouring genes using bedtools (see *materials and methods* for parameters) (Haberle et al. 2015).

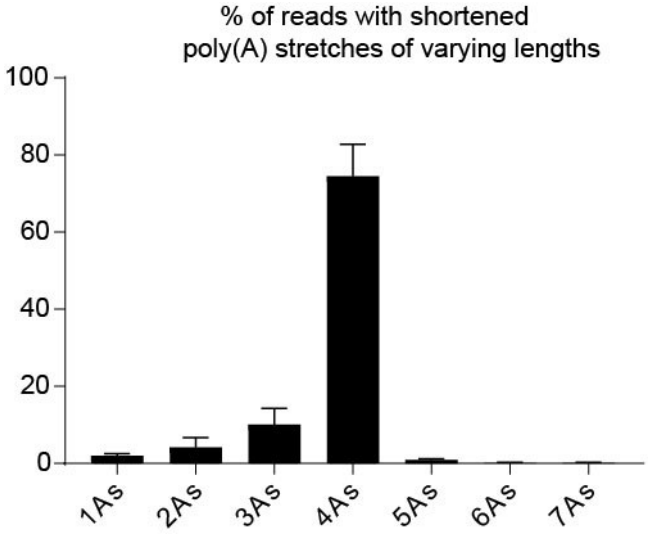


Figure 5.4 Most 3' reads have a shortened poly(A) stretch of 4 nt

Chart showing the percentage of 3' reads with shortened poly(A) stretches of varying lengths. After Illumina adapter removal, 3' reads were examined for streches of only As which were not part of the genomic template. Those poly(A) stretches represent the shortened tails generated by GsuI digestion during library preparation. Data are an average of Spo 0h, 4h and 7h libraries, n = 3 per time point.

Library	Total reads	Aligned reads	Usable reads (after filtering)	% Usable/total reads
TSS sequencing data				
Spo 0h A	35784104	12431650	12126942	33.89
Spo 0h B	25940057	9700310	9457384	36.46
Spo 0h C	27580633	6301480	6149852	22.30
Spo 4h A	26655946	9527689	9367489	35.14
Spo 4h B	24494040	8135973	7936260	32.40
Spo 4h C	27847638	6976456	6863022	24.64
Spo 7h A	29240621	13347793	13172602	45.05
Spo 7h B	35869437	13057454	12842173	35.80
Spo 7h C	28019795	6198718	6103581	21.78
No decapping	22570834	5806893	3223998	14.28
TES sequencing data				
Spo 0h A	24332405	20574829	5628655	23.13
Spo 0h B	36912893	26339095	6708138	18.17
Spo 0h C	28195108	22862180	6567717	23.29
Spo 4h A	29824624	25389592	7238117	24.27
Spo 4h B	35675773	27772364	8146997	22.84
Spo 4h C	26378081	20482870	5718101	21.68
Spo 7h A	31057024	26192352	7413071	23.87
Spo 7h B	32792623	25371161	7216917	22.01
Spo 7h C	41602086	33632310	9585671	23.04
RNA sequencing data				

Spo 0h A	41455554	40009399	34151068	82.38
Spo 0h B	40126646	38693504	32333519	80.58
Spo 0h C	40609556	39204071	33718988	83.03
Spo 4h A	34804342	33850902	30173407	86.69
Spo 4h B	39970278	38798560	34328595	85.89
Spo 4h C	36360183	35186841	31037398	85.36
Spo 7h A	43807146	42831237	38212968	87.23
Spo 7h B	47425906	46312952	41409264	87.31
Spo 7h C	50181260	49056362	43377211	86.44

Table 5.1 Read counts for libraries used in this chapter

5.3.4 TE-seq maps TSSs and TESs at nucleotide resolution

I examined if the trimmed TE-seq reads could identify the first transcribed nucleotide i.e. the TSS. Given the heterogeneity of 5' and 3' ends reported by Pelechano *et al.*, my initial approach was to examine the three IVT spike-ins which each have a single defined TSS and TES: pGIBS-LYS, pGIBS-PHE and pGIBS-THR (Pelechano et al. 2013). With the exception of the pGIBS-PHE IVT, >95% of reads mapping to other two spike-in sequences correctly identified the TSS **(Figure 5.5 A)**. Approximately 80% of the reads mapping to pGIBS-PHE correctly identified its TSS, with the remaining 20% scattered at different positions in the body of the transcript. Nonetheless, those scattered reads failed to pass the threshold in our pipeline to be called as a spurious alternative TSS peak **(Figure 5.6)**. In fact, the pipeline correctly identified a single TSS peak for each IVT at different selected time points in meiosis **(Figure 5.6)**. Importantly, I also analysed the samples which were not subjected to the decapping reaction **(Figure 5.6**, no decapping**)**. Any reads in the no decapping control represent artifactual, non-specific background ligation events. As expected, the no decapping controls showed signals scattered across the body of the transcripts, with no clear enrichment at a specific TSS **(Figure 5.6**, no decapping**)**.

I next examined if the trimmed TE-seq reads could identify the last transcribed nucleotide before the poly(A) tail i.e. the TES. Approximately 80% of all reads mapping to the three IVTs correctly identified their TESs, and this was increased to > 98% when the last 3 nucleotides were considered **(Figure 5.5 B and C)**. Regardless, our pipeline correctly clustered these signals as a single TES for

each IVT and identified the exact last nucleotide as the dominant signal in the TES cluster **(Figure 5.6)**.

Taken together, the IVT data suggest that the TE-seq protocol and our pipeline can reliably map TSSs and TESs at nucleotide resolution.

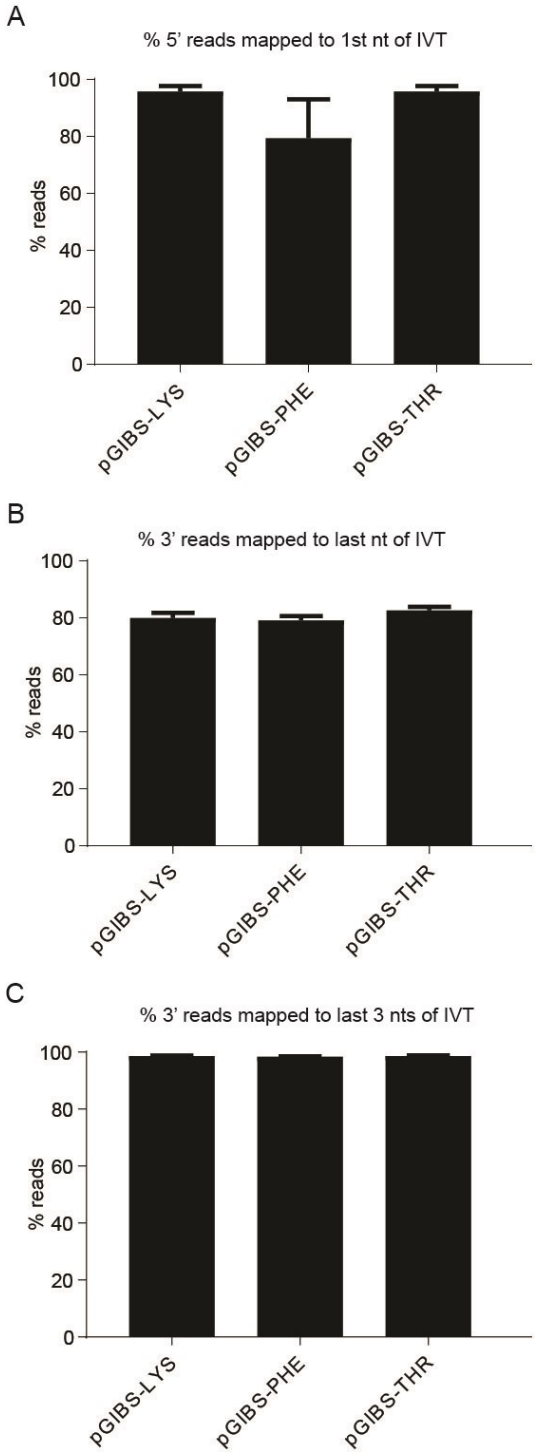


Figure 5.5 Identification of TSSs/TESs at nucleotide resolution from IVTs

(A) Bar chart showing the percentage of reads mapping to the TSS of each IVT.
(B) Bar chart showing the percentage of reads mapping to the TES of each IVT.
(C) Bar chart showing the percentage of reads mapping to the last 3 nucleotides of each IVT.
Six different libraries were averaged for the results in this figure, n = 6

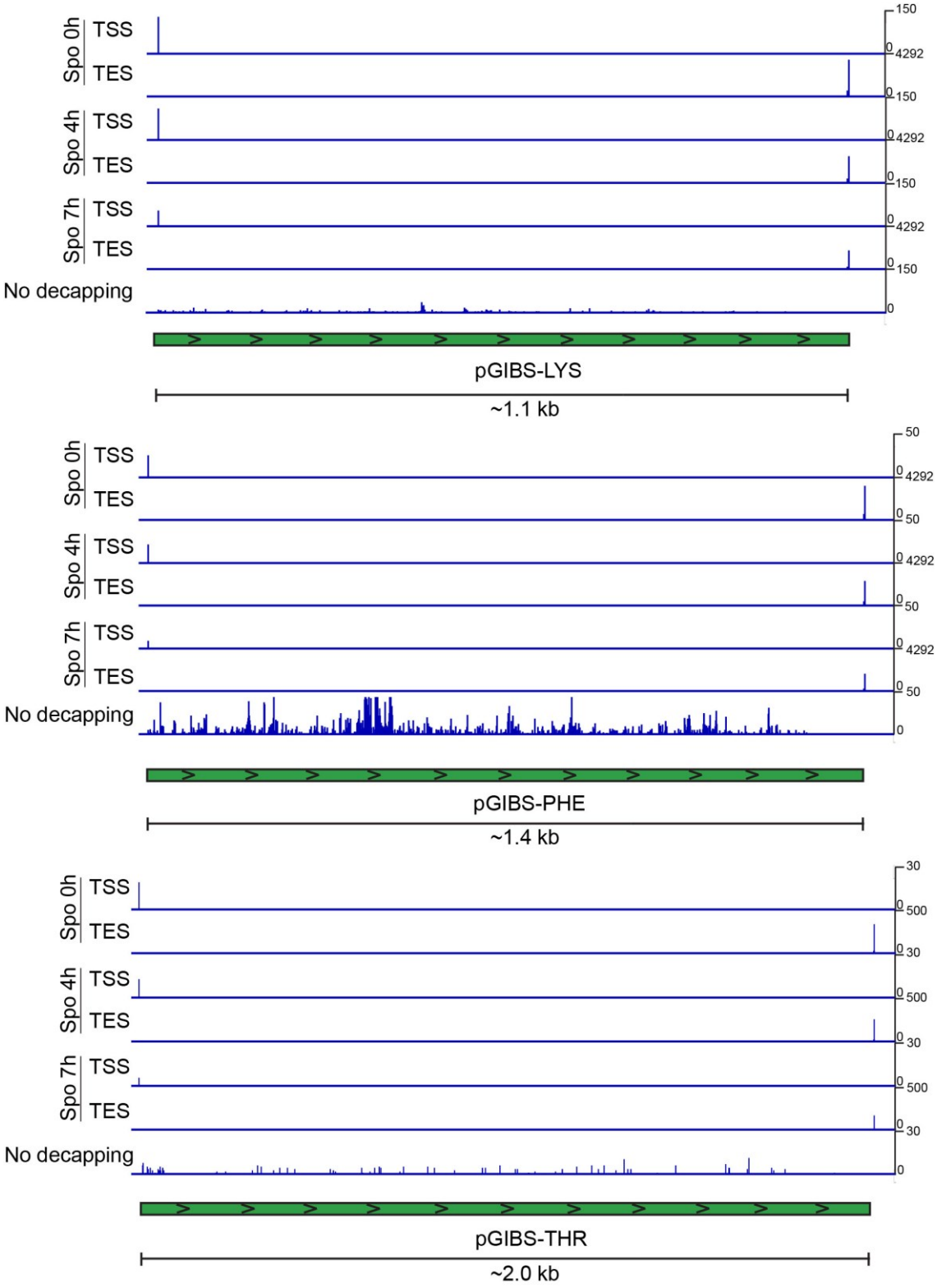
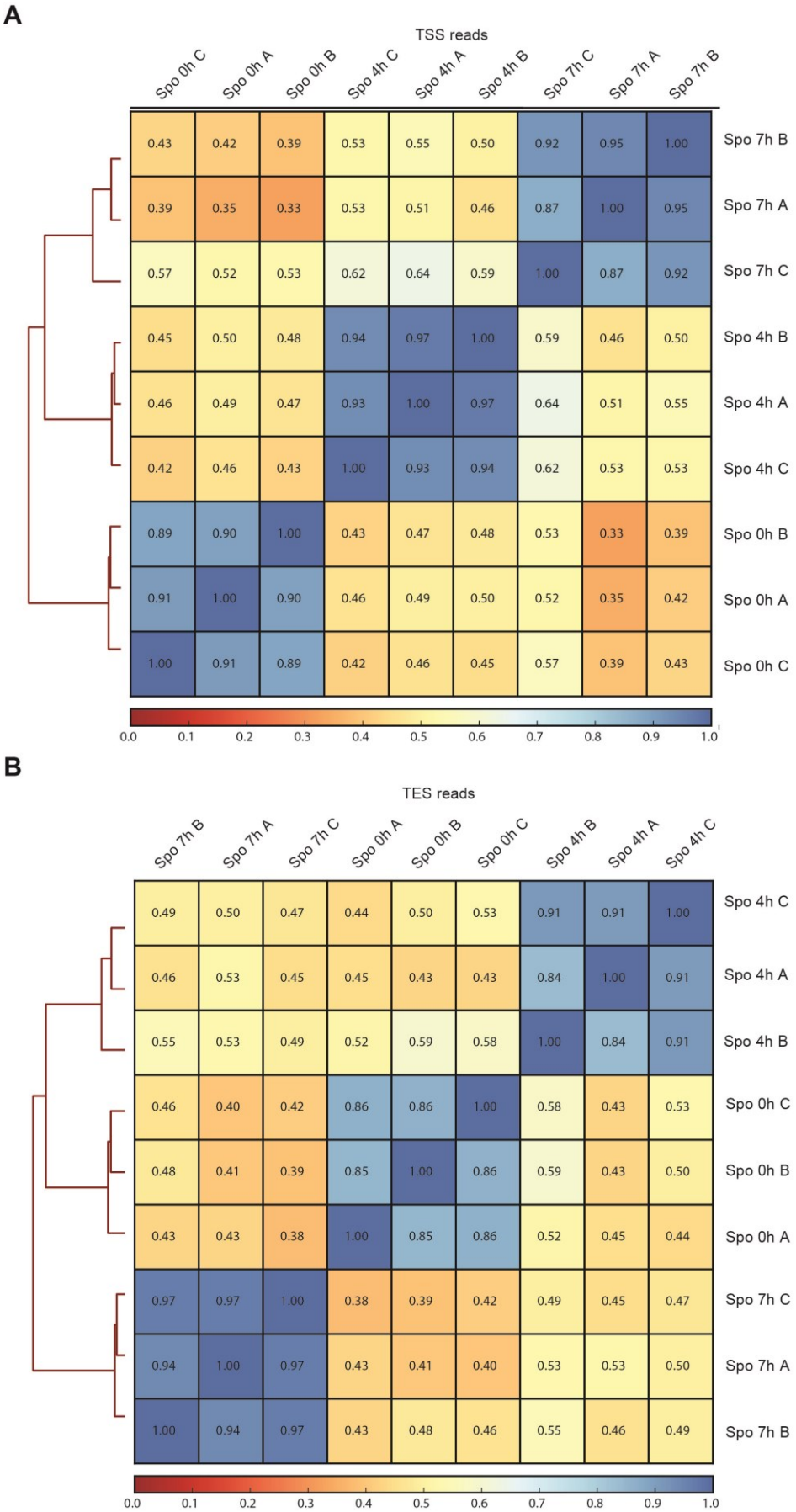


Figure 5.6 Visualization of TSS and TES IVT peaks on the Integrative Genome Viewer (IGV)

IGV tracks representing the TSS or TES peaks for each of the three IVTs, from different libraries (Spo 0h, 4h and 7h). After filtering and peak calling, only a single defined TSS/TES peak is seen for each IVT. As a control for 5' end sequencing, the non-decapping samples are also shown for comparison. Scales for TPM values and distance (kb) are shown. Data are an average of three independent repeats, n = 3.

5.3.5 TE-seq biological replicates are highly correlated

I next examined if TE-seq data was consistent between three independent time courses. Matched RNA samples harvested from pre-meiotic saturated cultures, early gametogenesis and mid-late gametogenesis were subjected to TE-seq and regular mRNA sequencing. A matrix summarizing Pearson correlation coefficients between different pairs of libraries was made for TSS reads **(Figure 5.7 A)**, TES reads **(Figure 5.7 B)** and mRNA sequencing reads **(Figure 5.7 C)**. In all cases, libraries corresponding to three biological repeats, each labelled A, B and C, were highly correlated for the same time points (0h, 4h or 7h).


A
TSS reads
Spo 0h C
Spo 0h A
Spo 0h B
Spo 4h C
Spo 4h A
Spo 4h B
Spo 7h C
Spo 7h A
Spo 7h B
0.43 0.42 0.39 0.53 0.55 0.50 0.92 0.95 1.00 Spo 7h B
0.39 0.35 0.33 0.53 0.51 0.46 0.87 1.00 0.95 Spo 7h A
0.57 0.52 0.53 0.62 0.64 0.59 1.00 0.87 0.92 Spo 7h C
0.45 0.50 0.48 0.94 0.97 1.00 0.59 0.46 0.50 Spo 4h B
0.46 0.49 0.47 0.93 1.00 0.97 0.64 0.51 0.55 Spo 4h A
0.42 0.46 0.43 1.00 0.93 0.94 0.62 0.53 0.53 Spo 4h C
0.89 0.90 1.00 0.43 0.47 0.48 0.53 0.33 0.39 Spo 0h B
0.91 1.00 0.90 0.46 0.49 0.50 0.52 0.35 0.42 Spo 0h A
1.00 0.91 0.89 0.42 0.46 0.45 0.57 0.39 0.43 Spo 0h C
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
B
TES reads
Spo 7h B
Spo 7h A
Spo 7h C
Spo 0h A
Spo 0h B
Spo 0h C
Spo 4h B
Spo 4h A
Spo 4h C
0.49 0.50 0.47 0.44 0.50 0.53 0.91 0.91 1.00 Spo 4h C
0.46 0.53 0.45 0.45 0.43 0.43 0.84 1.00 0.91 Spo 4h A
0.55 0.53 0.49 0.52 0.59 0.58 1.00 0.84 0.91 Spo 4h B
0.46 0.40 0.42 0.86 0.86 1.00 0.58 0.43 0.53 Spo 0h C
0.48 0.41 0.39 0.85 1.00 0.86 0.59 0.43 0.50 Spo 0h B
0.43 0.43 0.38 1.00 0.85 0.86 0.52 0.45 0.44 Spo 0h A
0.97 0.97 1.00 0.38 0.39 0.42 0.49 0.45 0.47 Spo 7h C
0.94 1.00 0.97 0.43 0.41 0.40 0.53 0.53 0.50 Spo 7h A
1.00 0.94 0.97 0.43 0.48 0.46 0.55 0.46 0.49 Spo 7h B
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

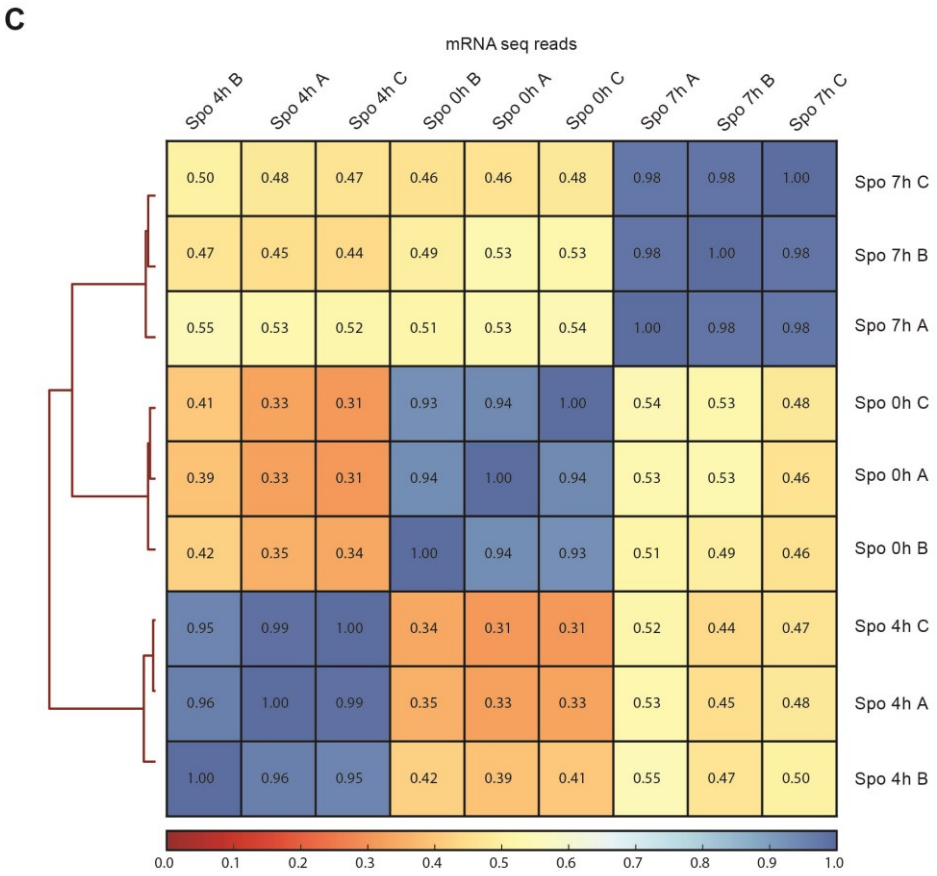


Figure 5.7 Biological replicates for TE-seq and RNA seq are highly reproducible

(A) Heatmap showing correlation between three biologically independent replicates (A, B and C) of different time points (Spo 0h, 4h and 7h) for TSS read counts. The Pearson correlation coefficient between any two libraries is indicated.
(B) Heatmap showing correlation between three biologically independent replicates (A, B and C) of different time points (Spo 0h, 4h and 7h) for TES read counts. The Pearson correlation coefficient between any two libraries is indicated.
(C) Heatmap showing correlation between three biologically independent replicates (A, B and C) of different time points (Spo 0h, 4h and 7h) for RNA seq reads. The Pearson correlation coefficient between any two libraries is indicated.
Clustering results and heatmaps were made by Cai, Li from the Luscombe lab.

5.3.6 Quantitative comparisons of TSS/TES usage

Given that TE-seq can identify TSSs/TESs reliably, I tested if this technique could accurately quantify the relative abundances of transcripts. Three different IVTs were used as spike-ins: pGIBS-LYS, pGIBS-PHE and pGIBS-THR. The IVT spike-ins were pooled at an approximate molar ratio of 25 pGIBS-LYS :5 pGIBS-PHE : pGIBS-THR. Therefore, approximately 80%, 15% and 5% of reads are

expected to map to pGIBS-LYS, pGIBS-PHE and pGIBS-THR respectively. Tag per million (TPM) values were obtained for each IVT in each library and the proportion of reads for each IVT was plotted **(Figure 5.8)**. The mean observed frequency for 5' reads were approximately 85%, 10% and 5% for pGIBS-LYS, pGIBS-PHE and pGIBS-THR respectively **(Figure 5.8 A)**. The mean observed frequency for 3' reads were approximately 75%, 21% and 5% for pGIBS-LYS, pGIBS-PHE and pGIBS-THR respectively **(Figure 5.8 B)**. Although the observed frequencies differ from the expected frequencies, there is little variance between different libraries with the same spike-ins. The differences in frequencies between the 5' and 3' reads for pGIBS-LYS and pGIBS-PHE could be caused by the sporadic failure of the *in vitro* transcription reaction to reach the ends of the templates. Consistent with this hypothesis, the non-decapping controls showed poorer coverage towards the 3' ends of the IVTs, perhaps due to fuzzy termination of transcription **(Figure 5.6)**.

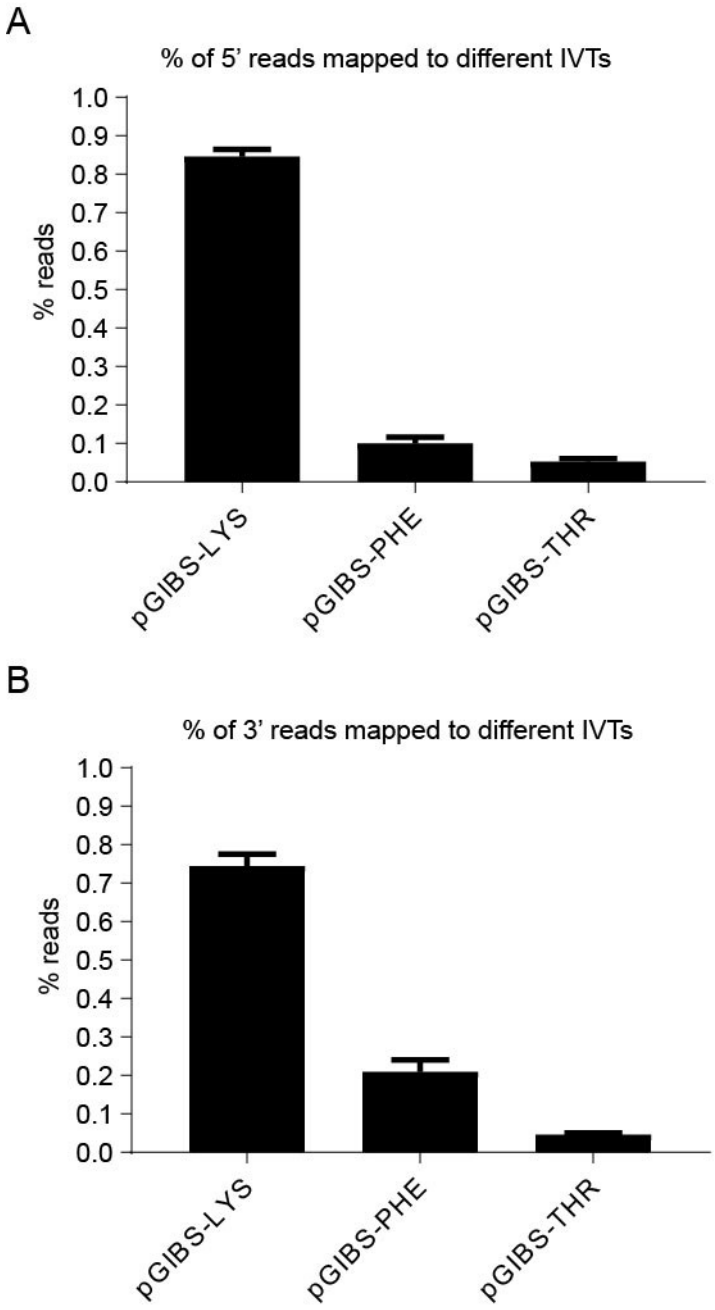


Figure 5.8 Quantifying TSSs/TESs usage from the IVT spike-ins

(A) The relative percentage of 5' reads mapped to the three different IVTs. The percentages were calculated by taking the TPM values for each IVT and dividing them by the sum of all TPM values for the IVTs. The mean and SEM are shown for each IVT, n = 6.
(B) The relative percentage of 3' reads mapped to the three different IVTs. The percentages were calculated by taking the TPM values for each IVT and dividing them by the sum of all TPM values for the IVTs. The mean and SEM are shown for each IVT, n = 6.
Six different libraries were analyzed for all the results in this figure, n = 6

Next, I examined if TSS, TES and RNA seq data correlated well with each other. I chose a selection of 1253 genes with only one matching TSS peak and one TES peak in saturated culture (Spo 0h). The $\log_{10}$(TPM) values of TESs and the $\log_{10}$(TPM) values of matching TSSs for 1253 genes were moderately correlated, with a Spearman rank correlation coefficient (r_s) of 0.66 **(Figure 5.9 A**, left**)**. The $\log_{10}$(TPM) values of TSSs and to a lesser extent TESs, were highly correlated with matched $\log_{10}$(FPKM) values from RNA seq, with r_s values of 0.86 and 0.75 respectively **(Figure 5.9 A**, middle and right**)**.

I also examined if TE-seq could accurately estimate changes in TSS/TES usage. To test this, I chose 208 early meiotic genes on the basis of them being up-regulated in early gametogenesis (Spo 4h) relative to the pre-meiotic state (Spo 0h); each of these 208 candidates also had only one annotated TSS and TES. The $\log_2$(Spo 4h/Spo 0h) fold change values for TESs were moderately correlated with matched TSSs, with a r_s value of 0.62 **(Figure 5.9 B**, left**)**. There was a stronger correlation between the $\log_2$(Spo 4h/Spo 0h) fold change values for TSSs or TESs and that of RNA seq, ,with r_s values of 0.76 and 0.71 respectively (**Figure 5.9 B**, middle and right**)**.

The outliers observed in the scatter plots prompted further investigation of individual examples in a genome browser. In some cases, poor correlation between TSS and TES reads were due to inconsistencies with existing ORF annotations. For example, TSS3015 and TES3366 were assigned to the *YEL034W* ORF. However, very few reads were associated with TES3366 despite relatively high TSS3015 signal. RNA seq data suggests that TSS3015 should be paired with TES3367 instead **(Figure 5.10 A**, left**)**. Perhaps the *YEL034W* and *YEL033W* ORFs belong to the same transcription unit in saturated phase cells. In another example, *YGR174C* ORF is assigned to TSS4876 and TES6195. However, RNA seq and TSS data suggest that the actual TSS used is TSS4875, which is within the mis-annotated ORF **(Figure 5.10 A**, right**)**.

In other cases, poor correlations between TSS and TES reads could be due to read-through transcription. TSS6285 and TES8005 were assigned to the *YJR107W* ORF. However, TSS6285 has fewer than proportional reads compared

to TES8005. RNA seq data suggests that transcription from the neighbouring *YJR106W* locus could overlap with the *YJR107W* locus and terminate at TES8005 **(Figure 5.10 B)**. If so, TSS6285 might have fewer reads due to read-through transcriptional interference, while TES8005 might have more reads contributed by longer 3'-extended *YJR106W* transcripts.

Imprecise or "fuzzy" transcription termination might also account for poor correlation between TSS and TES signals in some cases. For example, *YLR264W* has a single assigned TES9712, with fewer reads than expected possibly due to stochastic termination at earlier sites **(Figure 5.10 C**, left**)**. Many of these 3' reads were too diffuse to be called as TES peaks. Similarly, *YKL087C* has a single assigned TES9080. However, the broad distribution of TES reads as well as RNA seq signals suggest imprecise termination events **(Figure 5.10 C**, right**)**.

To summarize, TSS and TES signals are sometimes poorly correlated with each other due to different reasons that reflect the complicated organization of the gene-dense genome and overlapping TUs. Nonetheless, there is a sensible quantitative relationship between TSS, TES and RNA seq signals for most genes.

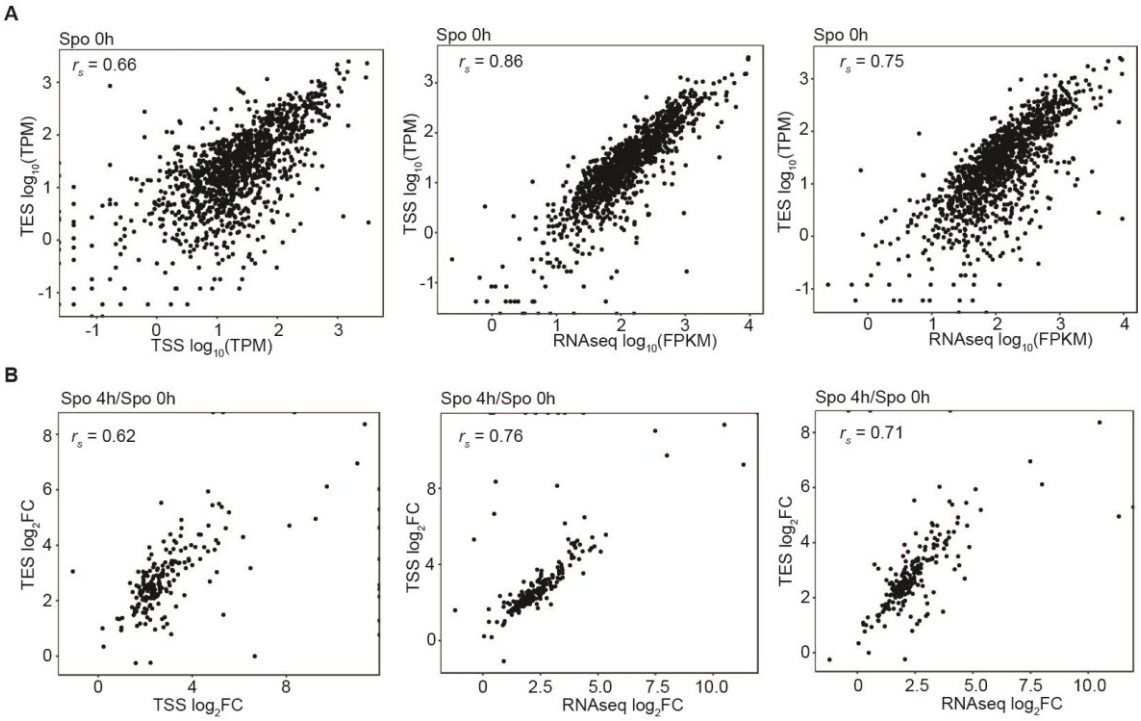


Figure 5.9 Quantifying TSS/TES correlations from genome wide data

(A) Left: scatter plot of $\log_{10}$(TPM) values of TESs against $\log_{10}$(TPM) values of TSSs for 1253 genes which had only one identified TSS and TES in pre-meiotic cells grown to saturation. Middle: scatter plot of $\log_{10}$(TPM) values of TSSs against $\log_{10}$(FPKM) values of RNAseq for the same set of genes. Right: scatter plot of $\log_{10}$(TPM) values of TESs against $\log_{10}$(FPKM) values of RNAseq for the same set of genes. The Spearman rank correlation coefficients (r_s) are given for each plot. Data are an average of three biological repeats, n = 3.
(B) Left: scatter plot of $\log_2$FC(Spo4h/Spo0h) values of TESs against $\log_2$FC values of TSSs for 208 early meiotic genes which had only one identified TSS and TES. Middle: scatter plot of $\log_2$FC values of TSSs against $\log_2$FC values of RNAseq for the same set of genes. Right: scatter plot of $\log_2$FC values of TESs against $\log_2$FC values of RNAseq for the same set of genes. The Spearman rank correlation coefficients (r_s) are given for each plot. Data are an average of three biological repeats, n = 3.

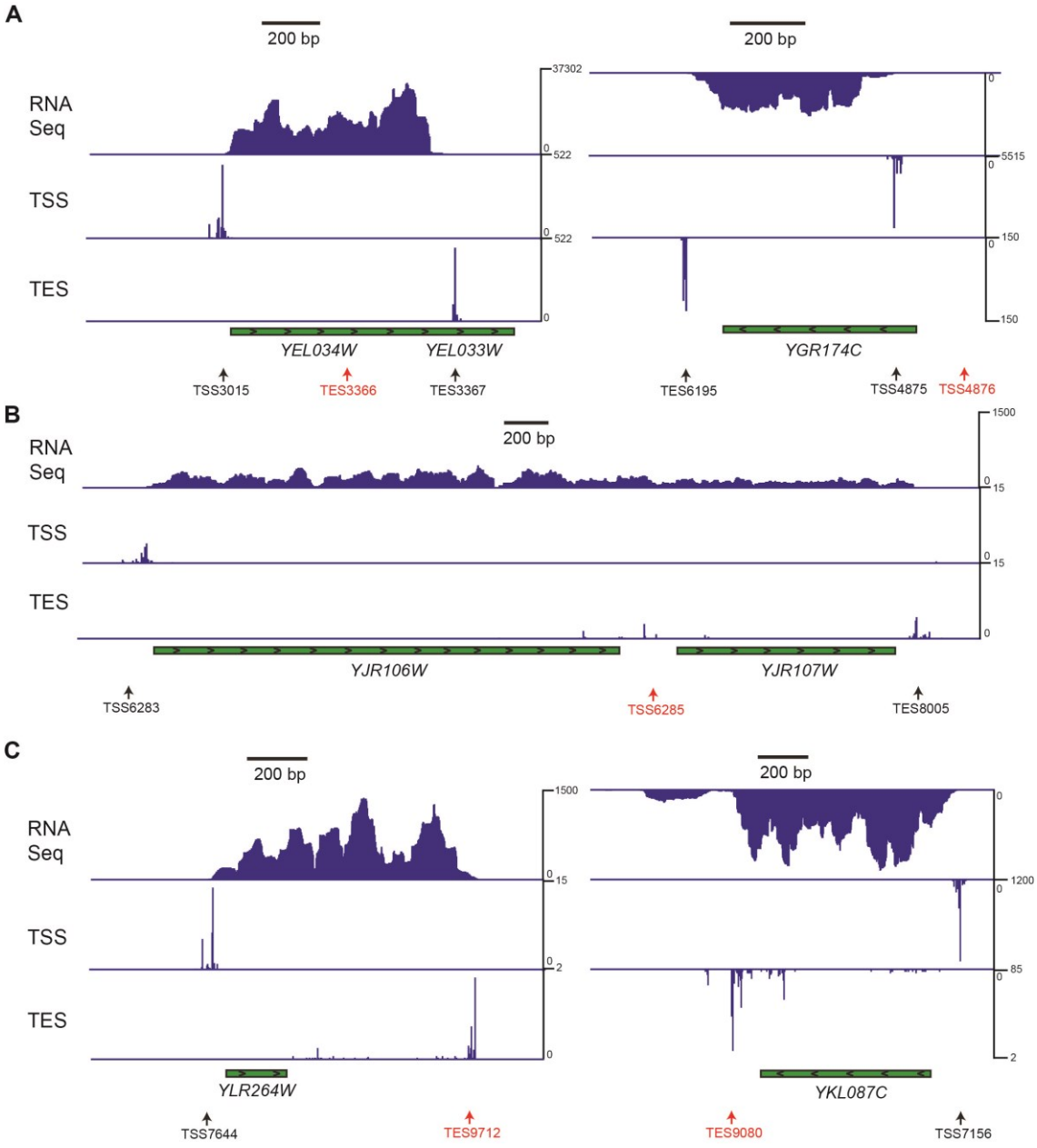


Figure 5.10 Visualizing "outlier" genes with poorly correlated TSS and TES reads

(A) Examples of mis-annotated ORFs resulting in poor correlation between TSS and TES signals. Left: The *YEL034W* ORF has a single TSS (TSS3015) and a wrongly assigned TES in red (TES3366). Very few reads were associated with TES3366. RNA seq data suggests that the actual TES is TES3367 which is in the middle of the neighbouring *YEL033W* ORF. Right: The *YGR174C* ORF has a single TES (TES6195) and a wrongly annotated TSS in red (TSS4876) with very few associated reads. RNA seq data suggests that the actual TSS is TSS4875. For simplicity, only features on the same strand (Watson or Crick) are shown. Scales showing values and distances (bp) are given. The data is an average of three independent repeats from the Spo 0h time point, n = 3.
(B) Example of read-through transcription resulting in poor correlation between TSS and TES signals. *YJR107W* has a single TSS shown in red (TSS6285) and a matching TES (TES8005). However TSS6285 has fewer than expected reads

possibly due to read-through transcription from the neighbouring *YJR106W* locus, as suggested by RNA seq. For simplicity, only features on the same strand (Watson or Crick) are shown. Scales showing values and distances (bp) are given. The data is an average of three independent repeats from the Spo 0h time point, n = 3.
(C) Examples of imprecise transcription termination resulting in poor correlation between TSS and TES signals. Left: *YLR264W* has a single annotated TES in red (TES9712), with fewer reads than expected possibly due to stochastic termination at earlier sites. Right: *YKL087C* has a single annotated TES in red (TES9080). However, the broad distribution of TES reads as well as RNA seq signals suggest imprecise termination events. For simplicity, only features on the same strand (Watson or Crick) are shown. Scales showing values and distances (bp) are given. The data is an average of three independent repeats from the Spo 0h time point, n = 3.

5.4 Discussion

In this chapter, I have described a procedure to profile the TSSs and TESs of mRNAs called TE-seq. TE-seq data can map TSSs and TESs at nucleotide resolution, allows for quantitative comparisons of TSS/TES usage and is reproducible across biological repeats. TE-seq, together with a rigorous pipeline to filter artifacts and assign peaks, is a useful approach to profile ncRNAs and transcript isoforms during budding yeast gametogenesis.

There are some advantages to profiling 5’ ends by TE-seq compared to other published methods such as 5’-CAGE and SMORE-Seq. TE-seq reads are much greater in length compared to 5’-CAGE tags, facilitating accurate mapping to the genome(Takahashi et al. 2012). While SMORE-seq is conceptually similar to TE-seq, the lack of an alkaline phosphatase treatment step in SMORE-seq means that intermediate degradation productions in addition to TSSs are also captured (Park et al. 2014).

Several different approaches exist for profiling the 3’ ends of mRNAs (Shepard et al. 2011;Zheng and Tian 2014;Lai et al. 2015;Wilkening et al. 2016). Some approaches such as PAL-seq yield information about the length of poly(A) tails, while and TAIL-seq gives poly(A) length together with the TES (Chang et al. 2014;Subtelny et al. 2014). However, those techniques are more complicated and unnecessary for the scope of my thesis, which is to profile TES usage. The PAS-

seq technique uses an anchored oligo d(T) primer for reverse transcription, and a custom oligo d(T) sequencing primer to avoid reading through the poly(A) tail(Shepard et al. 2011). However, imperfect oligo d(T) hybridization and the relatively long remaining A:T stretch can still lead to reduced sequencing quality and errors in TES identification. As a result, PAS-seq only identifies TESs within a range of +/- 9.4 nt (Shepard et al. 2011). TE-seq avoids this problem due to GsuI digestion which shortens the poly(A) stretch to around 4 nt in length. Wilkening *et al.* developed an orthogonal approach called 3'T-fill which accurately maps TESs (Wilkening et al. 2016). However unlike TE-seq, 3'T-fill has the disadvantage of requiring modifications to the Illumina clustering step, which necessitates additional coordination with external sequencing facilities to accommodate this custom procedure (Wilkening et al. 2016). As such, TE-seq is a relatively straightforward and accurate technique to profile the 3' ends of mRNAs.

Other approaches exist to profile both TSSs and TESs such as TIF-seq, GIS-PET, long read PacBio sequencing and nanopore sequencing. However in contrast to TIF-seq and GIS-PET, TE-seq does not rely on full length reverse transcription and can therefore quantify TSS/TES usage more reliably (Ng et al. 2005;Pelechano et al. 2014). Also, the complexity of the TIF-seq protocol (such as poor circularization of long mRNAs) results in much fewer usable reads and a poorer coverage compared to other 5' and 3' mapping approaches or TE-seq (Pelechano et al. 2014). The biases introduced due to differential ease of reverse transcription and amplification of full-length RNAs of varying lengths and base composition complicates quantitative analysis of TSS/TES usage from PacBio sequencing (Marinov 2017). Amplification-free direct RNA sequencing using nanopore technology is a promising future approach. However, the relatively poor sequencing depth due to technical inefficiencies still presents a challenge for comprehensive transcriptome profiling, especially in organisms with larger genomes (Marinov 2017). So far, TE-seq data show moderate to strong correlations between TSSs, TESs and RNA seq reads, facilitating quantitative analyses of gene regulation.

There are some important caveats when interpreting TE-seq data. One major limitation is that the 5' and 3' ends of any particular transcript are not matched. Therefore, any isoform prediction using TE-seq assumes that TES usage is stochastic, i.e. transcripts originating from a given TSS will use all available

downstream TESs proportionally. As such, novel isoform identification would be strengthened by complimentary techniques such as TIF seq or long read sequencing. It is also important to remember that transcriptional control is only one aspect of regulating gene expression. Integrating TE-seq with ribosome profiling data (rate of translation) and proteomic data (protein abundance) would provide a more complete understanding of how genes are regulated during cell fate transitions. Lastly, TE-seq only measures the levels of steady state RNAs, and cannot be used to study gene regulation by nascent transcription of unstable ncRNAs/isoforms.

In the next chapter, I use this optimized TE-seq protocol to study transcriptional regulation by ncRNAs and mRNA isoforms in a population of synchronously sporulating cells.

Chapter 6. Genome-wide analysis of gene regulation by mRNA isoforms and ncRNAs during gametogenesis

6.1 Abstract

Locus specific studies have shown that TSSs change in a stage specific manner throughout budding yeast gametogenesis. Till date, quantitative comparisons of TSS and TES usage throughout gametogenesis has been lacking. Here, I use TE-Seq to infer regulatory relationships between mRNA isoforms or ncRNAs on a genome wide level during gametogenesis. I show that there are over a thousand alternative TSSs which are upregulated in a developmental stage specific manner. The outcomes of overlapping upstream TSS transcription are complex, with some examples of repression, activation or more subtle effects on the coding TSS. TE-Seq analysis identified several genes whose repression by transcriptional interference depends on chromatin remodellers such as Set2, Set3 and Spt16. Finally, return to growth time courses suggest that gene repression by overlapping upstream mRNA isoforms or ncRNAs is highly dynamic. The results from this chapter reveal the complexities of gene regulation by mRNA isoforms and ncRNAs during gametogenesis on a genome-wide scale.

6.2 Introduction

In chapter 4, I described how transcription of the *NDC80*^{luti} RNA from an alternative upstream TSS led to repression of the *NDC80*^{ORF} coding TSS during early gametogenesis. Understanding how *NDC80* was regulated prompted me to ask how many other genes are regulated by luti RNA transcription during early or mid-late gametogenesis. What set of criteria can predict repression of the downstream coding TSS during the expression of an upstream, overlapping transcript? While northern blots could accurately distinguish between overlapping mRNA isoforms of different sizes, this low throughput technique is not suitable for genome-wide analyses. Furthermore, there has been no study focused on quantitative comparisons of TSS/TES usage to infer regulatory relationships

between mRNA isoforms or ncRNAs on a genome wide level during gametogenesis. These considerations motivated the development of an optimized approach to profile the 5' and 3' ends of mRNA isoforms and ncRNAs in chapter 5, called TE-seq. Here, I utilize TE-seq combined with mRNA seq to answer the abovementioned research questions.

Transcriptional interference (TI) of downstream TSSs by the expression of upstream overlapping isoforms/ncRNAs is also dependant on the H3K36 methyltransferase Set2, the histone deacetylase Set3 and the histone chaperone Spt16 (Hainer et al. 2011;van Werven et al. 2012;Chia et al. 2017). For example, transcriptional interference of *IME1* and *NDC80*^{ORF} is compromised in the *set2Δset3Δ* mutant. On the other hand, repression of *SER3* expression by transcription of the *SRG1* ncRNA depends on Spt16 but not Set2 or Set3. In this chapter, I utilize TE-Seq in mutant and control cells in early gametogenesis, to identify coding TSSs whose repression by TI is dependent on these chromatin modifiers.

TI of *NDC80*^{ORF} is rapidly reversible so that cells can resume the mitotic cell cycle if nutrients are re-introduced to the environment, prior to committing to meiotic divisions. How many genes are regulated in a dynamic manner similar to *NDC80*? To answer this question, TE-seq analysis was conducted on return to growth (RTG) time courses. Any gene which is regulated in a similar fashion to *NDC80* would show a rapid downregulation of an alternative upstream TSS, accompanied by an upregulation of the downstream coding TSS.

The results presented in this chapter are a genome-wide survey of gene regulation by mRNA isoforms and ncRNAs during budding yeast gametogenesis. Here, I show that more than 1000 genes have alternative TSSs which are upregulated by at least two fold during gametogenesis. Increased expression from some alternative TSSs were associated with repression while other cases were associated with activation or subtler changes in expression from the downstream coding TSSs. A relatively strong upstream promoter as well as a minimum inter-promoter distance were required for gene repression by mRNA isoforms or ncRNAs. In addition, the repression of some downstream coding TSSs also

required chromatin remodellers such as Set2, Set3 and Spt16. Lastly, most alternative TSSs associated with upstream repressive transcription are dynamically regulated during return to growth. These findings provide a genome-wide picture of how transcription of overlapping ncRNAs and mRNA isoforms contribute to the timely and dynamic regulation of genes during a developmental program.

6.3 Results

6.3.1 Applying TE-seq to a high resolution gametogenesis time course

Time course experiments were designed to obtain a high resolution profile of transcript isoforms and ncRNAs during gametogenesis **(Figure 6.1)**. In order to obtain a high level of synchrony throughout the yeast gametogenesis program, cells used here harboured both *IME1* fused to *CUP* promoter and *NDT80* expressed from the *GAL* promoter together with Gal4 fused to the estrogen receptor (*pCUP-IME1* and *GAL4.ER pGAL-NDT80*) (see chapter 3 for more details). I combined this genetic system as well as an optimized YPD to SPO synchronization procedure to ensure that cells initiated and underwent gametogenesis in a highly synchronous manner (see chapter 3 for more details). Controlling the induction of the Ime1 and Ndt80 transcription factors also allowed for a clear separation of the time course into the “pre-meiotic”, “early gametogenesis” and “mid-late gametogenesis” stages. Samples were also taken every hour from 2-9 h for high coverage of all transcription events throughout gametogenesis. The pre-meiotic time points were 0 and 2h. The early gametogenesis time points were from 3-6h and the mid-late gametogenesis time points were from 7-9h. Three different control samples were included to account for the effects of nutrient starvation on the transcriptome. Controls sampled at Spo 2h and Spo 3h (Spo 3M, no $CuSO_4$) represent the effects of nutrient starvation on the cells, without initiation of the gametogenesis program. On the other hand, the Spo 7h control (Spo 7M, no β-estradiol) represents the effects of nutrient starvation in cells arrested in meiotic prophase prior to meiotic divisions. Robust, stage-specific characterization and quantification of transcript isoforms and ncRNAs relies on a high degree of synchrony throughout gametogenesis. To ensure that this was the case, quality control measures were introduced for all time courses that were used in this study **(Figure 6.1)**. For the early meiotic time points, the synchrony of DNA

replication was assessed by flow cytometry. Upon induction of *IME1* at Spo 2h, cells initiated DNA replication and completed it within 1 hour **(Figure 6.2 A)**. For the mid-late meiotic time points, the kinetics of meiotic divisions were measured. After the addition of β-estradiol at Spo 6h, cells initiated meiotic divisions at Spo 7h and at least 90% of them complete two rounds of nuclear divisions by Spo 10h **(Figure 6.2 B-D).** Flow cytometry and nuclei counting data showed that each time point corresponded to the different stages of gametogenesis **(Figure 6.2)**. Spo 3h is prior to DNA replication, Spo 4h is during DNA replication in S phase, Spo 5-6h is during meiotic prophase, Spo 7h is prior to the onset of meiotic divisions, Spo 8h consists of cells mostly in meiosis I while Spo 9h consists of cells mostly in meiosis II. Taken together, the features and quality controls in these time course experiments facilitate reliable profiling of transcripts in a stage specific manner throughout gametogenesis.

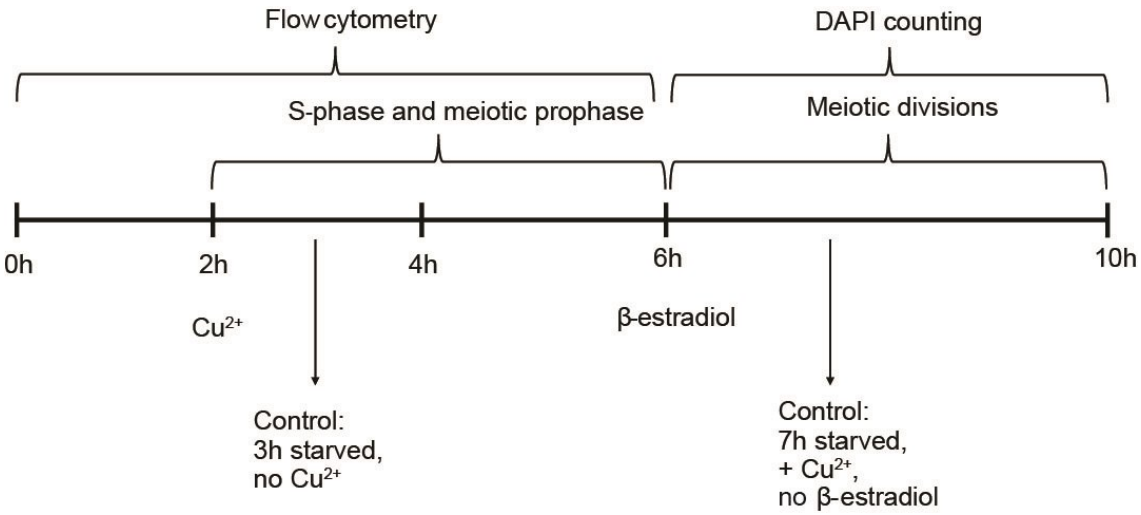


Figure 6.1 Schematic of a high resolution gametogenesis time course

For each time course experiment, diploid cells harbouring both *IME1* fused to *CUP* promoter and *NDT80* expressed from the *GAL* promoter together with Gal4 fused to the estrogen receptor (*pCUP-IME1* and *GAL4.ER pGAL-NDT80*) (FW2795) were grown in YPD overnight. Saturated cultures were pelleted by centrifugation, washed with sterile water and resuspended to a final OD_{600} of 2.5 in SPO; a portion of cells were also collected at this time point (0h). Cells were collected hourly from 2 to 9 h for mRNA seq and TE-seq analysis. 50 μM $CuSO_4$ was added 2 h after the cells were transferred to SPO to induce early gametogenesis. As a control, a separate flask of cells were not treated with $CuSO_4$ and were collected at the 3 h time point (Control: 3h starved). Between 2h and 6h, samples were also taken to assess the synchrony of S-phase by flow cytometry. 1 μM β-estradiol was added 6 h after transfer to SPO to initiate exit from meiotic prophase and commitment to meiotic divisions. As a control, a separate flask of

cells which were previously treated with $CuSO_4$, were not treated with β-estradiol and were collected at the 7 h time point (Control: 7h starved). Samples were also collected throughout the time course to measure the rate of meiotic divisions by counting DAPI masses. Three independent time course experiments were used in this chapter.

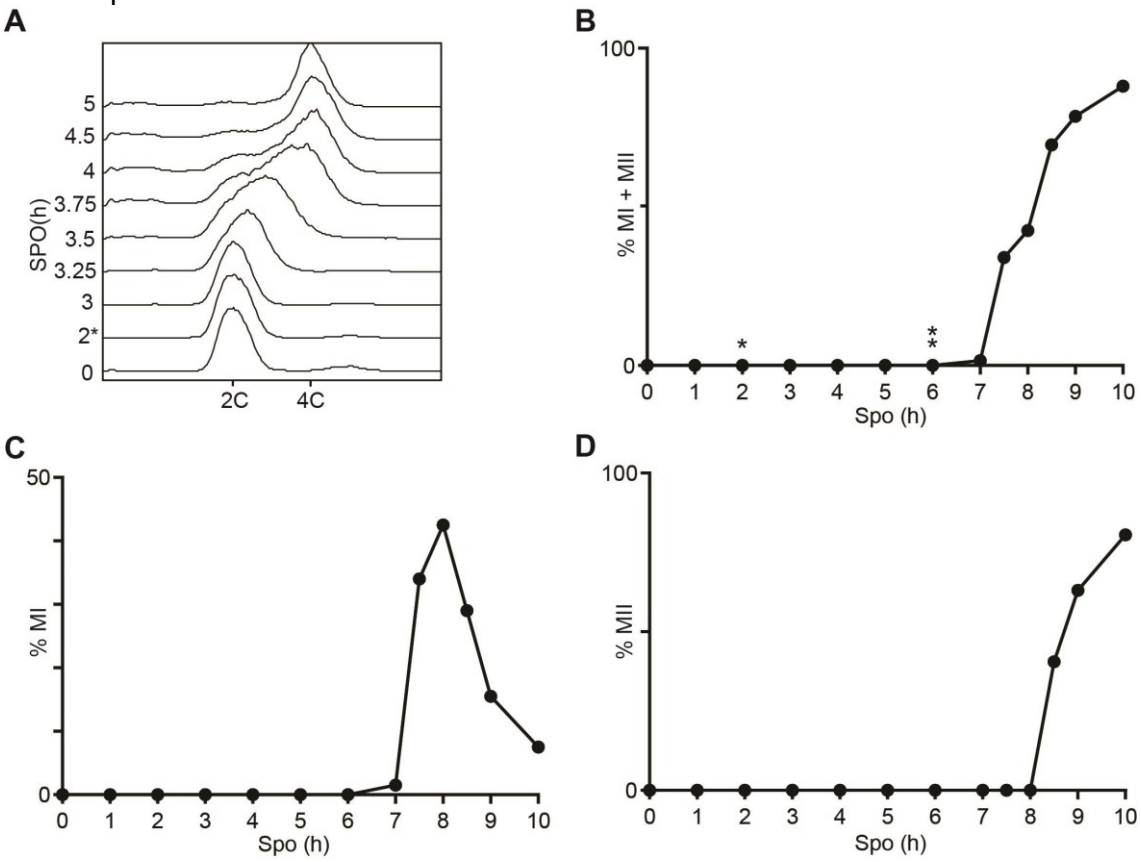


Figure 6.2 Assessing the synchrony of gametogenesis for each time course

(A) Flow cytometry analysis of DNA content of *pCUP-IME1* and *GAL4.ER pGAL-NDT80* cells (FW2795) that were induced to sporulate. Samples were taken at indicated time points, fixed, and DNA content was measured by propidium iodide staining; 50 μM $CuSO_4$ was added 2 h after the cells were transferred to SPO (*). At least 50,000 cells were analyzed at each time point. **(B)** Kinetics of meiotic divisions in the same cells (FW2795). Samples were taken at the indicated time point, fixed in ethanol, nuclei were stained with DAPI, and DAPI masses were counted. Cells that harboured two, three, or four DAPI masses were classified as cells undergoing meiosis I or meiosis II (% MI + MII). For each time point, at least 200 cells were counted. **(C)** Similar to B except that the percentages of bi-nucleate cells are shown (% MI). **(D)** Similar to B except that the percentages of tri-, and tetra-nucleate cells are shown (% MII). All panels in this figure are representative of at least three independent repeats, n = 3.

In order to increase the power of our study to detect differentially expressed TSSs and TESs throughout gametogenesis, I included three independent, biological repeats of the master time course as described above. I next examined if

the results were reproducible between the 3 repeats. After mapping the usable reads to the genome, the TSS, TES and regular mRNA seq reads for each repeat (A, B and C) clustered according to defined stages/time points **(Figure 6.3 A-C)**. The pre-meiotic stages Spo 0h, Spo 2h and the Spo 3M starvation control clustered as one group. The early meiotic time points up to meiotic prophase arrest, Spo 3-6h and the Spo 7M starvation control also clustered together. As expected, the mid-late meiotic time points, Spo 7-9h were distinct from the previous two clusters. My TE-seq and mRNA seq data show that the three, independent biological repeats were similar to each other and can be used for downstream analyses to detect differentially expressed TSSs/TESs with high confidence.

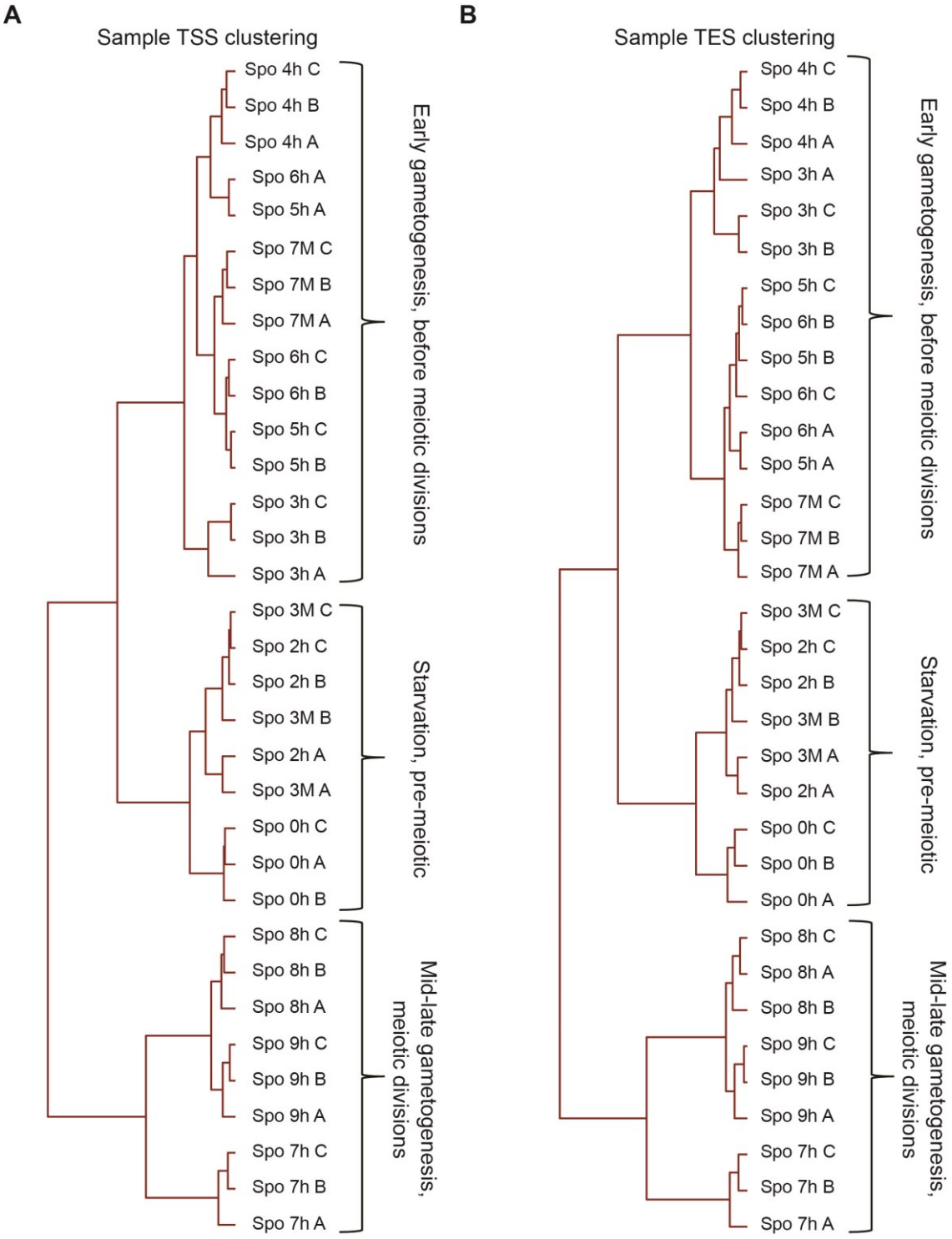

A
Sample TSS clustering
Spo 4h C
Spo 4h B
Spo 4h A
Spo 6h A
Spo 5h A
Spo 7M C
Spo 7M B
Spo 7M A
Spo 6h C
Spo 6h B
Spo 5h C
Spo 5h B
Spo 3h C
Spo 3h B
Spo 3h A
Spo 3M C
Spo 2h C
Spo 2h B
Spo 3M B
Spo 2h A
Spo 3M A
Spo 0h C
Spo 0h A
Spo 0h B
Spo 8h C
Spo 8h B
Spo 8h A
Spo 9h C
Spo 9h B
Spo 9h A
Spo 7h C
Spo 7h B
Spo 7h A
Early gametogenesis, before meiotic divisions
Starvation, pre-meiotic
Mid-late gametogenesis, meiotic divisions
B
Sample TES clustering
Spo 4h C
Spo 4h B
Spo 4h A
Spo 3h A
Spo 3h C
Spo 3h B
Spo 5h C
Spo 6h B
Spo 5h B
Spo 6h C
Spo 6h A
Spo 5h A
Spo 7M C
Spo 7M B
Spo 7M A
Spo 3M C
Spo 2h C
Spo 2h B
Spo 3M B
Spo 3M A
Spo 2h A
Spo 0h C
Spo 0h B
Spo 0h A
Spo 8h C
Spo 8h A
Spo 8h B
Spo 9h C
Spo 9h B
Spo 9h A
Spo 7h C
Spo 7h B
Spo 7h A
Early gametogenesis, before meiotic divisions
Starvation, pre-meiotic
Mid-late gametogenesis, meiotic divisions

Figure 6.3 TE-seq and mRNA seq data for the three independent time courses for are highly reproducible

(A) Clustering tree of TSS reads from different time points. A, B and C denote the three independent repeats of the time course experiment. The Spo 3M samples

refer to cells not treated with Cu^{2+} and collected at the 3 hr time point, as 3h starvation controls.The Spo 7M samples refer to cells treated with Cu^{2+}, but not treated with β-estradiol and collected at the 7 hr time point, as 7h starvation controls. The filtering, mapping and clustering of TE-seq reads were done by Cai, Li from the Luscombe lab. **(B)** Clustering tree of TES reads from different time points. The filtering, mapping and clustering of TE-seq reads were done by Cai, Li from the Luscombe lab. **(C)** Clustering tree of mRNA seq reads from different time points. The filtering, mapping and clustering of mRNA seq reads were done by Cai, Li from the Luscombe lab.

The transcriptional program of yeast gametogenesis is characterized by "waves" of expression of different genes at different stages (Primig et al. 2000). To assess whether TE-seq data matched mRNA seq data, I compared the TSS and TES signals of some known "early" and "mid-late" meiotic genes to that of mRNA seq. These meiotic genes were selected from an early microarray study by Primig *et al.* (Primig et al. 2000). I examined a subset of early meiotic genes including those involved in DNA replication (*RFA1* and *RFA2*), double strand break formation and meiotic recombination (*SPO11, REC102*, *REC104* and *MSH5*), double strand break repair (*RAD52* and *MEI5*), synaptonemal complex assembly and sister chromatid cohesion (*RED1, SCC2* and *SPO13*). The subset of mid-late genes included meiotic B-type cyclins (*CLB3* and *CLB4*), the anaphase promoting complex (*APC2, APC5* and *CDC27*), genes involved in regulating meiotic nuclear divisions (*PDS1, CDC5, CDC14, KAR1* and *SPO12*) and spore formation (*SPR1, SPR3, SPR28* and *SSP2*). As, expected, my mRNA seq data showed that the early meiotic genes were not expressed in pre-meiotic cells, but were rapidly activated during the early time points, Spo 3-6h **(Figure 6.4 A**, left**)**. Notably, *RFA1* and *RFA2* levels peaked at an earlier time (Spo 3h) compared to the other early meiotic genes, consistent with their role in DNA replication prior to the other early meiotic events like recombination. After a peak of expression in early meiosis, these genes were repressed or expressed at much lower levels during mid-late gametogenesis, Spo 7-9h **(Figure 6.4 A**, left**)**. Expression of most of the mid-late meiotic genes also followed a stereotypical pattern of expression, with little to no expression during the early time points and peak expression between Spo 7-9h **(Figure 6.4 A**, right**)**. The exceptions were *CDC14* and *PDS1,* perhaps because *CDC14* is further regulated by controlling its subcellular localization before meiotic divisions and *PDS1* is also involved in meiotic recombination (Cooper et al. 2009;Fox et al. 2017). Importantly, both TSS **(Figure 6.4 B)** and TES **(Figure 6.4 C)** data agree with mRNA seq data

(Figure 6.4 A), providing evidence that TE-seq can reliably quantify stage specific changes in gene expression. This supports the use of TE-seq data to measure differentially expressed TSSs/TES usage at different stages of gametogenesis.

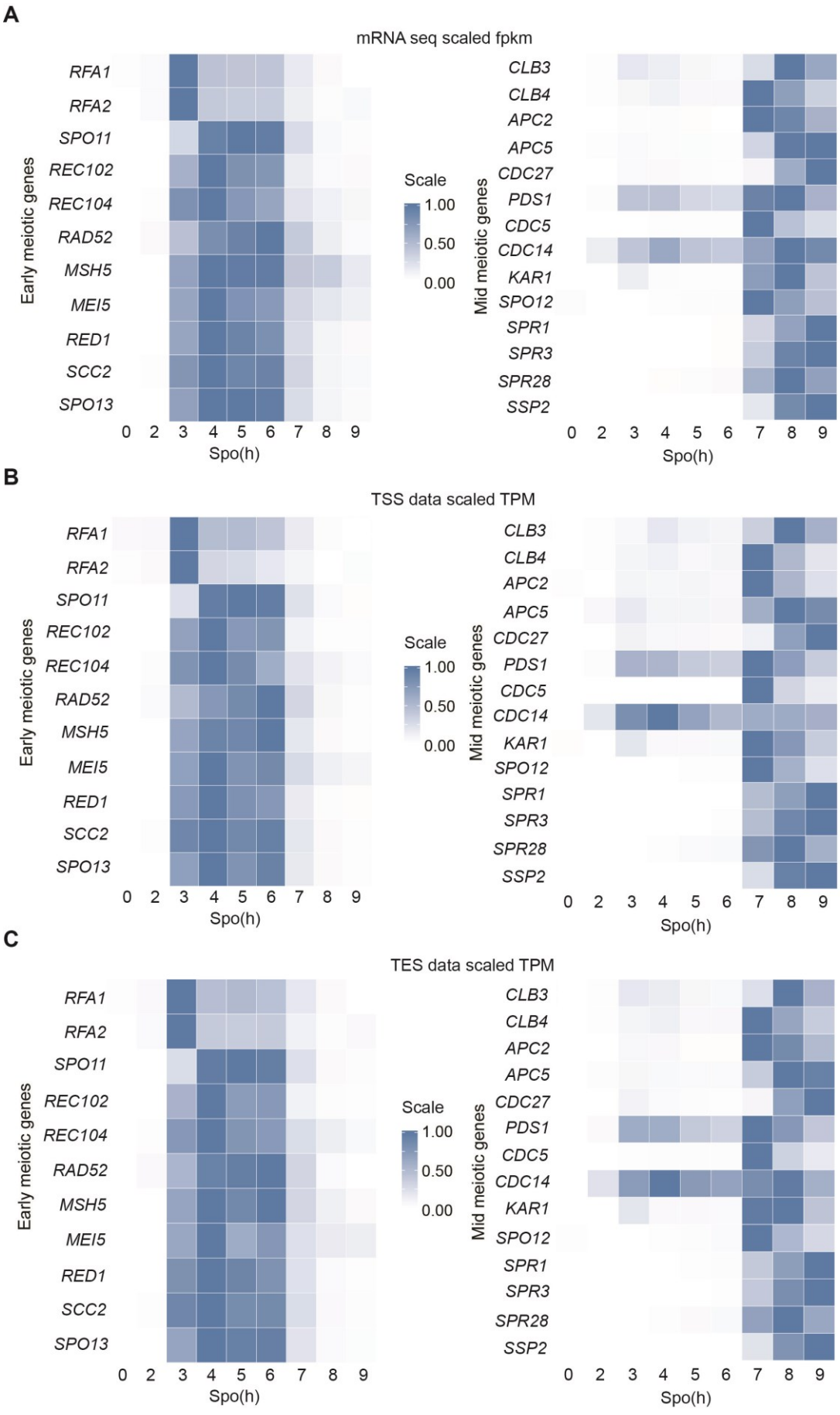

A
mRNA seq scaled fpkm
Early meiotic genes
RFA1
RFA2
SPO11
REC102
REC104
RAD52
MSH5
MEI5
RED1
SCC2
SPO13
Scale
1.00
0.50
0.00
Mid meiotic genes
CLB3
CLB4
APC2
APC5
CDC27
PDS1
CDC5
CDC14
KAR1
SPO12
SPR1
SPR3
SPR28
SSP2
0 2 3 4 5 6 7 8 9
Spo(h)
B
TSS data scaled TPM
C
TES data scaled TPM

Figure 6.4 TE-seq is consistent with mRNA seq for a set of known meiotic genes

(A) Expression heatmap of mRNA seq read counts of 11 early meiotic genes (left) and 14 mid-late meiotic genes (right) across different time points in gametogenesis. Early gametogenesis is represented by Spo 3-6 hrs. Mid-late gametogenesis is represented by Spo 7-9 hrs. Average read counts for each gene from 3 independent repeats for each time point, n = 3, were normalized between 0 and 1 across the time course using min-max normalization. **(B)** Same as A, except that TSS reads are represented. 3 independent repeats for each time point are represented, n = 3 **(C)** Same as A, except that TES reads are represented. 3 independent repeats for each time point are represented, n = 3

6.3.2 TE-seq identifies regulation by luti RNA transcription at the *NDC80* locus

Next, I further assessed the performance of TE-seq by analysing mRNA seq and TSS data at the *NDC80* locus at different time points in gametogenesis **(Figure 6.5)**. Consistent with my findings reported in chapter 4, Spo 0h (pre-meiotic) cells expressed *NDC80* from a proximal TSS, labelled $NDC80^{ORF}$ TSS. As expected, during Spo 4h (S phase) and 5h (meiotic prophase), *NDC80* was expressed from a distal TSS, labelled $NDC80^{luti}$ TSS and very few reads mapped to the $NDC80^{ORF}$ TSS. This is consistent with a switch to $NDC80^{luti}$ transcription and a repression of $NDC80^{ORF}$ transcription during prophase. At Spo 8h (meiotic divisions), $NDC80^{luti}$ transcription decreased while $NDC80^{ORF}$ was induced.

Analysis of the neighbouring *PAN6* locus also revealed that the *PAN6* gene was transcribed in the divergent direction upstream of *NDC80* **(Figure 6.5)**. In Spo 0h (pre-meiotic cells), *PAN6* levels were relatively low and an overlapping anti-sense transcript was detected. This *PAN6* anti-sense transcript might interfere with *PAN6* transcription, although this was not formally tested. During early gametogenesis, this anti-sense transcript was not expressed and *PAN6* levels increased. *PAN6* levels were correlated with $NDC80^{luti}$ levels and these two transcripts appeared to be controlled from a bi-directional promoter region. Interestingly, an upstream alternative TSS for *PAN6,* labelled *PAN6* alt. TSS was also observed during Spo 4h and 5h, and it was co-upregulated with the proximal *PAN6* TSS. It is not clear from TE-seq data alone if *PAN6* transcription during early gametogenesis had an effect on $NDC80^{luti}$ transcription. However, cloning the *NDC80* locus to an ectopic site did not appear to change the pattern of luti

expression during gametogenesis (Chen et al. 2017). Taken together, TE-seq accurately reflects the patterns of isoform regulation at a well-defined locus.

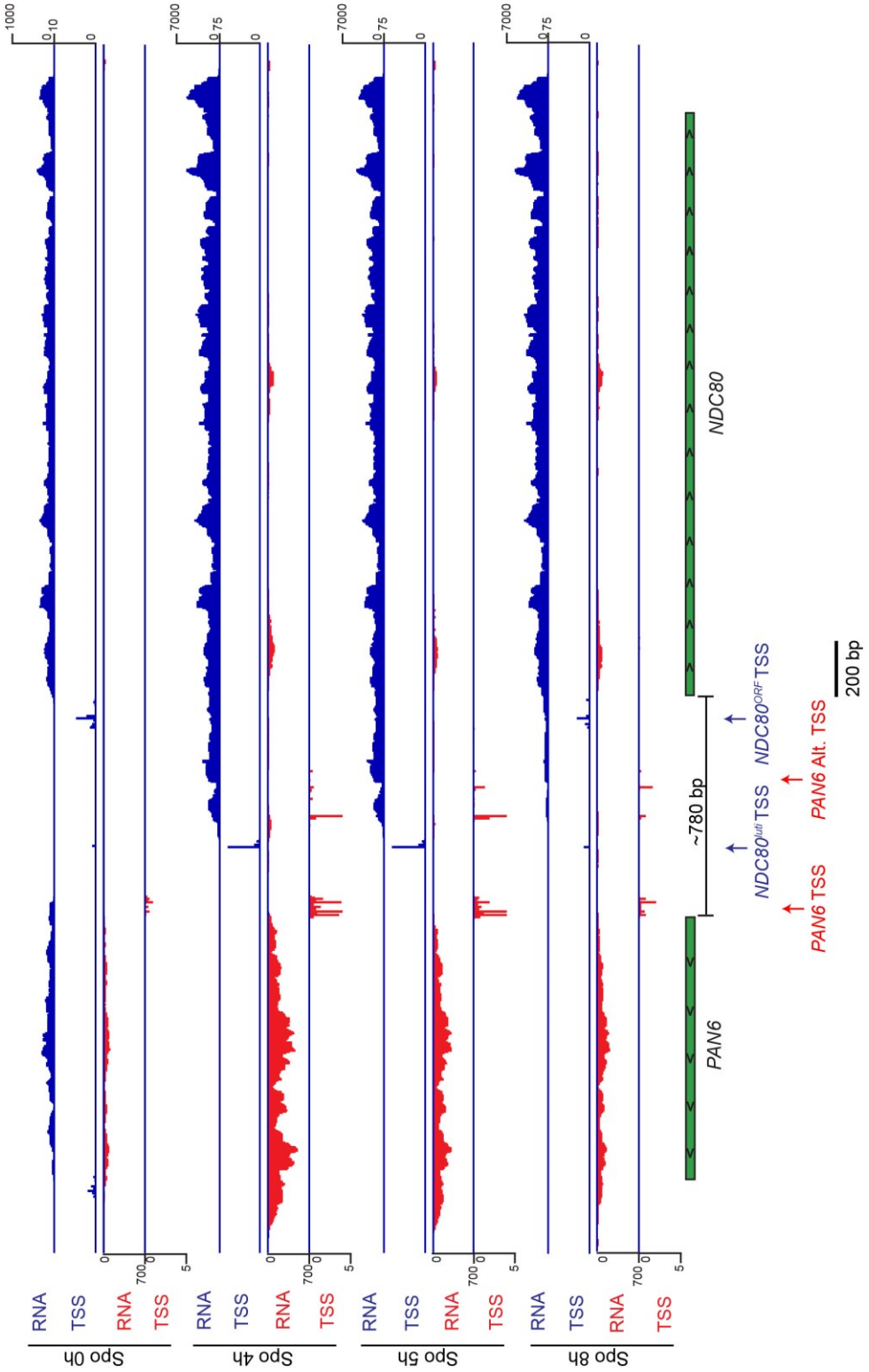


Figure 6.5 TE-seq analysis reveals alternative *NDC80* promoter usage in cells undergoing gametogenesis

mRNA-seq and TSS data at the *NDC80* and *PAN6* locus in *pCUP-IME1* and *GAL4.ER pGAL-NDT80* cells (FW2795) at different time points in a

gametogenesis time course. The time points are Spo 0h (pre-meiotic), Spo 4h (S phase), Spo 5h (prophase) and Spo 8h (meiotic divisions). mRNA seq FPKM values and TSS TPM values are plotted for both strands (red and blue). *NDC80* isoforms are transcribed in the "sense" orientation (blue). About 780 bp away, the neighbouring *PAN6* gene is transcribed in the "anti-sense" orientation (red). Canonical and alternative TSSs for both *NDC80* and *PAN6* are also labelled. Scales for values and distance (bp) are shown. Data shown here is an average of three independent repeats, n = 3.

TE-seq can also be used to analyse TES usage, as demonstrated for the *NDC80* locus **(Figure 6.6)**. Both *NDC80*^{ORF} and *NDC80*^{luti} predominantly use the same TES (compare Spo 0h and Spo 5h). This is consistent with *NDC80*^{luti} being a long transcript isoform that overlaps the *NDC80*^{ORF} TU. Interestingly, the presence of an intermediate TES peak within the *NDC80* ORF region suggests that a fraction of luti transcripts use this alternative TES. Taken together, TE-seq can be used to identify TSS and TES usage at a well-defined locus during gametogenesis.

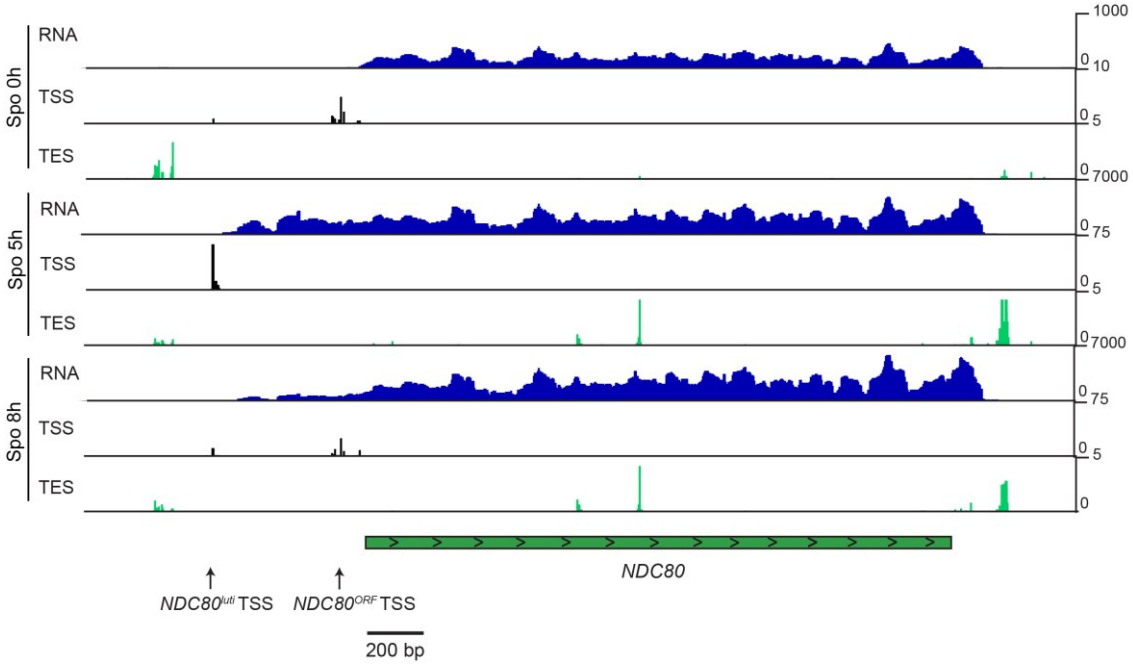


Figure 6.6 TE-seq analysis reveals *NDC80* TSS and TES usage in cells undergoing gametogenesis

mRNA-seq, TSS and TES data at the *NDC80* locus in *pCUP-IME1* and *GAL4.ER pGAL-NDT80* cells (FW2795) at different time points in a gametogenesis time course. The time points are Spo 0h (pre-meiotic), Spo 5h (prophase) and Spo 8h (meiotic divisions). mRNA seq FPKM values (blue), TSS TPM values (black) and TES TPM values (green) are plotted. For simplicity, only features on the "sense" (+) strand are shown. Canonical and alternative TSSs for *NDC80* are also labelled. Scales for values and distance (bp) are shown. Data shown here is an average of three independent repeats, n = 3.

6.3.3 TE-seq identifies stage specific upregulation of alternative TSSs and transcripts with longer 5' UTRs during gametogenesis

After establishing that the sequencing data were reliable and reproducible, I analysed the TE-seq data to identify alternative TSSs that are upregulated during gametogenesis. The initial steps of this analysis involved associating genes with a major coding TSS in pre-meiotic cells (Spo 0 and 2h time points). For a TSS to be classified as a "coding TSS" of a gene, it had fulfil certain criteria. First, the putative coding TSS had to be within 1000 nt upstream of the annotated ORF on the same DNA strand. Second, it had to have the largest number of reads amongst all possible TSSs for the same gene during the pre-meiotic time points. Third, each putative coding TSS can also only be assigned to at most one gene in the same orientation. All other possible TSSs of a particular gene were classed as "alternative TSSs" Using this criteria, 5279 genes were identified which had detectable expression in pre-meiotic cells, together with their major coding TSSs **(Figure 6.7 A)**. Pair-wise comparisons of meiotic time points to a reference time point were subsequently used for differential expression of TSSs (see *materials and methods* for details). Each time point corresponding to early gametogenesis (Spo 3-6h) was compared to pre-meiotic cells immediately prior to *IME1* induction (Spo 2h). Each time point corresponding to mid-late gametogenesis (Spo 7-9h) was compared to cells in meiotic prophase, prior to *NDT80* induction (Spo 6h). All results were also compared to the Spo 3M and Spo 7M samples to control for starvation signals. Out of 5279 genes, 1295 of them had at least one alternative TSS, distinct from the coding TSS; these alternative TSSs were upregulated during at least one meiotic time point (log2FC ≥ 1, false discovery rate, FDR < 0.05) **(Figure 6.7 A)**. Of these 1295 candidates, 870 genes had alternative TSSs upregulated during at least one time point in early gametogenesis, while 770 genes had alternative TSSs upregulated during at least one time point in mid-late gametogenesis. Members of two groups of genes were mostly distinct from each other, with 345 overlapping candidates **(Figure 6.7 B)**. Taken together, TE-seq shows that more than 1000 genes have meiosis specific alternative TSSs, which could be due to the transcription of a mRNA isoform or a proximal ncRNA.

I next examined how frequently do genes have multiple alternative TSSs and hence multiple putative isoforms during gametogenesis. Out of the 1295 genes identified above, over half of them (654) had one alternative TSS upregulated during early gametogenesis. 1885 alternative meiotic TSSs were associated with 1295 genes and most genes had three or fewer alternative meiotic TSSs specifically upregulated in either early or mid-late gametogenesis **(Figure 6.7 C and D)**. Most of these alternative TSSs were upregulated in distinct stages of gametogenesis (early or mid-late), with 380 of them showing co-regulation in both stages. These results suggest that the use of alternative TSSs during meiosis is a widespread phenomenon and occurs in a stage specific manner, expanding the diversity of transcripts produced at each loci.

Earlier studies in non-meiotic cells have shown that many ncRNAs are transcribed in an anti-sense orientation from coding genes; these divergent ncRNA transcripts arise due to promiscuous initiation from the NFR/NDR of bi-directional coding gene promoters (Neil et al. 2009;Xu et al. 2009;Yassour et al. 2010). Likewise, the expression of mRNA isoforms/ncRNAs from alternative meiotic TSSs could be driven by divergent transcription from the promoters of neighbouring genes. To test this, I examined the genomic features within 500 bp upstream of the 1885 upregulated meiotic TSSs. At both the early or mid-late time points, approximately half of the alternative TSSs did not have a neighbouring, divergent anti-sense TSS within 500 bp **(Figure 6.7 F and G)**. Within the group of alternative TSSs with a neighbouring divergent transcript, approximately 50-60% of them were coordinately upregulated with the neighbouring transcript. The remainder were upregulated independently of their divergent neighbours during meiosis. In summary, this shows that a large fraction of meiotic alternative TSSs are not products of bi-directional transcription from active promoters of neighbouring meiotic genes.

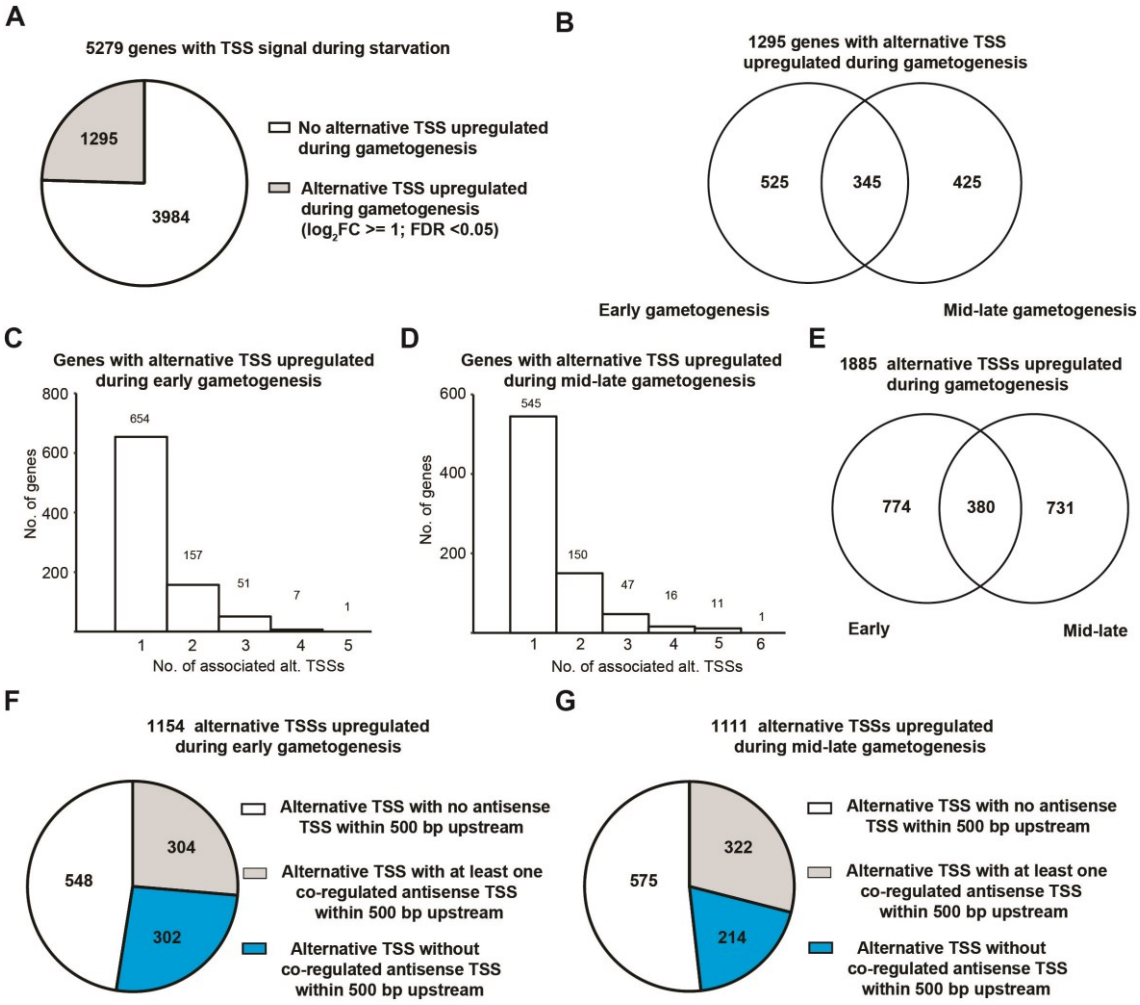


Figure 6.7 Thousands of genes have alternative TSSs upregulated during different stages of gametogenesis

(A) Proportion of genes with either none or at least one alternative TSS upregulated during gametogenesis. These 5279 genes were chosen on the basis of their being expressed in pre-meiotic saturated cultures (Spo 0h and 2h). TSSs were associated with a gene as long as they are in the same orientation, are within 1000 nt upstream of the annotated CDS and are not assigned to another gene on the same strand. The major coding TSS of a gene was defined as those with the highest expression in pre-meiotic cells. Alternative TSSs are by definition, called as different CAGEr peaks from the coding TSSs. Upregulation of a TSS was defined as those with a $\log_2$ fold change ≥ 1 and false discovery rate (FDR) < 0.05. For the early meiotic time points (Spo 3-6h), fold changes were calculated relative to Spo 2h. For the mid-late meiotic time points (Spo 7-9h), fold changes were calculated relative to Spo 6h. **(B)** Categorizing genes by the timing of when their alternative TSSs are upregulated during gametogenesis **(C)** Distribution of number of alternative TSSs associated to the same gene, during early gametogenesis. **(D)** Distribution of number of alternative TSSs associated to the same gene, during mid-late gametogenesis. **(E)** Categorizing alternative TSSs by the timing of when they are upregulated during gametogenesis. **(F)** Categorizing 1154 alternative TSSs upregulated during early gametogenesis, by a neighbouring gene feature or

activity. **(G)** Categorizing 1111 alternative TSSs upregulated during mid-late gametogenesis, by a neighbouring gene feature or activity.

My previous work in chapter 4 described a long transcript isoform of the kinetochore subunit Ndc80 (*NDC80*^{luti}). As its name suggests, *NDC80*^{luti} has a longer 5' UTR than the coding mRNA (*NDC80*^{ORF}), because luti RNA transcription is initiated from an alternative TSS upstream of the *NDC80*^{ORF} core promoter. I tested if transcription from alternative TSSs during meiosis tends to generate transcripts with longer 5' UTRs than their coding isoforms, similar to the *NDC80* example. Here, the 5' UTR length is defined as the distance in nucleotides (nt), from the start codon of a gene to the "peak" of a given TSS CAGEr tag. According to TE-seq data, the 5' UTR of *NDC80*^{luti} is 518 nt, while that of *NDC80*^{ORF} is 71 nt. Analysis of the aforementioned 5279 genes showed that the median 5' UTR length was 49 nt as measured from the coding TSS **(Figure 6.8 A)**. Most genes had a 5' UTR length lesser than or equal to 150 nt **(Figure 6.8 A)**. In contrast, the 5' UTR lengths of alternative meiotic isoforms was more distributed across larger sizes, with a median length of 187 and 197 nt for early and mid-late time points respectively **(Figure 6.8 B and D)**. To look for examples resembling *NDC80*^{luti}, the analysis was further restricted to upstream alternative TSSs. Most alternative TSSs were located upstream of the coding TSS, with 768 out of 1154 candidates during early meiosis and 818 out of 1111 candidates during mid-late meiosis. Alternative upstream isoforms transcribed during gametogenesis tend to have much longer 5' UTRs than their coding transcripts, with a median 5' UTR length of 314 and 273 nt for early and mid-late time points respectively **(Figure 6.8 C and E)**. Taken together, the data suggest that regulated extension of 5' UTR length occurs at numerous loci during gametogenesis.

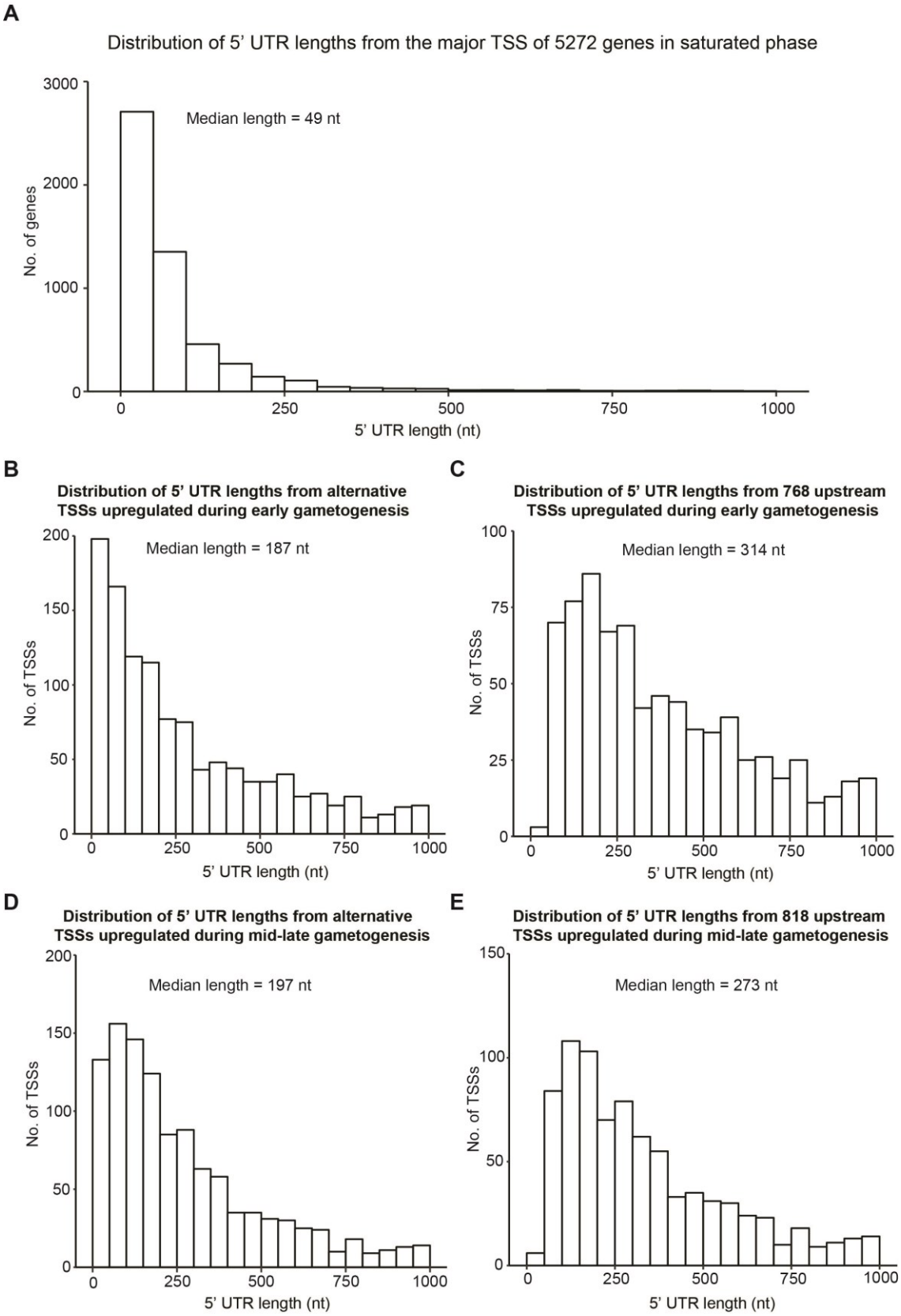


Figure 6.8 Alternative isoforms transcribed during gametogenesis tend to have longer 5' UTRs

(A) Distribution of 5’ UTR lengths of coding mRNAs from 5279 genes expressed in pre-meiotic saturated cultures (Spo 0h and 2h). 5’ UTR length here is defined as the distance in nt, from the start codon of the gene to the “peak” of a given TSS CAGEr tag. The major coding TSS of a gene was defined as those with the highest expression in pre-meiotic cells. The median 5’ UTR length is given in the figure. **(B)** Distribution of 5’ UTR lengths of alternative transcript isoforms or ncRNAs which are upregulated during early gametogenesis, relative to Spo 2h. **(C)** Distribution of 5’ UTR lengths of transcript isoforms or ncRNAs which are initiated from a TSS that is strictly upstream of the coding TSS and upregulated during early gametogenesis relative to Spo 2 h (pre-meiotic cells). **(D)** Distribution of 5’ UTR lengths of alternative transcript isoforms or ncRNAs which are upregulated during mid-late gametogenesis, relative to Spo 6h (meiotic prophase). **(E)** Distribution of 5’ UTR lengths of transcript isoforms or ncRNAs which are initiated from a TSS that is strictly upstream of the coding TSS and upregulated during mid-late gametogenesis relative to Spo 6h (meiotic prophase).

6.3.4 Gene repression by upstream transcripts tends to require relatively high levels of transcription and a minimum distance between two promoters

I next examined what set of criteria or features were common to upstream transcripts whose transcription was associated with repression of a downstream coding TSS. Previous studies on transcriptional interference have demonstrated that stronger upstream promoters repress downstream promoters with greater efficacy (Callen et al. 2004;Shearwin et al. 2005;Pande et al. 2018). To examine whether the levels of the upstream TSS repress downstream promoter activity, I estimated the relative upstream promoter strength by dividing the number of reads associated with the upstream TSS by the number of reads associated with the coding TSS, measured from the same time point **(Figure 6.9 A)**. Hence a value > 1 would mean that the read counts from the upstream TSS were greater than that from the coding TSS at the same time point. In this analysis, I focused on those genes which showed a greater than 2-fold repression of the coding TSS (log_2 fold change ≤ -1) when the alternative meiotic TSS was upregulated (log_2 fold change ≥ 1). Using the Spo 3h time point as an example, the relative upstream promoter strengths of each alternative TSS was calculated and grouped into four quartiles (Q_1 to Q_4). Q_1 denotes the range of values ≤ the 25th percentile of calculated values for the relative promoter strength at Spo 3h (smaller than 0.030). Q_2 denotes the range of values between the 25th and 50th percentile of relative promoter strength (between 0.030 and 0.14). Q_3 denotes the range of values between the 50th and 75th percentile of relative promoter strength (between 0.14 and 0.66). Q_4 denotes the range of values ≥ the 75th percentile of calculated values for the relative promoter strength at Spo 3h (above 0.66). The distribution of the log_2 fold change of the coding TSS from each of these four quartile groups were plotted **(Figure 6.9 B**, left**)**. All fold changes were derived from normalized read counts at Spo 3h relative to Spo 2h (pre-meiotic cells). A greater than 2-fold repression of the coding TSS (log_2 fold change ≤ -1) was not observed in most of the data points from the Q_1 to Q_3 groups **(Figure 6.9 B**, left**)**. On the other hand, alternative TSSs in the Q_4 group showed the greatest enrichment for repression of associated coding TSSs (log_2 fold change ≤ -1) **(Figure 6.9 B)**. Hence, increasing the relative upstream

promoter strength also increases the likelihood of repressing the downstream coding TSS **(Figure 6.9)**.

High relative promoter strength values alone could not exclude many cases where the coding TSS was not repressed (log_2 fold change > -1) **(Figure 6.9 B**, right**)**. Therefore, I looked for other criteria that would be biased for a greater than 2-fold repression of the coding TSS (log_2 fold change ≤ -1). By comparing changes in the Spo 3h time point relative to the pre-meiotic Spo 2h time point as an example, the log_2 fold change of the upstream TSS reads were grouped into four quartiles (Q_1 to Q_4). A box plot of the log_2 fold change of the coding TSS from each of these four quartile groups showed that this variable was a poor predictor of coding TSS repression, with similar medians and distribution of values across all groups **(Figure 6.9 C)**. I next examined if there was a minimum distance between the alternative and the coding TSS (difference in 5' UTR lengths) which could predict coding TSS repression. 5' UTR length difference was calculated by subtracting the 5' UTR length of the coding transcript from that of the upstream isoform/ncRNA, and values were grouped into four quartiles (Q_1 to Q_4). A box plot of the log_2 fold change of the coding TSS from each of these four quartile groups showed that 5' UTR differences in the second quartile and above (5' UTR difference ≥ 75 nt) were skewed towards negative values compared to that in the first quartile **(Figure 6.9 D)**. Consequently, the Q_2, Q_3 and Q_4 groups had a lower median than that of the Q_1 group. Hence, this analysis suggests that a 5' UTR difference ≥ 75 nt can also enrich for cases where the coding TSS is repressed by an upstream isoform/ncRNA.

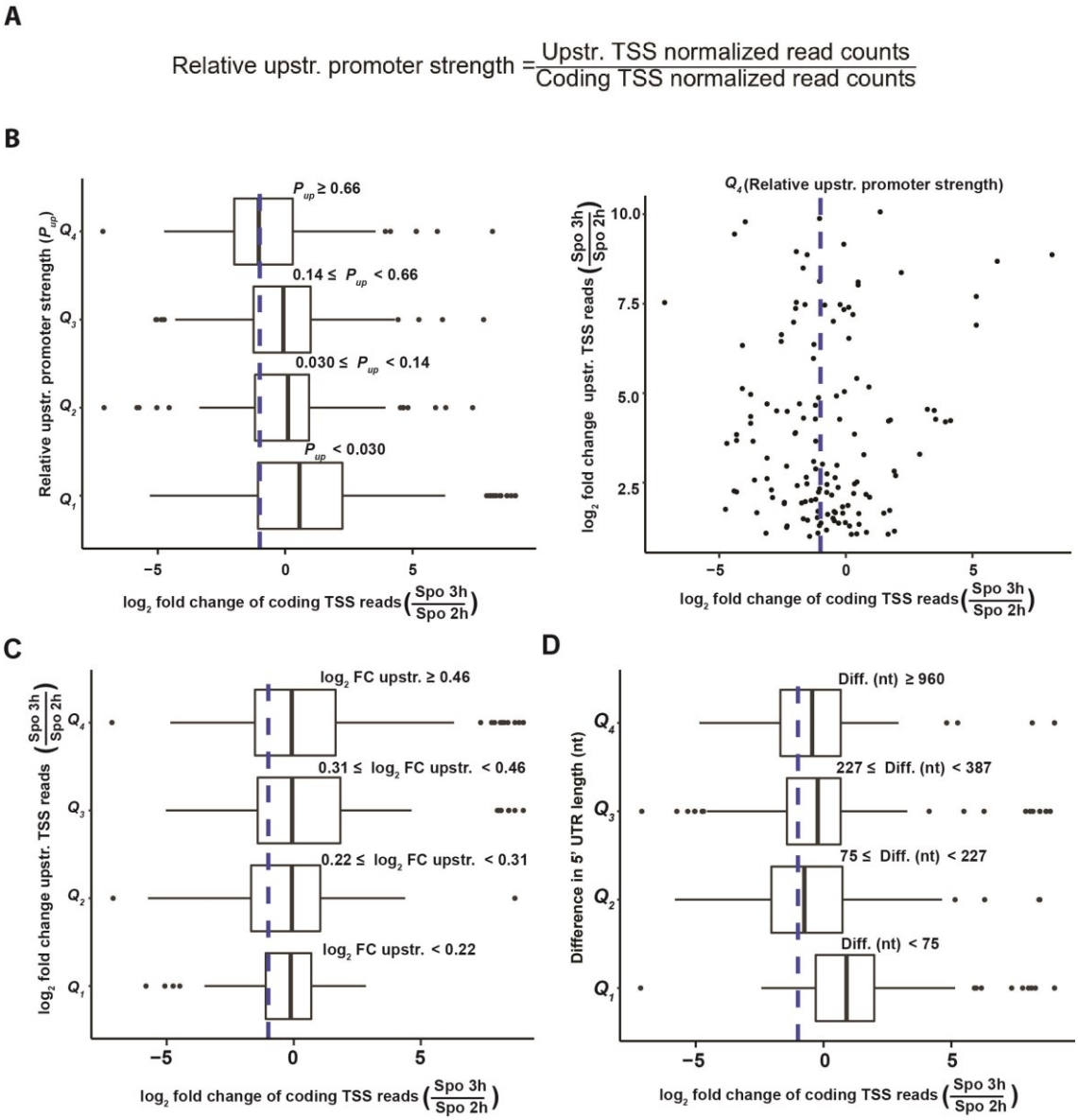


Figure 6.9 Relative promoter strength analysis of transcript isoforms and ncRNAs associated with coding TSS repression

(A) Formula to calculate relative upstream promoter strength. Normalized read counts associated with an upstream TSS were divided by normalized read counts associated with the coding TSS of the same gene, at the same time point. **(B) Left:** Box plots of the $\log_2$ fold change of coding TSS reads subsetted by quartiles of the relative upstream promoter strength, Q_1 to Q_4. The range of values for each quartile are displayed in the figure. All fold changes in this panel were derived from normalized read counts at Spo 3h relative to Spo 2h (pre-meiotic cells). Points to the left of the blue vertical dashed line are those whereby the $\log_2$ fold change of coding TSS reads was ≤ -1, representing a 2 fold or more downregulation of the gene. **Right:** Scatter plot of the $\log_2$ fold change of upstream TSS reads belonging to the 4th quartile of relative upstream promoter strength, against the $\log_2$ fold change of coding TSS reads from the same gene. Points to the left of the blue

vertical dashed line are those whereby the $\log_2$ fold change of coding TSS reads was ≤ -1, representing a 2 fold or more downregulation of the gene. **(C)** Box plots of the $\log_2$ fold change of coding TSS reads associated with different quartiles of the $\log_2$ fold change of upstream TSS reads associated with the same gene. The range of values for each quartile are displayed in the figure. All fold changes in this panel were derived from normalized read counts at Spo 3h relative to Spo 2h (pre-meiotic cells). Points to the left of the blue vertical dashed line are those whereby the $\log_2$ fold change of coding TSS reads was ≤ -1, representing a 2 fold or more downregulation of the gene. **(D)** Box plots of the $\log_2$ fold change of coding TSS reads associated with different quartiles of the difference in 5' UTR length. 5' UTR length difference was calculated by subtracting the 5' UTR length of the coding transcript from that of the alternative isoform/ncRNA. The range of values for each quartile are displayed in the figure. All fold changes in this panel were derived from normalized read counts at Spo 3h relative to Spo 2h (pre-meiotic cells). Points to the left of the blue vertical dashed line are those whereby the $\log_2$ fold change of coding TSS reads was ≤ -1, representing a 2 fold or more downregulation of the gene.

I next asked the question if the two combined criteria of relative upstream promoter strength ≥ 0.66 and a 5' UTR difference ≥ 75 nt, further improved the prediction whether the coding TSS is repressed by an upstream isoform/ncRNA. The differences in 5' UTR lengths were plotted against the $\log_2$ fold change of coding TSSs for Spo 3h vs 2h for the group of TSSs with a relative upstream promoter strength ≥ 0.66 **(Figure 6.10 A)**. I found that a 5' UTR difference ≥ 90 nt (instead of ≥ 75) between the two transcripts better enriched for alternative TSSs whose upregulation was associated with downregulation of the coding TSS **(Figure 6.10 A**, see points above horizontal blue line**)**. Conversely, closely spaced TSSs tend to have small effects on each other's activities or were coordinately upregulated **(Figure 6.10 A**, see points below horizontal blue line**)**. The *NDC80*^{luti} upstream alternative TSS (shown in red) was also identified, suggesting that these criteria could predict genes regulated by transcriptional interference **(Figure 6.10 A)**. In conclusion, during early meiosis, upstream transcripts which are associated with repression of downstream coding transcripts tend to have a relative upstream promoter strength ≥ 0.66 and a 5' UTR difference ≥ 90 nt.

I next asked the question if the same criteria could also be used to identify similar repressive upstream transcripts in mid-late gametogenesis. The differences in 5' UTR lengths were plotted against the $\log_2$ fold change of coding TSSs for Spo 8h vs 6h, for the group of mid-late alternative TSSs with a relative upstream promoter strength ≥ 0.66 **(Figure 6.10 B)**. To a lesser extent, these two criteria also

enriched for data points corresponding to repressive upstream transcripts **(Figure 6.10 B)**. Other studies have shown that *ORC1* mRNA expression is repressed by *ORC1*^{luti} transcription in mid-late gametogenesis (Cheng et al. 2018). TE-seq identified three distinct alternative upstream TSSs associated with *ORC1*^{luti} transcripts. These three *ORC1*^{luti} alternative TSSs were identified by my criteria and were highlighted in red, again indicating that these criteria could predict cases of transcriptional interference **(Figure 6.10 B)**. I expanded this analysis by selecting for alternative TSSs with a relative upstream promoter strength ≥ 0.66 and a 5' UTR difference ≥ 90 nt in all time points from Spo 3-9h. For the early meiotic time points (Spo 3-6h), changes in coding TSS reads were calculated with reference to the pre-meiotic Spo 2h time point. For the mid-late meiotic time points (Spo 7-9h), changes in coding TSS reads were calculated with reference to the Spo 6h time point, when cells were arrested in meiotic prophase. A box plot of the $\log_2$ fold change of the coding TSS from each time point in gametogenesis showed that the two combined criteria were good predictors of coding TSS repression in early gametogenesis and to a lesser extent in mid-late gametogenesis **(Figure 6.10 C)**. In all the early time points from Spo 3-6h, most alternative TSSs with a relative upstream promoter strength ≥ 0.66 and a 5' UTR difference ≥ 90 nt tended to have a greater than 2-fold repression of the coding TSS **(Figure 6.10 C**, left of the blue line**)**. For most mid-late alternative TSSs filtered by the same criteria above, there were no or subtle effects on the levels of the coding TSS at Spo 7h. At Spo 8h, more alternative TSSs showed at least a marginal repressive effect on the response of the coding TSS **(Figure 6.10 C**, left of the red line**)**. At Spo 9h, the repressive effect is more pronounced. This delay in repression at 7h was not due to delayed *NDT80* induction because these cells underwent meiotic divisions synchronously **(Figure 6.2 B-D)**. Perhaps this delay was due to slower induction of alternative TSSs from 7-9h, as because only 85 upstream alternative TSS-downstream coding TSS pairs were included in the 7h time point whereas 177 and 158 pairs were included in the 8h and 9h time points respectively **(Figure 6.10 C)**; in other words, there were fewer upstream alternative TSSs at 7h, in which their relative upstream promoter strength equalled or exceeded the cut off value of 0.66. Nonetheless, the data suggest that in both early and mid-late gametogenesis, upstream transcripts which are associated with repression of downstream coding

transcripts tend to have a relative upstream promoter strength ≥ 0.66 and a 5’ UTR difference ≥ 90 nt.

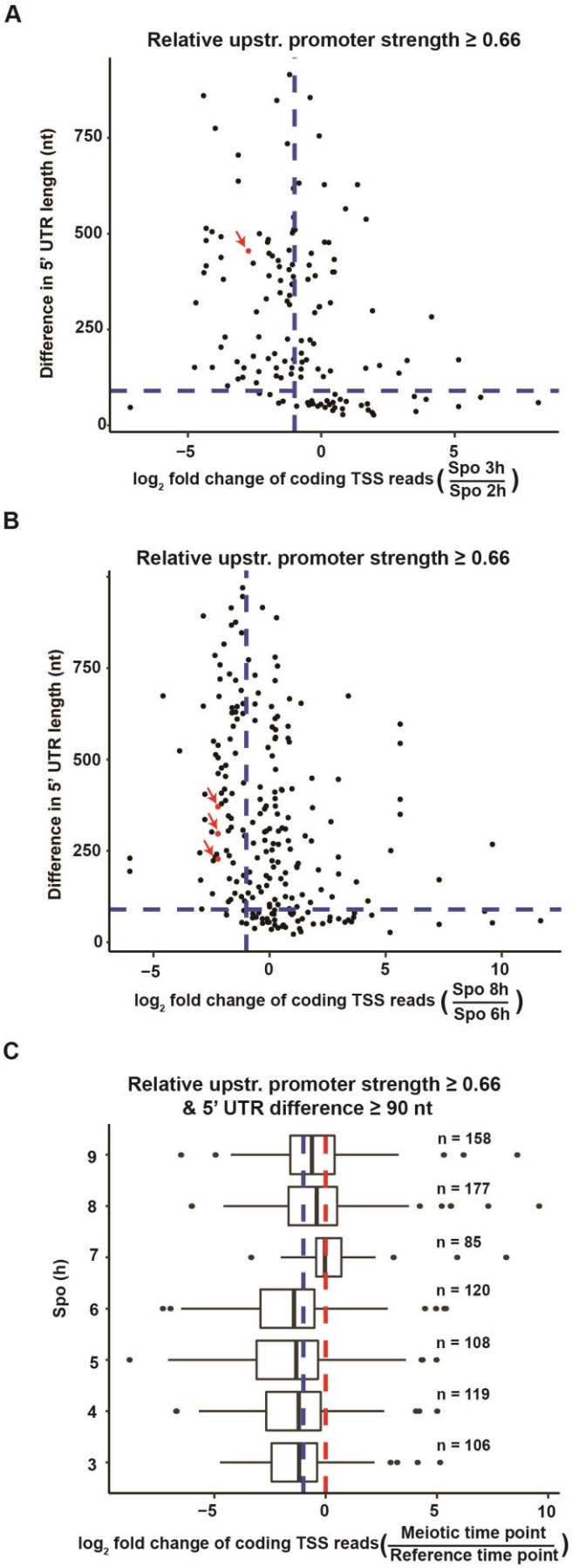

A
Relative upstr. promoter strength ≥ 0.66
Difference in 5' UTR length (nt)
750
500
250
0
−5
0
5
log2 fold change of coding TSS reads (Spo 3h / Spo 2h)
B
Relative upstr. promoter strength ≥ 0.66
Difference in 5' UTR length (nt)
750
500
250
0
−5
0
5
10
log2 fold change of coding TSS reads (Spo 8h / Spo 6h)
C
Relative upstr. promoter strength ≥ 0.66
& 5' UTR difference ≥ 90 nt
Spo (h)
9
8
7
6
5
4
3
n = 158
n = 177
n = 85
n = 120
n = 108
n = 119
n = 106
−5
0
5
10
log2 fold change of coding TSS reads (Meiotic time point / Reference time point)

Figure 6.10 Identifying common genomic features of repressive upstream transcripts

(A) Scatter plots of the difference in 5' UTR length, against the $\log_2$ fold change of coding TSS reads associated with the same gene. 5' UTR length difference was calculated by subtracting the 5' UTR length of the coding transcript from that of the alternative isoform/ncRNA. All fold changes in this panel were derived from normalized read counts at Spo 3h relative to Spo 2h (pre-meiotic cells). The data points in this scatter plot were subsetted from those genes which had a relative upstream promoter strength ≥ 0.66. Points to the left of the blue vertical dashed line are those whereby the $\log_2$ fold change of coding TSS reads was ≤ -1, representing a 2 fold or more downregulation of the gene. The horizontal dashed line marks the boundary where the 5' UTR difference equals 90 nt. The point marked in red corresponds to the *NDC80* locus. **(B)** Same as A, except that all fold changes in this panel were derived from normalized read counts at Spo 8h relative to Spo 6h (meiotic prophase). The 3 points marked in red correspond to the 3 upstream alternative TSSs at the *ORC1* locus. **(C)** Box plots showing the distribution of the $\log_2$ fold change of coding TSS reads of genes at different time points throughout gametogenesis. Candidates were included only if their relative upstream promoter strength was ≥ 0.66 and if the 5' UTR difference was ≥ 90 nt. The number of candidates (not the same as number of genes) for each time point are labelled in the figure. Values to the left of the red vertical dashed line are those whereby the $\log_2$ fold change of coding TSS reads was < 0. Values to the left of the blue vertical dashed line are those whereby the $\log_2$ fold change of coding TSS reads was ≤ -1, representing a 2 fold or more downregulation of the gene. For the early meiotic time points (Spo 3-6h), changes in coding TSS reads were calculated with reference to the pre-meiotic Spo 2h time point. For the mid-late meiotic time points (Spo 7-9h), changes in coding TSS reads were calculated with reference to the Spo 6h time point, when cells were arrested in meiotic prophase.

6.3.5 Expression of upstream transcripts has varying effects on the levels of coding transcripts

Besides repression, transcription of an upstream isoform or ncRNA could have different effects on the expression of the downstream TSS. Indeed, as shown in the previous section, a fraction of alternative TSSs with relative upstream promoter strength ≥ 0.66 and a 5' UTR difference ≥ 90 nt (hereafter defined as meiotic upstream transcripts or isoforms (MUTIs)) were not associated with repression of downstream coding transcripts **(Figure 6.10 C)**. To catalogue these different regulatory relationships in finer detail, all early or mid-late MUTIs were categorized into three main groups. MUTIs in the first group labelled "negative relationship" were included as long as there was a ≥ 2 fold decrease (FDR <0.05) in the coding TSS at least one matching time point throughout gametogenesis

(Table 8.1 and **Table 8.2)**. MUTIs in the second group labelled "positive relationship" were included as long as there was a ≥ 2 fold increase (FDR <0.05) in the coding TSS in at least one matching time point throughout gametogenesis **(Table 8.3** and **Table 8.4)**. MUTIs in the third group labelled "independent regulation" were included as long as there was no statistically significant change in the coding TSS (FDR > 0.05) in at least one matching time point throughout gametogenesis **(Table 8.5** and **Table 8.6)**. MUTIs could be long upstream transcript isoforms (luti RNAs) spanning the whole ORF, partially overlapping isoforms, or shorter promoter transcripts. In order to distinguish between these possibilities, the bedtools utility was used to identify the closest alternative TES (distinct from the canonical TES) to the alternative meiotic TSSs on the same strand (see column "nearest alternative TES ID" from **Table 8.1-8.8**). The distance between the alternative TSS and alternative TESs were also given in the same tables. In most examples, there were no alternative TESs and hence those alternative TSSs are likely to be luti RNAs using the same TES as the mRNA. In other examples such as the *MNN1* and *IML1* loci, the distance between the alternative TSS and the alternative TES was less than 500 nt, indicating that some of the RNAs could be relatively short promoter transcripts. In yet other examples such as *HMG1* and *ULP2*, the distance between the alternative TSS and the alternative TES was more than 900 nt. Perhaps some of these RNAs could be partially overlapping mRNA isoforms that terminate upstream of the canonical TES. In conclusion, various types of MUTIs have different effects on the expression of the downstream TSS.

To better visualize the regulatory effects of MUTI transcription, individual examples from each category were selected and the $\log_2$ normalized read counts of both the upstream TSS (red) and associated coding TSS (black) were plotted throughout gametogenesis **(Figure 6.11** and **Figure 6.12)**. Under the category: negative relationship, data from three example loci for early gametogenesis were plotted: *NDC80*, *SWI4* and *RAD16*. In these examples the MUTI is rapidly expressed after Spo 2h when *IME1* was induced **(Figure 6.11 A)**. While MUTI levels increased in the early time points (Spo 3-6h), levels of the associated coding TSS decreased by more than 2 fold. A similar pattern was observed in example loci under the same category in mid-late gametogenesis: *MCM5*, *ORC1* and *TAT2*

(Figure 6.12 A). In these examples, MUTI levels increase in the mid-late time points (Spo 7-9h) while the levels of the associated coding TSS decrease by more than 2 fold. For the next category: positive relationship, data from the *TIR4*, GAT1 and *MF(alpha)1* loci were plotted. In these examples, both the levels of the MUTI and the associated coding TSS increased in tandem by more than 2 fold during early meiosis **(Figure 6.11 B)**. Similarly for the *CLB6*, *DSE3* and *DIT2* loci, both the levels of the MUTI and the associated coding TSS increased in tandem by more than 2 fold during mid-late meiosis **(Figure 6.12 B)**. In the last category: independent regulation, data from the *DUG2*, *MOD5* and *YPK9* loci were plotted. For *DUG2*, levels of the coding TSS remained relatively unchanged despite upregulation of the MUTI **(Figure 6.11 C**, left**)**. At the *MOD5* and *YPK9* loci, levels of the coding TSS fall initially between 2-3h, before rising and remaining relatively constant despite high levels of MUTI **(Figure 6.11 C**, middle and right**)**. For the mid-late examples in this category, the *ORC3* coding TSS levels did not change substantially (less than 2 fold) despite rapid induction of the MUTI from 6-7h **(Figure 6.12 C**, left**)**. Levels of the *AIF1* and *ADH1* coding TSS appeared to fluctuate independent of the levels of their respective MUTIs from 6-9h **(Figure 6.12 C**, middle and right**)**. Taken together, MUTI transcription has diverse effects on mRNA expression and cannot be fully described in binary terms of repression or activation.

To verify some of the examples in the TE-seq data, northern blots experiments were conducted to probe for the mRNA(*) and the luti RNAs(**) for 4 candidates: *SWI4, RAD16*, *ORC3* and *MCM5* **(Figure 6.11 D** and **Figure 6.12 D)**. In agreement with the TE-seq data, the *RAD16*, *SWI4* and *MCM5* mRNAs were clearly downregulated upon expression of their respective luti RNAs. Changes in the levels of the *ORC3* mRNA were more subtle between 6-10 h, suggesting that *ORC3* luti RNA transcription neither represses nor promotes *ORC3* mRNA expression. In conclusion, both TE-Seq and northern blot data suggest that MUTI or luti RNA transcription can repress, activate or can have little effect on mRNA expression, depending on genomic context.

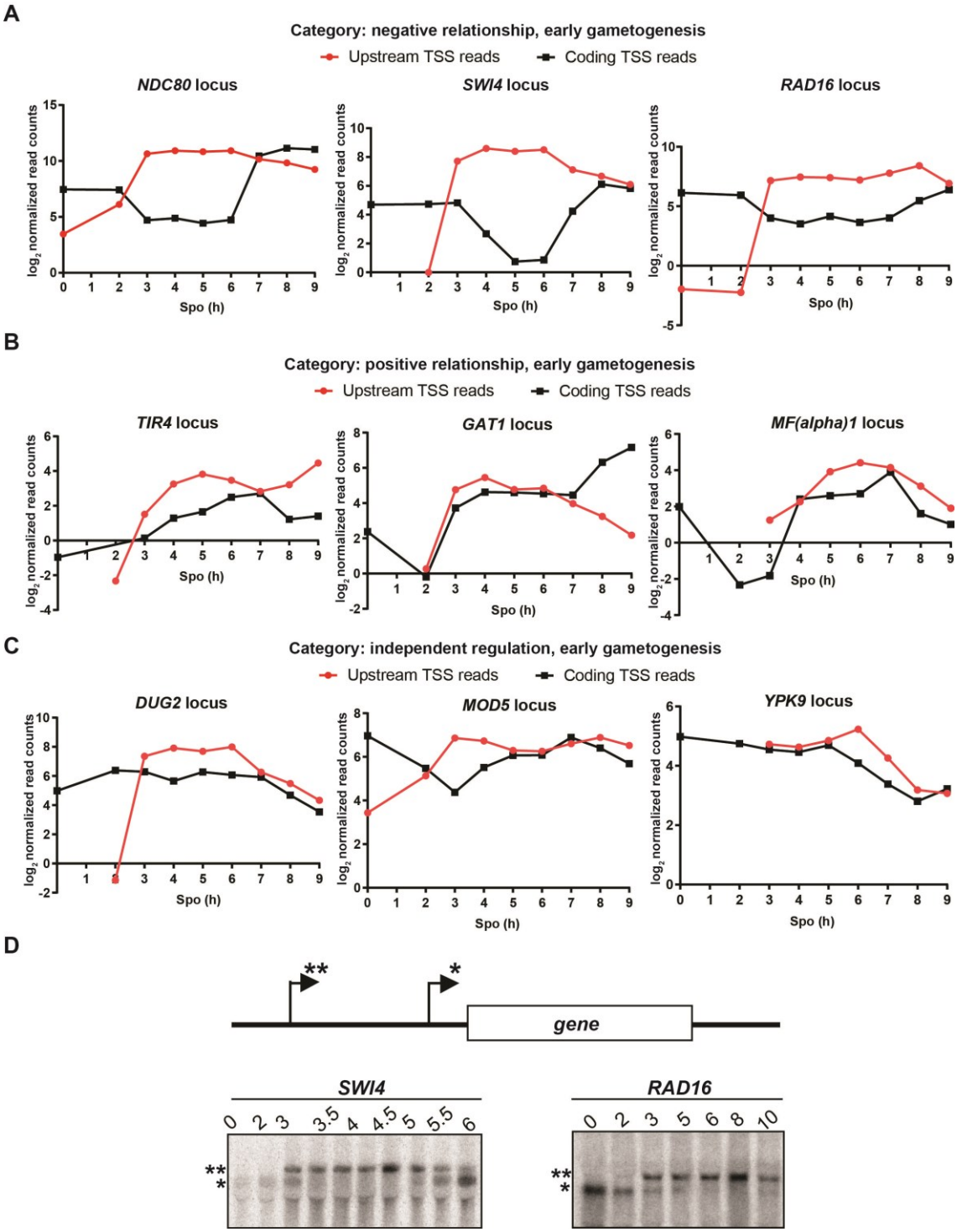


Figure 6.11 Categorizing early meiotic upstream transcripts by their relationships with associated coding transcripts

(A) Three examples where the levels of the coding TSS is negatively correlated with the levels of the upstream TSS during early gametogenesis (Spo 3-6 h): *NDC80*, SWI4 and *RAD16*. Plots represent the $\log_2$ values of normalized read counts from DESeq2 of both the upstream TSS (red) and associated coding TSS (black) throughout the entire gametogenesis time course. If a gene had more than one upstream TSS, the one with the highest average expression was chosen. If a

data point is not shown on the graph, it is because there were no detectable reads. **(B)** Same as A, except that the three examples are where the levels of the coding TSS is positively correlated with the levels of the upstream TSS during early gametogenesis: *TIR4, GAT1* and *MF(alpha1)*. **(C)** Same as A, except that the three examples are where the levels of the coding TSS is not correlated with changes in the levels of the upstream TSS during early gametogenesis: *DUG2*, *MOD5* and *YPK1*. **(D)** Northern blots of two example loci: *SWI4* and *RAD16* throughout gametogenesis. ** denotes the long upstream transcript isoform while * denotes the shorter coding mRNA.

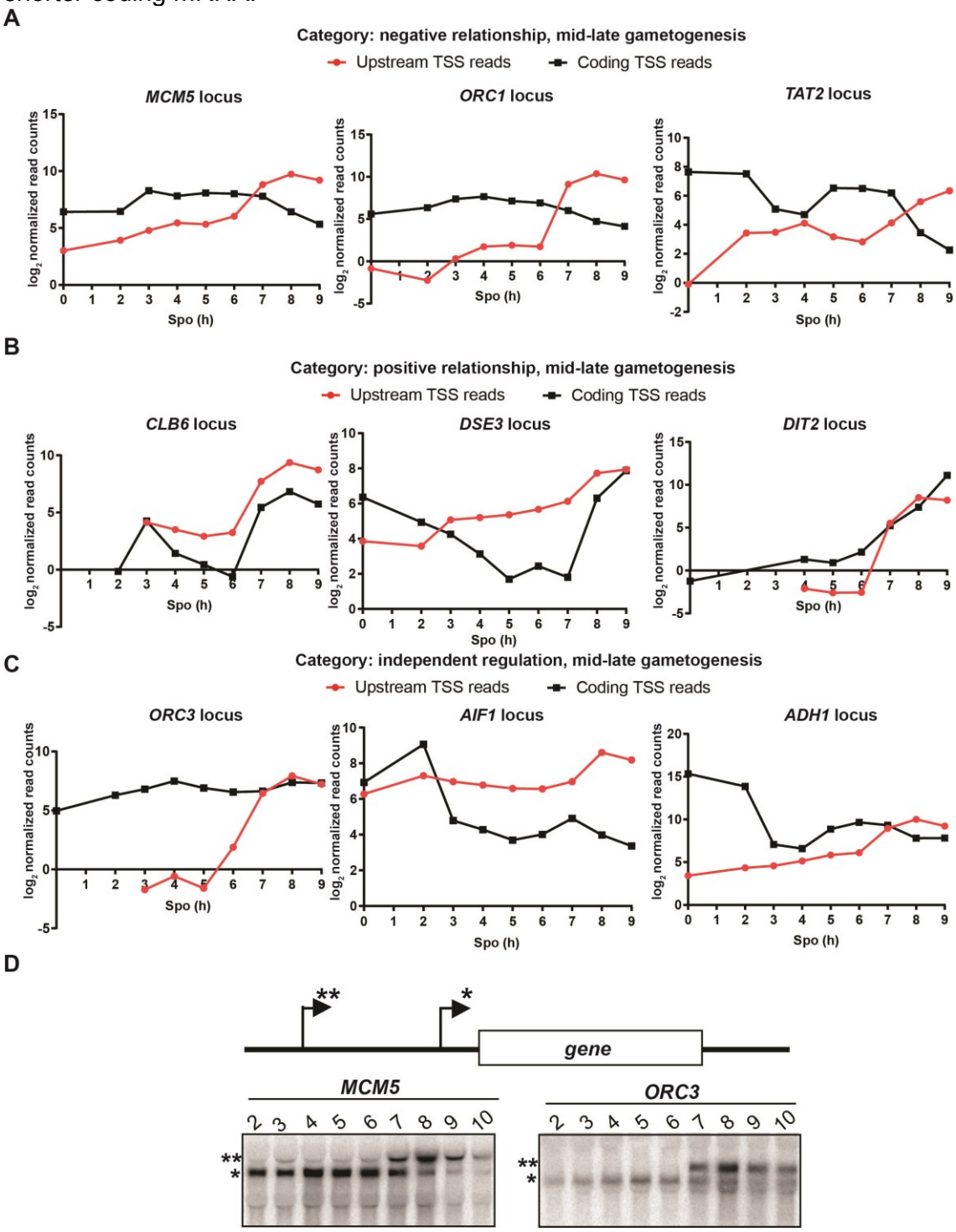


Figure 6.12 Categorizing mid-late meiotic upstream transcripts by their relationships with associated coding transcripts

(A) Three examples where the levels of the coding TSS is negatively correlated with the levels of the upstream TSS during mid-late gametogenesis (Spo 7-9 h): *MCM5*, *ORC1* and *TAT2*. Plots represent the $\log_2$ values of normalized read counts of both the upstream TSS (red) and associated coding TSS (black) throughout the entire gametogenesis time course. If a gene had more than one upstream TSS, the one with the highest average expression was chosen. If a data point is not shown on the graph, it is because there were no detectable reads. **(B)** Same as A, except that the three examples are where the levels of the coding TSS is positively correlated with the levels of the upstream TSS during mid-late gametogenesis: *CLB6, DSE3* and *DIT2*. **(C)** Same as A, except that the three examples are where the levels of the coding TSS is not correlated with changes in the levels of the upstream TSS during mid-late gametogenesis: *ORC3*, *AIF1* and *ADH1*. **(D)** Northern blots of two example loci: *MCM5* and *ORC3* throughout gametogenesis. ** denotes the long upstream transcript isoform while * denotes the shorter coding mRNA.

6.3.6 Applying TE-seq to profile meiotic prophase in mutants lacking Set2, Set3 or Spt16

Previous studies have established that the H3K36 methyltransferase Set2, the histone deacetylase Set3 and the histone chaperone Spt16 play a role in the repression of some coding mRNAs such as *NDC80*, *IME1* and *SER3* via transcriptional interference(Hainer et al. 2011;van Werven et al. 2012;Chia et al. 2017). However, the extent to which these chromatin remodellers are involved in genome-wide gene regulation by transcript isoforms and ncRNAs during gametogenesis is unknown. To answer this question, I applied TE-seq to control and mutant cells to profile their transcriptomes in meiotic prophase. In the first approach, control, *set2Δ*, *set3Δ* and *set2Δset3Δ* mutants in the *pCUP-IME1* and *GAL4.ER pGAL-NDT80* background were induced to initiate gametogenesis. Two time points were sampled for the control and mutant cells: a pre-meiotic time point (Spo 0h) and a meiotic prophase time point (Spo 6h). In the absence of β-estradiol, cells will arrest in meiotic prophase due to lack of Ndt80. Flow cytometry analysis showed that the control and mutant cells completed DNA replication by Spo 6h, as expected for cells in meiotic prophase **(Figure 6.13 A)**. Given that Spt16 is essential, depletion of this chaperone was achieved by a different approach. Endogenous *SPT16* was replaced with an *SPT16* allele with a C-terminal V5 epitope tag and an auxin-induced degron (AID) domain. These

mutant cells also had the *pCUP-TIR1* allele, allowing for its controlled induction. In the presence of the auxin class of plant hormones, the Tir1 ligase catalyses the poly-ubiquitylation of the AID degron on Spt16 (Nishimura et al. 2009). This modification targets Spt16 for degradation by the proteasome. The important features of this degron system are that it allows for viable growth before depletion of Spt16 and that the protein can be degraded rapidly and efficiently. To demonstrate this, most of the Spt16 is depleted within 1 h of treating log phase mutant cells with Cu^{2+} and the auxin indole-3-acetic-acid (IAA), **(Figure 6.13 B)**. Prior to meiotic entry, Spt16 levels were reduced at Spo 2h before the addition of Cu^{2+} and IAA, which caused further depletion of Spt16 **(Figure 6.13 B)**. Regardless, it is clear that Spt16 can be depleted in mutant cells before early gametogenesis. As a control, another strain was made with the *SPT16-AID* allele but without the *TIR1* gene. Both control and mutant cells also had the *pCUP-IME1* and *GAL4.ER pGAL-NDT80* alleles. A Spo 0h time point was collected to represent pre-meiotic cells before Spt16 depletion **(Figure 6.13 B** and **Figure 6.13 C)**. 50 µM $CuSO_4$ and 500 µM IAA were added 2 hr after the cells were transferred to SPO to induce *IME1* and to deplete any remaining Spt16 simultaneously. Control and mutant cells arrested in meiotic prophase were collected at Spo 6h **(Figure 6.13 C)**. Flow cytometry data verified that the depletion of Spt16 did not prevent mutant cells from progressing through DNA replication to meiotic prophase after a delay **(Figure 6.13 D)**. By Spo 6h, both mutant and control cells had completed DNA replication and were arrested in meiotic prophase. In summary, cells with the *set2Δ*, *set3Δ* and *SPT16-AID* alleles can be arrested at meiotic prophase, harvested and subjected to TE-Seq analysis.

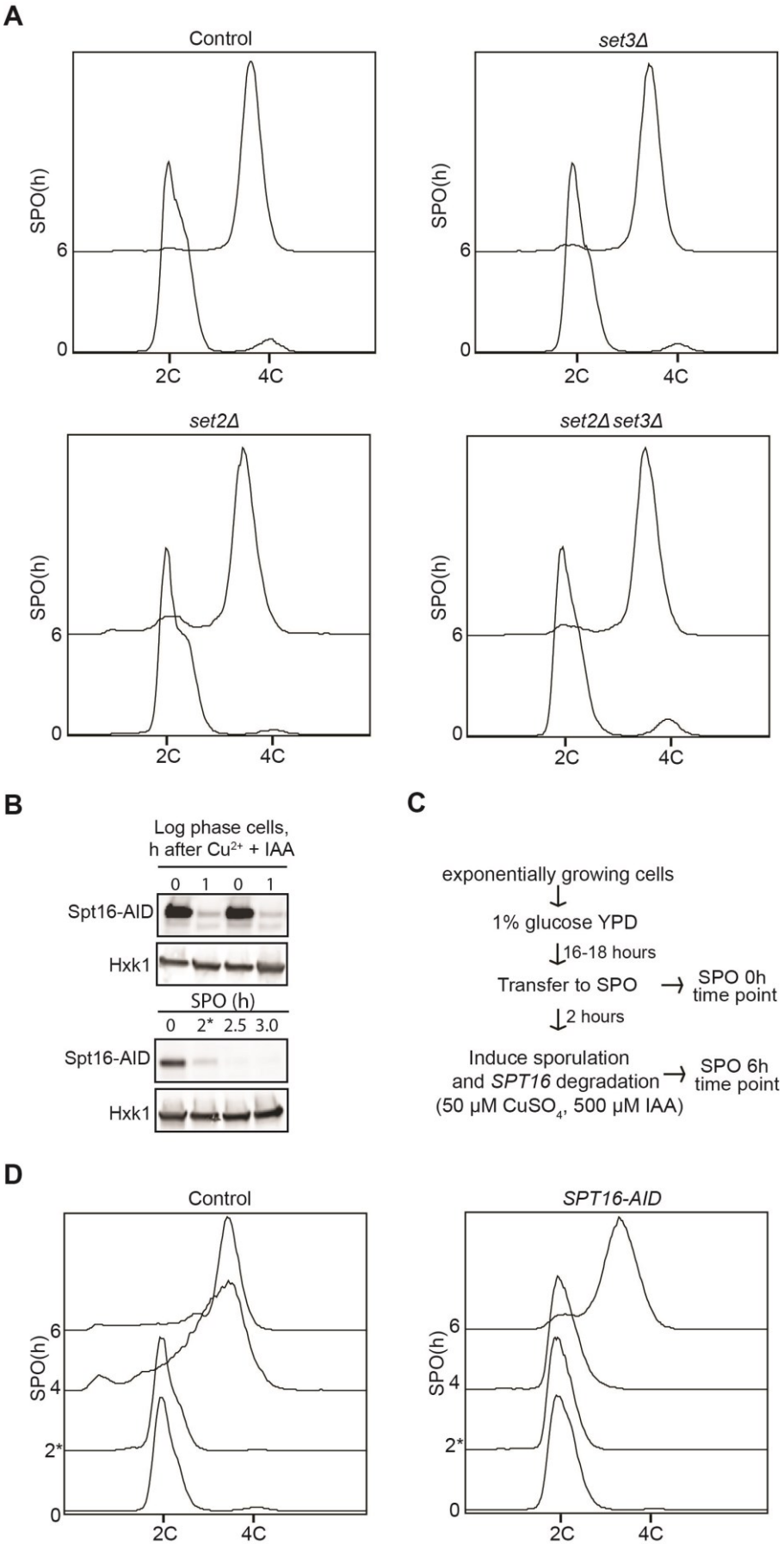
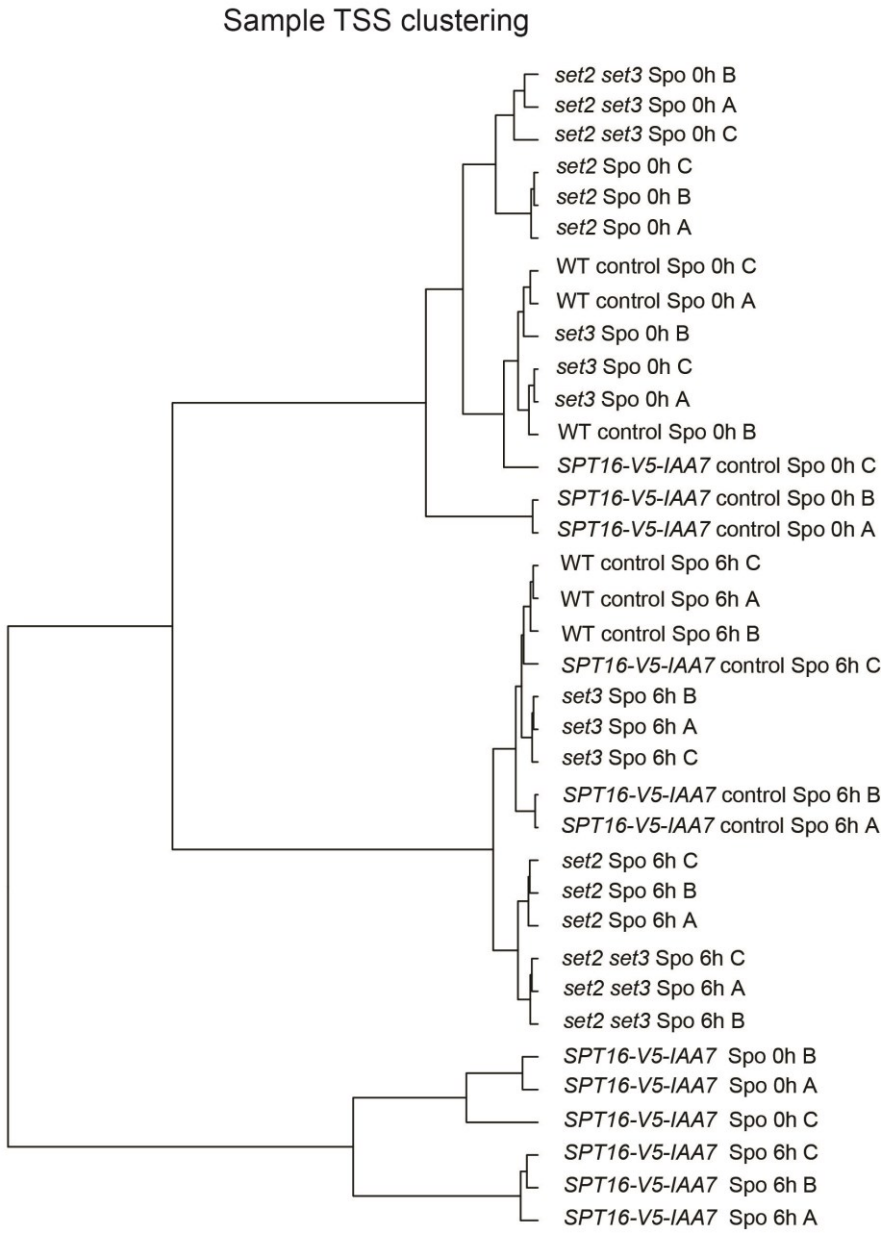

A
Control
set3Δ
set2Δ
set2Δ set3Δ
SPO(h)
6
0
2C
4C
B
Log phase cells,
h after Cu2+ + IAA
0 1 0 1
Spt16-AID
Hxk1
SPO (h)
0 2* 2.5 3.0
Spt16-AID
Hxk1
C
exponentially growing cells
1% glucose YPD
16-18 hours
Transfer to SPO
SPO 0h
time point
2 hours
Induce sporulation
and SPT16 degradation
(50 μM CuSO4, 500 μM IAA)
SPO 6h
time point
D
Control
SPT16-AID
SPO(h)
6
4
2*
0
2C
4C

Figure 6.13 *set2Δ, set3Δ, set2Δset3Δ* and *SPT16-AID* mutants initiate gametogenesis and complete DNA replication

(A) Flow cytometry analysis of DNA content of control (FW2795), *set3Δ* (FW5770), *set2Δ* (FW5767) and *set2Δset3Δ* cells (FW2912). All these cells also had the *pCUP-IME1* and *GAL4.ER pGAL-NDT80* alleles. Samples were taken at indicated time points, fixed, and DNA content was measured by propidium iodide staining; 50 μM $CuSO_4$ was added 2 h after the cells were transferred to SPO (*). At least 50,000 cells were analyzed at each time point. Results are representative of at least three independent repeats, n = 3. **(B)** Western blot showing degradation of Spt16-AID protein in log phase (top) and when shifting from saturated culture to SPO medium (bottom). Cells used here had the *SPT16-AID*, *pCUP-TIR1*, *pCUP-IME1* and *GAL4.ER pGAL-NDT80* alleles (FW6083). For the SPO time points, expression of *IME1* and degradation of tagged Spt16 were induced at the same time at 2 h (*) by the addition of 50 μM $CuSO_4$ and 500 μM IAA. Results are representative of two independent repeats, n = 2. **(C)** Flowchart for initiating gametogenesis and degrading tagged Spt16 in cells. Diploid cells were grown to saturation for in YPD and collected for the 0 hr time point. Remaining cells were then pelleted by centrifugation, washed with sterile water and resuspended to a final OD_{600} of 2.5 in SPO; 50 μM $CuSO_4$ and 500 μM IAA were added 2 h after the cells were transferred to SPO to induce *IME1* and to degrade tagged Spt16. By Spo 6h, cells are arrested in meiotic prophase and are collected for another time point. **(D)** Flow cytometry analysis of DNA content of control (FW6109) and *SPT16-AID* mutant cells (FW6083). All these cells had the *SPT16-AID, pCUP-IME1* and *GAL4.ER pGAL-NDT80* alleles. The only difference between the control and mutants is that the control cells lack the *pCUP-TIR1* gene. Samples were taken at indicated time points, fixed, and DNA content was measured by propidium iodide staining; 50 μM $CuSO_4$ and 500 μM IAA were added 2 h after the cells were transferred to SPO (*). At least 50,000 cells were analyzed at each time point. Results are representative of at least three independent repeats, n = 3.

In order to detect differentially expressed TSSs in mutants relative to the controls during meiotic prophase, three independent, biological repeats of the time courses were conducted. I next examined if the results for TSS data and mRNA seq were reproducible between the 3 repeats **(Figure 6.14 A and B)**. The Spo 0h samples comprising of pre-meiotic cells for both sets of controls (WT control and *SPT16-AID* control), as well as those for the set3Δ mutants clustered together. The *set2Δ* and *set2Δset3Δ* mutant Spo 0h samples clustered together, while the *SPT16-AID* mutant Spo 0h samples formed a separate cluster. A similar pattern was observed for the Spo 6h meiotic prophase samples. In summary, both TE-seq and mRNA seq data show that the repeats for the mutant and control samples were reproducible for the two time points.

A


Sample TSS clustering
set2 set3 Spo 0h B
set2 set3 Spo 0h A
set2 set3 Spo 0h C
set2 Spo 0h C
set2 Spo 0h B
set2 Spo 0h A
WT control Spo 0h C
WT control Spo 0h A
set3 Spo 0h B
set3 Spo 0h C
set3 Spo 0h A
WT control Spo 0h B
SPT16-V5-IAA7 control Spo 0h C
SPT16-V5-IAA7 control Spo 0h B
SPT16-V5-IAA7 control Spo 0h A
WT control Spo 6h C
WT control Spo 6h A
WT control Spo 6h B
SPT16-V5-IAA7 control Spo 6h C
set3 Spo 6h B
set3 Spo 6h A
set3 Spo 6h C
SPT16-V5-IAA7 control Spo 6h B
SPT16-V5-IAA7 control Spo 6h A
set2 Spo 6h C
set2 Spo 6h B
set2 Spo 6h A
set2 set3 Spo 6h C
set2 set3 Spo 6h A
set2 set3 Spo 6h B
SPT16-V5-IAA7 Spo 0h B
SPT16-V5-IAA7 Spo 0h A
SPT16-V5-IAA7 Spo 0h C
SPT16-V5-IAA7 Spo 6h C
SPT16-V5-IAA7 Spo 6h B
SPT16-V5-IAA7 Spo 6h A

B

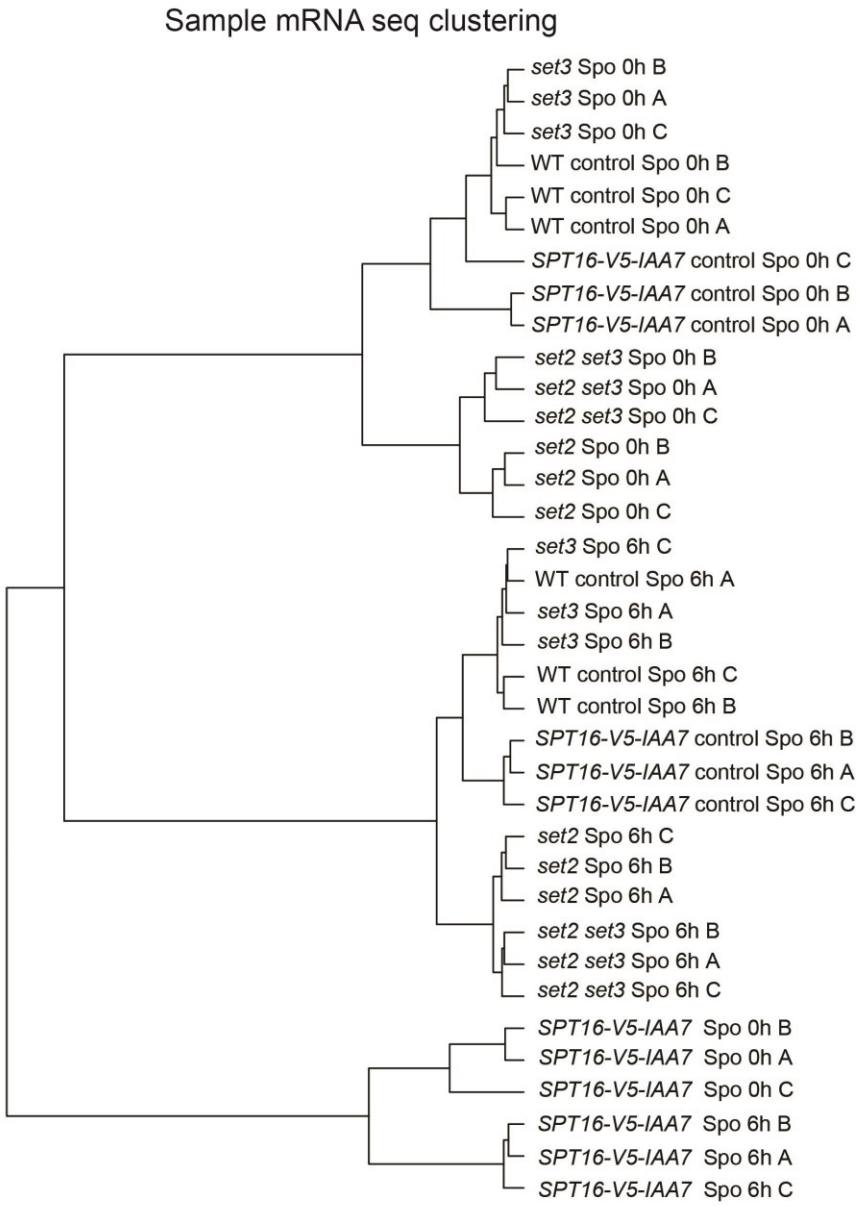


Figure 6.14 TE-seq and mRNA seq data for the mutants and their controls are highly reproducible

(A) Clustering tree of TSS reads from different time points. A, B and C denote the three independent repeats of the time course experiment. The filtering, mapping and clustering of TE-seq reads were done by Cai, Li from the Luscombe lab. **(B)** Clustering tree of mRNA seq reads from different time points. The filtering, mapping and clustering of mRNA-seq reads were done by Cai, Li from the Luscombe lab.

6.3.6 The *set2Δset3Δ* and *SPT16-AID* mutants show defects in genes repressed by upstream transcription during meiotic prophase

The extent to which Set2, Set3 and Spt16 are involved in genome-wide gene regulation by transcript isoforms and ncRNAs during gametogenesis is unknown. To investigate this, I compared mutant samples relative to their controls during meiotic prophase at Spo 6h. In order to find examples of de-repression of the coding TSS, I focused on the list of 114 early meiotic upstream alternative TSSs associated with repression of the coding TSS in wildtype cells **(Table 8.1)**. These 114 candidates were further examined for a greater than 2 fold up-regulation of the coding TSS (FDR < 0.05) in the mutants relative to the controls during meiotic prophase at Spo 6h. No targets were identified for the *set3Δ* single mutant. 4 genes were de-repressed in the *set2Δ* single mutant **(Table 8.7)**. Importantly, de-repression of the coding TSS could not be explained by downregulation (FDR <0.05) of the upstream alternative TSS in the mutants relative to the control **(Table 8.7**, see last column**)**. 9 genes were found to be de-repressed in the *set2Δset3Δ* mutant under this criteria and most of the alternative TSSs were not downregulated in the mutants relative to the control as well **(Table 8.8**, see last column**)**. *NDC80* was not selected by this criteria although the results in chapter 4 showed that the coding TSS was de-repressed in the *set2Δset3Δ* double mutant. A closer inspection of the fold change values revealed that the *NDC80*^{ORF} coding TSS was upregulated by approximately 1.5 fold in the double mutant relative to the WT (FDR > 0.05) and hence failed to meet the stringent selection criteria. In contrast, 37 genes were found to be de-repressed in the *SPT16-AID* mutant under this criteria, including the *NDC80*^{ORF} coding TSS **(Table 8.9)**. Among these 37 genes, 20 of them have alternative TSSs which were downregulated in the mutants relative to the control **(Table 8.9**, see last column**)**. In these examples, de-repression can be explained in part, by reduced upstream transcription. In conclusion, the repression of several genes by MUTI/luti RNA transcription is dependent on these chromatin remodellers under stringent selection criteria.

6.3.7 Expression of most meiotic upstream transcripts is dynamic

My analysis of the *NDC80*^{luti} transcript in chapter 4 showed that luti-mediated repression is rapidly reversible when cells were re-exposed to nutrient rich media before commitment to meiotic divisions. However, the extent to which all MUTI/luti RNA transcription is dynamically regulated upon return to growth (RTG) is unknown. To test this, 3 repeats of a return to growth time course experiment were conducted using cells in the *pCUP-IME1* and *GAL4.ER pGAL-NDT80* background **(Figure 6.15 A)**. In brief, cells were induced to initiate gametogenesis and arrested in meiotic prophase by withholding β-estradiol from the medium. Subsequently, cells were re-exposed to YPD and samples were collected at 6h (prior to RTG) and at 4 time points after RTG. These 3 independent, biological repeats were examined for reproducibility. After mapping the usable reads to the genome, the TSS and regular mRNA seq reads for each repeat (A, B and C) clustered according to defined RTG time points **(Figure 6.15 B-C)**. In conclusion, both the TE-seq and mRNA seq data show that the three, independent biological repeats for the RTG time course were similar to each other and can all be used for downstream analyses.

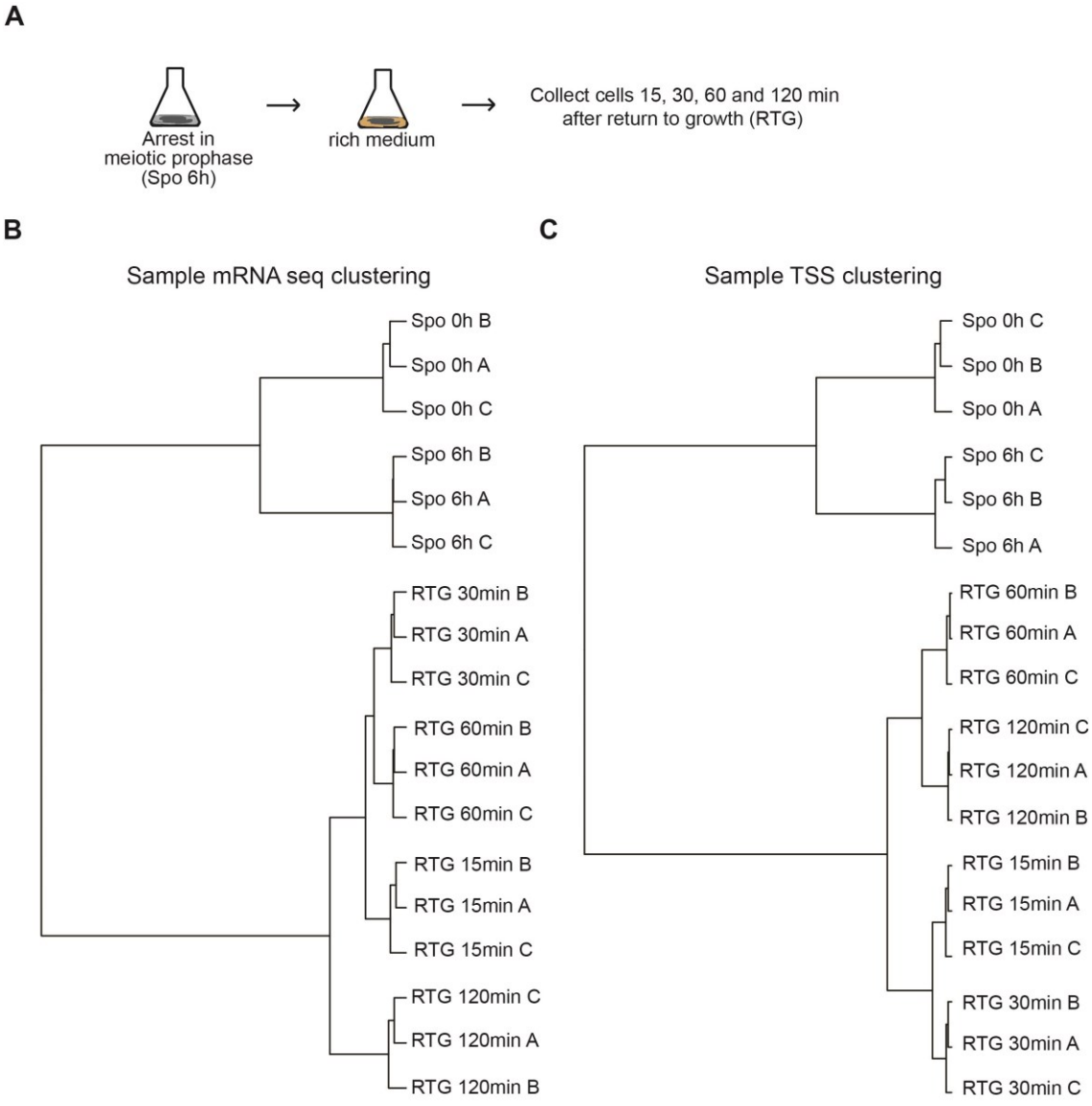


Figure 6.15 Applying TE-seq to "return to growth" time courses

(A) Schematic overview of the experimental set-up for the return to growth experiments. The diploid cells used in these experiments harboured both *IME1* fused to *CUP* promoter and *NDT80* expressed from the *GAL* promoter together with Gal4 fused to the estrogen receptor (*pCUP-IME1* and *GAL4.ER pGAL-NDT80*) (FW2795) **(B)** Clustering tree of mRNA seq reads from different time points. A, B and C denote the three independent repeats of the time course experiment. The filtering, mapping and clustering of mRNA seq reads were done by Cai, Li from the Luscombe lab. **(C)** Clustering tree of TSS reads from different time points. The filtering, mapping and clustering of TE-seq reads were done by Cai, Li from the Luscombe lab.

Next, I identified differentially expressed TSSs in the RTG time points relative to Spo 6h, meiotic prophase. Out of the 114 alternative upstream TSSs which were upregulated during early gametogenesis, 102 cases were downregulated by 2 fold or more (FDR <0.05) within 2 hours of RTG **(Figure 6.16 A)**. In this subset of 102 alternative upstream TSSs, 79 examples were rapidly repressed within 30 minutes of RTG **(Figure 6.16 B)**. In this subset of 79 rapidly repressed upstream TSSs, 68 examples were also associated with upregulation of the downstream TSS within 30 minutes of RTG **(Figure 6.16 C)**. Almost the same number of downstream TSSs were upregulated within 60 and 120 minutes of RTG (69 and 70 respectively), suggesting that the vast majority of these loci were dynamically regulated. The full list of early meiotic "repressed" genes and the effect of RTG on the upstream and coding TSSs are given in **Table 8.10**. To further illustrate the dynamic regulation of alternative upstream TSSs and their downstream coding TSSs during RTG, $\log_2$ normalized read counts were plotted against different time points for three loci: *NDC80, RAD16* and *SWI4* **(Figure 6.16 D)**. In these examples, the number of upstream TSS reads (red) decreased rapidly within 15 mins of RTG, accompanied by an upregulation of the coding TSS (black). Taken together, the data suggest that the transcription of most MUTI/luti RNAs is rapidly shut down upon RTG and as a consequence, the formerly repressed coding TSSs are induced again to meet the needs of the cell.

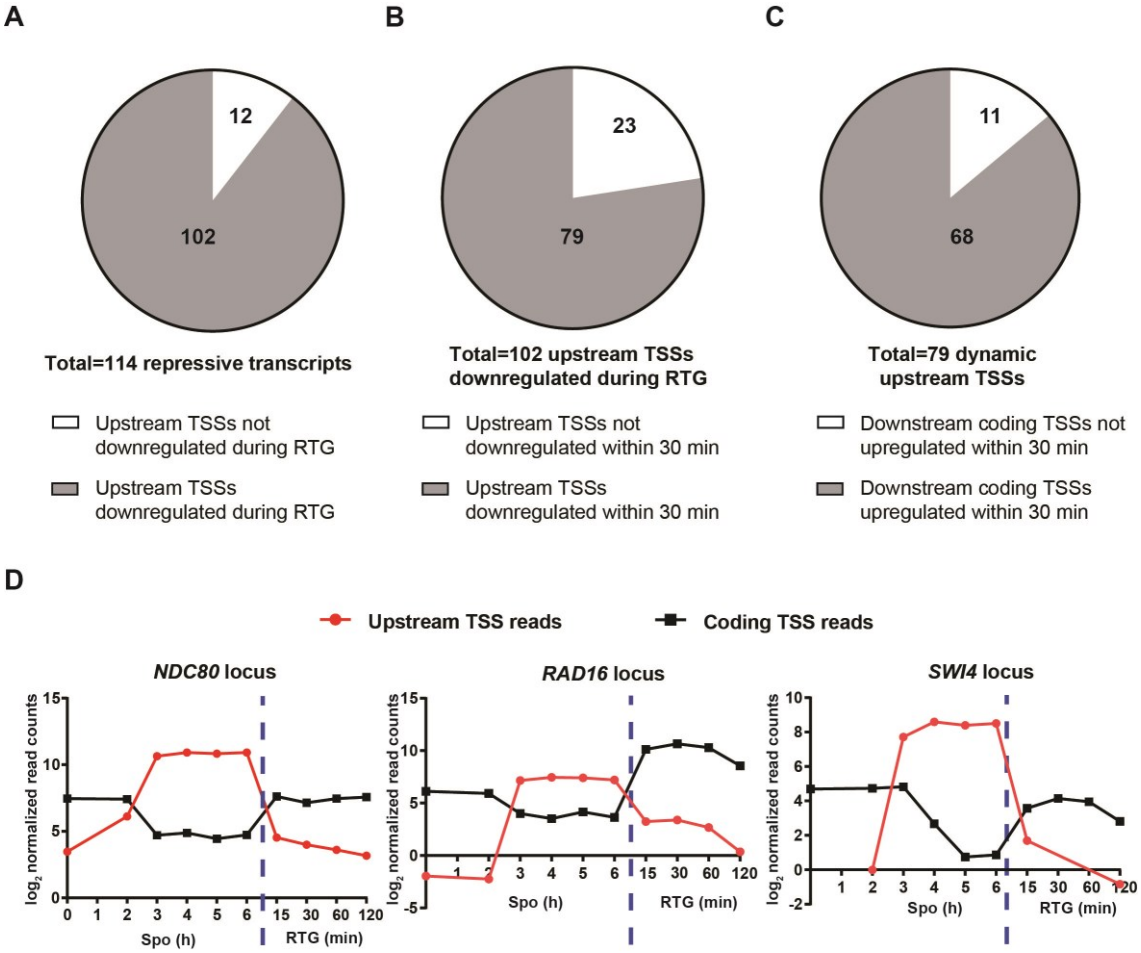


Figure 6.16 Most upstream TSSs are repressed and their associated coding TSSs are upregulated rapidly after RTG

(A) Proportion of upstream TSSs that are downregulated within 2 hours of RTG. Each of these 114 upstream TSSs were upregulated during early gametogenesis and were correlated with downregulation of associated coding TSSs. **(B)** Proportion of upstream TSSs that are downregulated within 30 mins of RTG, indicating dynamic regulation. Each of these 102 upstream TSSs were shown to be downregulated within 2 hours of RTG in panel A. **(C)** Proportion of upstream TSSs that are downregulated within 30 mins of RTG accompanied by an upregulation of associated coding TSSs. Each of these 79 upstream TSSs were shown to be dynamically regulated during RTG in panel B. **(D)** Three examples where both the upstream TSS and the coding TSS is dynamically regulated upon RTG: *NDC80*, SWI4 and *RAD16*. Plots represent the $\log_2$ values of normalized read counts of both the upstream TSS (red) and associated coding TSS (black) throughout the time course. If a gene had more than one upstream TSS, the one with the highest average expression was chosen. If a data point is not shown on the graph, it is because there were no detectable reads.

6.4 Discussion

In this chapter, I have used TE-Seq and mRNA seq to profile the TSSs and TESs of mRNAs in synchronously sporulating cells. By comparing TSS/TES usage throughout a high resolution gametogenesis time course, I have identified numerous alternative upstream TSSs that are induced by either Ime1 or Ndt80 during early or mid-late time points respectively. When comparing TE-seq data between samples, one key consideration is how best to identify and quantify differential TSS/TES usage during gametogenesis. Here, I have used DESeq2 as a "between sample" normalisation and comparison method to obtain lists of differentially changing TSSs in my time-course data (refer to Chapter 2 section 2.12.4 in this thesis for a description of the DESeq2 approach) (Love et al. 2014). My current approach makes a crucial assumption for DESeq2 that the expression levels of most genes are not changing significantly throughout gametogenesis. However, gametogenesis in budding yeast takes place in the context of starvation and meiotic signalling pathways, resulting in global changes in gene expression (Primig et al. 2000). The impact of global changes in gene expression is to mask true variation in TSS/TES usage between samples and to reduce the reliability of fold change comparisons. Multiple studies have addressed similar issues by "spiking in" external transcripts of known sequence and abundance to the same number of cells for each library (Loven et al. 2012;Booth et al. 2016). In principle, the proportion and abundance of these "spike in" transcripts are invariant across libraries. Thus, spike-ins serve as an external calibration standard to normalise samples for DESeq2 analysis. The quality of my TE-seq analysis could be further improved by incorporating spike-in controls in future work.

One important conclusion in this study is that transcription of meiotic upstream transcripts or isoforms (MUTIs) did not necessarily lead to repression of the downstream coding gene as in the *NDC80* example. Gene repression by MUTI/luti RNA transcription tended to require a strong upstream promoter relative to the downsteam one. Indeed, most repressive examples of MUTIs had a relative upstream promoter strength of at least 0.66, meaning that there are at least two-thirds as many upstream transcripts as there are coding mRNAs, to achieve a

greater than 2-fold repression of the downstream coding TSS. This is in line with other studies showing that transcriptional interference is most effective if the upstream promoter is stronger relative to the downstream one (Callen et al. 2004;Shearwin et al. 2005;Pande et al. 2018). If the alternative upstream promoter is too weak, then relatively low levels of transcription are insufficient to repress a highly active downstream promoter. In addition, the TE-seq data identified many examples or time points in which MUTI/luti RNA transcription had little or no effect on the levels of the coding TSS. This is despite a relatively high upstream promoter strength. Why were these downstream promoters refractory to transcriptional interference? Perhaps the answer lies in the concentration of transcription factors driving expression of the coding mRNA at different time points during gametogenesis. Past a certain threshold, high levels of coding transcription can out-compete and bypass TI at physiological levels of MUTI/luti RNA transcription. This hypothesis also explains why Ndt80 dependent expression of *NDC80*^{ORF} could still occur in a robust manner prior to the onset of meiotic divisions, despite high levels of *NDC80*^{luti} transcription (see chapter 4 for more details). This observation further highlights an advantage of dynamic gene regulation by MUTI/luti RNA transcription. A critical cell-cycle gene (e.g. *NDC80*^{ORF}) can be rapidly produced after meiotic prophase, leading to a timely increase in its protein levels when required by the cell.

Interestingly, TE-seq data revealed that a fraction of MUTI/luti RNAs were associated with a greater than 2 fold upregulation of the downstream coding TSS. How might overlapping ncRNA transcription lead to greater induction of the downstream gene? One recent genome-wide study in budding yeast showed that overlapping anti-sense transcription is associated with increased histone acetylation, reduced H3K36me3 and increased histone turnover in the promoters of their overlapping sense partners (Murray et al. 2015). Perhaps MUTI/luti RNA transcription can also establish a dynamic chromatin environment and prime the activation of sense gene promoters. In these instances, upstream transcription transiently increases chromatin accessibility; open chromatin is then rapidly bound by relatively high concentrations of transcription factor to drive coding gene expression. Work done in other model systems such as *Drosophila* showed that ncRNA transcription can activate neighbouring genes by displacing repressor

proteins or by antagonizing their binding to chromatin (Schmitt et al. 2005). It should be noted that TE-seq genome-wide conclusions were mostly supported by correlative measurements of different TSSs. These conclusions would be strengthened by cloning candidate promoters identified in this chapter and measuring the effects of MUTI/luti transcription on a fluorescent reporter such as GFP. Nonetheless, these genome-wide analyses suggest a model in which gene regulation by upstream transcription can fine-tune coding gene transcription depending on the relative upstream promoter strength and the local concentration of transcription factors.

Another interesting feature identified in this study was that repression by MUTI/luti RNA transcription usually required a minimal distance between the two promoters (≥ 90 nt). If the promoters were too closely spaced, then upstream transcription did not repress coding gene expression. What might be an explanation for this observation? In my model of TI, MUTI/luti RNA transcription increases the nucleosome occupancy at the downstream promoter which represses the gene. Perhaps in cases where the inter-promoter distance is too short (< 90 nt), there is insufficient space for the occlusion of the nucleosome free/depleted region (NFR/NDR) at the downstream promoter. Consistent with this hypothesis, the minimum NFR size in yeast is approximately 80 bp (Zaugg and Luscombe 2012). Furthermore, a genome-wide study showed that shorter distances between two tandem promoters were associated with a more strongly defined NDR at the downstream promoter (Woo and Li 2011). In such a scenario, the +1 nucleosome of the upstream promoter also acts as a strongly positioned -1 nucleosome of the downstream coding promoter. Taken together, these observations could explain why repression by MUTI/luti RNA transcription usually required both a relatively high upstream promoter strength as well as a minimal distance between the two promoters.

The data from this study also identified some genes whose repression by an overlapping upstream transcript was compromised in mutant cells lacking chromatin remodellers like Set2, Set3 and Spt16. Unexpectedly, *NDC80*^{ORF} was not shortlisted as target for Set2 and Set3 regulation, in contradiction of the findings in chapter 4. One explanation for this difference could be fluctuations in *NDC80*^{ORF}

levels during early gametogenesis (Spo 3-6h). These fluctuations were also observed in the *NDC80* northern blot experiments **(Figure 4.8** and **Figure 4.9)** Given that my TE-seq mutant data was only limited to one specific time point in early gametogenesis, there was insufficient statistical power to conclude that there is a de-repression of $NDC80^{ORF}$ in the *set2Δset3Δ* double mutant. Indeed, the $NDC80^{ORF}$ coding TSS was upregulated by approximately 1.5 fold in the double mutant relative to the WT but the FDR failed to meet the required stringency. Nonetheless, the data provide several examples where the repression of several genes by MUTI/luti RNA transcription was dependent on these chromatin remodellers. Also, the use of DESeq2 without external spike-in controls could have masked true biological variation in libraries due to global changes in gene expression. For mutant data, this problem could be further exacerbated by widespread initiation of cryptic transcripts and other transcription defects (Kaplan et al. 2003;Carrozza et al. 2005b). Further work is required to improve TE-seq analysis for mutant data.

Lastly, TE-seq data from the RTG time courses showed that most MUTI/luti RNAs are dynamically regulated during gametogenesis. In more than half of the examples, MUTI/luti RNA expression was lost and the coding mRNA was rapidly induced when cells in meiotic prophase were returned to nutrient rich YPD. The reversibility of repression of numerous genes allows cells to quickly adjust cell fate decisions (mitosis or meiosis) to sudden changes in the environment.

In conclusion, the results from this chapter shed light on the complexities of gene regulation by mRNA isoforms and ncRNAs during gametogenesis. Further research directions as well as the relevance to higher eukaryotes will be discussed in the next chapter.

Chapter 7. Final Discussion and Conclusions

This research focused on understanding how genes are regulated by mRNA isoforms and ncRNAs throughout a developmental program. I have chosen budding yeast gametogenesis as a model developmental program due to its relatively small genome and its experimental tractability. I developed an optimized approach to synchronize budding yeast gametogenesis, facilitating stage specific analysis of the transcriptome. By focusing my initial efforts on just the *NDC80* locus, I identified the mechanism by which the transcription of a **l**ong **u**ndecodable **t**ranscript **i**soform (*NDC80*^{luti}) repressed the coding mRNA during early gametogenesis. This repressive mechanism required the H3K36 methyltransferase Set2 and the Set3 histone de-acetylase. Furthermore, *NDC80*^{luti} mediated repression could be rapidly reversed depending on the needs of the cell. From a single locus approach, I next looked for genome-wide evidence for gene regulation by luti RNA or overlapping ncRNA transcription. In order to achieve this objective, I optimized an approach called transcript end sequencing (TE-seq) to profile the 5' and 3' ends of mRNAs in a quantitative manner. By applying TE-seq to a high resolution gametogenesis time course, I identified over a thousand examples of alternative TSSs which were upregulated at specific stages of this developmental program. Most of these alternative TSSs were from meiotic upstream transcripts or isoforms (MUTIs). I further categorized these alternative TSSs by their effects on the levels of the downstream coding TSS and showed diverse regulatory outcomes associated with the transcription of mRNA isoforms/ncRNAs. TE-seq was also used to identify examples of transcriptional interference which were dependent on the chromatin remodellers Set2, Set3 and Spt16. Finally, I demonstrated that most MUTIs are dynamically regulated upon return to growth (RTG). In this chapter, I review my results, answer the research questions (**bolded** for clarity) set out in my introduction chapter (see chapter 1, section 1.8), suggest areas of future exploration and discuss the relevance of my data to higher eukaryotes.

7.1 Developing an optimized approach to synchronize budding yeast gametogenesis

Established protocols for synchronous entry into gametogenesis typically involve culturing the cells in pre-sporulation media containing nitrogen and acetate without glucose (BYTA), before shifting to sporulation medium (SPO) (Falk et al. 2010). Even with these culture conditions, cell-to-cell variability in *IME1* levels still results in a relatively asynchronous gametogenesis (Nachman et al. 2007). My data presented in chapter 3 showed that timely induction of *IME1* from the copper inducible *CUP* promoter was sufficient to drive cells to undergo gametogenesis synchronously. In my optimized sporulation protocol, cells could be grown to saturation in rich medium (YPD) and directly shifted to SPO, before inducing *IME1*. The advantages of my protocol included a shorter work flow and the ease of obtaining more cellular material from saturated cultures grown in YPD than in BYTA (see discussion in chapter 3 for more details). Furthermore, I showed that inducible *IME1* and *NDT80* expression systems could be combined to improve sporulation synchrony. Controlling the timing of *IME1* and *NDT80* expression enabled the dissection of gametogenesis into the early and mid-late stages respectively. This method was subsequently used in later chapters to profile transcript isoforms and ncRNAs in a stage-specific manner.

7.2 Gene repression by the transcription of an mRNA isoform *NDC80^{luti}* during early gametogenesis

Ndc80 is a component of the kinetochore, a structure which mediates attachment of spindle microtubules to the centromere (Ciferri et al. 2007). As such, Ndc80 is essential for chromosome segregation during nuclear divisions(Ciferri et al. 2007). Ndc80 levels must be repressed in meiotic prophase to establish the meiosis I chromosome segregation pattern (Miller et al. 2012;Meyer et al. 2015). In the presence of a spindle, mis-expression of Ndc80 during meiotic prophase causes aberrant separation of sister chromatids during meiosis I (Miller et al. 2012). Thus, the dynamic control of Ndc80 expression is critical for meiotic divisions. The transcriptional regulation of *NDC80* during early gametogenesis was not well understood. **One research question was to ask if budding yeast cells regulate**

gene expression during gametogenesis by overlapping RNA isoform or ncRNA transcription. The work presented in chapter 4 of this thesis showed that the coding mRNA *NDC80*^{ORF} was repressed during meiotic prophase by the transcription of a 5'-extended mRNA isoform *NDC80*^{luti}. Transcription of *NDC80*^{luti} from an upstream alternative promoter increased nucleosome occupancy and established a repressive chromatin environment at the downstream *NDC80*^{ORF} promoter. **What is the molecular mechanism by which mRNA isoform or ncRNA transcription regulate gene expression and do chromatin remodellers play a role?** For *NDC80* at least, this mechanism depended on both the H3K36 methyltransferase Set2 and the histone de-acetylase Set3. ChIP data indicated that the H3K4me2 and H3K36me3 modifications overlapped in the *NDC80*^{ORF} promoter, although it is unclear if these marks occurred on the same nucleosome as a repressive combinatorial histone code. Nonetheless, both the Set1/Set3C and Set2/Rpd3S pathways appear to be redundant in establishing a repressive chromatin state at the *NDC80*^{ORF} promoter. This is consistent with their well characterized roles in regulating gene expression via overlapping ncRNA transcription (see discussion in chapter 4 for more details).

One key conclusion drawn from studying the *NDC80* locus was that gene repression by *NDC80*^{luti} transcription is dynamic and rapidly reversible. Firstly, *NDC80*^{ORF} was induced after meiotic prophase by the Ndt80 meiotic transcription factor despite ongoing *NDC80*^{luti} transcription. Secondly, ectopic expression of Ndt80 could bypass *NDC80*^{luti} induced repression, suggesting that repression was not refractory to increasing levels of *NDC80*^{ORF} transcription. This is possible because the Ndt80 binding site in the *NDC80*^{ORF} promoter was not occluded by nucleosomes even during *NDC80*^{luti} transcription. Thus, transcriptional interference by *NDC80*^{luti} together with the retained accessibility of chromatin to some transcription factors, could be a strategy for repressing gene expression while still "poising" it for rapid activation during development. Dynamic regulation by *NDC80*^{luti} transcription could also be observed in RTG time course experiments. When cells arrested in meiotic prophase were returned to nutrient rich YPD, *NDC80*^{luti} expression was lost and *NDC80*^{ORF} was rapidly induced. Interestingly, levels of the repressive H3K36me3 in the *NDC80*^{ORF} promoter were rapidly reduced within 15 min of RTG. This was a surprising result due to the role of H3K36me3 in

reducing histone turnover and the reported half-life of H3K36me3 being relatively long at 60 minutes (Sein et al. 2015). The mechanism for the rapid loss of H3K36me3 at the *NDC80*^{ORF} promoter upon RTG is not understood. Perhaps the act of *NDC80*^{ORF} transcription leads to the eviction of H3K36me3 modified nucleosomes and their subsequent replacement with un-methylated nucleosomes. These observations also support the hypothesis that histone modifications reportedly associated with "transcriptional memory" act to reinforce rather than to prevent transcriptional changes.

The work presented in this thesis showed that gene regulation by transcriptional interference is not binary with an on or off state, but could tune expression of the coding mRNA. Varying the strength of the *NDC80*^{luti} promoter allowed for scalable transcriptional repression of *NDC80*^{ORF} (see discussion in chapter 4 for more details). This feature might also be exploited to achieve different regulatory outcomes at other loci, depending on the needs of the cell. For instance, transcription driven by a weaker luti RNA promoter might only serve to dampen but not completely repress coding mRNA transcription.

Another unresolved question remains about the regulation of *NDC80*^{ORF} during gametogenesis. My work has emphasized that it is the act of *NDC80*^{luti} transcription, rather than the *NDC80*^{luti} transcript per se which has a regulatory function. However, it is still possible that nascent *NDC80*^{luti} RNA contributes to repression of *NDC80*^{ORF}. Both Set1 and Set2 have been reported to bind nascent RNA, and this interaction facilitates recruitment and stabilization of these histone methyltransferases during transcription elongation (Battaglia et al. 2017;Sayou et al. 2017). One possible follow-up experiment would be to knock-in a RNA de-stabilizing sequence such as the hammerhead ribozyme to trigger cleavage of the nascent *NDC80*^{luti} RNA (Samarsky et al. 1999). It would be informative to test if *NDC80*^{luti} transcription is still sufficient for gene repression under these conditions.

7.3 Developing TE-seq to profile the 5' and 3' ends of mRNAs

My findings at the *NDC80* locus prompted me to investigate how many genes are regulated by luti mRNA transcription during gametogenesis. Bulk mRNA seq

data cannot accurately determine what isoforms are transcribed from a locus nor distinguish the transcription start sites (TSSs) or transcript end sites (TES) of interleaved, alternative transcripts. Therefore to answer my research question, I optimized TE-seq to profile both TSS and TES usage quantitatively in chapter 5 of this thesis. The TE-seq protocol generated reproducible data and could map TSSs and TESs at single nucleotide resolution. While TE-seq is a useful tool to quantitatively profile the 5' and 3' ends of mRNAs, there were numerous examples in the data set showing poor correlation between TSS and TES reads. Many poorly correlated TSS and TES pairs were due to inconsistencies with existing ORF annotations as well as ambiguities in the computational assignment of TSS/TES peaks to unannotated mRNA isoforms or ncRNAs. One possible way of addressing these issues would be to complement TE-seq data with transcript isoform sequencing (TIF-seq), to improve transcript annotation and to facilitate the assignment of TSS/TESs to specific transcripts (Pelechano et al. 2014) (see discussion in chapter 5 for more details).

7.4 Alternative upstream TSSs identified during gametogenesis and their regulatory features

In chapter 6 of this thesis, I used TE-Seq and mRNA seq to profile the TSSs and TESs of mRNAs in synchronously sporulating cells. **How many examples of repressive, alternative TSSs (similar to *NDC80*^{luti}) can be observed across genome during budding yeast gametogenesis?** By making quantitative comparisons of TSS/TES usage throughout a high resolution gametogenesis time course, I identified over 1000 alternative upstream TSSs that were induced in the presence of either Ime1 or Ndt80 during early or mid-late time points respectively. 114 and 93 upstream alternative TSS-downstream coding TSSs pairs were identified which were characterized by repression of the downstream coding TSS (similar to *NDC80* regulation) during early and mid-late gametogenesis respectively **(Table 8.1** and **Table 8.2)**. I also identified genomic features which selected for coding mRNAs negatively regulated by the transcription of meiotic upstream transcripts or isoforms (MUTIs), similar to the *NDC80* example. **What were the criteria common to genes repressed by transcriptional interference?** From my data, gene repression by MUTI/luti RNA transcription tended to require a relatively

strong upstream promoter such that there are at least two-thirds as many MUTI/luti transcripts as there are coding transcripts (relative upstream promoter strength ≥ 0.66) and a minimum inter-promoter distance of 90 nt. While these values were estimates from the particular techniques used in this study, it is clear that gene repression by *NDC80*^{luti} and other cases requires relatively strong upstream transcription from an alternative promoter that is not too closely located to the coding promoter. These findings agree with other studies investigating the mechanisms of transcriptional interference (TI)(Callen et al. 2004;Shearwin et al. 2005;Pande et al. 2018) (see discussion in chapter 6 for more details).

What are the other effects of overlapping isoform or ncRNA transcription on downstream genes, apart from repression? My TE-seq data set showed that the expression of many genes was either upregulated or refractory to MUTI/luti RNA transcription. This presents a complex picture where upstream, overlapping transcription does not necessarily cause TI, but can tune or activate gene expression. The poising or activation of gene expression by the transcription of upstream ncRNAs has been described in other model systems such as *fbp1* expression during glucose starvation in fission yeast and *HSPC152* expression in human melanoma cell lines (Bonen et al. 2016;Takemata et al. 2016). It is still unclear what differentiates these activated genes from the others that are repressed by TI. Perhaps the downstream promoters of these activated genes have nucleosome depleted regions (NDRs) with sequences which are not conducive for nucleosome formation (Raveh-Sadka et al. 2012). In this scenario, upstream overlapping transcription cannot establish a repressive chromatin state. In some cases, MUTI transcription might dislodge or occlude transcriptional repressors, thus leading to gene activation. It would be informative to test this hypothesis by MNase H3 ChIP of candidate activated genes during gametogenesis.

TE-seq was also applied to mutant cells in meiotic prophase to identify examples of TI in which gene repression depended on the Set2, Set3 or Spt16 chromatin remodellers. **What extent do chromatin remodellers play a role in TI mediated gene repression during budding yeast gametogenesis?** Very few genes were identified by our approach as regulated by the histone

methyltransferase Set2 and the histone deacetylase Set3 during early meiosis. On the other hand, many cryptic transcripts were de-repressed during early meiosis in the *set2Δ*, *set3Δ* or *set2Δ set3Δ* mutants such as those found at the *SAP190* and *EAP1* locus (data not shown). Perhaps the cell is more reliant on other pathways (e.g. Spt16 and/or Spt6) to suppress coding gene promoters (distinct from cryptic promoters) by transcriptional interference. A follow up experiment to test this might be to do ChIP-seq analysis for H3K36me3 and acetylated histones across the genome during gametogenesis. Perhaps H3K36me3 is not deposited at many downsteam coding promoters in the wake of MUTI/luti RNA transcription. Many more downstream coding TSSs were found to be regulated by the Spt16 histone chaperone. This underscores the importance of nucleosome deposition as a key mechanism for suppressing genes by TI. For example, *SER3* repression by *SRG1* ncRNA transcription, requires Spt16 but not H3K36me3 deposited by Set2 nor H3K4me2 deposited by Set1 (Hainer and Martens 2011;Hainer et al. 2011). In approximately half of the cases reported in **Table 8.9**, MUTI/luti transcription was also reduced; this explains why the coding TSS was upregulated in the Spt16 depletion mutant but does not exclude the importance of the histone chaperone function of Spt16 in depositing nucleosomes at the coding gene promoter during early meiosis. A follow up experiment to test this might be to do MNase-seq analysis during gametogenesis. As a caveat, the examples shortlisted in chapter 6 are likely to be an underestimate of the number of genes regulated by these factors. Indeed, *NDC80* was not shortlisted because its coding TSS was only upregulated by approximately 1.5 fold in the *set2Δ set3Δ* double mutant relative to the WT control (FDR > 0.05) and hence failed to meet the stringent selection criteria. One reason for why subtler fold changes were not deemed statistically significant could be due to the sampling of only a single meiotic time point for the mutants and the stringency of statistical testing. Further analyses are required to dissect the role of these chromatin remodellers in genome-wide gene repression by transcriptional interefence.

Interestingly, TE-Seq data from RTG experiments strongly suggest that most genes repressed by MUTI/luti RNA transcription are dynamically regulated similar to that described for *NDC80*. Within 30 minutes of RTG, the number of reads mapping to the upstream alternative TSS fell, while the number of reads mapping

to the coding TSS rose. How might this reversibility be achieved during RTG? First, MUTI/luti expression rapidly declines, perhaps due to active degradation or depletion of nuclear Ime1. Under nutrient rich conditions, *IME1* transcription shuts down and Ime1 protein translocates to the cytoplasm (Colomina et al. 2003;van Werven and Amon 2011). In addition, Ime1 has a half-life of only a few minutes, facilitating rapid clearance of Ime1 protein during return-to-growth (Chia et al. 2017). Subsequently, mitotic transcriptional activators and other chromatin remodelers stimulate nucleosome eviction in the coding gene promoter when cells resume the mitotic cell cycle. In line with this idea, levels of the repressive H3K36me3 in the *NDC80*^{ORF} promoter were rapidly lost when the repression is reversed during RTG **(Figure 4.11 E)**. The reversible repression of many genes by controlling MUTI/luti RNA transcription allows cells to quickly adjust their transcriptional program (mitosis or meiosis) in response to sudden changes in the environment.

Why is it advantageous for the cell to regulate gene expression via the transcription of overlapping mRNA isoforms or ncRNAs? As mentioned above, the reversibility of MUTI/luti RNA mediated repression allows the cell to adjust its transcriptional program to sudden changes in the environment. Furthermore, MUTI/luti RNA transcription is an evolutionarily simple mechanism to regulate gene expression. This system co-opts existing transcription factor networks such as Ime1 and Ndt80 to achieve either gene activation or gene repression in different contexts. This system only requires the evolution of transcription factor binding sites and promoter sequences upstream of another target promoter (Tresenrider and Unal 2018). Developmentally coordinated tuning, activation or repression of different sets of genes can thus be achieved by the activity of a stage specific transcription factor (Tresenrider and Unal 2018) (also see discussion in chapter 4 for more details).

7.5 mRNA isoforms and their translational control

I focused on the transcriptional control of genes in my thesis. However, post-transcriptional regulatory mechanisms are equally important in determining gene

expression output. The most comprehensive genome-wide attempt to relate mRNA isoforms to protein levels in budding yeast gametogenesis till date is from Cheng *et al*., (an updated dataset from the original paper published by Brar *et al*.) in which a combination of mRNA seq, ribosome profiling and mass spectrometry was used to infer "protein-level changes driven by transcript isoform switching" (Brar et al. 2012;Cheng et al. 2018). Cheng *et al*. found 624 genes with putative meiotic uORFs which were poorly translated compared to their canonical counterparts. The 624 transcripts were short-listed based on examples with a weak positive or negative correlation (<0.4) between mRNA seq reads across annotated ORFs and protein levels. (Cheng et al. 2018). Poor translatability of mRNA isoforms like *NDC80*^{luti} is due to upstream open reading frames (uORFs) which titrate ribosomes away from the canonical ORF (Brar et al. 2012;Cheng et al. 2018). Further analysis of gene regulation by mRNA isoforms should integrate these ribosome footprint data to identify more examples similar to *NDC80* regulation. As an example, mRNA and TSS data from my work demonstrate that the appearance of the *RAD16* long isoform during gametogenesis is correlated with reduced levels of *RAD16* mRNA **(Figure 7.1)**. Concurrently, ribosome footprinting data suggests that this *RAD16* long isoform is translationally repressed, based on lack of ribosome protected reads downstream of the *RAD16* upstream alternative TSS **(Figure 7.1)**.

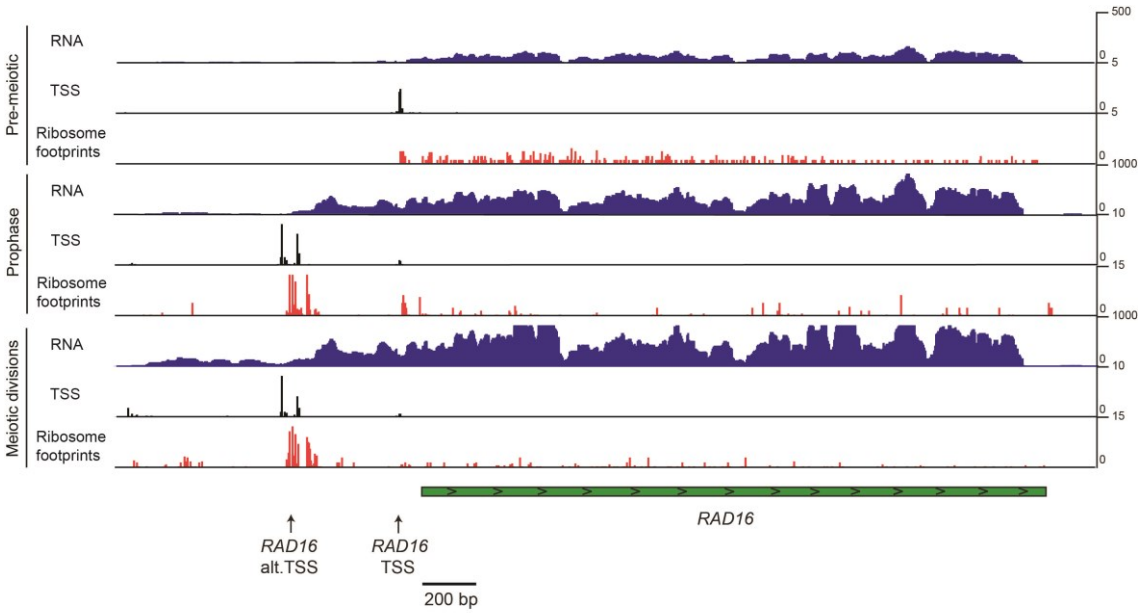


Figure 7.1 mRNA, TSS and ribosome footprinting data at the *RAD16* locus in pre-meiotic cells and during gametogenesis.

mRNA-seq and TSS data at the *RAD16* locus in *pCUP-IME1* and *GAL4.ER pGAL-NDT80* cells (FW2795) at different time points in a gametogenesis time course. Ribosome profiling data from Brar *et al.* 2012 were also incorporated (Brar et al. 2012). Ribosome profiling data corresponding to pre-meiotic, meiotic recombination and metaphase II stages were matched with Spo 0h, Spo 5h and Spo 8h for mRNA seq and TSS data respectively. mRNA seq FPKM values (blue), TSS TPM values (black) and ribosome profiling RPKM values (red) are plotted. For simplicity, only features on the "sense" (+) strand are shown. Canonical and alternative TSSs for *RAD16* are also labelled. Scales for values and distance (bp) are shown. Data shown here is an average of three independent repeats, n = 3. Alignment of ribosome footprinting data in this work was done by Cai, Li from the Luscombe lab.

Intriguingly, my data set showed that numerous mRNA isoforms of a gene co-exist during gametogenesis **(Table 8.3-8.6)**. Perhaps in some of these examples, these alternative mRNA isoforms with lower translational efficiencies might co-exist with the canonical mRNAs to increase protein levels incrementally. This could be a way of regulating gene expression separate from straightforward transcriptional repression of the coding mRNA. For example, mRNA and TSS data from my work suggest that the *VPS30* long isoform co-exists with the canonical *VPS30* mRNA during gametogenesis. Importantly, ribosome protected reads were observed downstream of the *VPS30* alternative TSS and across the ORF **(Figure 7.2)**. Only one upstream ORF could be found in the 5' leader sequence of the putative *VPS30* long isoform (data not shown); it is still unclear if this longer isoform can still be translated, albeit at lower efficiencies. More work and further analysis are required to test if mRNA isoforms can tune gene expression in this manner.

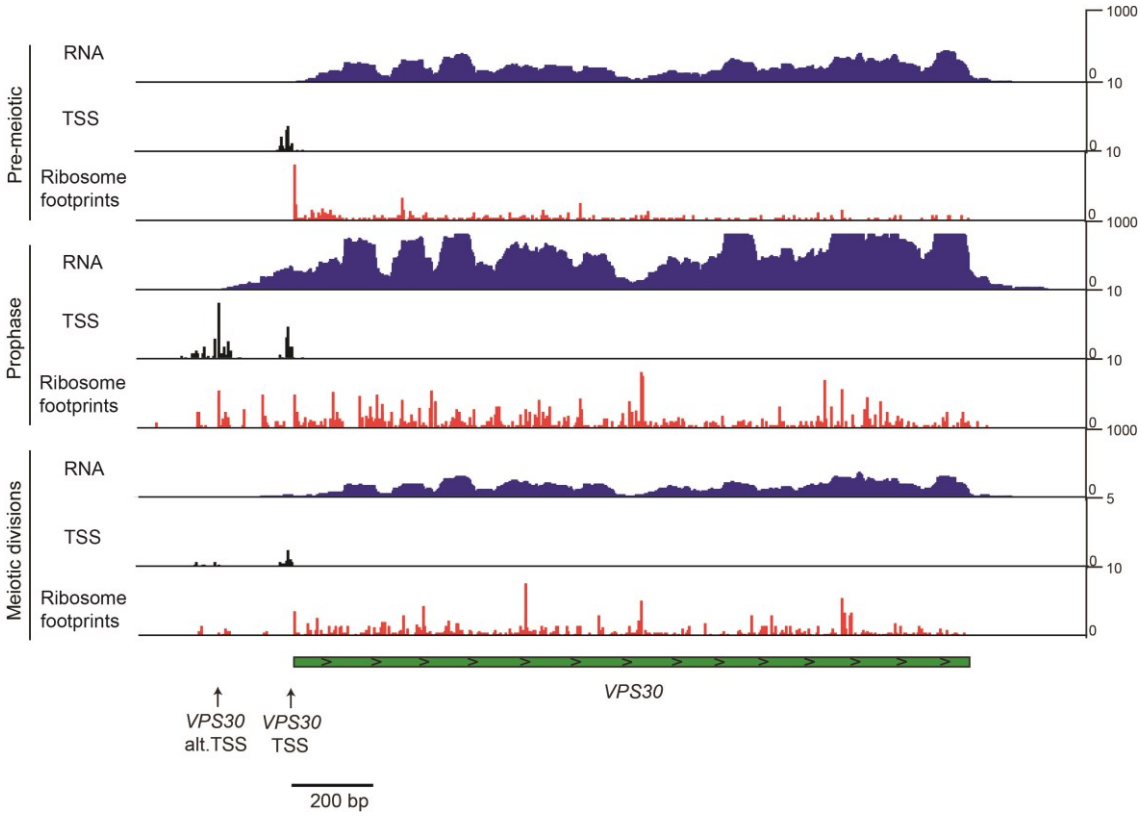


Figure 7.2 mRNA, TSS and ribosome footprinting data at the *VPS30* locus in pre-meiotic cells and during gametogenesis

mRNA-seq and TSS data at the *VPS30* locus in *pCUP-IME1* and *GAL4.ER pGAL-NDT80* cells (FW2795) at different time points in a gametogenesis time course. Ribosome profiling data from Brar *et al.* 2012 were also incorporated (Brar et al. 2012). Ribosome profiling data corresponding to pre-meiotic, meiotic recombination and metaphase II stages were matched with Spo 0h, Spo 5h and Spo 8h for mRNA seq and TSS data respectively. mRNA seq FPKM values (blue), TSS TPM values (black) and ribosome profiling RPKM values (red) are plotted. For simplicity, only features on the "sense" (+) strand are shown. Canonical and alternative TSSs for *VPS30* are also labelled. Scales for values and distance (bp) are shown. Data shown here is an average of three independent repeats, n = 3. Alignment of ribosome footprinting data in this work was done by Cai, Li from the Luscombe lab.

Notably, 261 genes were common between my list for those repressed by an upstream transcript **(Table 8.1-8.2)** and the dataset of 624 genes reported by Cheng *et al* (Cheng et al. 2018). This discrepancy is likely due to the different methods used in both studies to estimate the levels of overlapping transcript isoforms. For instance, mRNA seq data cannot distinguish all signals from overlapping transcripts with degenerate ends. Consequentially, a poor correlation between mRNA:protein levels could either be due to increased isoform abundance

with little changes in protein, or an actual *NDC80*^{luti} type of regulation whereby levels of both the coding isoform and protein decrease. Furthermore, negative associations between footprint signals of uORFs (presumably from a luti transcript) and canonical ORFs could be either due to a decrease in canonical mRNA abundance or from a decrease in ribosomes associated with similar amounts of transcripts (Ingolia 2014). As such, TE-seq provides a complimentary approach to understanding transcriptional control of genes involving ncRNA or isoform transcription during a developmental program. Although outside the immediate scope of this thesis, TE-seq data sets can also be combined with ribosome profiling and proteomic data from Cheng *et al*. for a complete view of transcriptional and translational control throughout different stages of gametogenesis.

7.6 Relevance to higher eukaryotes

In higher eukaryotes including human cells, a wide range of 5' extended, overlapping mRNA isoforms or ncRNAs are also expressed often in a cell type-specific manner although their regulatory significance remains unexplored (Wang et al. 2008;Aanes et al. 2013;Brown et al. 2014). While not formally investigated in this thesis, another source of upstream overlapping transcription is by read-through transcription from a neighbouring gene. Indeed, studies conducted in human cell lines, flies and budding yeast provide evidence for pervasive read-through transcription past the TES of most genes (Churchman and Weissman 2011;Kwak et al. 2013;Mayer et al. 2015;Nojima et al. 2015). TI by overlapping upstream transcripts might be wide-spread in mammalian cells. For example, interferon-γ leads to repression of the immune gene *MICA* via the transcription of an overlapping 5'- extended *MICA* isoform in human primary arterial endothelial cells (Lin et al. 2018). More recently, Pande *et al.* characterized numerous small promoter transcripts in 8 human cell lines, whose expression interfered with downstream coding genes (Pande et al. 2018). Thus, studying how upstream transcription regulates genes during budding yeast gametogenesis could uncover general principles of TI during development. Further insights from the budding yeast developmental model could also help answer how genes are poised or activated by overlapping, upstream transcription instead of being repressed by TI.

7.7 Future directions

Here, I will suggest and describe future directions for my work, categorized into different experimental aims/questions.

7.7.1 Validating the model of *NDC80* transcriptional regulation

The data presented in this thesis is consistent with a model whereby *NDC80*^{luti} transcription interferes with the downstream *NDC80*^{ORF} promoter during meiotic prophase. One proposed repressive mechanism involves the co-transcriptional deposition of H3K4me2 and H3K36me3, followed by the recruitment of HDACs to the *NDC80*^{ORF} promoter **(Figure 4.15)**.

One limitation of my current data and model is that I have not provided direct evidence for the recruitment of HDACs by H3K4me2 or H3K36me3, nor have I measured levels of acetylated histones at the *NDC80* locus. Important follow up experiments could include ChIP for acetylated histones, Set3 and Rpd3S subunits (e.g. Rco1) in control, *set1Δ* and *set2Δ* cells during gametogenesis. If my model is correct, then Set3 and Rco1 recruitment and a reduction in histone acetylation should be observed at the *NDC80*^{ORF} promoter during prophase. Conversely, the loss of H3K4 methylation in *set1Δ* mutants and the loss of H3K36 methylation in *set2Δ* mutants should reduce the recruitment of HDACs and interfere with de-acetylation at the *NDC80*^{ORF} promoter during prophase. These experiments are required to provide stronger evidence for my proposed model.

Furthermore, a *rco1Δ set3Δ* mutant should show increased acetylated histone ChIP signals in the *NDC80*^{ORF} promoter, even when *NDC80*^{luti} is transcribed. The *rco1Δ set3Δ* mutant should phenocopy the *set2Δ set3Δ* mutant by de-repressing *NDC80*^{ORF} in meiotic prophase, as measured by northern and western blotting. H3 MNase ChIP assays at the NDC80^{ORF} promoter can be done in the *rco1Δ set3Δ* mutant to measure changes in nucleosome occupancy arising due to absence of HDACs. As a parallel approach, lysine to arginine mutant histone alleles (e.g. H3K14R or H4K16R) can be used to abolish acetylation at these tail residues (Ferrari and Strubin 2015). Lastly, histone exchange assays

can be used to test if histone turnover is reduced at the *NDC80*^{ORF} promoter during meiotic prophase in both the *set2Δ set3Δ* mutant and the *rco1Δ set3Δ* mutant. These experiments could be useful for dissecting the role of histone de-acetylation in nucleosomal dynamics and structure during *NDC80*^{luti} transcription.

7.7.2 Addressing global changes in gene expression for TE-seq analysis

Gametogenesis in budding yeast takes place in the context of starvation and meiotic signalling pathways, resulting in global changes in gene expression (Primig et al. 2000). True variation in TSS/TES usage between samples might be masked by global changes in gene expression and thus confound DESeq2 results. One solution for this issue is to spike in external transcripts of known sequence and abundance to the same number of budding yeast cells at each time point. In principle, these spike-ins should serve as an invariant external standard for calculation of normalisation factors by DESeq2 (Love et al. 2014;Chen et al. 2015).

Different spike-ins can be considered for my TE seq data set in budding yeast, such as a commercially available mixture of External RNA Control Consortium (ERCC) transcripts, fission yeast RNA or even whole fission yeast cells (Lemire et al. 2011;Malabat et al. 2015;Booth et al. 2016). Using ERCC spike-ins is not recommended for my TE-seq protocol because they were found to vary between technical repeats due to inefficient capture during poly(A)+ selection (Qing et al. 2013). With this in mind, fission yeast RNA or whole fission yeast cells should be used. Care must be taken to ensure that the same amount of spike-ins are added to the same starting number of budding yeast cells. Re-making TE-seq libraries with spike-ins from selected time-points could help to identify bona fide control genes with relatively stable expression across gametogenesis. These control genes can then be used to retrospectively normalise existing TE-seq data (Chen et al. 2015).

However, the use of spike-ins relies on a crucial assumption that technical factors (e.g. RNA extraction efficiencies) affect the the spike-ins to the same extent as endogenous transcripts (Risso et al. 2014). It is not clear if this assumption

holds true for samples taken during budding yeast gametogenesis. In late gametogenesis, spore wall formation and spore packaging could reduce the efficiency of RNA extraction (Ares 2012). As a consequence, the ratio of spike-ins to endogenous RNA could vary between pre-meiotic cells and mid-late stage meiotic cells. This would lead to incorrect normalisation solely based on spike-ins. Further work needs to be done to test if spike-ins can provide robust normalisation for meiotic cells.

Alternatively, some newer computational approaches like *moose*^{2} and median Composition-Decomposition (MedianCD) have been developed to normalise datasets without external spike-ins, even in a scenarios with widespread changes in gene expression (Berghoff et al. 2017;Roca et al. 2017). The *moose*^{2} method predicts a set of control genes based on raw counts, which can then be used as endogenous standards to normalise libraries (Berghoff et al. 2017). Likewise, the MedianCD method uses ANOVA and post-hoc statistical testing to identify genes which do not vary between conditions, and subsequently uses the identified set of control genes for “between sample” normalisation (Roca et al. 2017). The MedianCD method is especially useful because it does not require prior knowledge of control genes (Roca et al. 2017). These approaches should be tested if they can improve TE-seq analysis and identify differentially expressed TSS/TES tags more robustly.

7.7.3 Isoform detection by integrating TIF seq data

One major limitation of TE-seq is that the 5’ and 3’ ends of any particular transcript are not matched. Therefore, any isoform prediction using TE-seq assumes that TES usage is stochastic, i.e. transcripts originating from a given TSS will use all available downstream TESs proportionally. As such, novel isoform identification would be strengthened by TIF seq, which maps paired 5’ and 3’ ends of all mRNAs (Pelechano et al. 2014).

7.7.4 Integrating TE-seq analysis with ribosome profiling and proteomic data

TE-seq, together with existing ribosome profiling data sets from Brar *et al.* could be used to categorize mRNA isoforms by their involvement in transcriptional regulation, as well as their translational capacities (Brar et al. 2012). Perhaps some isoforms are not translated such as *NDC80*^{luti}, while others might be translated at different efficiencies depending on their 5' leader sequences. Mass spectrometry proteomic data from the Brar group would also be important in correlating transcript isoform levels with protein levels (Cheng et al. 2018). Although merely speculative at this juncture, some 5' extended transcript isoforms could code for proteins with different N-termini, similar to that observed in plants (Ushijima et al. 2017). To summarize, TE-seq together with ribosome profiling data (rate of translation) and proteomic data (protein abundance) would provide a more complete understanding of how genes are regulated during cell fate transitions.

7.7.5 Validating findings from TE-seq data

An important caveat of interpreting TE-seq data is that all the trends described in chapter 6 are correlative. Further experiments should be done to bolster the conclusions drawn from genome-wide analysis. One possible approach is to clone candidate sequences containing the alternative upstream promoter and the downstream coding promoter in frame with an unstable fluorescent protein, such as *eGFP* fused to the *CLN2 PEST* degron (Salama et al. 1994). Due to the high turnover rates of the reporter, changes in the fluorescent signal would be a proxy for transcriptional regulation during gametogenesis. Hypotheses about gene regulation by MUTIs/luti RNAs can then be further tested by mutating the upstream and/or downstream promoters. Microscope images of cells expressing fluorescent reporters would also complement population-based TE-seq data with measurements from individual cells.

7.7.6 Identifying other regulatory transcripts during gametogenesis

It should be noted that the work in this thesis focused on overlapping, upstream transcripts during gametogenesis. Another class of transcripts observed

in my TE-seq data set were overlapping anti-sense transcripts. For example, an anti-sense transcript was initiated from the 3' end of the *PUS7* gene specifically during mid-late gametogenesis from Spo 7-9h (data not shown). Different studies in have shown that overlapping anti-sense transcripts can either repress, activate or tune gene expression depending on the context. (Hongay et al. 2006;Murray et al. 2015;Huber et al. 2016). For example, overlapping sense and anti-sense transcription mediates reciprocal activation and repression in some tandemly arranged genes during different phases of the yeast metabolic cycle (Nguyen et al. 2014). The functions of the meiotic anti-sense transcripts observed in my data set should be further examined in future work, perhaps using fluorescent protein reporter constructs as well. Another class of transcripts observed in my data set were intragenic transcripts which initiated from within gene bodies in wildtype cells during meiosis. For example, intragenic transcripts in the same orientation as the coding transcript, were observed at the *NDI1* and *SSE2* loci in the early and mid-late gametogenesis time points respectively. The functions of these intragenic transcripts are not well understood. A recent study showed that an intragenic *MRK1* transcript isoform encoding an N-terminally truncated protein isoform was made during mid-late gametogenesis; this protein isoform was found to promote efficient sporulation in mutants depleted of *RIM11*, a homolog of *MRK1* with redundant functions (Zhou et al. 2017). Perhaps some intragenic transcripts identified by TE-seq can encode protein isoforms important for meiosis. The functions of the meiotic intragenic transcripts observed in my data set should be further examined in future work, perhaps by tagging various candidate ORFs and looking for differently sized proteins. These observations reinforce the exciting idea that cells adopt different strategies to adjust gene expression outputs during a developmental program like gametogenesis.

7.7.7 Extending TE-seq to mammalian cell differentiation

Notwithstanding the usefulness of the budding yeast model, it is important to extend my findings by applying TE-seq to differentiating mammalian cells. Gene regulation by the transcription of mRNA isoforms or ncRNAs during mammalian cell differentiation could involve more complex molecular mechanisms which do not exist in budding yeast such as RNAi and DNA methylation (Han et al. 2007;Bao et

al. 2015). A recent study in mouse ES cells showed that H3K36me3 recruits Dnmt3b to methylate DNA in gene bodies to suppress cryptic transcription (Neri et al. 2017). Perhaps some coding mRNAs could be regulated in a similar fashion by the deposition of H3K36me3 in their promoters as a consequence of upstream transcription. Future work examining different models from yeast to humans will give a fuller picture of how mRNA isoforms and ncRNAs regulate gene expression.

Chapter 8. Appendix

8.1 Supplemental figures

8.1.1 ChIP data normalised to *HMR*

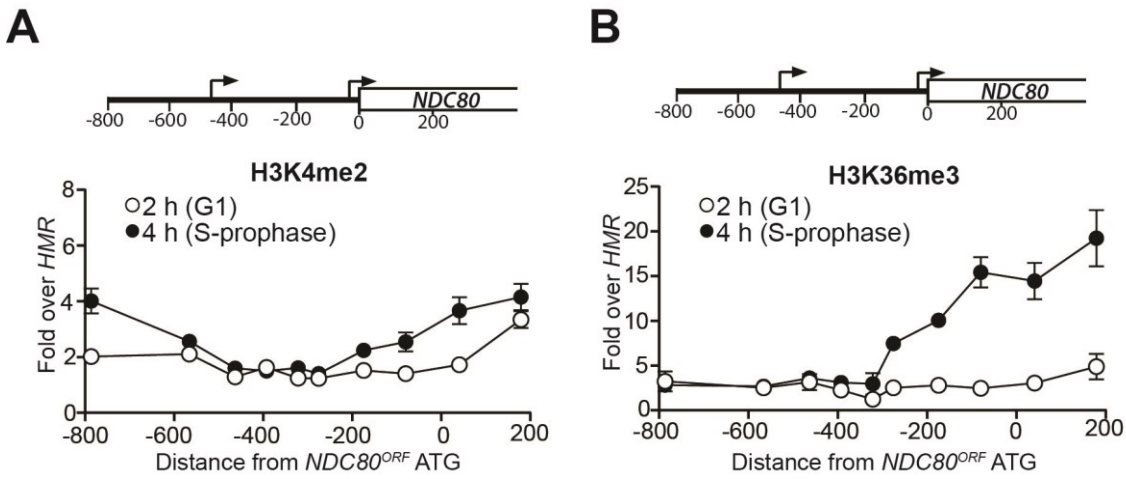
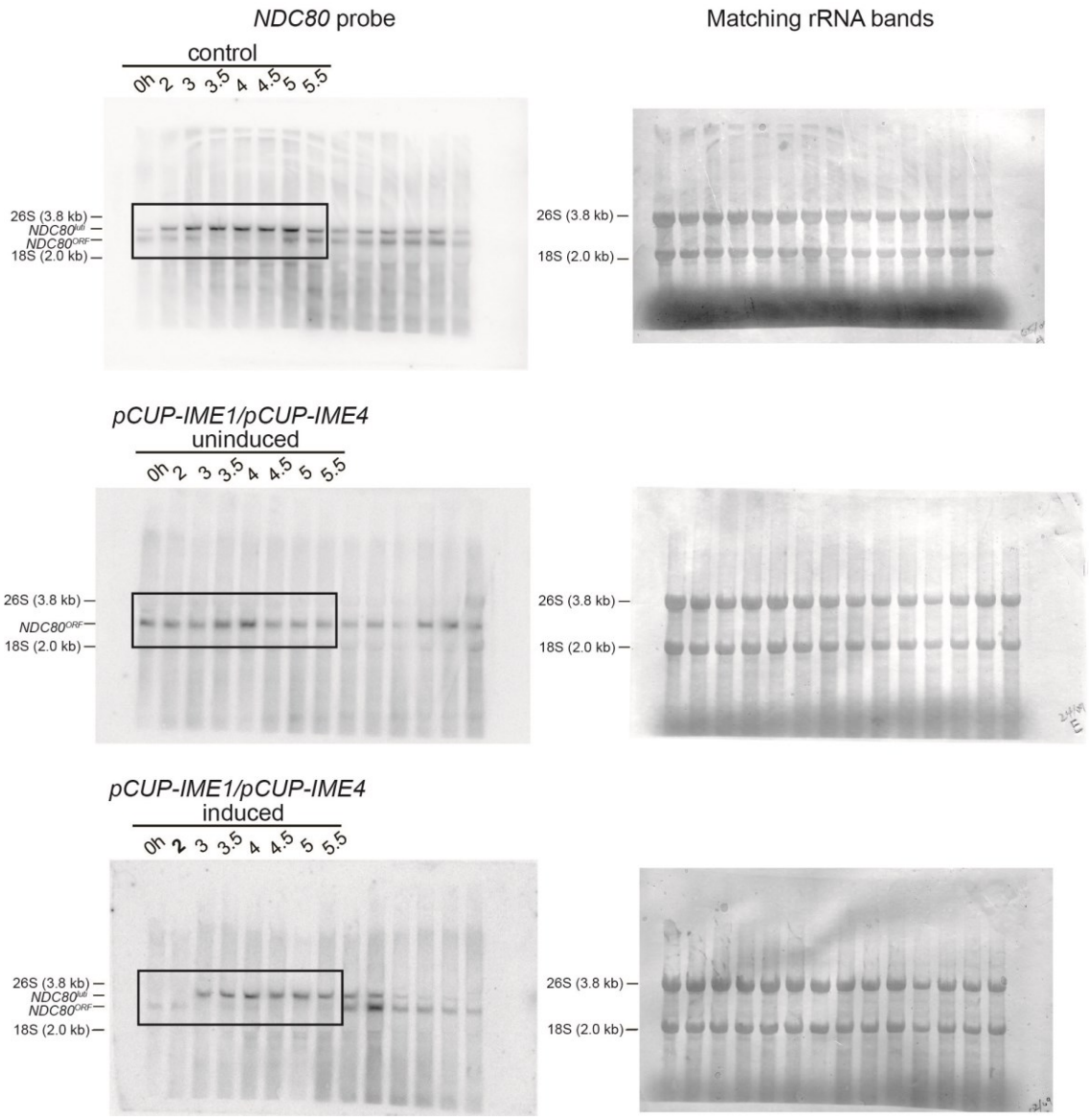


Figure 8.1 Transcription of *NDC80^{luti}* promotes H3K4me2 and H3K36me3 in the promoter and 5' region of *NDC80^{ORF}*

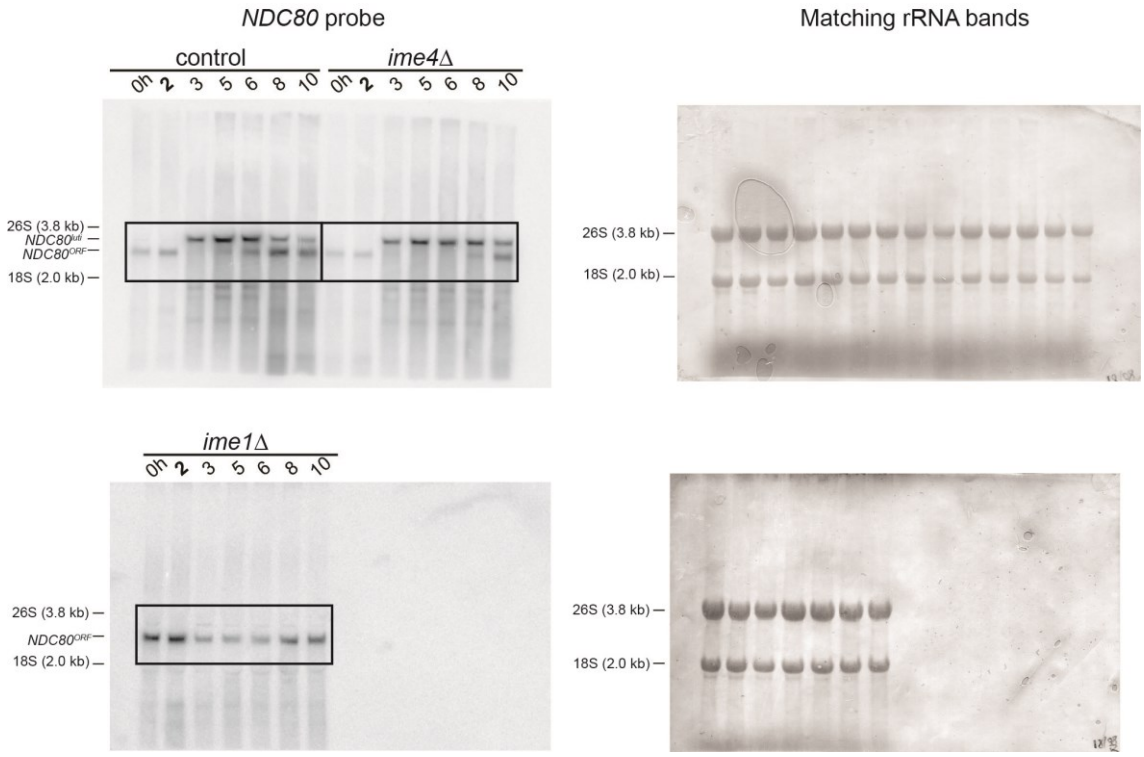
(A) Wildtype (FW1902) and *set1Δ* (FW3033) cells harbouring the *pCUP-IME1*/*pCUP-IME4* alleles were induced to undergo meiosis synchronously. Samples for chromatin immunoprecipitation were taken at two hours (2 hr (pre-meiotic), no *NDC80^{luti}* transcription) and four hours in sporulation medium (SPO) (4 hr (S + prophase), *NDC80^{luti}* transcription). Cells were fixed with formaldehyde, chromatin extracts were prepared and H3K4me2 enriched fragments were immunoprecipitated using magnetic Prot A beads coupled with anti-H3K4me2 antibodies. The recovered DNA fragments were quantified by qPCR using ten different primer pairs scanning the *NDC80* locus. The midpoint position of each primer pair is indicated in the x-axis. The mean enrichment for each primer pair and the SEM from three independent experiments, n = 3, is displayed. The signals were normalized to the signal from the silent mating-type locus (*HMR*). **(B)** Similar to panel A, except that histone H3 lysine 36 trimethylation (H3K36me3) abundance was determined by ChIP. Wildtype (FW1902) and *set2Δ* (FW1472) cells harbouring the *pCUP-IME1*/*pCUP-IME4* alleles were used for the analysis. The mean enrichment for each primer pair and the SEM from three independent experiments, n = 3, is displayed.

8.1.2 Uncropped northern blots

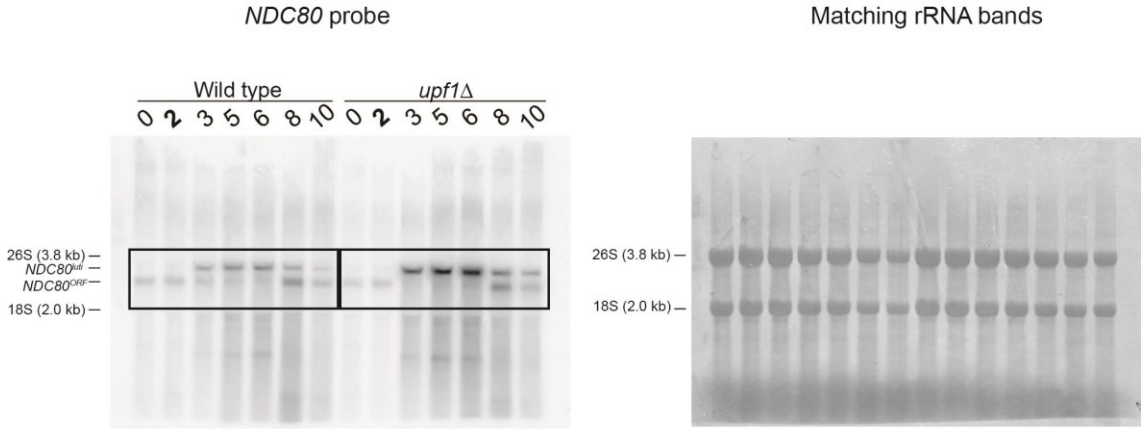
Uncropped northerns from Figure 4.1B



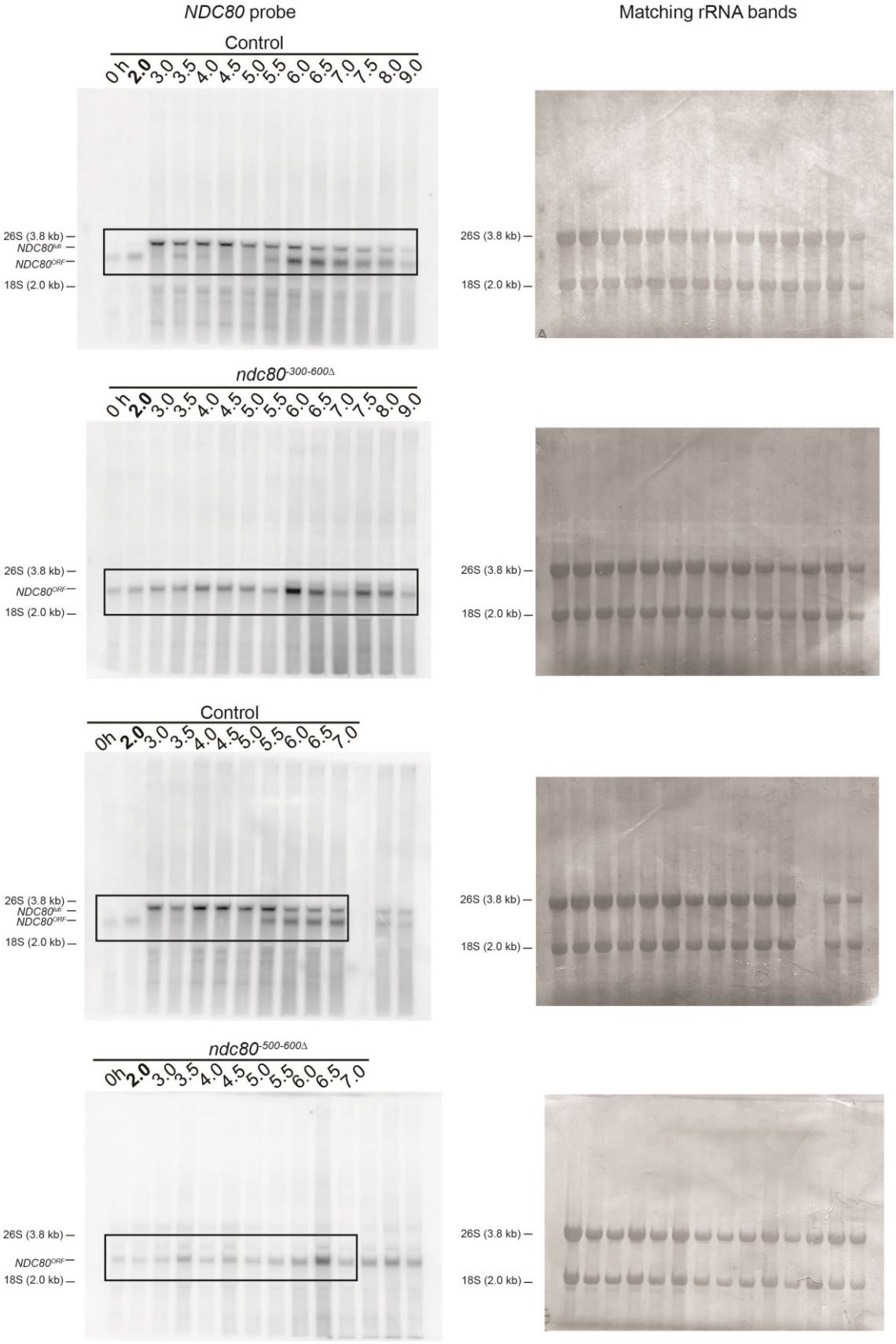
Uncropped northerns from Figure 4.1C



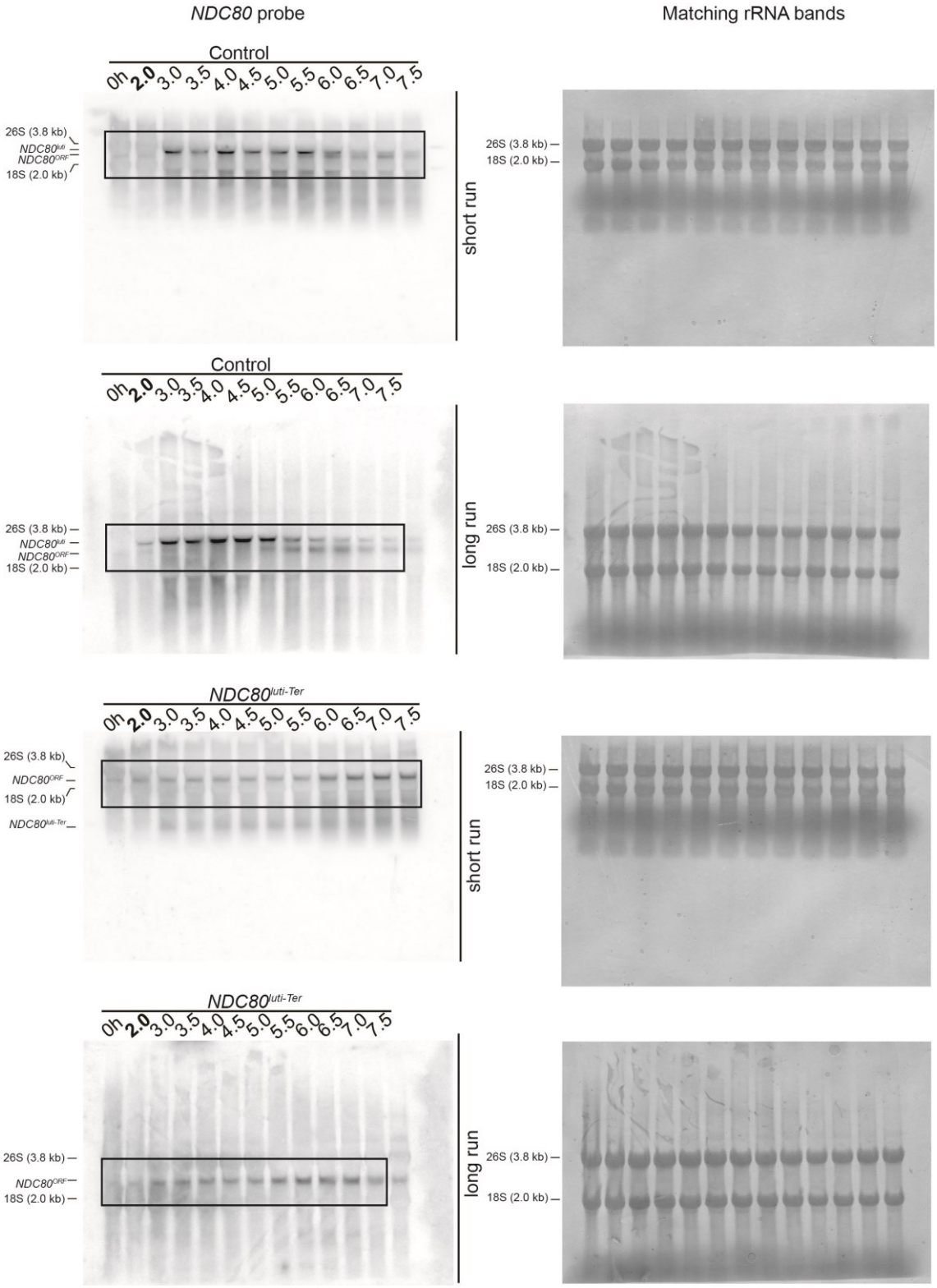
Uncropped northerns from Figure 4.2A



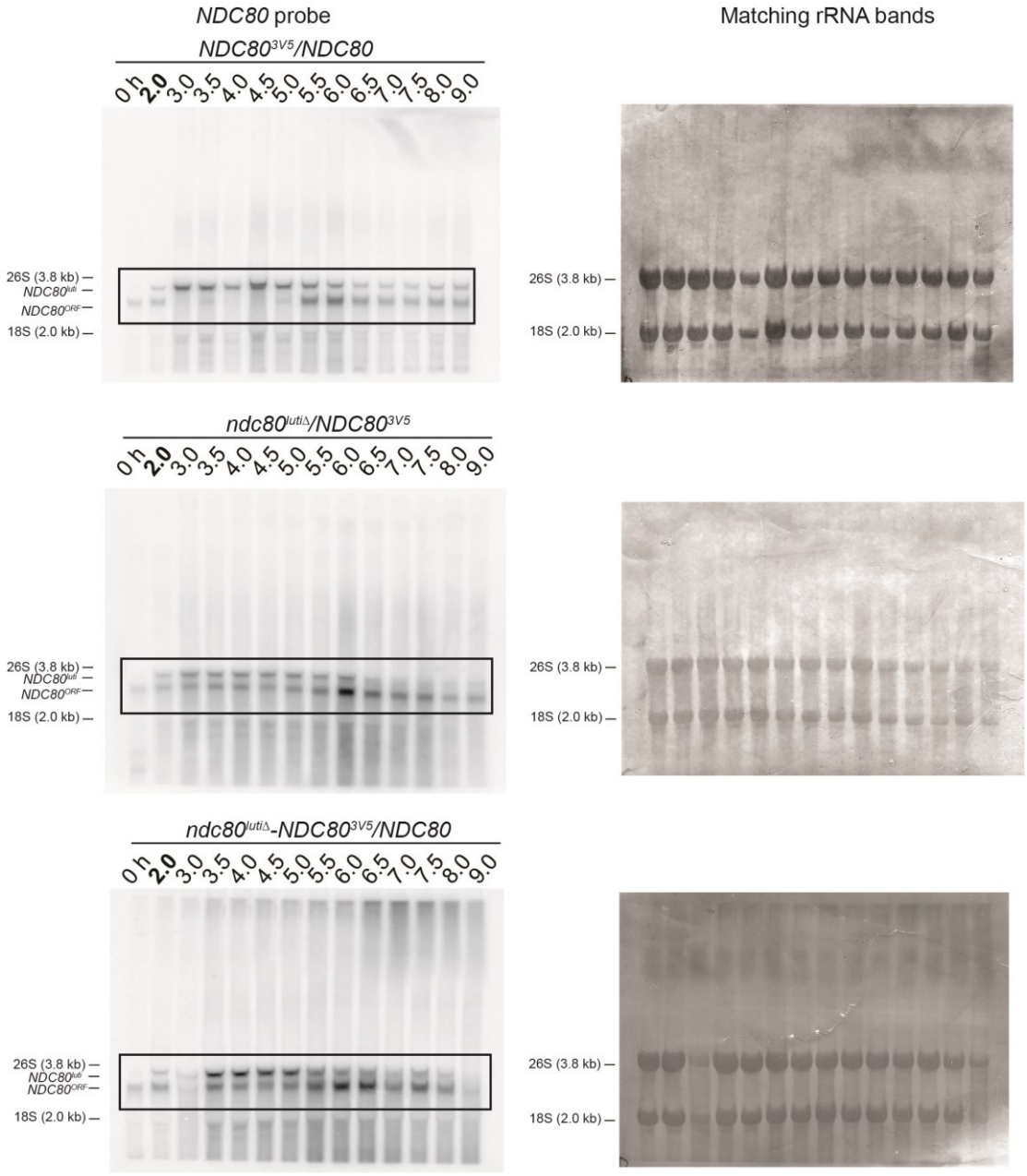
Uncropped northerns from Figure 4.4 A and B



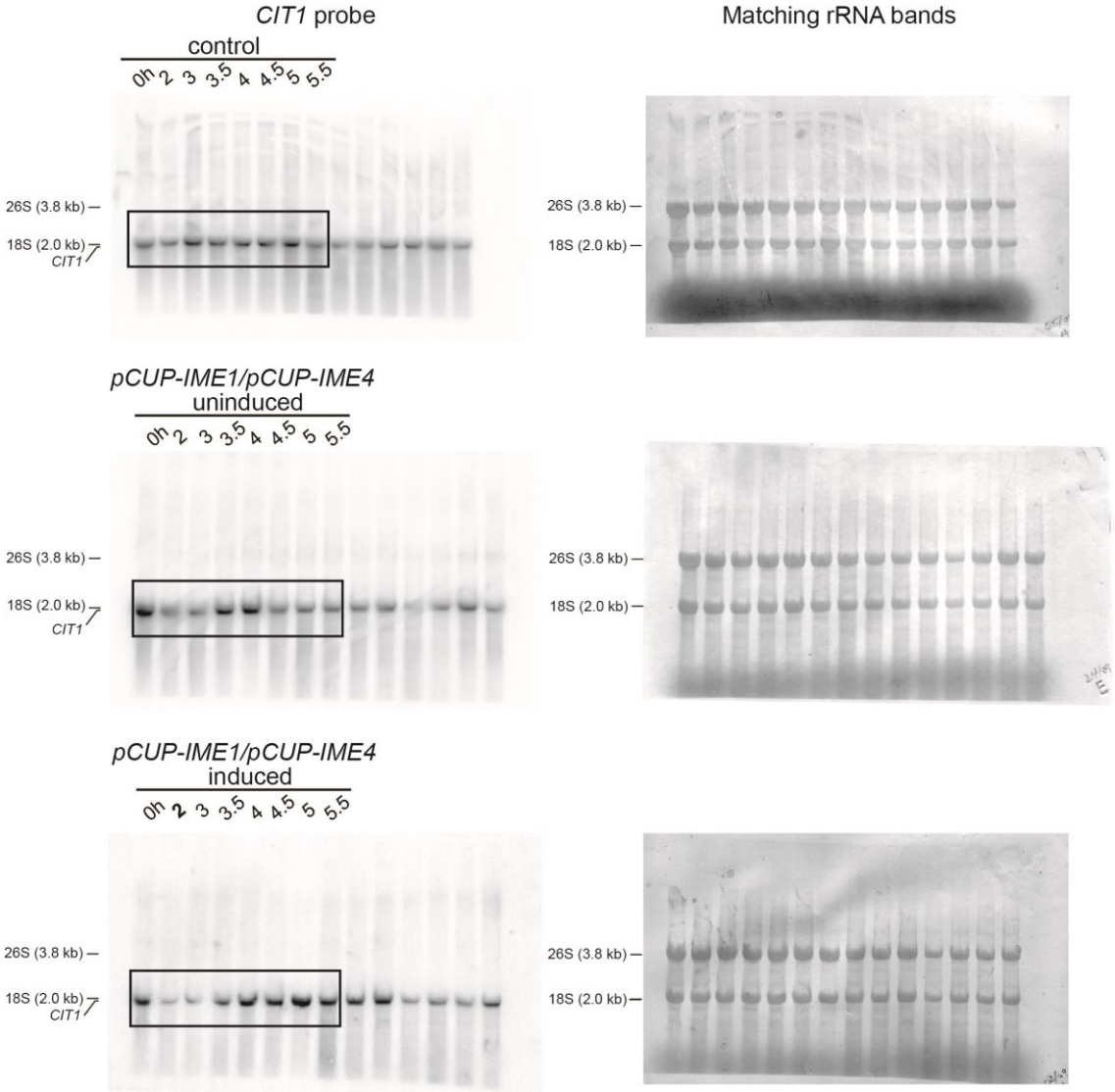
Uncropped northerns from Figure 4.4C



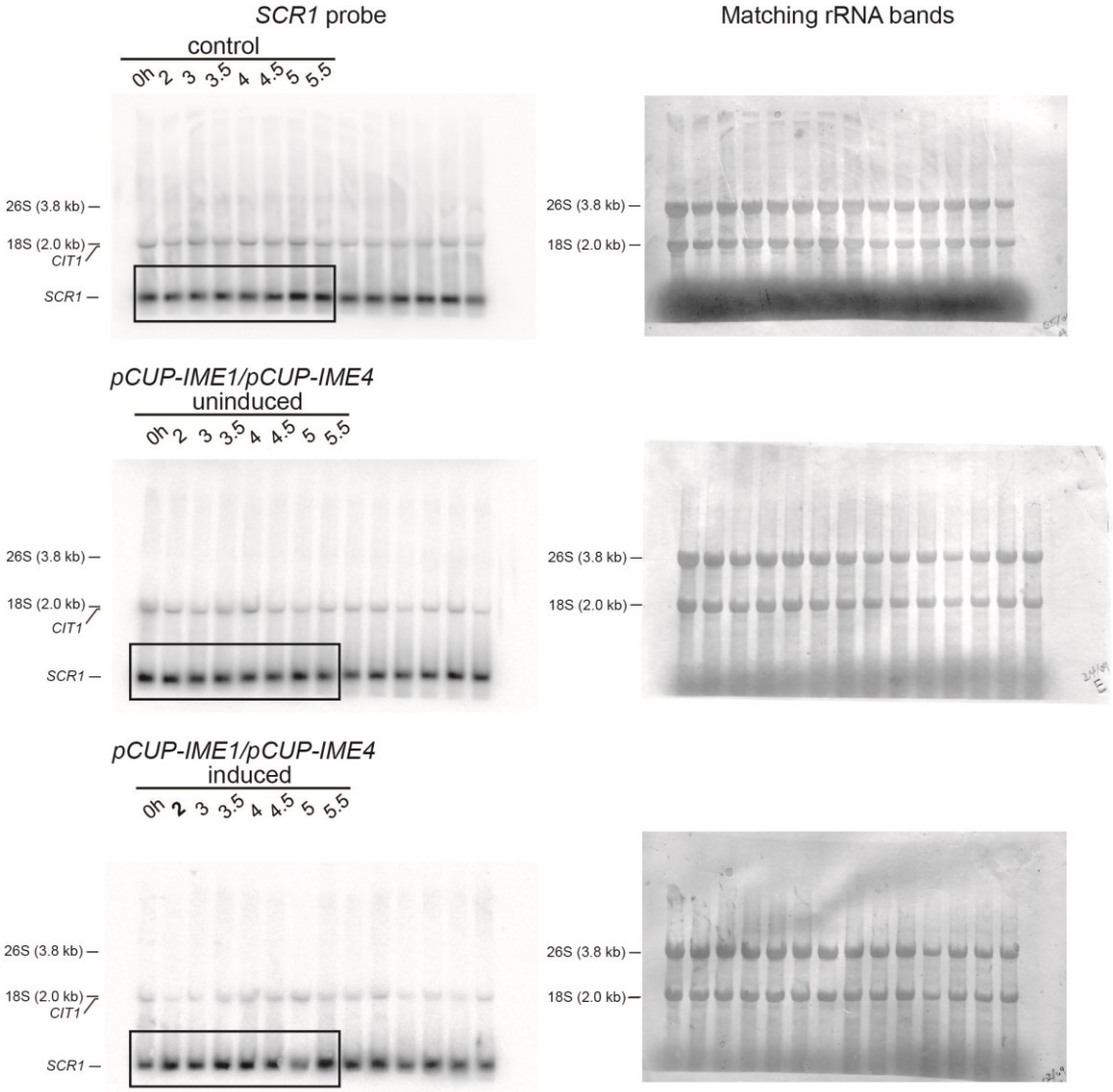
Uncropped northerns from Figure 4.5



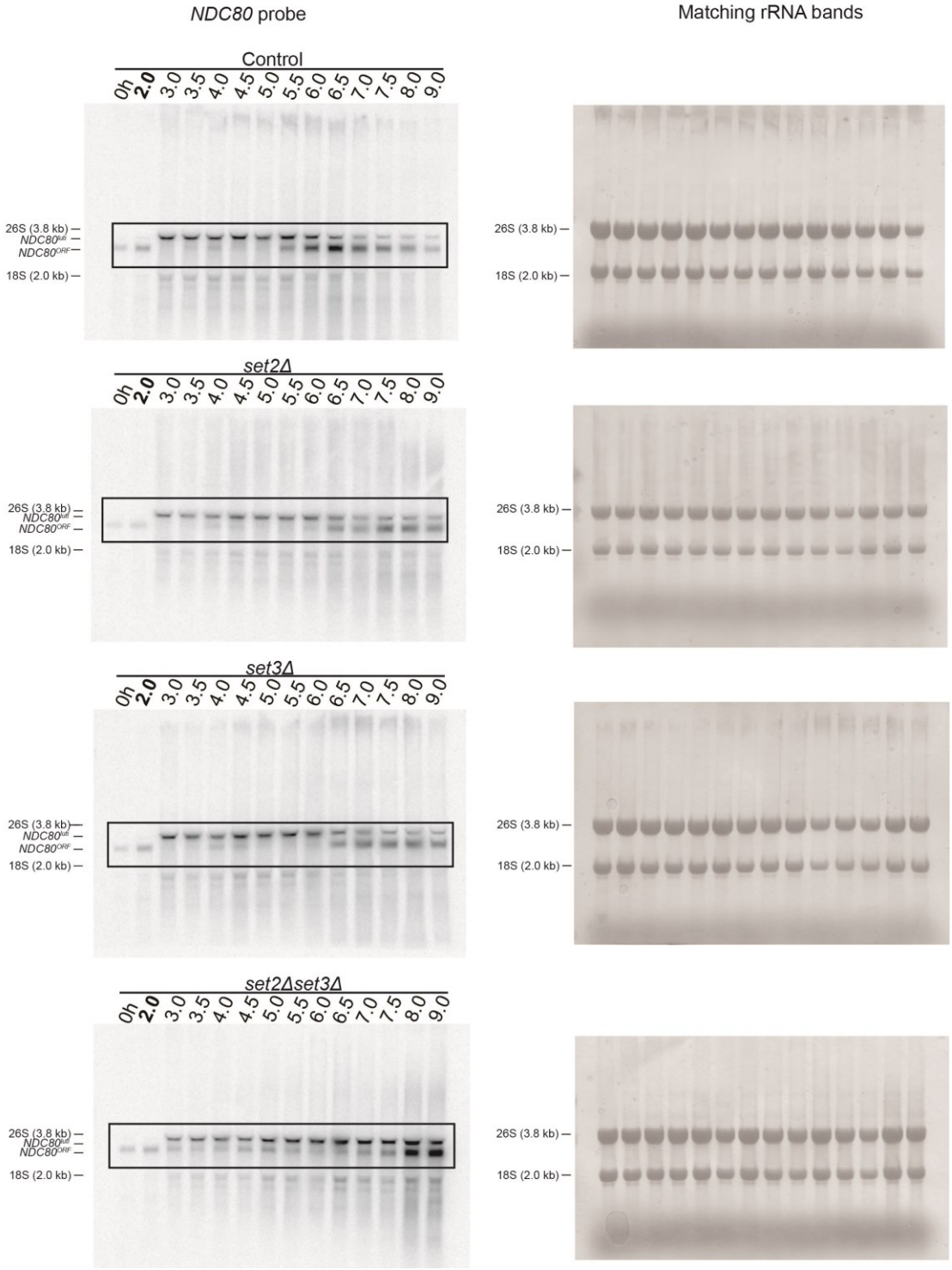
Uncropped northerns from Figure 4.7A



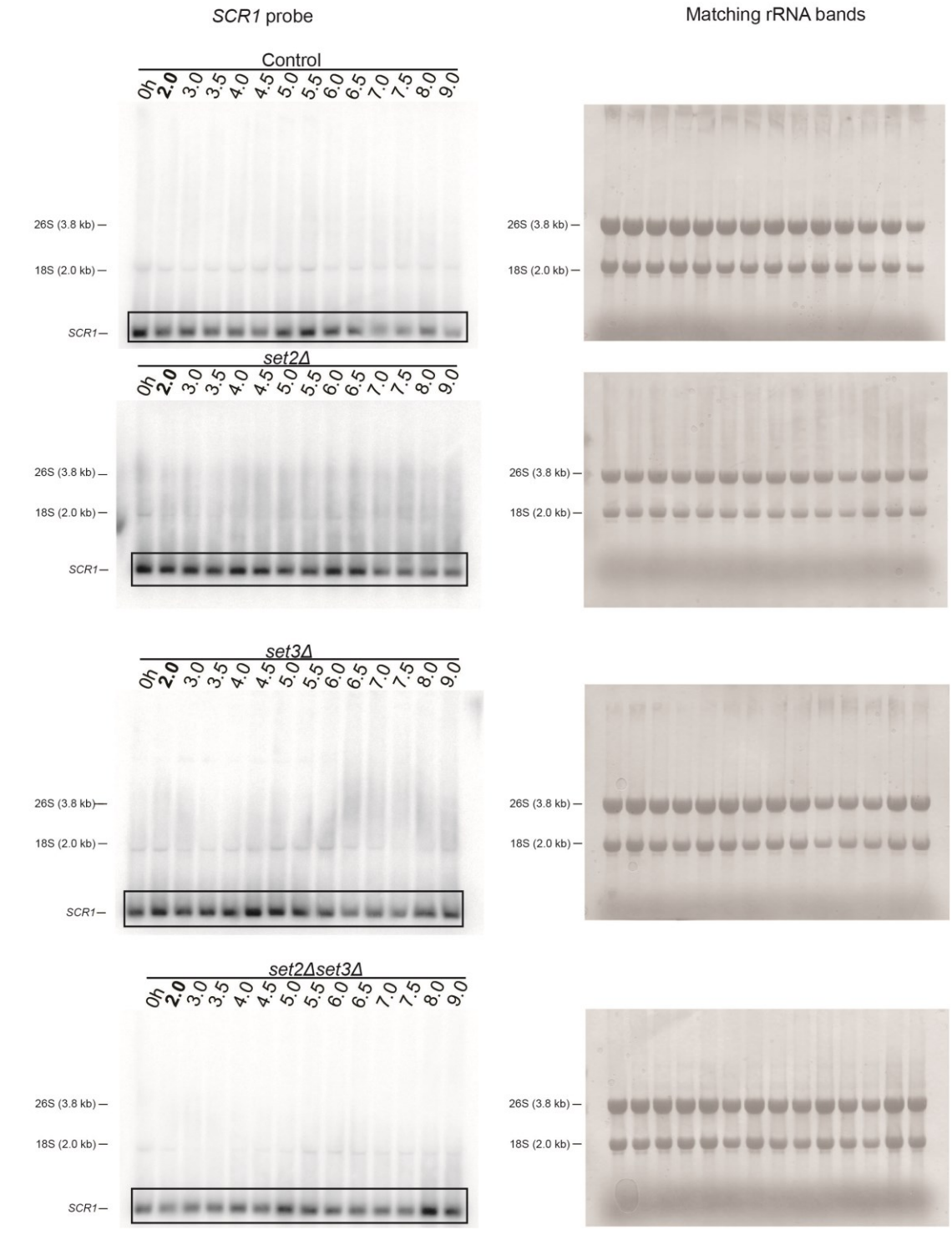
Uncropped northerns from Figure 4.7A



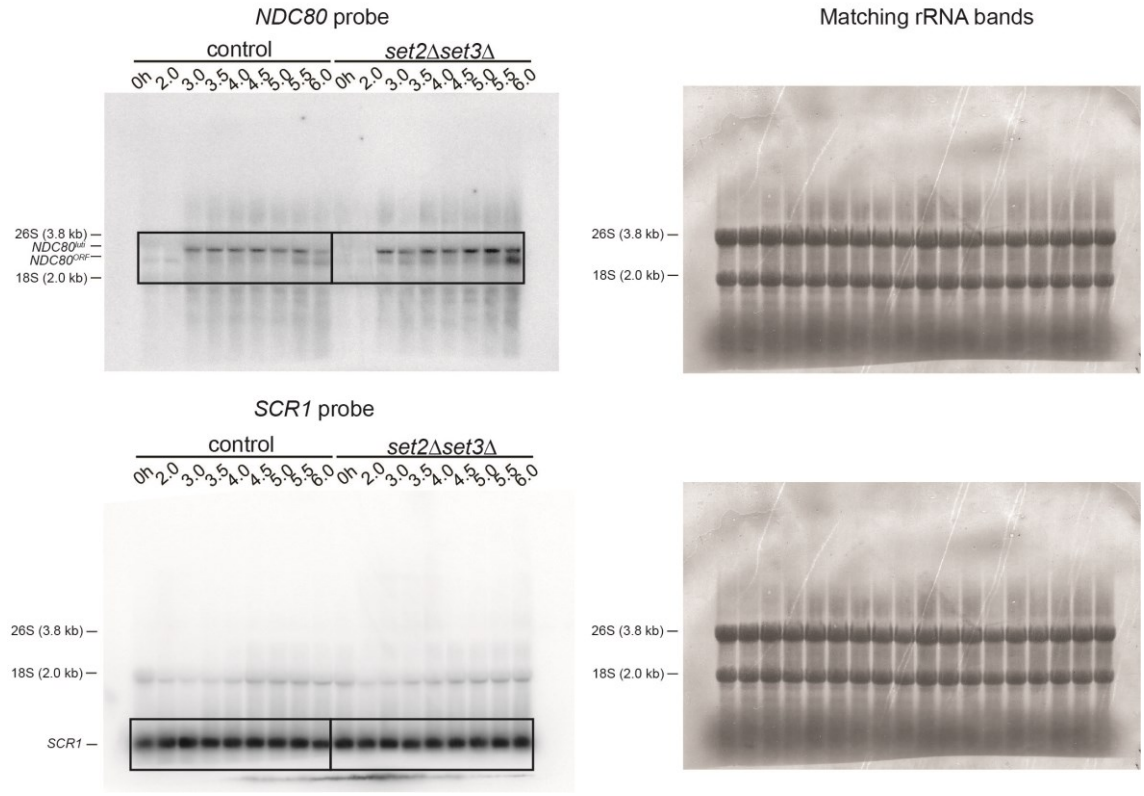
Uncropped northerns from Figure 4.8A



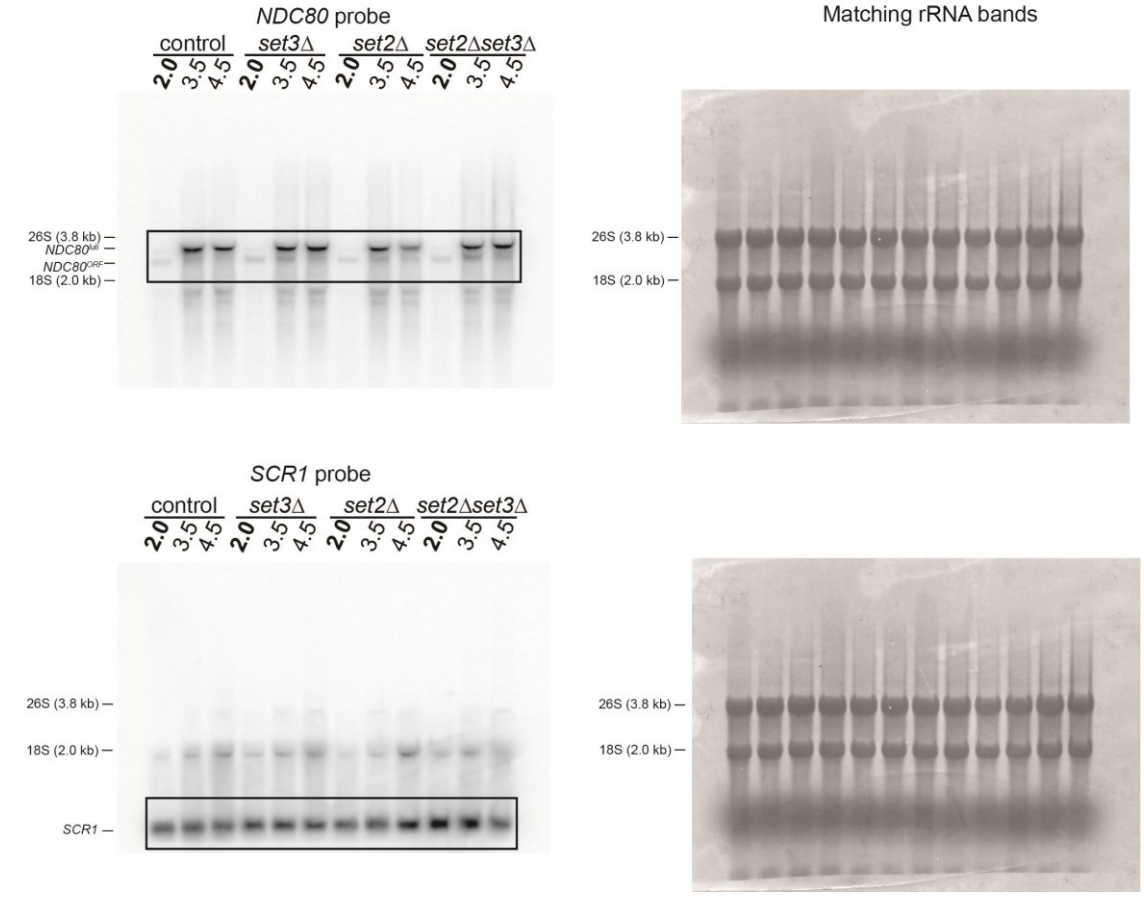
Uncropped northerns from Figure 4.8A



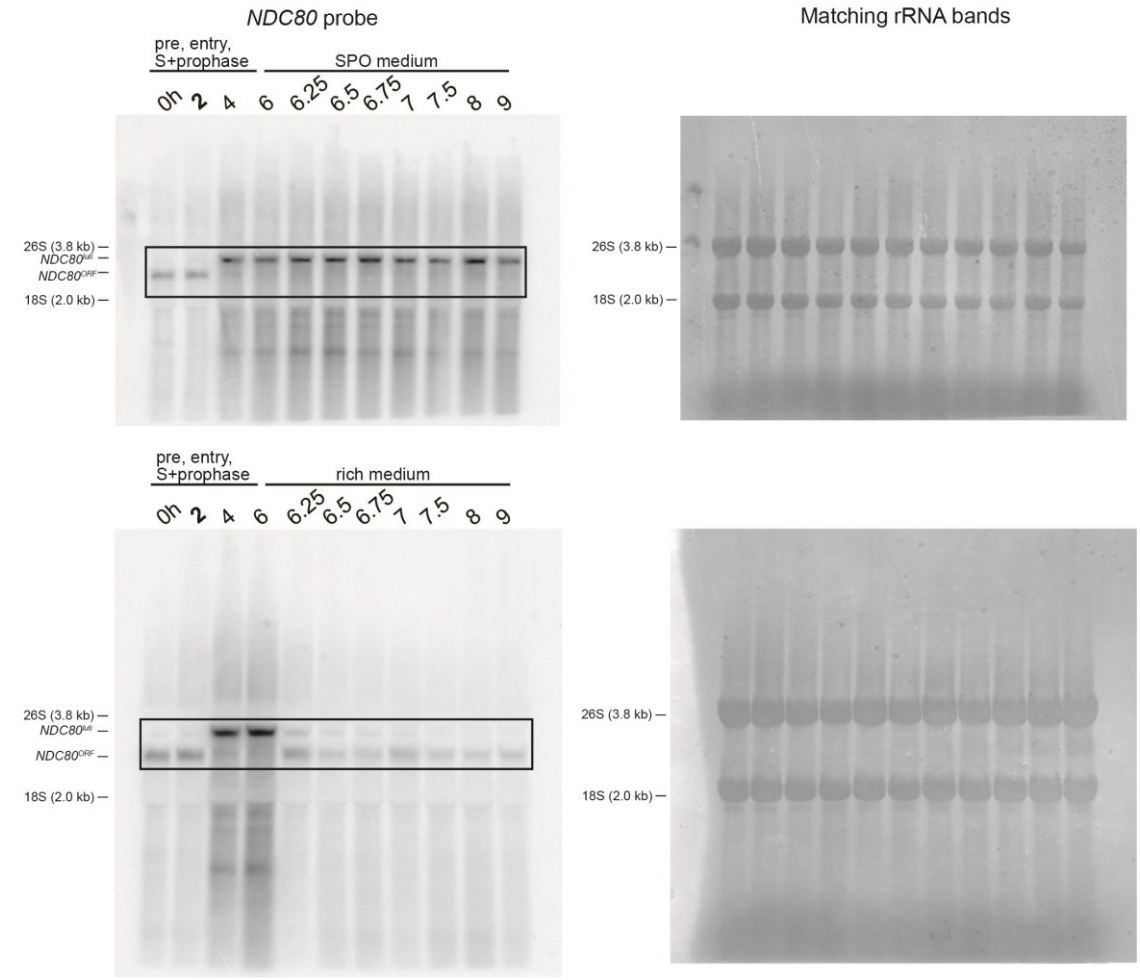
Uncropped northerns from Figure 4.9D



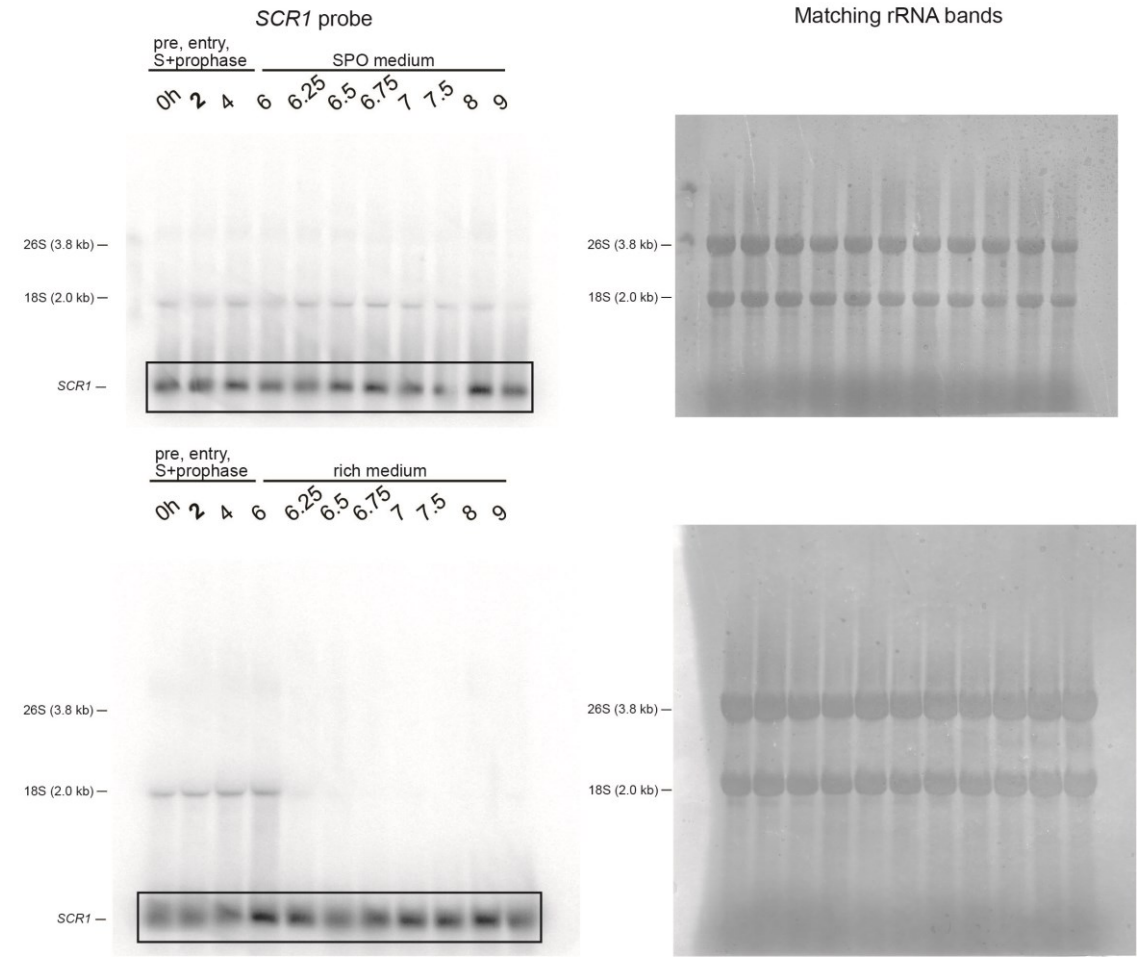
Uncropped northerns from Figure 4.9G



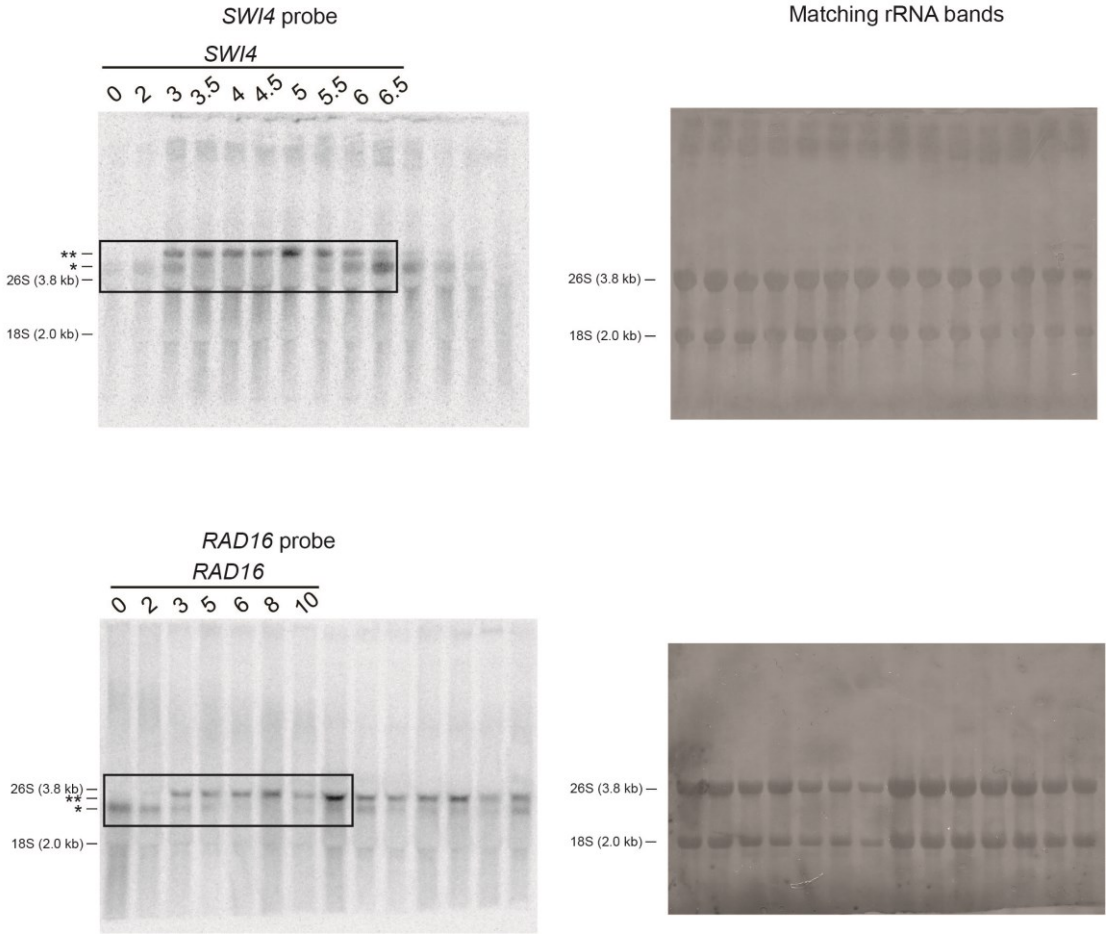
Uncropped northerns from Figure 4.11B



Uncropped northerns from Figure 4.11B



Uncropped northerns from Figure 6.11D



Uncropped northerns from Figure 6.12D

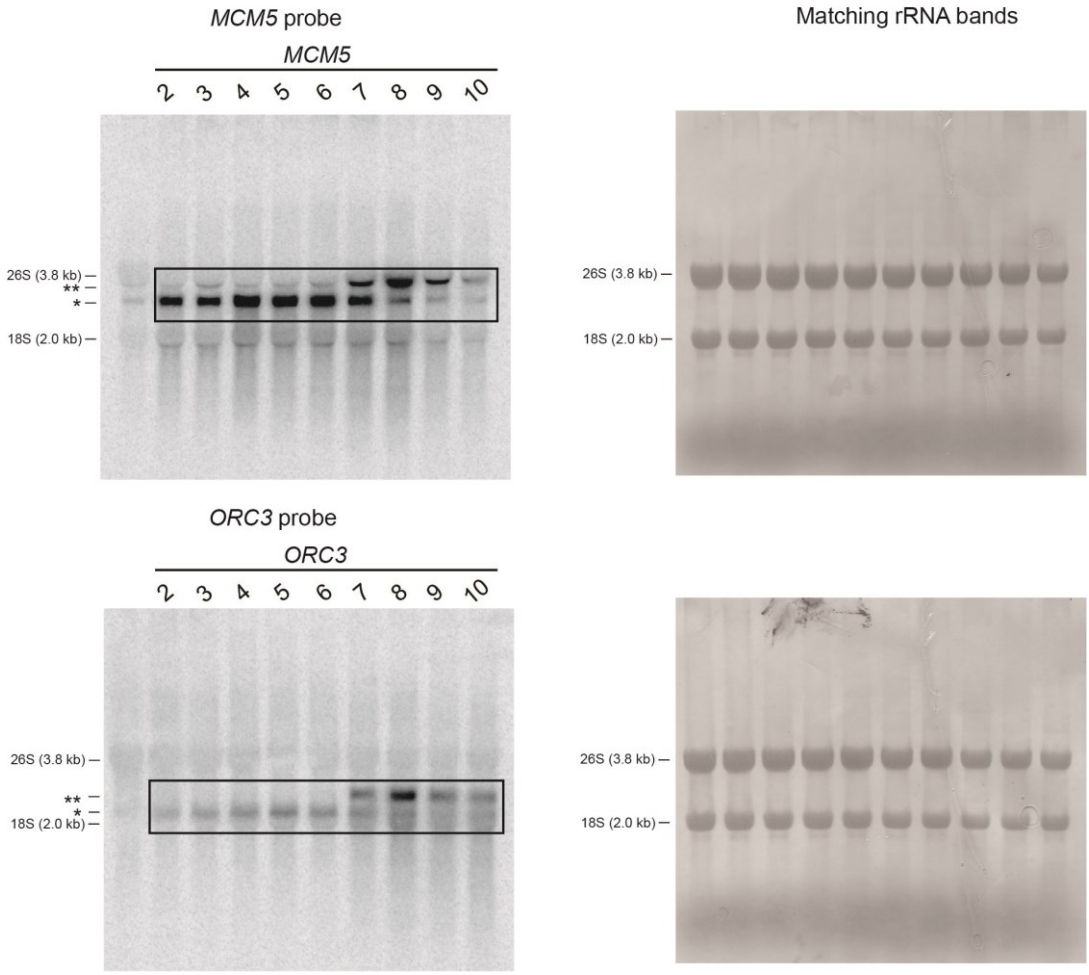
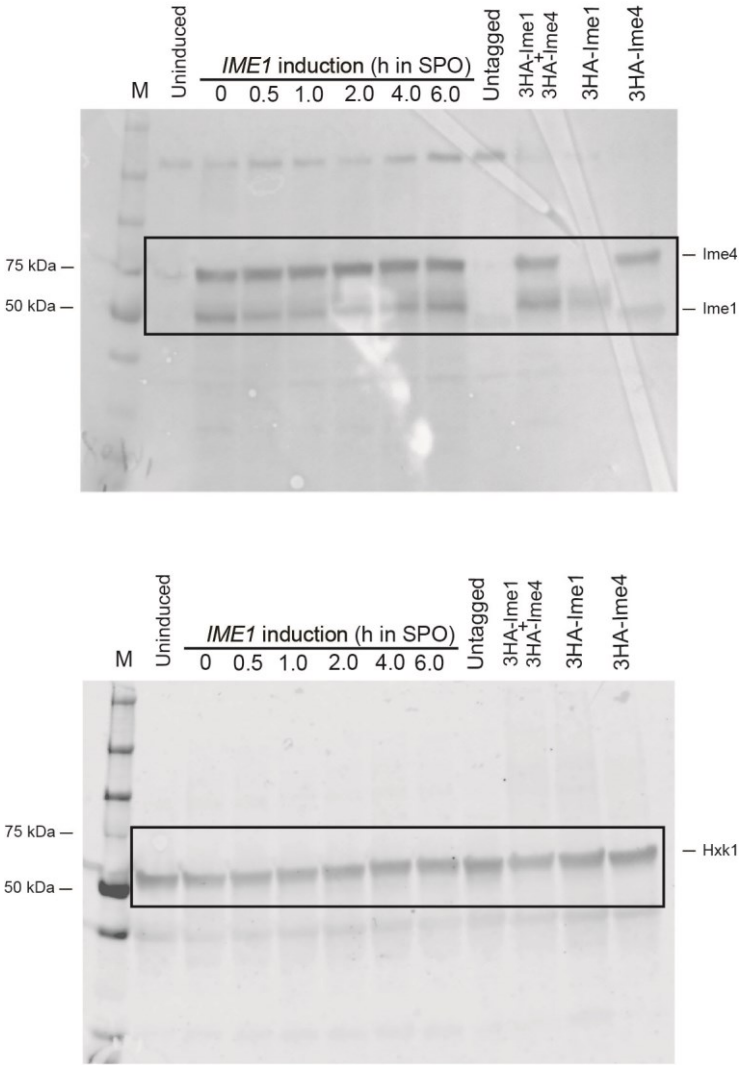


Figure 8.2 Uncropped northern blots corresponding to main figures

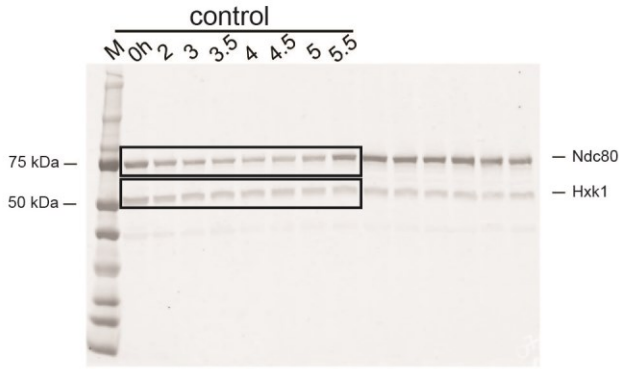
Uncropped northern blots showing the positions of the 26S and 18S rRNA bands relative to the bands of interest. The black boxes indicate the approximate area which was cropped for presentation in the main figures. For reference, the rRNA bands are displayed at the side. Please refer to corresponding legends in the main figures for further details. Note that Figure 4.7 A is the same northern as Figure 4.1 B.

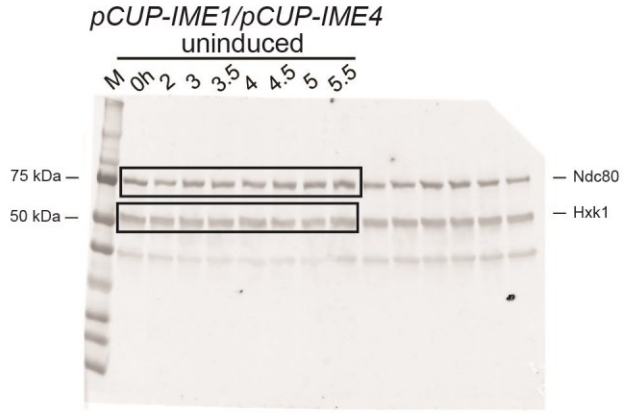
8.1.3 Uncropped western blots

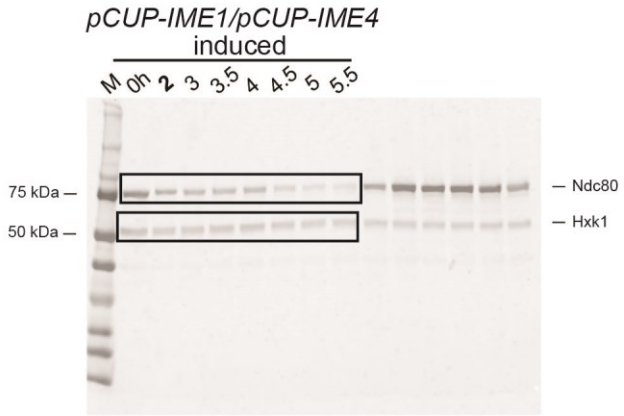
Uncropped western blots for Figure 3.1C



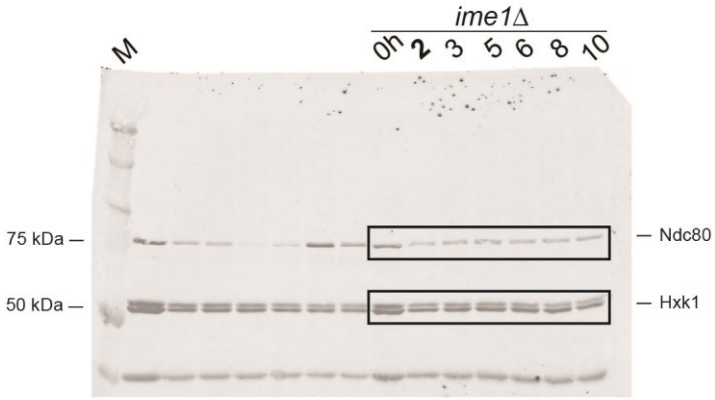
Uncropped western blots for Figure 4.1B

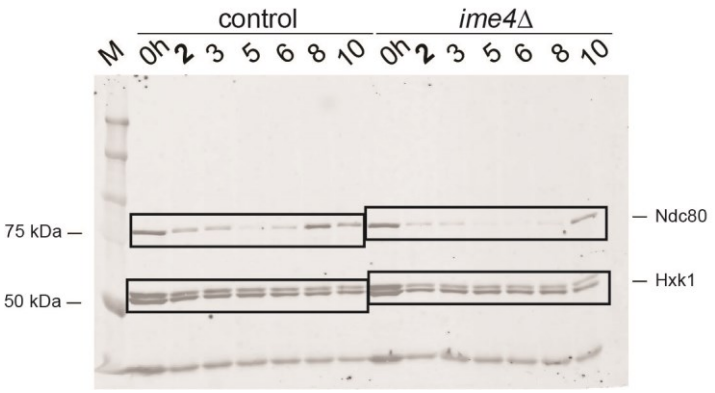




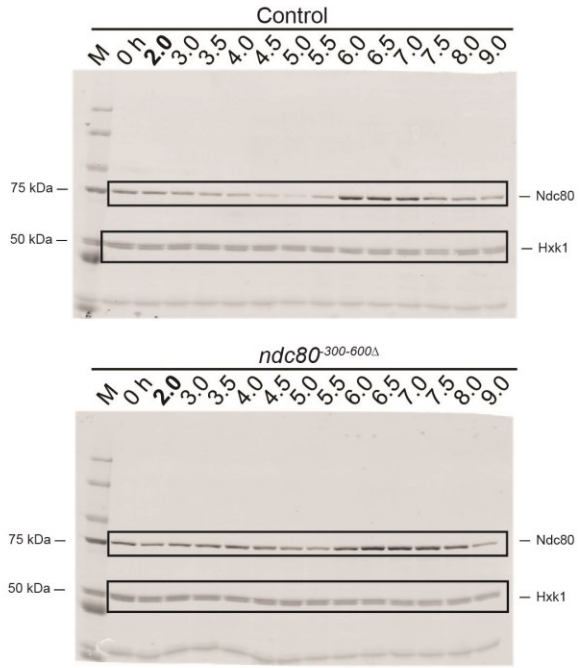


Uncropped western blots for Figure 4.1C

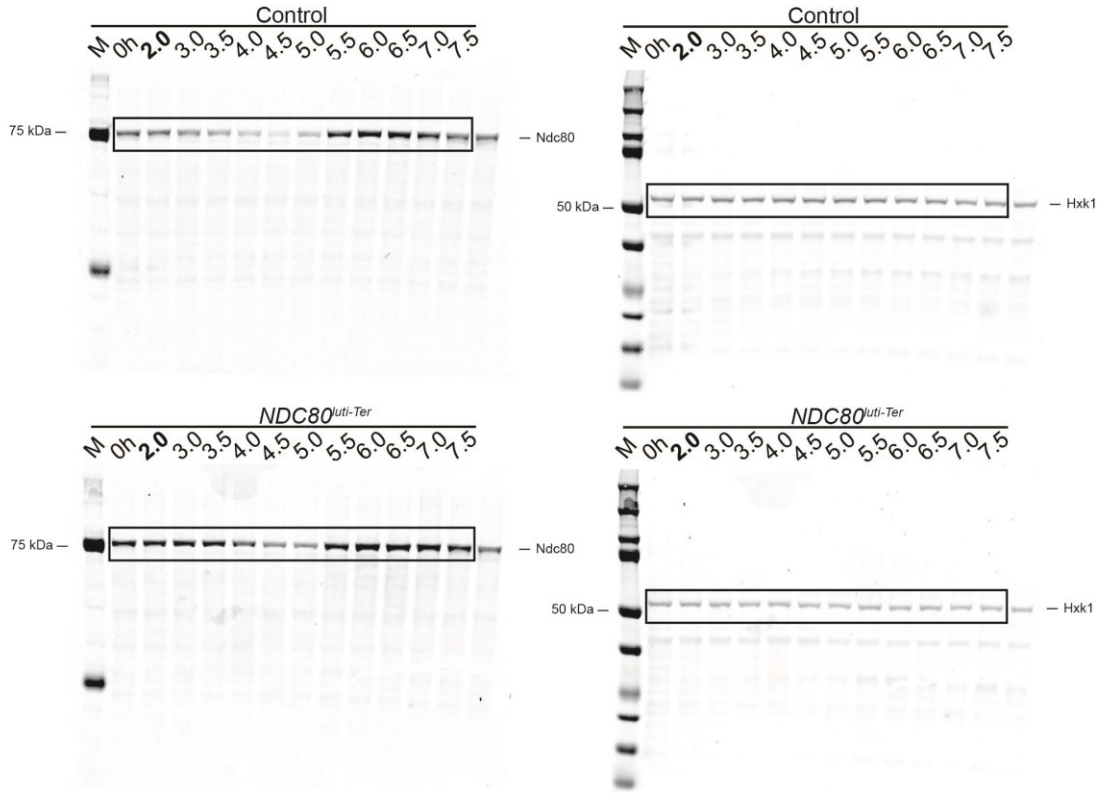




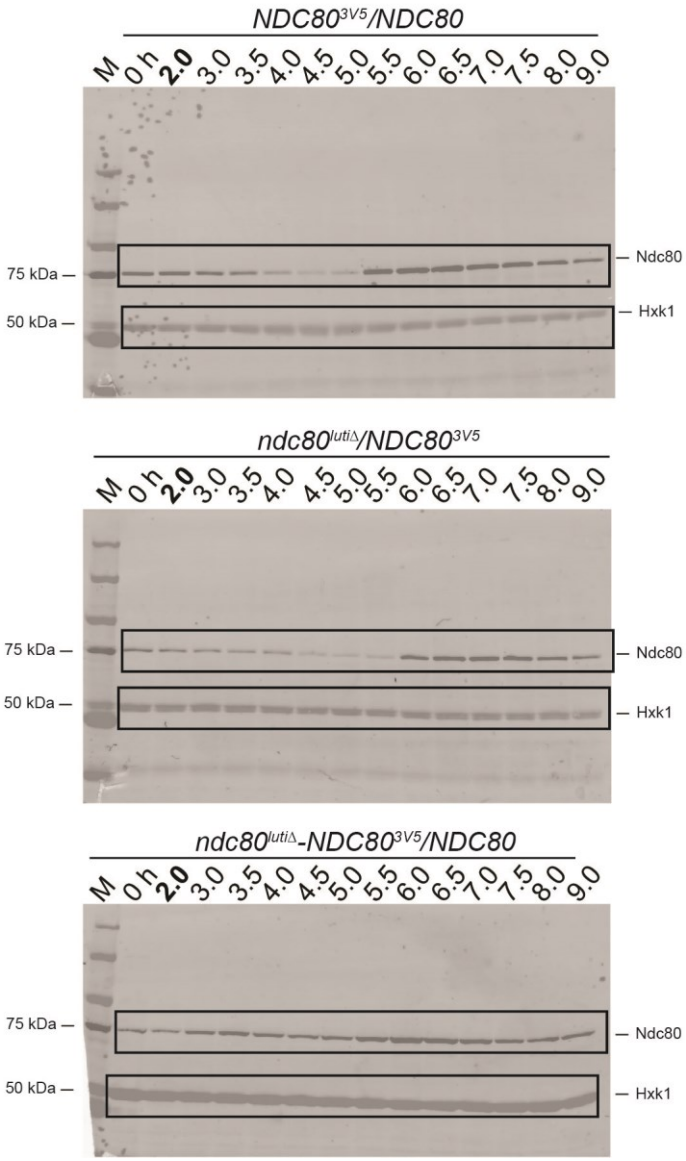
Uncropped western blots for Figure 4.4A



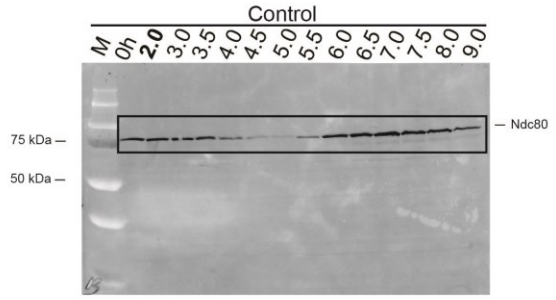
Uncropped western blots for Figure 4.4C

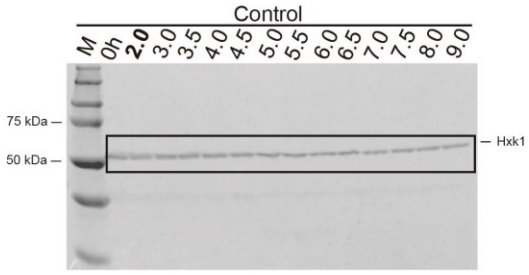


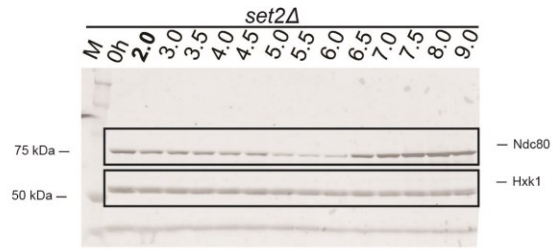
Uncropped western blots for Figure 4.5

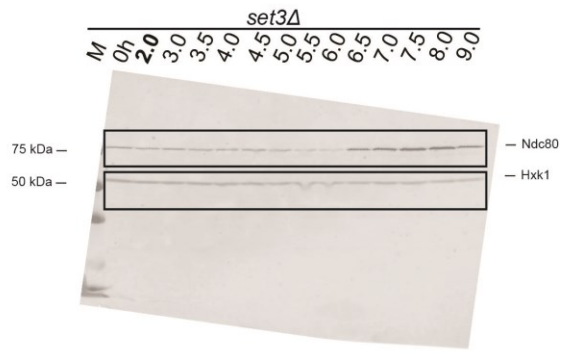


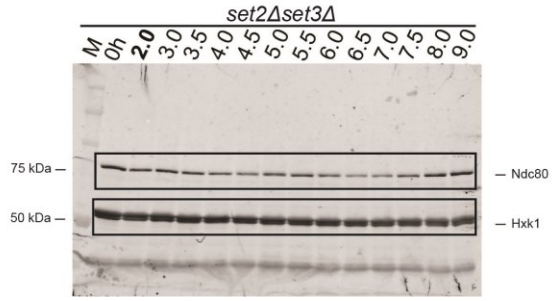
Uncropped western blots for Figure 4.8A



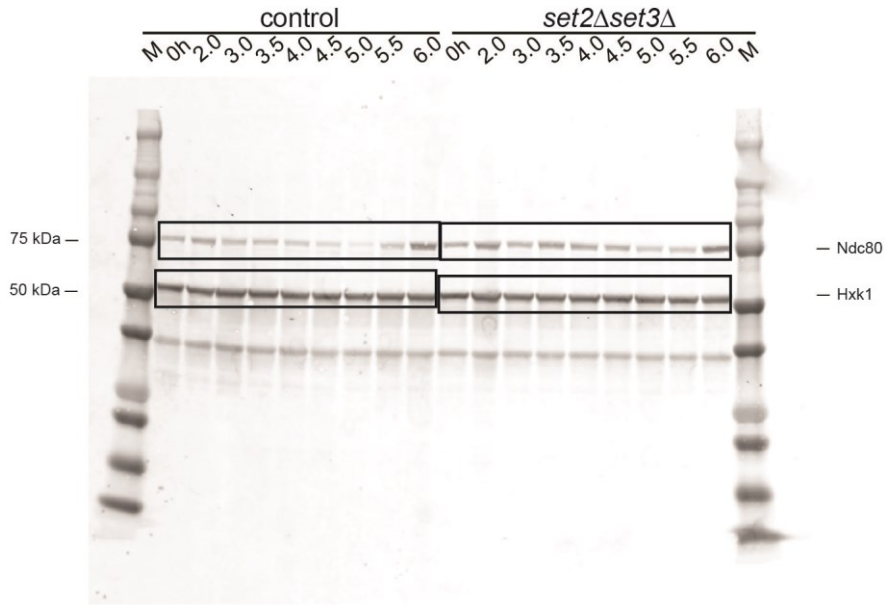




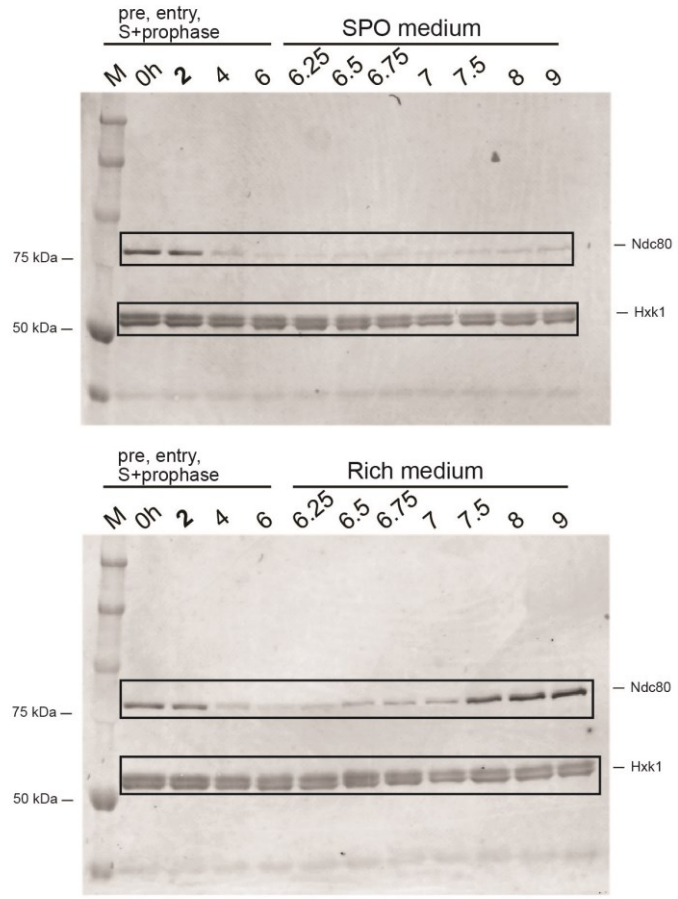




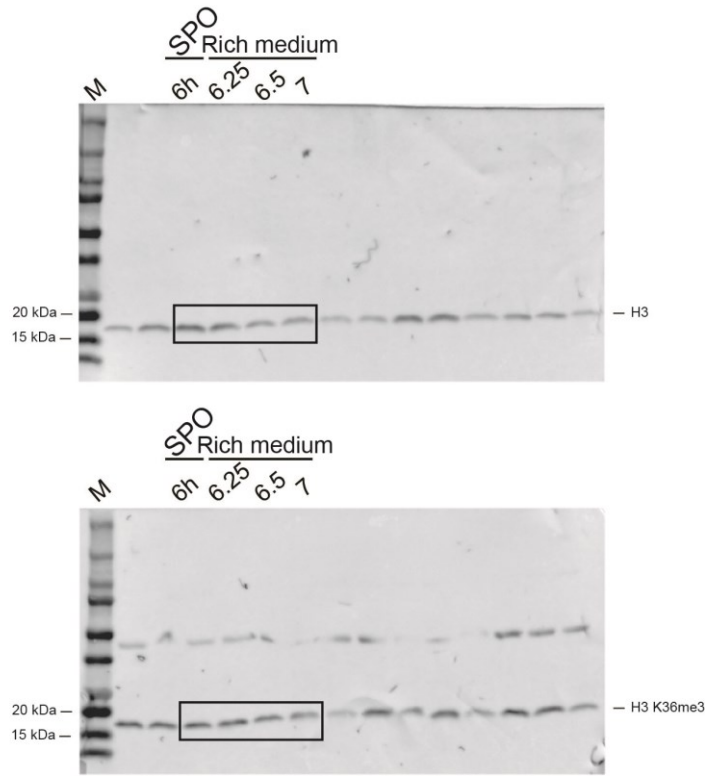
Uncropped western blots for Figure 4.9D



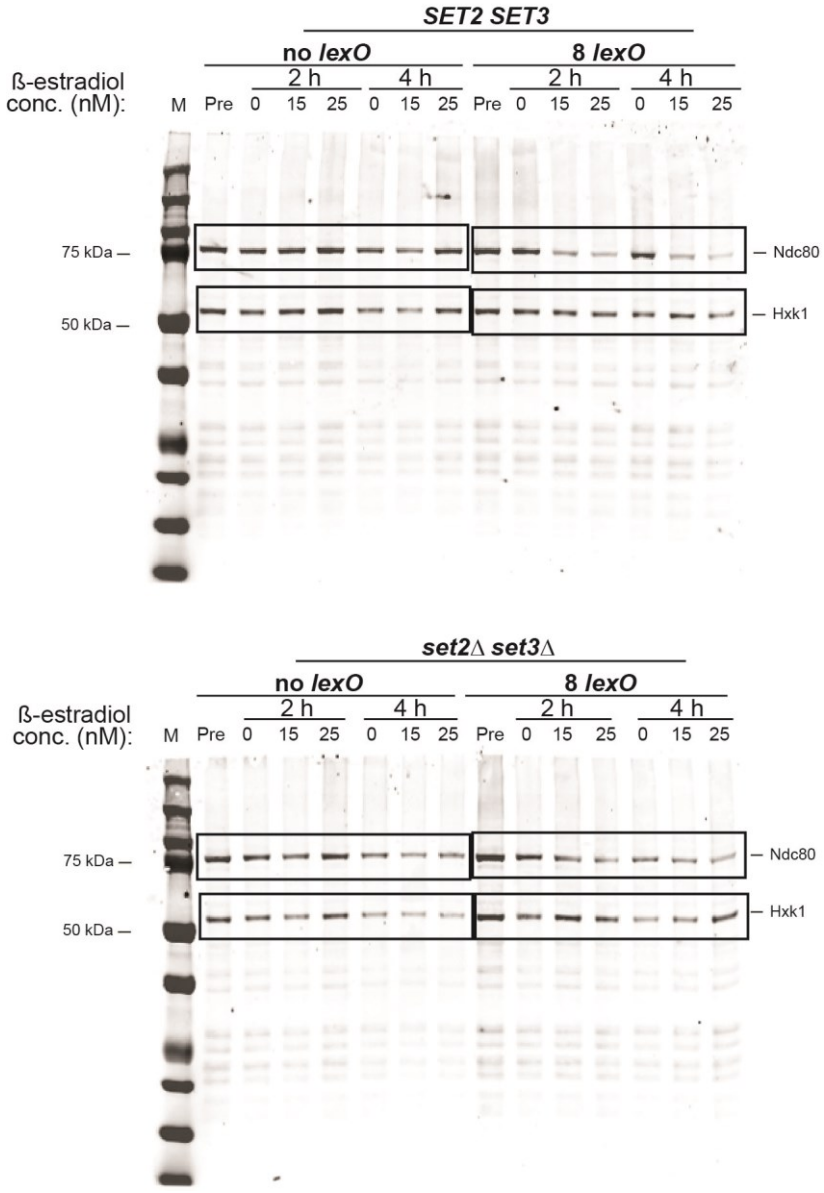
Uncropped western blots for Figure 4.11B



Uncropped western blots for Figure 4.11D



Uncropped western blots for Figure 4.13B



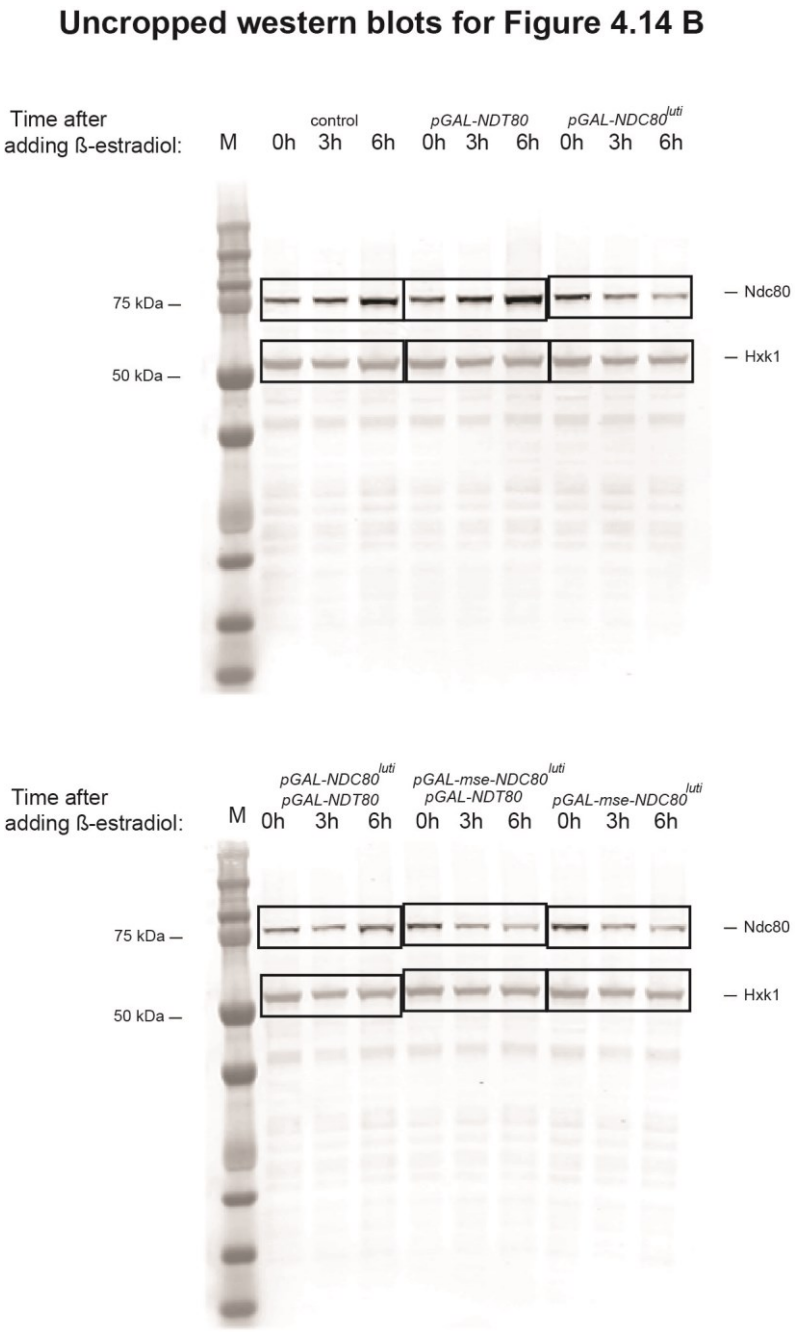


Figure 8.3 Uncropped western blots matching the main figures

Uncropped western blots showing the bands of interest together with the marker lane (M). Black boxes are the approximate areas which were cropped for presentation in the main figures. Please refer to the legends in the main figures for more details. Note that Figure 4.7 A has the same western as Figure 4.1 B.

8.2 Tables of TSSs and TESs

Gene	Systematic name	Upstream TSS ID	Difference in 5' UTR length (nt)	Minimum log2FC of coding TSS	Nearest alternative TES ID	Distance between upstream TSS and alternative TES (nt)
SVS1	YPL163C	TSS11988-interQ90	151	-8.819487661	NA	NA
DAL1	YIR027C	TSS5971-interQ90	223	-7.078774172	NA	NA
CLB1	YGR108W	TSS4280-interQ90	268	-6.972418602	NA	NA
HMG1	YML075C	TSS8930-interQ90	860	-6.968690722	TES11346-interQ90	926
PCL9	YDL179W	TSS1410-interQ90	113	-5.929958648	NA	NA
CLN1	YMR199W	TSS8727-interQ90	482	-5.872004914	NA	NA
PIR1	YKL164C	TSS7079-interQ90	414	-5.708990308	NA	NA
IRC7	YFR055W	TSS3713-interQ90	91	-5.696569731	NA	NA
FMP48	YGR052W	TSS4217-interQ90	416	-5.59228794	NA	NA
CLN1	YMR199W	TSS8726-interQ90	514	-5.592267957	NA	NA
PUT4	YOR348C	TSS11284-interQ90	653	-5.406080084	TES14278-interQ90	580
APL4	YPR029C	TSS12204-interQ90	381	-5.13441155	TES15521-interQ90	1025
MNN1	YER001W	TSS3064-interQ90	506	-5.117565693	TES3928-interQ90	482
YOR1	YGR281W	TSS4478-interQ90	627	-4.832681516	NA	NA
FRE1	YLR214W	TSS7584-interQ90	151	-4.748881282	NA	NA
MMS1	YPR164W	TSS11811-interQ90	447	-4.738172282	TES14968-interQ90	1113
MMS1	YPR164W	TSS11812-interQ90	398	-4.738172282	TES14968-interQ90	1041
CLB2	YPR119W	TSS11764-interQ90	320	-4.697995726	NA	NA
ATO3	YDR384C	TSS2826-interQ90	513	-4.595728548	NA	NA
SKM1	YOL113W	TSS10215-interQ90	150	-4.550020607	NA	NA
BOI1	YBL085W	TSS145-interQ90	775	-4.378788893	NA	NA
CSI2	YOL007C	TSS10901-interQ90	180	-4.364344802	NA	NA
CLN2	YPL256C	TSS11886-interQ90	652	-4.347135447	TES15057-interQ90	601
YOR020W-A	YOR020W-A	TSS10364-interQ90	505	-4.331256999	TES13137-interQ90	439
MES1	YGR264C	TSS5001-interQ90	478	-4.267524482	TES6322-interQ90	2874
BUL2	YML111W	TSS8361-interQ90	705	-4.128241278	TES10652-interQ90	2120
BUL2	YML111W	TSS8362-interQ90	637	-4.128241278	TES10652-interQ90	2052
SWI4	YER111C	TSS3489-interQ90	628	-4.066786994	NA	NA
POP4	YBR257W	TSS500-interQ90	492	-3.998531295	NA	NA
POP4	YBR257W	TSS501-interQ90	438	-3.998531295	NA	NA
POP4	YBR257W	TSS502-interQ90	204	-3.998531295	NA	NA
IML1	YJR138W	TSS6309-interQ90	597	-3.845703085	TES8042-interQ90	393
AUA1	YFL010W-A	TSS3643-interQ90	666	-3.74693577	TES4659-interQ90	596
KNS1	YLL019C	TSS7886-interQ90	174	-3.706728468	NA	NA
KNS1	YLL019C	TSS7887-interQ90	231	-3.706728468	NA	NA
MNN4	YKL201C	TSS7034-interQ90	230	-3.606754014	NA	NA
MNN4	YKL201C	TSS7033-interQ90	194	-3.606754014	NA	NA
MNN4	YKL201C	TSS7032-interQ90	103	-3.606754014	NA	NA
CMR3	YPR013C	TSS12183-interQ90	489	-3.594081328	TES15494-interQ90	494
CAN1	YEL063C	TSS3301-interQ90	167	-3.562183612	NA	NA
CIS1	YLR346C	TSS8252-interQ90	103	-3.511015837	TES10486-interQ90	39
RNR1	YER070W	TSS3150-interQ90	169	-3.488842925	TES4042-interQ90	2874
ALY2	YJL084C	TSS6471-interQ90	430	-3.417095142	TES8236-interQ90	2257
DSE3	YOR264W	TSS10599-interQ90	165	-3.317483949	NA	NA
GIP3	YPL137C	TSS12028-interQ90	125	-3.298852856	NA	NA
RRT6	YGL146C	TSS4574-interQ90	121	-3.256195338	NA	NA
RPL28	YGL103W	TSS4035-interQ90	605	-3.155201401	NA	NA
GDH2	YDL215C	TSS2218-interQ90	166	-3.153299608	NA	NA
NDC80	YIL144W	TSS5564-interQ90	455	-3.001494305	TES7126-interQ90	1277
YGR053C	YGR053C	TSS4755-interQ90	330	-2.830055039	NA	NA
RPO41	YFL036W	TSS3609-interQ90	580	-2.764062489	TES4617-interQ90	3209
RPO41	YFL036W	TSS3610-interQ90	539	-2.764062489	TES4617-interQ90	3168
RPO41	YFL036W	TSS3611-interQ90	485	-2.764062489	TES4617-interQ90	3103
CDC60	YPL160W	TSS11445-interQ90	442	-2.654224478	TES14533-interQ90	256
RTT10	YPL183C	TSS11967-interQ90	423	-2.648241203	TES15170-interQ90	2337
MTR2	YKL186C	TSS7059-interQ90	111	-2.431109803	NA	NA
YPR159C-A	YPR159C-A	TSS12327-interQ90	296	-2.428470036	NA	NA
RAD16	YBR114W	TSS365-interQ90	449	-2.427212408	NA	NA
RAD16	YBR114W	TSS366-interQ90	390	-2.427212408	NA	NA
GSC2	YGR032W	TSS4180-interQ90	187	-2.395413071	NA	NA
MIT1	YEL007W	TSS3052-interQ90	231	-2.346399631	NA	NA
MIT1	YEL007W	TSS3053-interQ90	140	-2.335794762	NA	NA
TAX4	YJL083W	TSS6108-interQ90	449	-2.332424444	NA	NA
SSH1	YBR283C	TSS1014-interQ90	500	-2.316037611	TES1298-interQ90	2173

MCD1	YDL003W	TSS1579-interQ90	545	-2.260067418	TES1980-interQ90	525
MCD1	YDL003W	TSS1580-interQ90	111	-2.260067418	TES1980-interQ90	88
ZPS1	YOL154W	TSS10171-interQ90	154	-2.210253051	NA	NA
YPS1	YLR120C	TSS8022-interQ90	339	-2.154766988	NA	NA
HUA1	YGR268C	TSS5008-interQ90	365	-1.99903382	NA	NA
STP4	YDL048C	TSS2396-interQ90	888	-1.997108023	TES2998-interQ90	597
ECM38	YLR299W	TSS7666-interQ90	346	-1.916362412	TES9745-interQ90	275
STL1	YDR536W	TSS2173-interQ90	339	-1.807064743	TES2719-interQ90	373
STL1	YDR536W	TSS2174-interQ90	315	-1.807064743	TES2719-interQ90	336
STL1	YDR536W	TSS2172-interQ90	915	-1.807064743	TES2719-interQ90	949
ULP2	YIL031W	TSS5678-interQ90	129	-1.751272229	TES7261-interQ90	986
GLC3	YEL011W	TSS3042-interQ90	150	-1.682685567	NA	NA
RNR4	YGR180C	TSS4884-interQ90	848	-1.667595009	NA	NA
PKH2	YOL100W	TSS10234-interQ90	478	-1.637741373	TES12971-interQ90	602
PKH2	YOL100W	TSS10235-interQ90	414	-1.637741373	TES12971-interQ90	553
MDJ2	YNL328C	TSS9788-interQ90	368	-1.573314772	NA	NA
MBA1	YBR185C	TSS908-interQ90	378	-1.559881264	NA	NA
FLR1	YBR008C	TSS684-interQ90	490	-1.522014823	TES895-interQ90	54
YBR201C-A	YBR201C-A	TSS924-interQ90	124	-1.507950821	TES1202-interQ90	502
RNR4	YGR180C	TSS4883-interQ90	181	-1.507063341	NA	NA
CMR3	YPR013C	TSS12182-interQ90	436	-1.493766149	TES15494-interQ90	423
FPS1	YLL043W	TSS7364-interQ90	795	-1.470926579	NA	NA
FPS1	YLL043W	TSS7365-interQ90	502	-1.470926579	NA	NA
FPS1	YLL043W	TSS7366-interQ90	419	-1.470926579	NA	NA
SUR2	YDR297W	TSS1946-interQ90	134	-1.422363894	NA	NA
MAK21	YDR060W	TSS1670-interQ90	510	-1.42020874	NA	NA
MAK21	YDR060W	TSS1671-interQ90	469	-1.42020874	NA	NA
TRA1	YHR099W	TSS5160-interQ90	457	-1.40885702	TES6568-interQ90	2339
TRA1	YHR099W	TSS5161-interQ90	406	-1.40885702	TES6568-interQ90	2259
RIF2	YLR453C	TSS8334-interQ90	225	-1.34024858	NA	NA
TNA1	YGR260W	TSS4463-interQ90	350	-1.316082526	TES5631-interQ90	350
HSL1	YKL101W	TSS6795-interQ90	95	-1.297863022	NA	NA
PRR2	YDL214C	TSS2221-interQ90	324	-1.278635334	NA	NA
CAX4	YGR036C	TSS4735-interQ90	126	-1.276781044	NA	NA
OXA1	YER154W	TSS3252-interQ90	735	-1.269230686	NA	NA
SKI7	YOR076C	TSS11007-interQ90	632	-1.266947194	NA	NA
YAR1	YPL239W	TSS11363-interQ90	477	-1.262781505	NA	NA
YOR238W	YOR238W	TSS10575-interQ90	170	-1.241471744	NA	NA
YOR238W	YOR238W	TSS10576-interQ90	143	-1.241471744	NA	NA
FUN19	YAL034C	TSS95-interQ90	588	-1.119109418	NA	NA
RRP5	YMR229C	TSS9254-interQ90	855	-1.114600082	NA	NA
SUR1	YPL057C	TSS12109-interQ90	150	-1.097539951	NA	NA
LIN1	YHR156C	TSS5496-interQ90	294	-1.080593035	NA	NA
AVL9	YLR114C	TSS8016-interQ90	477	-1.064211813	NA	NA
MOD5	YOR274W	TSS10609-interQ90	618	-1.043897229	NA	NA
MOD5	YOR274W	TSS10610-interQ90	388	-1.043897229	NA	NA
VPS30	YPL120W	TSS11494-interQ90	543	-1.035455021	TES14593-interQ90	2256
VPS30	YPL120W	TSS11496-interQ90	172	-1.035455021	TES14593-interQ90	1893
MLP1	YKR095W	TSS7006-interQ90	510	-1.033800584	NA	NA
CFT2	YLR115W	TSS7503-interQ90	381	-1.026216541	TES9539-interQ90	633

Table 8.1 List of genes and alternative upstream TSSs in which the associated coding TSS is repressed in at least one time point during early gametogenesis

This list is sorted according to the $\log_2$ fold change of the coding TSS, from the lowest value (most repressed) to the highest (less repressed). A nearest alternative TES ID is also given in cases where the upstream transcript could terminate at a different site; the distance in nt spanning the upstream TSS and the alternative TES is also given.

Gene	Systematic name	Upstream TSS ID	Difference in 5' UTR length (nt)	Minimum log2FC of coding TSS	Nearest alternative TES ID	Distance between upstream TSS and alternative TES (nt)
CMR3	YPR013C	TSS12184-interQ90	524	-6.501718171	TES15494-interQ90	529
MNN4	YKL201C	TSS7033-interQ90	194	-6.01972606	TES8920-interQ90	3846
MNN4	YKL201C	TSS7034-interQ90	230	-6.01972606	TES8920-interQ90	3926
YHB1	YGR234W	TSS4425-interQ90	674	-4.949298138	TES5584-interQ90	609
FAS1	YKL182W	TSS6710-interQ90	245	-4.260930203	TES8545-interQ90	367
VHR2	YER064C	TSS3435-interQ90	297	-4.191375496	NA	NA
TAT2	YOL020W	TSS10314-interQ90	170	-3.954952453	NA	NA
AUA1	YFL010W-A	TSS3643-interQ90	666	-3.346505594	TES4659-interQ90	596
YGL039W	YGL039W	TSS4115-interQ90	484	-3.101253914	TES5196-interQ90	39
RKI1	YOR095C	TSS11018-interQ90	91	-2.918690278	NA	NA
ARO4	YBR249C	TSS977-interQ90	646	-2.853478594	TES1255-interQ90	213
ARO4	YBR249C	TSS978-interQ90	893	-2.853478594	TES1255-interQ90	480
ORC1	YML065W	TSS8405-interQ90	371	-2.787420772	NA	NA
ORC1	YML065W	TSS8406-interQ90	297	-2.787420772	NA	NA
ORC1	YML065W	TSS8407-interQ90	228	-2.787420772	NA	NA
BAP2	YBR068C	TSS775-interQ90	336	-2.787352039	TES1013-interQ90	147
BAP2	YBR068C	TSS776-interQ90	405	-2.787352039	TES1013-interQ90	208
UTP6	YDR449C	TSS2881-interQ90	462	-2.727524123	TES3702-interQ90	717
UTP6	YDR449C	TSS2882-interQ90	505	-2.727524123	TES3702-interQ90	760
UTP6	YDR449C	TSS2883-interQ90	539	-2.727524123	TES3702-interQ90	791
MCM5	YLR274W	TSS7653-interQ90	628	-2.651138795	NA	NA
ECM38	YLR299W	TSS7666-interQ90	346	-2.638839879	TES9745-interQ90	275
ECM38	YLR299W	TSS7667-interQ90	111	-2.638839879	TES9745-interQ90	31
FAA4	YMR246W	TSS8772-interQ90	133	-2.493185172	NA	NA
YER053C-A	YER053C-A	TSS3416-interQ90	302	-2.491475749	TES4396-interQ90	339
PST1	YDR055W	TSS1664-interQ90	233	-2.461241327	NA	NA
RGI1	YER067W	TSS3142-interQ90	223	-2.424544988	NA	NA
YGR067C	YGR067C	TSS4771-interQ90	259	-2.414438254	NA	NA
GCV2	YMR189W	TSS8710-interQ90	550	-2.410558708	NA	NA
NMA1	YLR328W	TSS7700-interQ90	785	-2.350695261	NA	NA
KRS1	YDR037W	TSS1631-interQ90	454	-2.34879951	NA	NA
KRS1	YDR037W	TSS1632-interQ90	388	-2.34879951	NA	NA
ATO3	YDR384C	TSS2826-interQ90	513	-2.321767594	NA	NA
PRS2	YER099C	TSS3479-interQ90	241	-2.276989869	NA	NA
NKP2	YLR315W	TSS7685-interQ90	876	-2.179070267	NA	NA
NKP2	YLR315W	TSS7686-interQ90	717	-2.179070267	NA	NA
NKP2	YLR315W	TSS7687-interQ90	630	-2.179070267	NA	NA
YIR014W	YIR014W	TSS5726-interQ90	673	-2.174894139	NA	NA
YAP1802	YGR241C	TSS4964-interQ90	734	-2.164609364	NA	NA
RPL4b	YDR012W	TSS1606-interQ90	816	-2.151958884	TES2001-interQ90	657
MSA1	YOR066W	TSS10405-interQ90	759	-2.1375755	NA	NA
MSA1	YOR066W	TSS10406-interQ90	720	-2.1375755	NA	NA
YAR1	YPL239W	TSS11363-interQ90	477	-2.070389707	NA	NA
YAR1	YPL239W	TSS11364-interQ90	408	-2.070389707	NA	NA
YAR1	YPL239W	TSS11365-interQ90	379	-2.070389707	NA	NA
FIT2	YOR382W	TSS10729-interQ90	291	-2.046866207	NA	NA
ASN1	YPR145W	TSS11788-interQ90	847	-2.009215724	NA	NA
IPI3	YNL182C	TSS9914-interQ90	419	-1.936068627	NA	NA
CLN2	YPL256C	TSS11885-interQ90	607	-1.909034349	TES15057-interQ90	527
CLN2	YPL256C	TSS11886-interQ90	652	-1.909034349	TES15057-interQ90	601
CLN2	YPL256C	TSS11887-interQ90	731	-1.909034349	TES15057-interQ90	662
KGD1	YIL125W	TSS5579-interQ90	125	-1.907362485	NA	NA
ADH1	YOL086C	TSS10817-interQ90	251	-1.852307046	NA	NA
URA10	YMR271C	TSS9302-interQ90	519	-1.816092112	TES11794-interQ90	198
URA10	YMR271C	TSS9303-interQ90	566	-1.816092112	TES11794-interQ90	245
NAN1	YPL126W	TSS11490-interQ90	305	-1.776908849	TES14579-interQ90	2526
SUP35	YDR172W	TSS1813-interQ90	237	-1.740783849	TES2250-interQ90	2376
EFB1	YAL003W	TSS47-interQ90	174	-1.729214134	NA	NA
COS111	YBR203W	TSS446-interQ90	591	-1.703727545	NA	NA
FTR1	YER145C	TSS3532-interQ90	556	-1.699847749	NA	NA
GLG1	YKR058W	TSS6968-interQ90	117	-1.673391852	NA	NA

STL1	YDR536W	TSS2172-interQ90	915	-1.653893276	TES2719-interQ90	949
STL1	YDR536W	TSS2173-interQ90	339	-1.653893276	TES2719-interQ90	373
STL1	YDR536W	TSS2174-interQ90	315	-1.653893276	TES2719-interQ90	336
FCY2	YER056C	TSS3424-interQ90	868	-1.642042678	TES4404-interQ90	669
SOM1	YEL059C-A	TSS3305-interQ90	407	-1.617466249	NA	NA
CMK2	YOL016C	TSS10892-interQ90	642	-1.613907977	NA	NA
TPO1	YLL028W	TSS7376-interQ90	645	-1.612771344	TES9376-interQ90	560
TPO1	YLL028W	TSS7377-interQ90	611	-1.612771344	TES9376-interQ90	509
GDH2	YDL215C	TSS2218-interQ90	166	-1.59685272	NA	NA
AMN1	YBR158W	TSS400-interQ90	217	-1.570319765	NA	NA
PTK1	YKL198C	TSS7041-interQ90	134	-1.492935338	NA	NA
RFT1	YBL020W	TSS213-interQ90	517	-1.473311902	NA	NA
BNA1	YJR025C	TSS6569-interQ90	103	-1.403489234	NA	NA
RAV1	YJR033C	TSS6574-interQ90	682	-1.400546198	NA	NA
RAD2	YGR258C	TSS4993-interQ90	946	-1.343172695	NA	NA
RAD2	YGR258C	TSS4994-interQ90	970	-1.343172695	NA	NA
SVS1	YPL163C	TSS11988-interQ90	151	-1.3163607	TES15195-interQ90	154
MIC60	YKR016W	TSS6931-interQ90	689	-1.314395413	NA	NA
RPS6b	YBR181C	TSS904-interQ90	645	-1.273094244	NA	NA
PUT4	YOR348C	TSS11282-interQ90	307	-1.191652255	TES14278-interQ90	230
PUT4	YOR348C	TSS11283-interQ90	626	-1.191652255	TES14278-interQ90	553
PUT4	YOR348C	TSS11284-interQ90	653	-1.191652255	TES14278-interQ90	580
CMG1	YLR271W	TSS7649-interQ90	155	-1.188984494	NA	NA
HRK1	YOR267C	TSS11189-interQ90	164	-1.179796365	NA	NA
HRK1	YOR267C	TSS11190-interQ90	192	-1.179796365	NA	NA
YER076C	YER076C	TSS3449-interQ90	255	-1.179330191	NA	NA
GIS4	YML006C	TSS9003-interQ90	328	-1.156508555	TES11432-interQ90	261
LSB6	YJL100W	TSS6091-interQ90	436	-1.137768449	TES7772-interQ90	234
RPL14a	YKL006W	TSS6911-interQ90	183	-1.120284244	TES8796-interQ90	1042
VID24	YBR105C	TSS823-interQ90	113	-1.074516769	NA	NA
DPB11	YJL090C	TSS6463-interQ90	122	-1.055745019	TES8232-interQ90	366
STP4	YDL048C	TSS2396-interQ90	888	-1.037018134	TES2998-interQ90	597

Table 8.2 List of genes and alternative upstream TSSs in which the associated coding TSS is repressed in at least one time point during mid-late gametogenesis

This list is sorted according to the $\log_2$ fold change of the coding TSS, from the lowest value (most repressed) to the highest (less repressed). A nearest alternative TES ID is also given in cases where the upstream transcript could terminate at a different site; the distance in nt spanning the upstream TSS and the alternative TES is also given.

Gene	Systematic name	Upstream TSS ID	Difference in 5' UTR length (nt)	Maximum log2FC of coding TSS	Nearest alternative TES ID	Distance between upstream TSS and alternative TES (nt)
FLO5	YHR211W	TSS74-interQ90	266	5.410035075	NA	NA
TIR4	YOR009W	TSS10346-interQ90	136	5.333487248	NA	NA
CLB6	YGR109C	TSS4806-interQ90	171	5.148365226	NA	NA
GAT1	YFL021W	TSS3621-interQ90	283	5.015369634	TES4634-interQ90	676
MF(ALPHA)1	YPL187W	TSS11421-interQ90	142	4.465377442	TES14502-interQ90	751
RNR1	YER070W	TSS3150-interQ90	169	3.217562711	TES4042-interQ90	2874
YCL021W-A	YCL021W-A	TSS1072-interQ90	538	3.044802314	TES1353-interQ90	1018
YKL071W	YKL071W	TSS6835-interQ90	554	2.79164442	TES8697-interQ90	686
ITR2	YOL103W	TSS10227-interQ90	156	2.628553213	NA	NA
YJL045W	YJL045W	TSS6144-interQ90	400	2.57683015	TES7851-interQ90	445
YLR460C	YLR460C	TSS1349-interQ90	149	2.458648433	NA	NA
CLB6	YGR109C	TSS4807-interQ90	394	2.326954652	NA	NA
TNA1	YGR260W	TSS4463-interQ90	350	2.093482609	TES5631-interQ90	350
HXT10	YFL011W	TSS3638-interQ90	493	1.942290684	TES4654-interQ90	773
LCB2	YDR062W	TSS1676-interQ90	299	1.918803954	NA	NA
SPO22	YIL073C	TSS5864-interQ90	565	1.694347369	NA	NA
YDR445C	YDR445C	TSS2877-interQ90	346	1.622316506	NA	NA
LSM4	YER112W	TSS3201-interQ90	628	1.357013627	TES4109-interQ90	302
BDF2	YDL070W	TSS1520-interQ90	279	1.303449928	NA	NA
GAL2	YLR081W	TSS7472-interQ90	127	1.215964035	TES9503-interQ90	174
ALT2	YDR111C	TSS2556-interQ90	213	1.043604876	NA	NA
HAS1	YMR290C	TSS9323-interQ90	320	1.027006519	TES11818-interQ90	277

Table 8.3 List of genes and alternative upstream TSSs in which the associated coding TSS is upregulated in at least one time point during early gametogenesis

This list is sorted according to the $\log_2$ fold change of the coding TSS, from the high value (greatest degree of transcriptional activation) to the lowest (lowest degree of transcriptional activation). A nearest alternative TES ID is also given in cases where the upstream transcript could terminate at a different site; the distance in nt spanning the upstream TSS and the alternative TES is also given.

Gene	Systematic name	Upstream TSS ID	Difference in 5' UTR length (nt)	Maximum log2FC of coding TSS	Nearest alternative TES ID	Distance between upstream TSS and alternative TES (nt)
CLB1	YGR108W	TSS4280-interQ90	268	9.595487694	NA	NA
CLB6	YGR109C	TSS4806-interQ90	171	7.305475699	TES6119-interQ90	150
YLR012C	YLR012C	TSS7913-interQ90	350	5.626209659	NA	NA
YLR012C	YLR012C	TSS7914-interQ90	391	5.626209659	NA	NA
YLR012C	YLR012C	TSS7915-interQ90	544	5.626209659	NA	NA
YLR012C	YLR012C	TSS7916-interQ90	597	5.626209659	NA	NA
DSE3	YOR264W	TSS10599-interQ90	165	5.313781114	NA	NA
DIT2	YDR402C	TSS2838-interQ90	251	5.215887756	TES3635-interQ90	1736
PCL9	YDL179W	TSS1410-interQ90	113	4.236975004	NA	NA
FUI1	YBL042C	TSS617-interQ90	674	3.391534269	TES818-interQ90	2364
SRL1	YOR247W	TSS10586-interQ90	130	3.274541659	NA	NA
GPG1	YGL121C	TSS4602-interQ90	97	3.095910316	TES5848-interQ90	503
GSC2	YGR032W	TSS4180-interQ90	187	2.971246563	NA	NA
MRL1	YPR079W	TSS11726-interQ90	446	2.964205553	TES14867-interQ90	1720
MRL1	YPR079W	TSS11727-interQ90	233	2.964205553	TES14867-interQ90	1497
YBR201C-A	YBR201C-A	TSS924-interQ90	124	2.696269377	TES1202-interQ90	502
HXK2	YGL253W	TSS3877-interQ90	92	2.652956361	NA	NA
CYC8	YBR112C	TSS833-interQ90	330	2.638735629	TES1086-interQ90	1755
YPR127W	YPR127W	TSS11772-interQ90	144	2.222556236	NA	NA
PDC6	YGR087C	TSS4788-interQ90	148	2.169057258	NA	NA
HRR25	YPL204W	TSS11403-interQ90	201	2.153102005	NA	NA
PHM7	YOL084W	TSS10248-interQ90	198	2.103915819	NA	NA
SPB1	YCL054W	TSS1039-interQ90	716	2.019627383	NA	NA
SPB1	YCL054W	TSS1041-interQ90	557	2.019627383	NA	NA
CDC14	YFR028C	TSS3827-interQ90	277	1.864482754	NA	NA
MDJ2	YNL328C	TSS9788-interQ90	368	1.848572488	TES12397-interQ90	1130
RAD16	YBR114W	TSS365-interQ90	449	1.82579385	NA	NA
PCK1	YKR097W	TSS7013-interQ90	216	1.807262506	NA	NA
RAS2	YNL098C	TSS9999-interQ90	138	1.805429895	NA	NA
SSA4	YER103W	TSS3189-interQ90	109	1.760039413	NA	NA
CLB5	YPR120C	TSS12275-interQ90	128	1.712194243	TES15618-interQ90	812
GLG1	YKR058W	TSS6968-interQ90	117	1.573786326	NA	NA
DIG2	YDR480W	TSS2118-interQ90	291	1.531207086	TES2657-interQ90	99
TPO1	YLL028W	TSS7377-interQ90	611	1.510251844	TES9376-interQ90	509
MPT5	YGL178W	TSS3953-interQ90	548	1.464475932	TES5032-interQ90	317
DGR2	YKL121W	TSS6774-interQ90	654	1.354991864	TES8631-interQ90	398
CIS1	YLR346C	TSS8252-interQ90	103	1.344989672	NA	NA
GPX2	YBR244W	TSS484-interQ90	310	1.31208952	NA	NA
GPX2	YBR244W	TSS487-interQ90	125	1.31208952	NA	NA
CSH1	YBR161W	TSS404-interQ90	215	1.215330418	NA	NA
CBT1	YKL208W	TSS6680-interQ90	205	1.17994971	NA	NA
MGS1	YNL218W	TSS9460-interQ90	208	1.171756011	TES11985-interQ90	2183
RRT6	YGL146C	TSS4574-interQ90	121	1.081308233	TES5815-interQ90	1246
YCR006C	YCR006C	TSS1248-interQ90	418	1.014185629	NA	NA

Table 8.4 List of genes and alternative upstream TSSs in which the associated coding TSS is upregulated in at least one time point during mid-late gametogenesis

This list is sorted according to the $\log_2$ fold change of the coding TSS, from the high value (greatest degree of transcriptional activation) to the lowest (lowest degree of transcriptional activation). A nearest alternative TES ID is also given in cases

where the upstream transcript could terminate at a different site; the distance in nt spanning the upstream TSS and the alternative TES is also given.

Gene	Systematic name	Upstream TSS ID	Difference in 5' UTR length (nt)	Nearest alternative TES ID	Distance between upstream TSS and alternative TES (nt)
DUG2	YBR281C	TSS1009-interQ90	253	TES1295-interQ90	496
DUG2	YBR281C	TSS1010-interQ90	309	TES1295-interQ90	552
PKH2	YOL100W	TSS10234-interQ90	478	TES12971-interQ90	602
YOR238W	YOR238W	TSS10575-interQ90	170	NA	NA
DSE3	YOR264W	TSS10599-interQ90	165	NA	NA
MOD5	YOR274W	TSS10609-interQ90	618	NA	NA
MOD5	YOR274W	TSS10610-interQ90	388	NA	NA
MOD5	YOR274W	TSS10611-interQ90	347	NA	NA
YPK9	YOR291W	TSS10627-interQ90	390	NA	NA
IZH4	YOL101C	TSS10802-interQ90	755	NA	NA
SKI7	YOR076C	TSS11007-interQ90	632	NA	NA
MSC6	YOR354C	TSS11292-interQ90	310	TES14299-interQ90	235
MF(ALPHA)1	YPL187W	TSS11421-interQ90	142	TES14502-interQ90	751
VPS30	YPL120W	TSS11494-interQ90	543	TES14593-interQ90	2256
VPS30	YPL120W	TSS11495-interQ90	227	TES14593-interQ90	1952
VPS30	YPL120W	TSS11496-interQ90	172	TES14593-interQ90	1893
CLB2	YPR119W	TSS11764-interQ90	320	NA	NA
DSS4	YPR017C	TSS12193-interQ90	223	NA	NA
YCR006C	YCR006C	TSS1248-interQ90	418	NA	NA
SLM5	YCR024C	TSS1287-interQ90	171	NA	NA
MCD1	YDL003W	TSS1579-interQ90	545	TES1980-interQ90	525
MCD1	YDL003W	TSS1580-interQ90	111	TES1980-interQ90	88
MAK21	YDR060W	TSS1670-interQ90	510	NA	NA
MAK21	YDR060W	TSS1671-interQ90	469	NA	NA
GIS1	YDR096W	TSS1736-interQ90	373	NA	NA
YDR222W	YDR222W	TSS1870-interQ90	147	NA	NA
STL1	YDR536W	TSS2173-interQ90	339	TES2719-interQ90	373
STL1	YDR536W	TSS2174-interQ90	315	TES2719-interQ90	336
ALT2	YDR111C	TSS2556-interQ90	213	NA	NA
RNR1	YER070W	TSS3150-interQ90	169	TES4042-interQ90	2874
GIP2	YER054C	TSS3420-interQ90	147	NA	NA
SWI4	YER111C	TSS3489-interQ90	628	NA	NA
SRM1	YGL097W	TSS4048-interQ90	608	TES5119-interQ90	400
CAX4	YGR036C	TSS4735-interQ90	126	NA	NA
CLB6	YGR109C	TSS4806-interQ90	171	NA	NA
TRA1	YHR099W	TSS5160-interQ90	457	TES6568-interQ90	2339
TRA1	YHR099W	TSS5161-interQ90	406	TES6568-interQ90	2259
NSG1	YHR133C	TSS5466-interQ90	894	TES6984-interQ90	841
NSG1	YHR133C	TSS5467-interQ90	940	TES6984-interQ90	880
LIN1	YHR156C	TSS5496-interQ90	294	NA	NA
SPO22	YIL073C	TSS5864-interQ90	565	NA	NA
TAX4	YJL083W	TSS6108-interQ90	449	NA	NA
YJL045W	YJL045W	TSS6144-interQ90	400	TES7851-interQ90	445
BNA1	YJR025C	TSS6569-interQ90	103	TES8356-interQ90	684
MLP1	YKR095W	TSS7006-interQ90	510	NA	NA
GAL2	YLR081W	TSS7472-interQ90	127	TES9503-interQ90	174
CFT2	YLR115W	TSS7503-interQ90	381	TES9539-interQ90	633
CMG1	YLR271W	TSS7649-interQ90	155	NA	NA
NMA1	YLR328W	TSS7700-interQ90	785	NA	NA
AVL9	YLR114C	TSS8016-interQ90	477	NA	NA
NAM2	YLR382C	TSS8274-interQ90	188	NA	NA
NAM2	YLR382C	TSS8275-interQ90	221	NA	NA
RRN11	YML043C	TSS8966-interQ90	400	NA	NA
RRN11	YML043C	TSS8967-interQ90	433	NA	NA

YBR201C-A	YBR201C-A	TSS924-interQ90	124	TES1202-interQ90	502
RRP5	YMR229C	TSS9254-interQ90	855	NA	NA
HAS1	YMR290C	TSS9323-interQ90	320	TES11818-interQ90	277
HAS1	YMR290C	TSS9324-interQ90	380	TES11818-interQ90	384
TIM23	YNR017W	TSS9719-interQ90	387	NA	NA
AAD14	YNL331C	TSS9784-interQ90	293	NA	NA
MDJ2	YNL328C	TSS9788-interQ90	368	NA	NA

Table 8.5 List of genes and alternative upstream TSSs in which the associated coding TSS is neither up- nor downregulated in at least one time point during early gametogenesis

In this list, changes in the levels of coding TSS are independent or not correlated with changes in the levels of the upstream TSS during the early gametogenesis time points. A nearest alternative TES ID is also given in cases where the upstream transcript could terminate at a different site; the distance in nt spanning the upstream TSS and the alternative TES is also given.

Gene	Systematic name	Upstream TSS ID	Difference in 5' UTR length (nt)	Nearest alternative TES ID	Distance between upstream TSS and alternative TES (nt)
AIF1	YNR074C	TSS10165-interQ90	104	NA	NA
PHM7	YOL084W	TSS10248-interQ90	198	NA	NA
SPB1	YCL054W	TSS1038-interQ90	780	NA	NA
SPB1	YCL054W	TSS1039-interQ90	716	NA	NA
SPB1	YCL054W	TSS1040-interQ90	612	NA	NA
SPB1	YCL054W	TSS1041-interQ90	557	NA	NA
MOD5	YOR274W	TSS10609-interQ90	618	NA	NA
ADH1	YOL086C	TSS10817-interQ90	251	NA	NA
RPB2	YOR151C	TSS11064-interQ90	352	TES14006-interQ90	54
RPB2	YOR151C	TSS11065-interQ90	391	TES14006-interQ90	93
RPB2	YOR151C	TSS11066-interQ90	425	NA	NA
HRK1	YOR267C	TSS11189-interQ90	164	NA	NA
HRK1	YOR267C	TSS11190-interQ90	192	NA	NA
PUT4	YOR348C	TSS11284-interQ90	653	TES14278-interQ90	580
YAR1	YPL239W	TSS11363-interQ90	477	NA	NA
YAR1	YPL239W	TSS11364-interQ90	408	NA	NA
YAR1	YPL239W	TSS11365-interQ90	379	NA	NA
MRL1	YPR079W	TSS11727-interQ90	233	TES14867-interQ90	1497
CLN2	YPL256C	TSS11885-interQ90	607	TES15057-interQ90	527
CLN2	YPL256C	TSS11886-interQ90	652	TES15057-interQ90	601
CLN2	YPL256C	TSS11887-interQ90	731	TES15057-interQ90	662
YPL222C-A	YPL222C-A	TSS11933-interQ90	117	NA	NA
SVS1	YPL163C	TSS11988-interQ90	151	TES15195-interQ90	154
YCR006C	YCR006C	TSS1248-interQ90	418	NA	NA
BDF2	YDL070W	TSS1520-interQ90	279	NA	NA
KRS1	YDR037W	TSS1631-interQ90	454	NA	NA
GIS1	YDR096W	TSS1736-interQ90	373	NA	NA
GIS1	YDR096W	TSS1737-interQ90	304	NA	NA
GIS1	YDR096W	TSS1738-interQ90	248	NA	NA
GIS1	YDR096W	TSS1739-interQ90	182	NA	NA
NUM1	YDR150W	TSS1791-interQ90	411	TES2217-interQ90	2096
MRPL28	YDR462W	TSS2095-interQ90	589	TES2635-interQ90	118
MRPL28	YDR462W	TSS2096-interQ90	510	TES2635-interQ90	32
DIG2	YDR480W	TSS2118-interQ90	291	TES2657-interQ90	99
GRH1	YDR517W	TSS2154-interQ90	916	NA	NA
STL1	YDR536W	TSS2172-interQ90	915	TES2719-interQ90	949
STL1	YDR536W	TSS2173-interQ90	339	TES2719-interQ90	373
STL1	YDR536W	TSS2174-interQ90	315	TES2719-interQ90	336
GDH2	YDL215C	TSS2218-interQ90	166	NA	NA
STP4	YDL048C	TSS2396-interQ90	888	TES2998-interQ90	597
UTP6	YDR449C	TSS2882-interQ90	505	TES3702-interQ90	760
HSP31	YDR533C	TSS2977-interQ90	270	NA	NA
MIT1	YEL007W	TSS3052-interQ90	231	TES3918-interQ90	1075
RGI1	YER067W	TSS3142-interQ90	223	NA	NA
ALD5	YER073W	TSS3153-interQ90	452	TES4047-interQ90	466
ALD5	YER073W	TSS3154-interQ90	135	TES4047-interQ90	156
YER121W	YER121W	TSS3210-interQ90	91	TES4130-interQ90	4425
SOM1	YEL059C-A	TSS3305-interQ90	407	NA	NA
ACA1	YER045C	TSS3401-interQ90	351	TES4380-interQ90	217
FCY2	YER056C	TSS3424-interQ90	868	TES4404-interQ90	669

YER076C	YER076C	TSS3449-interQ90	255	NA	NA
PRS2	YER099C	TSS3479-interQ90	241	NA	NA
AUA1	YFL010W-A	TSS3643-interQ90	666	TES4659-interQ90	596
IES1	YFL013C	TSS3767-interQ90	420	TES4820-interQ90	235
CDC14	YFR028C	TSS3827-interQ90	277	NA	NA
HXK2	YGL253W	TSS3877-interQ90	92	NA	NA
MPT5	YGL178W	TSS3953-interQ90	548	TES5032-interQ90	317
AMN1	YBR158W	TSS400-interQ90	217	NA	NA
RIM8	YGL045W	TSS4106-interQ90	246	NA	NA
PUF4	YGL014W	TSS4133-interQ90	341	NA	NA
FMP48	YGR052W	TSS4217-interQ90	416	TES5328-interQ90	332
FMP48	YGR052W	TSS4218-interQ90	313	TES5328-interQ90	202
GPC1	YGR149W	TSS4342-interQ90	271	NA	NA
YCH1	YGR203W	TSS4395-interQ90	591	NA	NA
YCH1	YGR203W	TSS4396-interQ90	558	NA	NA
GCN5	YGR252W	TSS4452-interQ90	561	TES5613-interQ90	6
GCN5	YGR252W	TSS4453-interQ90	473	NA	NA
COS111	YBR203W	TSS446-interQ90	591	NA	NA
RRT6	YGL146C	TSS4574-interQ90	121	TES5815-interQ90	1246
GPX2	YBR244W	TSS484-interQ90	310	NA	NA
GPX2	YBR244W	TSS485-interQ90	190	NA	NA
RSR1	YGR152C	TSS4855-interQ90	364	NA	NA
GTO1	YGR154C	TSS4858-interQ90	311	TES6176-interQ90	1363
GPX2	YBR244W	TSS486-interQ90	159	NA	NA
GPX2	YBR244W	TSS487-interQ90	125	NA	NA
RAD2	YGR258C	TSS4994-interQ90	970	NA	NA
HUA1	YGR268C	TSS5007-interQ90	302	NA	NA
HUA1	YGR268C	TSS5008-interQ90	365	NA	NA
SNF6	YHL025W	TSS5042-interQ90	268	NA	NA
OSH7	YHR001W	TSS5067-interQ90	206	NA	NA
YHR097C	YHR097C	TSS5439-interQ90	106	NA	NA
ECM21	YBL101C	TSS555-interQ90	268	TES746-interQ90	1632
ECM21	YBL101C	TSS556-interQ90	304	TES746-interQ90	1702
KGD1	YIL125W	TSS5579-interQ90	125	NA	NA
YIR014W	YIR014W	TSS5726-interQ90	673	NA	NA
TMA108	YIL137C	TSS5792-interQ90	440	NA	NA
TMA108	YIL137C	TSS5793-interQ90	877	NA	NA
TMA108	YIL137C	TSS5794-interQ90	913	NA	NA
DPB11	YJL090C	TSS6463-interQ90	122	TES8232-interQ90	366
RAV1	YJR033C	TSS6574-interQ90	682	NA	NA
CBT1	YKL208W	TSS6680-interQ90	205	NA	NA
DGR2	YKL121W	TSS6774-interQ90	654	TES8631-interQ90	398
RPL14a	YKL006W	TSS6911-interQ90	183	TES8796-interQ90	1042
GLG1	YKR058W	TSS6968-interQ90	117	NA	NA
ECM4	YKR076W	TSS6991-interQ90	175	NA	NA
ECM4	YKR076W	TSS6992-interQ90	97	NA	NA
MTR2	YKL186C	TSS7060-interQ90	758	TES8946-interQ90	252
SRP21	YKL122C	TSS7121-interQ90	582	NA	NA
MRS4	YKR052C	TSS7306-interQ90	325	NA	NA
REG2	YBR050C	TSS734-interQ90	138	NA	NA
ORC3	YLL004W	TSS7405-interQ90	278	NA	NA
FRE1	YLR214W	TSS7583-interQ90	393	NA	NA
NRG2	YBR066C	TSS762-interQ90	175	NA	NA
NRG2	YBR066C	TSS763-interQ90	223	NA	NA
NRG2	YBR066C	TSS764-interQ90	253	NA	NA

CMG1	YLR271W	TSS7649-interQ90	155	NA	NA
MCM5	YLR274W	TSS7653-interQ90	628	NA	NA
ECM38	YLR299W	TSS7667-interQ90	111	TES9745-interQ90	31
NKP2	YLR315W	TSS7687-interQ90	630	NA	NA
KNS1	YLL019C	TSS7886-interQ90	174	NA	NA
PUF3	YLL013C	TSS7894-interQ90	599	NA	NA
CIS1	YLR346C	TSS8252-interQ90	103	NA	NA
NAM2	YLR382C	TSS8274-interQ90	188	NA	NA
RIF2	YLR453C	TSS8335-interQ90	533	NA	NA
ORC1	YML065W	TSS8406-interQ90	297	NA	NA
ORC1	YML065W	TSS8407-interQ90	228	NA	NA
YET2	YMR040W	TSS8521-interQ90	103	TES10847-interQ90	395
SNZ1	YMR096W	TSS8585-interQ90	773	NA	NA
YMR155W	YMR155W	TSS8667-interQ90	219	NA	NA
CLN1	YMR199W	TSS8727-interQ90	482	NA	NA
POP7	YBR167C	TSS887-interQ90	164	NA	NA
POP7	YBR167C	TSS890-interQ90	373	NA	NA
GIS4	YML006C	TSS9003-interQ90	328	TES11432-interQ90	261
RPS6b	YBR181C	TSS904-interQ90	645	NA	NA
YBR201C-A	YBR201C-A	TSS924-interQ90	124	TES1202-interQ90	502
FET4	YMR319C	TSS9357-interQ90	336	NA	NA
TOS6	YNL300W	TSS9383-interQ90	756	TES11896-interQ90	715
ARO4	YBR249C	TSS977-interQ90	646	TES1255-interQ90	213
RAS2	YNL098C	TSS9999-interQ90	138	NA	NA

Table 8.6 List of genes and alternative upstream TSSs in which the associated coding TSS is neither up- nor downregulated in at least one time point during mid-late gametogenesis

In this list, changes in the levels of coding TSS are independent or not correlated with changes in the levels of the upstream TSS during the mid-late gametogenesis time points. A nearest alternative TES ID is also given in cases where the upstream transcript could terminate at a different site; the distance in nt spanning the upstream TSS and the alternative TES is also given.

Gene	Systematic name	Upstream TSS ID	log2FC of coding TSS (mutant/control)	Upstream TSS downregulated in mutant relative to control?
HMG1	YML075C	TSS8930-interQ90	2.854178008	N
ULP2	YIL031W	TSS5678-interQ90	1.493679498	N
RAD16	YBR114W	TSS365-interQ90	1.431677747	N
RAD16	YBR114W	TSS366-interQ90	1.431677747	N
SSH1	YBR283C	TSS1014-interQ90	1.290727022	N
SSH1	YBR283C	TSS1014-interQ90	1.290727022	N

Table 8.7 List of genes where the levels of the coding TSS is de-repressed in the *set2Δ* mutant relative to the control

This list is sorted according to the $\log_2$ fold change of the coding TSS, from the highest value (greatest extent of de-repression) to the lowest. All values were calculated by comparing the mutant to its control at Spo 6h (meiotic prophase).

Gene	Systematic Nme	Upstream TSS ID	log2FC of coding TSS (mutant/control)	Upstream TSS downregulated in mutant relative to control?
DAL1	YIR027C	TSS5971-interQ90	2.788773798	N
IRC7	YFR055W	TSS3713-interQ90	2.356559932	N
ZPS1	YOL154W	TSS10171-interQ90	1.621942014	Y
CAN1	YEL063C	TSS3301-interQ90	1.536058609	N
PKH2	YOL100W	TSS10234-interQ90	1.429215261	N
PKH2	YOL100W	TSS10235-interQ90	1.429215261	N
AVL9	YLR114C	TSS8016-interQ90	1.306844502	N
MNN4	YKL201C	TSS7032-interQ90	1.215254405	N
MNN4	YKL201C	TSS7033-interQ90	1.215254405	Y
MNN4	YKL201C	TSS7034-interQ90	1.215254405	N
ECM38	YLR299W	TSS7666-interQ90	1.102554889	N
KNS1	YLL019C	TSS7886-interQ90	1.097704407	N
KNS1	YLL019C	TSS7887-interQ90	1.097704407	N

Table 8.8 List of genes where the levels of the coding TSS is de-repressed in the *set2Δ set3Δ* mutant relative to the control

This list is sorted according to the $\log_2$ fold change of the coding TSS, from the highest value (greatest extent of de-repression) to the lowest. All values were calculated by comparing the mutant to its control at Spo 6h (meiotic prophase).

Gene	Systematic name	Upstream TSS ID	log2FC of coding TSS (mutant/control)	Upstream TSS downregulated in mutant relative to control?
SVS1	YPL163C	TSS11988-interQ90	9.80553259	Y
SVS1	YPL163C	TSS11988-interQ90	9.80553259	Y
CLB1	YGR108W	TSS4280-interQ90	8.910251713	N
PUT4	YOR348C	TSS11284-interQ90	8.42841721	N
RNR1	YER070W	TSS3150-interQ90	8.379149984	Y
IRC7	YFR055W	TSS3713-interQ90	8.143145295	N
MCD1	YDL003W	TSS1579-interQ90	7.751313565	N
MCD1	YDL003W	TSS1580-interQ90	7.751313565	N
PIR1	YKL164C	TSS7079-interQ90	7.604498905	Y
STL1	YDR536W	TSS2172-interQ90	7.5423782	N
STL1	YDR536W	TSS2173-interQ90	7.5423782	N
STL1	YDR536W	TSS2174-interQ90	7.5423782	N
PCL9	YDL179W	TSS1410-interQ90	7.252961512	Y
ZPS1	YOL154W	TSS10171-interQ90	6.409781576	Y
DAL1	YIR027C	TSS5971-interQ90	6.035633182	N
FRE1	YLR214W	TSS7584-interQ90	5.117677921	Y
APL4	YPR029C	TSS12204-interQ90	4.83405694	N
CLN2	YPL256C	TSS11886-interQ90	4.577790276	N
RPL28	YGL103W	TSS4035-interQ90	4.192601684	Y
SUR2	YDR297W	TSS1946-interQ90	4.147877931	N
HUA1	YGR268C	TSS5008-interQ90	4.044992922	Y
STP4	YDL048C	TSS2396-interQ90	3.965434307	N
PRR2	YDL214C	TSS2221-interQ90	3.639861895	N
CDC60	YPL160W	TSS11445-interQ90	3.28568649	Y
ULP2	YIL031W	TSS5678-interQ90	3.117475943	N
ECM38	YLR299W	TSS7666-interQ90	2.624411946	Y
SSH1	YBR283C	TSS1014-interQ90	2.562888883	Y
YOR020W-A	YOR020W-A	TSS10364-interQ90	2.554161908	Y
ALY2	YJL084C	TSS6471-interQ90	2.493625393	Y
YPR159C-A	YPR159C-A	TSS12327-interQ90	2.489794189	N
TNA1	YGR260W	TSS4463-interQ90	2.365277803	N
YAR1	YPL239W	TSS11363-interQ90	2.270691828	N
RNR4	YGR180C	TSS4883-interQ90	2.261686483	N
RNR4	YGR180C	TSS4884-interQ90	2.261686483	N
ULP2	YIL031W	TSS5678-interQ90	2.212559908	N
NDC80	YIL144W	TSS5564-interQ90	2.14418951	Y
YBR201C-A	YBR201C-A	TSS924-interQ90	2.023466769	Y
MNN1	YER001W	TSS3064-interQ90	2.0219894	Y
RAD16	YBR114W	TSS365-interQ90	1.760005209	Y
RAD16	YBR114W	TSS366-interQ90	1.760005209	N
FUN19	YAL034C	TSS95-interQ90	1.735628564	Y
CIS1	YLR346C	TSS8252-interQ90	1.215198744	Y
FMP48	YGR052W	TSS4217-interQ90	1.192283885	N
PKH2	YOL100W	TSS10234-interQ90	1.048323096	Y
PKH2	YOL100W	TSS10235-interQ90	1.048323096	N

Table 8.9 List of genes where the levels of the coding TSS is de-repressed in the *SPT16-AID* mutant relative to the control

This list is sorted according to the log_2 fold change of the coding TSS, from the highest value (greatest extent of de-repression) to the lowest. All values were calculated by comparing the mutant to its control at Spo 6h (meiotic prophase).

Gene	Systematic name	Upstream TSS ID	Upstream TSS downregulated during RTG?	Coding TSS upregulated during RTG?	Upstream TSS downregulated within 15 min of RTG?	Upstream TSS downregulated within 30 min of RTG?	Coding TSS upregulated within 15 min of RTG?	Coding TSS upregulated within 30 min of RTG?
SVS1	YPL163C	TSS11988-interQ90	Y	Y	Y	Y	Y	Y
DAL1	YIR027C	TSS5971-interQ90	N	N	N	N	N	N
CLB1	YGR108W	TSS4280-interQ90	Y	Y	Y	Y	Y	Y
HMG1	YML075C	TSS8930-interQ90	Y	Y	Y	Y	Y	Y
PCL9	YDL179W	TSS1410-interQ90	Y	Y	N	N	N	N
CLN1	YMR199W	TSS8727-interQ90	Y	Y	Y	Y	N	Y
PIR1	YKL164C	TSS7079-interQ90	Y	Y	Y	Y	Y	Y
IRC7	YFR055W	TSS3713-interQ90	N	Y	N	N	Y	Y
FMP48	YGR052W	TSS4217-interQ90	Y	N	Y	Y	N	N
CLN1	YMR199W	TSS8726-interQ90	Y	Y	Y	Y	N	Y
PUT4	YOR348C	TSS11284-interQ90	Y	N	Y	Y	N	N
APL4	YPR029C	TSS12204-interQ90	Y	Y	Y	Y	Y	Y
MNN1	YER001W	TSS3064-interQ90	Y	Y	N	N	Y	Y
YOR1	YGR281W	TSS4478-interQ90	Y	Y	Y	Y	Y	Y
FRE1	YLR214W	TSS7584-interQ90	Y	Y	N	N	Y	Y
MMS1	YPR164W	TSS11811-interQ90	Y	Y	Y	Y	N	N
MMS1	YPR164W	TSS11812-interQ90	Y	Y	Y	Y	N	N
CLB2	YPR119W	TSS11764-interQ90	Y	Y	Y	Y	N	N
ATO3	YDR384C	TSS2826-interQ90	Y	Y	Y	N	Y	Y
SKM1	YOL113W	TSS10215-interQ90	Y	Y	Y	Y	N	N
BOI1	YBL085W	TSS145-interQ90	Y	Y	Y	Y	Y	Y
CSI2	YOL007C	TSS10901-interQ90	Y	Y	N	N	Y	Y
CLN2	YPL256C	TSS11886-interQ90	Y	Y	Y	N	Y	Y
YOR020W-A	YOR020W-A	TSS10364-interQ90	Y	N	Y	Y	N	N
MES1	YGR264C	TSS5001-interQ90	Y	Y	Y	Y	Y	Y
BUL2	YML111W	TSS8361-interQ90	Y	Y	Y	Y	Y	Y
BUL2	YML111W	TSS8362-interQ90	Y	Y	Y	Y	Y	Y
SWI4	YER111C	TSS3489-interQ90	Y	Y	Y	Y	Y	Y
POP4	YBR257W	TSS501-interQ90	Y	Y	Y	Y	Y	Y
POP4	YBR257W	TSS502-interQ90	Y	Y	Y	Y	Y	Y
POP4	YBR257W	TSS500-interQ90	Y	Y	Y	Y	Y	Y
IML1	YJR138W	TSS6309-interQ90	Y	Y	Y	Y	N	Y
AUA1	YFL010W-A	TSS3643-interQ90	Y	N	Y	Y	N	N
KNS1	YLL019C	TSS7886-interQ90	Y	Y	N	N	Y	N
KNS1	YLL019C	TSS7887-interQ90	Y	Y	Y	N	Y	N
MNN4	YKL201C	TSS7034-interQ90	Y	Y	Y	Y	Y	Y
MNN4	YKL201C	TSS7033-interQ90	Y	Y	N	N	Y	Y
MNN4	YKL201C	TSS7032-interQ90	N	Y	N	N	Y	Y
CMR3	YPR013C	TSS12183-interQ90	Y	Y	N	N	Y	Y
CAN1	YEL063C	TSS3301-interQ90	N	Y	N	N	N	Y
CIS1	YLR346C	TSS8252-interQ90	Y	N	Y	Y	N	N
RNR1	YER070W	TSS3150-interQ90	N	Y	N	N	Y	Y
ALY2	YJL084C	TSS6471-interQ90	Y	Y	Y	Y	Y	Y
DSE3	YOR264W	TSS10599-interQ90	Y	N	N	N	N	N
GIP3	YPL137C	TSS12028-interQ90	Y	Y	Y	N	Y	Y
RRT6	YGL146C	TSS4574-interQ90	Y	Y	N	N	Y	N
RPL28	YGL103W	TSS4035-interQ90	Y	Y	Y	Y	Y	Y
GDH2	YDL215C	TSS2218-interQ90	Y	N	Y	Y	N	N
NDC80	YIL144W	TSS5564-interQ90	Y	Y	Y	Y	Y	Y
YGR053C	YGR053C	TSS4755-interQ90	Y	N	Y	Y	N	N
RPO41	YFL036W	TSS3611-interQ90	Y	Y	Y	Y	Y	Y
RPO41	YFL036W	TSS3609-interQ90	Y	Y	Y	Y	Y	Y
RPO41	YFL036W	TSS3610-interQ90	Y	Y	Y	Y	Y	Y
CDC60	YPL160W	TSS11445-interQ90	Y	Y	Y	Y	Y	Y
RTT10	YPL183C	TSS11967-interQ90	Y	Y	Y	Y	Y	Y
MTR2	YKL186C	TSS7059-interQ90	Y	N	N	N	N	N
YPR159C-A	YPR159C-A	TSS12327-interQ90	Y	N	Y	Y	N	N
RAD16	YBR114W	TSS365-interQ90	Y	Y	Y	Y	Y	Y
RAD16	YBR114W	TSS366-interQ90	Y	Y	Y	Y	Y	Y
GSC2	YGR032W	TSS4180-interQ90	Y	N	Y	N	N	N
MIT1	YEL007W	TSS3052-interQ90	Y	Y	Y	Y	Y	Y
MIT1	YEL007W	TSS3053-interQ90	Y	Y	N	N	Y	Y
TAX4	YJL083W	TSS6108-interQ90	Y	Y	N	N	N	N

SSH1	YBR283C	TSS1014-interQ90	Y	Y	Y	Y	Y	Y
MCD1	YDL003W	TSS1580-interQ90	N	Y	N	N	Y	Y
MCD1	YDL003W	TSS1579-interQ90	Y	Y	N	N	Y	Y
ZPS1	YOL154W	TSS10171-interQ90	Y	Y	Y	Y	Y	N
YPS1	YLR120C	TSS8022-interQ90	Y	N	Y	Y	N	N
HUA1	YGR268C	TSS5008-interQ90	Y	N	Y	Y	N	N
STP4	YDL048C	TSS2396-interQ90	Y	Y	Y	N	Y	Y
ECM38	YLR299W	TSS7666-interQ90	Y	N	Y	Y	N	N
STL1	YDR536W	TSS2173-interQ90	Y	N	Y	Y	N	N
STL1	YDR536W	TSS2174-interQ90	Y	N	Y	Y	N	N
STL1	YDR536W	TSS2172-interQ90	N	N	N	N	N	N
ULP2	YIL031W	TSS5678-interQ90	Y	Y	Y	Y	N	Y
GLC3	YEL011W	TSS3042-interQ90	Y	N	Y	Y	N	N
RNR4	YGR180C	TSS4884-interQ90	Y	Y	Y	Y	Y	Y
PKH2	YOL100W	TSS10234-interQ90	Y	N	Y	Y	N	N
PKH2	YOL100W	TSS10235-interQ90	Y	N	Y	Y	N	N
MDJ2	YNL328C	TSS9788-interQ90	Y	N	Y	Y	N	N
MBA1	YBR185C	TSS908-interQ90	Y	Y	Y	Y	Y	Y
FLR1	YBR008C	TSS684-interQ90	N	Y	N	N	N	N
YBR201C-A	YBR201C-A	TSS924-interQ90	Y	N	Y	Y	N	N
RNR4	YGR180C	TSS4883-interQ90	N	Y	N	N	Y	Y
CMR3	YPR013C	TSS12182-interQ90	Y	Y	Y	Y	Y	Y
FPS1	YLL043W	TSS7364-interQ90	Y	Y	Y	Y	Y	Y
FPS1	YLL043W	TSS7365-interQ90	Y	Y	Y	Y	Y	Y
FPS1	YLL043W	TSS7366-interQ90	Y	Y	Y	Y	Y	Y
SUR2	YDR297W	TSS1946-interQ90	N	Y	N	N	Y	Y
MAK21	YDR060W	TSS1671-interQ90	Y	Y	N	Y	Y	Y
MAK21	YDR060W	TSS1670-interQ90	Y	Y	Y	Y	Y	Y
TRA1	YHR099W	TSS5161-interQ90	Y	Y	Y	Y	Y	Y
TRA1	YHR099W	TSS5160-interQ90	Y	Y	Y	Y	Y	Y
RIF2	YLR453C	TSS8334-interQ90	Y	N	Y	Y	N	N
TNA1	YGR260W	TSS4463-interQ90	Y	Y	Y	Y	Y	Y
HSL1	YKL101W	TSS6795-interQ90	N	Y	N	N	Y	Y
PRR2	YDL214C	TSS2221-interQ90	Y	N	Y	N	N	N
CAX4	YGR036C	TSS4735-interQ90	N	Y	N	N	N	Y
OXA1	YER154W	TSS3252-interQ90	Y	Y	Y	Y	Y	Y
SKI7	YOR076C	TSS11007-interQ90	Y	N	Y	Y	N	N
YAR1	YPL239W	TSS11363-interQ90	Y	Y	N	N	Y	Y
YOR238W	YOR238W	TSS10575-interQ90	Y	N	Y	Y	N	N
YOR238W	YOR238W	TSS10576-interQ90	Y	N	N	Y	N	N
FUN19	YAL034C	TSS95-interQ90	Y	N	Y	Y	N	N
RRP5	YMR229C	TSS9254-interQ90	Y	Y	N	N	Y	Y
SUR1	YPL057C	TSS12109-interQ90	Y	N	N	N	N	N
LIN1	YHR156C	TSS5496-interQ90	Y	Y	Y	Y	N	Y
AVL9	YLR114C	TSS8016-interQ90	Y	Y	Y	Y	N	N
MOD5	YOR274W	TSS10610-interQ90	Y	Y	Y	Y	Y	Y
MOD5	YOR274W	TSS10609-interQ90	Y	Y	Y	Y	Y	Y
VPS30	YPL120W	TSS11494-interQ90	Y	N	Y	Y	N	N
VPS30	YPL120W	TSS11496-interQ90	Y	N	Y	Y	N	N
MLP1	YKR095W	TSS7006-interQ90	Y	Y	Y	Y	N	N
CFT2	YLR115W	TSS7503-interQ90	Y	Y	Y	Y	N	N

Table 8.10 List of early meiotic "repressed" genes and the effect of RTG on the upstream and coding TSSs

The genes in this list have an upstream alternative TSS which was upregulated and an associated coding TSS which was downregulated in at least one time point during early gametogenesis. Columns 4-9 provide information about the behaviors of the alternative TSSs and coding TSSs upon RTG.

8.3 Copyright permissions

This research in chapter 3 of this thesis has been published in G3 (Bethesda) and has been modified for this results chapter (Chia and van Werven 2016). Chia M, van Werven FJ. Temporal Expression of a Master Regulator Drives Synchronous Sporulation in Budding Yeast. *G3: Genes|Genomes|Genetics*. 2016;6(11):3553-3560. doi:10.1534/g3.116.034983. *G3* articles are published as open-access articles distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This research in chapter 4 has been published in two separate papers in eLife and has been modified for this results chapter (Chen et al. 2017;Chia et al. 2017). Chen J, Tresenrider A, Chia M, McSwiggen DT, Spedale G, Jorgensen V, Liao H, van Werven FJ, Unal E. 2017. Kinetochore inactivation by expression of a repressive mRNA. Elife 6:e27417; Chia M, Tresenrider A, Chen J, Spedale G, Jorgensen V, Unal E, van Werven FJ. 2017. Transcription of a 5' extended mRNA isoform directs dynamic chromatin changes and interference of a downstream promoter. Elife 6:e27420. The articles and journal content published by eLife on the eLife Sites are licensed under a Creative Commons Attribution license (also known as a CC-BY license, https://creativecommons.org/licenses/by/4.0/). This means that you are free to use, reproduce and distribute the articles and related content (unless otherwise noted), for commercial and noncommercial purposes, subject to citation of the original source in accordance with the CC-BY license.

Reference List

1 Aanes, H., Ostrup, O., Andersen, I. S., Moen, L. F., Mathavan, S., Collas, P.,Alestrom, P. Differential transcript isoform usage pre- and post-zygotic genome activation in zebrafish. *Bmc Genomics* **14**, 331, doi:10.1186/1471-2164-14-331 (2013).

2 Abarrategui, I., Krangel, M. S. Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination. *The EMBO journal* **26**, 4380-4390, doi:10.1038/sj.emboj.7601866 (2007).

3 Acosta, I., Ontoso, D.,San-Segundo, P. A. The budding yeast polo-like kinase Cdc5 regulates the Ndt80 branch of the meiotic recombination checkpoint pathway. *Molecular biology of the cell* **22**, 3478-3490, doi:10.1091/mbc.E11-06-0482 (2011).

4 Acquaviva, L., Szekvolgyi, L., Dichtl, B., Dichtl, B. S., de La Roche Saint Andre, C., Nicolas, A.,Geli, V. The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science* **339**, 215-218, doi:10.1126/science.1225739 (2013).

5 Adjalley, S. H., Chabbert, C. D., Klaus, B., Pelechano, V.,Steinmetz, L. M. Landscape and Dynamics of Transcription Initiation in the Malaria Parasite Plasmodium falciparum. *Cell Rep* **14**, 2463-2475, doi:10.1016/j.celrep.2016.02.025 (2016).

6 Ali, M., Rincon-Arano, H., Zhao, W., Rothbart, S. B., Tong, Q., Parkhurst, S. M., Strahl, B. D., Deng, L. W., Groudine, M.,Kutateladze, T. G. Molecular basis for chromatin binding and regulation of MLL5. *Proc Natl Acad Sci U S A* **110**, 11296-11301, doi:10.1073/pnas.1310156110 (2013).

7 Ansari, A., Hampsey, M. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev* **19**, 2969-2978, doi:10.1101/gad.1362305 (2005).

8 Ard, R., Allshire, R. C. Transcription-coupled changes to chromatin underpin gene silencing by transcriptional interference. *Nucleic Acids Res* **44**, 10619-10630, doi:10.1093/nar/gkw801 (2016).

9 Ard, R., Tong, P.,Allshire, R. C. Long non-coding RNA-mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. *Nature communications* **5**, 5576, doi:10.1038/ncomms6576 (2014).

10 Ares, M. Isolation of total RNA from yeast cell cultures. *Cold Spring Harb Protoc* **2012**, 1082-1086, doi:10.1101/pdb.prot071456 (2012).

11 Arribere, J. A., Gilbert, W. V. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome research* **23**, 977-987, doi:10.1101/gr.150342.112 (2013).

12 Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**, 25-29, doi:10.1038/75556 (2000).

13 Bannister, A. J., Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395, doi:10.1038/cr.2011.22 (2011).

14 Bao, X., Wu, H., Zhu, X., Guo, X., Hutchins, A. P., Luo, Z., Song, H., Chen, Y., Lai, K., Yin, M., Xu, L. et al. The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. *Cell Res* **25**, 80-92, doi:10.1038/cr.2014.165 (2015).

15 Battaglia, S., Lidschreiber, M., Baejen, C., Torkler, P., Vos, S. M.,Cramer, P. RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases. *Elife* **6**, doi:10.7554/eLife.25637 (2017).

16 Becker, E., Liu, Y., Lardenois, A., Walther, T., Horecka, J., Stuparevic, I., Law, M. J., Lavigne, R., Evrard, B., Demougin, P., Riffle, M. et al. Integrated RNA- and protein profiling of fermentation and respiration in diploid budding yeast provides insight into nutrient control of cell growth and development. *J Proteomics* **119**, 30-44, doi:10.1016/j.jprot.2015.01.015 (2015).

17 Belotserkovskaya, R., Oh, S., Bondarenko, V. A., Orphanides, G., Studitsky, V. M.,Reinberg, D. FACT facilitates transcription-dependent nucleosome alteration. *Science* **301**, 1090-1093, doi:10.1126/science.1085703 (2003).

18 Belotserkovskii, B. P., Soo Shin, J. H.,Hanawalt, P. C. Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Res* **45**, 6589-6599, doi:10.1093/nar/gkx403 (2017).

19 Benjamin, K. R., Zhang, C., Shokat, K. M.,Herskowitz, I. Control of landmark events in meiosis by the CDK Cdc28 and the meiosis-specific kinase Ime2. *Genes Dev* **17**, 1524-1539, doi:10.1101/gad.1101503 (2003).

20 Berchowitz, L. E., Gajadhar, A. S., van Werven, F. J., De Rosa, A. A., Samoylova, M. L., Brar, G. A., Xu, Y., Xiao, C., Futcher, B., Weissman, J. S., White, F. M. et al. A developmentally regulated translational control pathway establishes the meiotic chromosome segregation pattern. *Genes Dev* **27**, 2147-2163, doi:10.1101/gad.224253.113 (2013).

21 Berghoff, B. A., Karlsson, T., Kallman, T., Wagner, E. G. H.,Grabherr, M. G. RNA-sequence data normalization through in silico prediction of reference genes: the bacterial response to DNA damage as case study. *BioData Min* **10**, 30, doi:10.1186/s13040-017-0150-8 (2017).

22 Berretta, J., Pinskaya, M.,Morillon, A. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. *Genes Dev* **22**, 615-626, doi:10.1101/gad.458008 (2008).

23 Bird, A. J., Gordon, M., Eide, D. J.,Winge, D. R. Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts. *The EMBO journal* **25**, 5726-5734, doi:10.1038/sj.emboj.7601453 (2006).

24 Bodi, Z., Button, J. D., Grierson, D.,Fray, R. G. Yeast targets for mRNA methylation. *Nucleic Acids Res* **38**, 5327-5335, doi:10.1093/nar/gkq266 (2010).

25 Bonen, H., Kol, N., Shomron, N., Leibowitz-Amit, R., Quagliata, L., Lorber, T., Sidi, Y.,Avni, D. Promoter-Associated RNAs Regulate HSPC152 Gene Expression in Malignant Melanoma. *Noncoding RNA* **2**, doi:10.3390/ncrna2030007 (2016).

26 Booth, G. T., Wang, I. X., Cheung, V. G.,Lis, J. T. Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome research* **26**, 799-811, doi:10.1101/gr.204578.116 (2016).

27 Borde, V., Robine, N., Lin, W., Bonfils, S., Geli, V.,Nicolas, A. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *The EMBO journal* **28**, 99-111, doi:10.1038/emboj.2008.257 (2009).

28 Bordoy, A. E., Varanasi, U. S., Courtney, C. M.,Chatterjee, A. Transcriptional Interference in Convergent Promoters as a Means for Tunable Gene Expression. *ACS synthetic biology* **5**, 1331-1341, doi:10.1021/acssynbio.5b00223 (2016).

29 Borner, G. V., Cha, R. S. Induction and Analysis of Synchronous Meiotic Yeast Cultures. *Cold Spring Harb Protoc* **2015**, 908-913, doi:10.1101/pdb.prot085035 (2015).

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

31 Bowdish, K. S., Yuan, H. E.,Mitchell, A. P. Analysis of RIM11, a yeast protein kinase that phosphorylates the meiotic activator IME1. *Mol Cell Biol* **14**, 7909-7919 (1994).
32 Bowdish, K. S., Yuan, H. E.,Mitchell, A. P. Positive control of yeast meiotic genes by the negative regulator UME6. *Mol Cell Biol* **15**, 2955-2961 (1995).
33 Brar, G. A., Yassour, M., Friedman, N., Regev, A., Ingolia, N. T.,Weissman, J. S. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**, 552-557, doi:10.1126/science.1215110 (2012).
34 Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y., Winston, F.,Allis, C. D. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev* **15**, 3286-3295, doi:10.1101/gad.940201 (2001).
35 Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W., Wen, J., Park, S., Suzuki, A. M., Wan, K. H. et al. Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**, 393-399, doi:10.1038/nature12962 (2014).
36 Bumgarner, S. L., Dowell, R. D., Grisafi, P., Gifford, D. K.,Fink, G. R. Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proc Natl Acad Sci U S A* **106**, 18321-18326, doi:10.1073/pnas.0909641106 (2009).
37 Buratowski, S., Kim, T. The role of cotranscriptional histone methylations. *Cold Spring Harbor symposia on quantitative biology* **75**, 95-102, doi:10.1101/sqb.2010.75.036 (2010).
38 Bushnell, D. A., Kornberg, R. D. Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A* **100**, 6969-6973, doi:10.1073/pnas.1130601100 (2003).
39 Callen, B. P., Shearwin, K. E.,Egan, J. B. Transcriptional interference between convergent promoters caused by elongation over the promoter. *Molecular cell* **14**, 647-656, doi:10.1016/j.molcel.2004.05.010 (2004).
40 Camblong, J., Beyrouthy, N., Guffanti, E., Schlaepfer, G., Steinmetz, L. M.,Stutz, F. Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. *Genes Dev* **23**, 1534-1545, doi:10.1101/gad.522509 (2009).
41 Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G.,Stutz, F. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. *Cell* **131**, 706-717, doi:10.1016/j.cell.2007.09.014 (2007).
42 Carlile, T. M., Amon, A. Meiosis I is established through division-specific translational control of a cyclin. *Cell* **133**, 280-291, doi:10.1016/j.cell.2008.02.032 (2008).
43 Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., Izawa, M., Muramatsu, M. et al. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* **37**, 327-336, doi:10.1006/geno.1996.0567 (1996).
44 Carrozza, M. J., Florens, L., Swanson, S. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P.,Workman, J. L. Stable incorporation of sequence specific repressors Ash1 and Ume6 into the Rpd3L complex. *Biochimica et biophysica acta* **1731**, 77-87; discussion 75-76, doi:10.1016/j.bbaexp.2005.09.005 (2005a).
45 Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P.,Workman, J. L. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581-592, doi:10.1016/j.cell.2005.10.023 (2005b).

46 Carter, R., Drouin, G. Structural differentiation of the three eukaryotic RNA polymerases. *Genomics* **94**, 388-396, doi:10.1016/j.ygeno.2009.08.011 (2009).
47 Castelnuovo, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F.,Zenklusen, D. Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat Struct Mol Biol* **20**, 851-858, doi:10.1038/nsmb.2598 (2013).
48 Chang, H., Lim, J., Ha, M.,Kim, V. N. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Molecular cell* **53**, 1044-1052, doi:10.1016/j.molcel.2014.02.007 (2014).
49 Chang, J., Vancura, A. Analysis of SUC2 promoter structure by nucleosome scanning. *Methods Mol Biol* **809**, 321-333, doi:10.1007/978-1-61779-376-9_22 (2012).
50 Chedin, F. Nascent Connections: R-Loops and Chromatin Patterning. *Trends Genet* **32**, 828-838, doi:10.1016/j.tig.2016.10.002 (2016).
51 Chen, J., Tresenrider, A., Chia, M., McSwiggen, D. T., Spedale, G., Jorgensen, V., Liao, H., van Werven, F. J.,Unal, E. Kinetochore inactivation by expression of a repressive mRNA. *Elife* **6**, doi:10.7554/eLife.27417 (2017).
52 Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W.,Tyler, J. K. The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Mol Cell Biol* **36**, 662-667, doi:10.1128/MCB.00970-14 (2015).
53 Chen, X. F., Kuryan, B., Kitada, T., Tran, N., Li, J. Y., Kurdistani, S., Grunstein, M., Li, B.,Carey, M. The Rpd3 core complex is a chromatin stabilization module. *Curr Biol* **22**, 56-63, doi:10.1016/j.cub.2011.11.042 (2012).
54 Cheng, Z., Otto, G. M., Powers, E. N., Keskin, A., Mertins, P., Carr, S. A., Jovanovic, M.,Brar, G. A. Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis. *Cell* **172**, 910-923 e916, doi:10.1016/j.cell.2018.01.035 (2018).
55 Cheung, V., Chua, G., Batada, N. N., Landry, C. R., Michnick, S. W., Hughes, T. R.,Winston, F. Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces cerevisiae genome. *PLoS Biol* **6**, e277, doi:10.1371/journal.pbio.0060277 (2008).
56 Chia, M., Tresenrider, A., Chen, J., Spedale, G., Jorgensen, V., Unal, E.,van Werven, F. J. Transcription of a 5' extended mRNA isoform directs dynamic chromatin changes and interference of a downstream promoter. *Elife* **6**, doi:10.7554/eLife.27420 (2017).
57 Chia, M., van Werven, F. J. Temporal Expression of a Master Regulator Drives Synchronous Sporulation in Budding Yeast. *G3 (Bethesda)* **6**, 3553-3560, doi:10.1534/g3.116.034983 (2016).
58 Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E.,Chang, H. Y. Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416, doi:10.1016/j.cell.2015.03.025 (2015).
59 Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O.,Herskowitz, I. The transcriptional program of sporulation in budding yeast. *Science* **282**, 699-705 (1998).
60 Chu, S., Herskowitz, I. Gametogenesis in yeast is regulated by a transcriptional cascade dependent on Ndt80. *Molecular cell* **1**, 685-696 (1998).
61 Churchman, L. S., Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368-373, doi:10.1038/nature09652 (2011).
62 Ciferri, C., Musacchio, A.,Petrovic, A. The Ndc80 complex: hub of kinetochore activity. *FEBS letters* **581**, 2862-2869, doi:10.1016/j.febslet.2007.05.012 (2007).
63 Clancy, M. J., Shambaugh, M. E., Timpte, C. S.,Bokar, J. A. Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N6-

methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Res* **30**, 4509-4518 (2002).

64 Codon, A. C., Gasent-Ramirez, J. M.,Benitez, T. Factors Which Affect the Frequency of Sporulation and Tetrad Formation in Saccharomyces cerevisiae Baker's Yeasts. *Appl Environ Microbiol* **61**, 1677 (1995).

65 Colomina, N., Gari, E., Gallego, C., Herrero, E.,Aldea, M. G1 cyclins block the Ime1 pathway to make mitosis and meiosis incompatible in budding yeast. *The EMBO journal* **18**, 320-329, doi:10.1093/emboj/18.2.320 (1999).

66 Colomina, N., Liu, Y., Aldea, M.,Gari, E. TOR regulates the subcellular localization of Ime1, a transcriptional activator of meiotic development in budding yeast. *Mol Cell Biol* **23**, 7415-7424 (2003).

67 Cooper, K. F., Mallory, M. J., Guacci, V., Lowe, K.,Strich, R. Pds1p is required for meiotic recombination and prophase I progression in Saccharomyces cerevisiae. *Genetics* **181**, 65-79, doi:10.1534/genetics.108.095513 (2009).

68 Dallas, A., Vlassov, A. V.,Kazakov, S. A. in *Artificial Nucleases* 10.1007/978-3-642-18510-6_6 (ed Marina A. Zenkova) 61-88 (Springer Berlin Heidelberg, 2004).

69 David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W.,Steinmetz, L. M. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A* **103**, 5320-5325, doi:10.1073/pnas.0601091103 (2006).

70 Deutschbauer, A. M., Davis, R. W. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nature genetics* **37**, 1333-1340, doi:10.1038/ng1674 (2005).

71 Dickinson, J. R., Dawes, I. W., Boyd, A. S.,Baxter, R. L. 13C NMR studies of acetate metabolism during sporulation of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **80**, 5847-5851 (1983).

72 Dirick, L., Goetsch, L., Ammerer, G.,Byers, B. Regulation of meiotic S phase by Ime2 and a Clb5,6-associated kinase in Saccharomyces cerevisiae. *Science* **281**, 1854-1857 (1998).

73 Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C. et al. Landscape of transcription in human cells. *Nature* **489**, 101-108, doi:10.1038/nature11233 (2012).

74 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M.,Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

75 Du, H. N., Fingerman, I. M.,Briggs, S. D. Histone H3 K36 methylation is mediated by a trans-histone methylation pathway involving an interaction between Set2 and histone H4. *Genes Dev* **22**, 2786-2798, doi:10.1101/gad.1700008 (2008).

76 Duina, A. A., Miller, M. E.,Keeney, J. B. Budding yeast for budding geneticists: a primer on the Saccharomyces cerevisiae model system. *Genetics* **197**, 33-48, doi:10.1534/genetics.114.163188 (2014).

77 Edmunds, J. W., Mahadevan, L. C.,Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *The EMBO journal* **27**, 406-420, doi:10.1038/sj.emboj.7601967 (2008).

78 Eilbeck, K., Lewis, S. E., Mungall, C. J., Yandell, M., Stein, L., Durbin, R.,Ashburner, M. The Sequence Ontology: a tool for the unification of genome annotations. *Genome biology* **6**, R44, doi:10.1186/gb-2005-6-5-r44 (2005).

79 Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D. K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., Sato, H. et al. Human Spt6 stimulates

transcription elongation by RNA polymerase II in vitro. *Mol Cell Biol* **24**, 3324-3336 (2004).

80 Fabrega, C., Shen, V., Shuman, S.,Lima, C. D. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Molecular cell* **11**, 1549-1561 (2003).

81 Falk, J. E., Chan, A. C., Hoffmann, E.,Hochwagen, A. A Mec1- and PP4-dependent checkpoint couples centromere pairing to meiotic recombination. *Developmental cell* **19**, 599-611, doi:10.1016/j.devcel.2010.09.006 (2010).

82 Feng, J., Gan, H., Eaton, M. L., Zhou, H., Li, S., Belsky, J. A., MacAlpine, D. M., Zhang, Z.,Li, Q. Noncoding Transcription Is a Driving Force for Nucleosome Instability in spt16 Mutant Cells. *Mol Cell Biol* **36**, 1856-1867, doi:10.1128/MCB.00152-16 (2016).

83 Ferrari, P., Strubin, M. Uncoupling histone turnover from transcription-associated histone H3 modifications. *Nucleic Acids Res* **43**, 3972-3985, doi:10.1093/nar/gkv282 (2015).

84 Fox, C., Zou, J., Rappsilber, J.,Marston, A. L. Cdc14 phosphatase directs centrosome re-duplication at the meiosis I to meiosis II transition in budding yeast. *Wellcome Open Res* **2**, 2, doi:10.12688/wellcomeopenres.10507.2 (2017).

85 Frenk, S., Oxley, D.,Houseley, J. The nuclear exosome is active and important during budding yeast meiosis. *PloS one* **9**, e107648, doi:10.1371/journal.pone.0107648 (2014).

86 Friedlander, G., Joseph-Strauss, D., Carmi, M., Zenvirth, D., Simchen, G.,Barkai, N. Modulation of the transcription regulatory program in yeast cells committed to sporulation. *Genome biology* **7**, R20, doi:10.1186/gb-2006-7-3-r20 (2006).

87 Friesen, H., Hepworth, S. R.,Segall, J. An Ssn6-Tup1-dependent negative regulatory element controls sporulation-specific expression of DIT1 and DIT2 in Saccharomyces cerevisiae. *Mol Cell Biol* **17**, 123-134 (1997).

88 Furlong, E. E. The importance of being specified: cell fate decisions and their role in cell biology. *Molecular biology of the cell* **21**, 3797-3798, doi:10.1091/mbc.E10-05-0436 (2010).

89 Gardini, A. Global Run-On Sequencing (GRO-Seq). *Methods Mol Biol* **1468**, 111-120, doi:10.1007/978-1-4939-4035-6_9 (2017).

90 Gelfand, B., Mead, J., Bruning, A., Apostolopoulos, N., Tadigotla, V., Nagaraj, V., Sengupta, A. M.,Vershon, A. K. Regulated antisense transcription controls expression of cell-type-specific genes in yeast. *Mol Cell Biol* **31**, 1701-1709, doi:10.1128/MCB.01071-10 (2011).

91 Gerke, J. P., Chen, C. T.,Cohen, B. A. Natural isolates of Saccharomyces cerevisiae display complex genetic variation in sporulation efficiency. *Genetics* **174**, 985-997, doi:10.1534/genetics.106.058453 (2006).

92 Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I.,Chedin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular cell* **45**, 814-825, doi:10.1016/j.molcel.2012.01.017 (2012).

93 Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J. et al. Life with 6000 genes. *Science* **274**, 546, 563-547 (1996).

94 Gong, Q. H., McDowell, J. C.,Dean, A. Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene in vivo by 5' hypersensitive site 2 of the beta-globin locus control region. *Mol Cell Biol* **16**, 6055-6064 (1996).

95 Govind, C. K., Qiu, H., Ginsburg, D. S., Ruan, C., Hofmeyer, K., Hu, C., Swaminathan, V., Workman, J. L., Li, B.,Hinnebusch, A. G. Phosphorylated Pol

II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes. *Molecular cell* **39**, 234-246, doi:10.1016/j.molcel.2010.07.003 (2010).
96 Granot, D., Margolskee, J. P.,Simchen, G. A long region upstream of the IME1 gene regulates meiosis in yeast. *Mol Gen Genet* **218**, 308-314 (1989).
97 Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R.,Fraser, P. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Molecular cell* **5**, 377-386 (2000).
98 Groh, M., Gromak, N. Out of balance: R-loops in human disease. *PLoS Genet* **10**, e1004630, doi:10.1371/journal.pgen.1004630 (2014).
99 Gubler, U., Hoffman, B. J. A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263-269, doi:https://doi.org/10.1016/0378-1119(83)90230-5 (1983).
100 Guillemette, B., Drogaris, P., Lin, H. H., Armstrong, H., Hiragami-Hamada, K., Imhof, A., Bonneil, E., Thibault, P., Verreault, A.,Festenstein, R. J. H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation. *PLoS Genet* **7**, e1001354, doi:10.1371/journal.pgen.1001354 (2011).
101 Guttmann-Raviv, N., Martin, S.,Kassir, Y. Ime2, a meiosis-specific kinase in yeast, is required for destabilization of its transcriptional activator, Ime1. *Mol Cell Biol* **22**, 2047-2056 (2002).
102 Haberle, V., Forrest, A. R., Hayashizaki, Y., Carninci, P.,Lenhard, B. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res* **43**, e51, doi:10.1093/nar/gkv054 (2015).
103 Hainer, S. J., Martens, J. A. Identification of histone mutants that are defective for transcription-coupled nucleosome occupancy. *Mol Cell Biol* **31**, 3557-3568, doi:10.1128/MCB.05195-11 (2011).
104 Hainer, S. J., Pruneski, J. A., Mitchell, R. D., Monteverde, R. M.,Martens, J. A. Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev* **25**, 29-40, doi:10.1101/gad.1975011 (2011).
105 Hampsey, M., Reinberg, D. Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* **113**, 429-432 (2003).
106 Han, J., Kim, D.,Morris, K. V. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci U S A* **104**, 12422-12427, doi:10.1073/pnas.0701635104 (2007).
107 Hao, N., Palmer, A. C., Ahlgren-Berg, A., Shearwin, K. E.,Dodd, I. B. The role of repressor kinetics in relief of transcriptional interference between convergent promoters. *Nucleic Acids Res* **44**, 6625-6638, doi:10.1093/nar/gkw600 (2016).
108 Harigaya, Y., Tanaka, H., Yamanaka, S., Tanaka, K., Watanabe, Y., Tsutsumi, C., Chikashige, Y., Hiraoka, Y., Yamashita, A.,Yamamoto, M. Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* **442**, 45-50, doi:10.1038/nature04881 (2006).
109 Hartley, P. D., Madhani, H. D. Mechanisms that specify promoter nucleosome location and identity. *Cell* **137**, 445-458, doi:10.1016/j.cell.2009.02.043 (2009).
110 Hartzog, G. A., Wada, T., Handa, H.,Winston, F. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes Dev* **12**, 357-369 (1998).
111 Hassan, A. H., Neely, K. E.,Workman, J. L. Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. *Cell* **104**, 817-827 (2001).

112 He, F., Li, X., Spatrick, P., Casillo, R., Dong, S.,Jacobson, A. Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Molecular cell* **12**, 1439-1452 (2003).

113 He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N.,Kinzler, K. W. The antisense transcriptomes of human cells. *Science* **322**, 1855-1857, doi:10.1126/science.1163853 (2008).

114 Hepworth, S. R., Friesen, H.,Segall, J. NDT80 and the meiotic recombination checkpoint regulate expression of middle sporulation-specific genes in Saccharomyces cerevisiae. *Mol Cell Biol* **18**, 5750-5761 (1998).

115 Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C. S., Shibata, T.,Ohta, K. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* **456**, 130-134, doi:10.1038/nature07348 (2008).

116 Hobson, D. J., Wei, W., Steinmetz, L. M.,Svejstrup, J. Q. RNA polymerase II collision interrupts convergent transcription. *Molecular cell* **48**, 365-374, doi:10.1016/j.molcel.2012.08.027 (2012).

117 Hongay, C. F., Grisafi, P. L., Galitski, T.,Fink, G. R. Antisense transcription controls cell fate in Saccharomyces cerevisiae. *Cell* **127**, 735-745, doi:10.1016/j.cell.2006.09.038 (2006).

118 Honigberg, S. M., Purnapatre, K. Signal pathway integration in the switch from the mitotic cell cycle to meiosis in yeast. *Journal of cell science* **116**, 2137-2147, doi:10.1242/jcs.00460 (2003).

119 Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J. Y., Yehia, G.,Tian, B. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nature methods* **10**, 133-139, doi:10.1038/nmeth.2288 (2013).

120 Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D.,Vogelauer, M. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Molecular cell* **32**, 685-695, doi:10.1016/j.molcel.2008.09.027 (2008).

121 Howe, F. S., Fischl, H., Murray, S. C.,Mellor, J. Is H3K4me3 instructive for transcription activation? *Bioessays* **39**, 1-12, doi:10.1002/bies.201600095 (2017).

122 Huber, F., Bunina, D., Gupta, I., Khmelinskii, A., Meurer, M., Theer, P., Steinmetz, L. M.,Knop, M. Protein Abundance Control by Non-coding Antisense Transcription. *Cell Rep* **15**, 2625-2636, doi:10.1016/j.celrep.2016.05.043 (2016).

123 Hurt, J. A., Robertson, A. D.,Burge, C. B. Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome research* **23**, 1636-1650, doi:10.1101/gr.157354.113 (2013).

124 Illingworth, R. S., Bird, A. P. CpG islands--'a rough guide'. *FEBS letters* **583**, 1713-1720, doi:10.1016/j.febslet.2009.04.012 (2009).

125 Inai, T., Yukawa, M.,Tsuchiya, E. Interplay between chromatin and trans-acting factors on the IME2 promoter upon induction of the gene at the onset of meiosis. *Mol Cell Biol* **27**, 1254-1263, doi:10.1128/MCB.01661-06 (2007).

126 Ingolia, N. T. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet* **15**, 205-213, doi:10.1038/nrg3645 (2014).

127 Ito, T., Miura, F.,Onda, M. Unexpected complexity of the budding yeast transcriptome. *IUBMB Life* **60**, 775-781, doi:10.1002/iub.121 (2008).

128 Jacquier, A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet* **10**, 833-844, doi:10.1038/nrg2683 (2009).

129 Jambhekar, A., Amon, A. Control of meiosis by respiration. *Curr Biol* **18**, 969-975, doi:10.1016/j.cub.2008.05.047 (2008).

130 Jona, G., Choder, M.,Gileadi, O. Glucose starvation induces a drastic reduction in the rates of both transcription and degradation of mRNA in yeast. *Biochimica et biophysica acta* **1491**, 37-48 (2000).
131 Joshi, A. A., Struhl, K. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Molecular cell* **20**, 971-978, doi:10.1016/j.molcel.2005.11.021 (2005).
132 Kane, S. M., Roth, R. Carbohydrate metabolism during ascospore development in yeast. *J Bacteriol* **118**, 8-14 (1974).
133 Kaplan, C. D., Laprade, L.,Winston, F. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301**, 1096-1099, doi:10.1126/science.1087374 (2003).
134 Kaplan, T., Liu, C. L., Erkmann, J. A., Holik, J., Grunstein, M., Kaufman, P. D., Friedman, N.,Rando, O. J. Cell cycle- and chaperone-mediated regulation of H3K56ac incorporation in yeast. *PLoS Genet* **4**, e1000270, doi:10.1371/journal.pgen.1000270 (2008).
135 Kassir, Y., Granot, D.,Simchen, G. IME1, a positive regulator gene of meiosis in S. cerevisiae. *Cell* **52**, 853-862 (1988).
136 Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S.,Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207, doi:10.1093/bioinformatics/btq351 (2010).
137 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., Boone, C. et al. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**, 593-605, doi:10.1016/j.cell.2005.10.025 (2005).
138 Kim Guisbert, K. S., Zhang, Y., Flatow, J., Hurtado, S., Staley, J. P., Lin, S.,Sontheimer, E. J. Meiosis-induced alterations in transcript architecture and noncoding RNA expression in S. cerevisiae. *RNA* **18**, 1142-1153, doi:10.1261/rna.030510.111 (2012).
139 Kim, H. J., Seol, J. H., Han, J. W., Youn, H. D.,Cho, E. J. Histone chaperones regulate histone exchange during transcription. *The EMBO journal* **26**, 4467-4474, doi:10.1038/sj.emboj.7601870 (2007).
140 Kim, J. H., Lee, B. B., Oh, Y. M., Zhu, C., Steinmetz, L. M., Lee, Y., Kim, W. K., Lee, S. B., Buratowski, S.,Kim, T. Modulation of mRNA and lncRNA expression dynamics by the Set2-Rpd3S pathway. *Nature communications* **7**, 13534, doi:10.1038/ncomms13534 (2016).
141 Kim, T., Buratowski, S. Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* **137**, 259-272, doi:10.1016/j.cell.2009.02.045 (2009).
142 Kim, T., Xu, Z., Clauder-Munster, S., Steinmetz, L. M.,Buratowski, S. Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. *Cell* **150**, 1158-1169, doi:10.1016/j.cell.2012.08.016 (2012).
143 Kirmizis, A., Santos-Rosa, H., Penkett, C. J., Singer, M. A., Vermeulen, M., Mann, M., Bahler, J., Green, R. D.,Kouzarides, T. Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* **449**, 928-932, doi:10.1038/nature06160 (2007).
144 Komarnitsky, P., Cho, E. J.,Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**, 2452-2460 (2000).
145 Koster, M. J., Yildirim, A. D., Weil, P. A., Holstege, F. C.,Timmers, H. T. Suppression of intragenic transcription requires the MOT1 and NC2 regulators of TATA-binding protein. *Nucleic Acids Res* **42**, 4220-4229, doi:10.1093/nar/gkt1398 (2014).

146 Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C. L., Pugh, B. F.,Korber, P. Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell* **167**, 709-721 e712, doi:10.1016/j.cell.2016.09.045 (2016).
147 Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S.,Greenblatt, J. F. RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Mol Cell Biol* **22**, 6979-6992 (2002).
148 Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D. P., Beattie, B. K., Emili, A., Boone, C., Shilatifard, A. et al. Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Mol Cell Biol* **23**, 4207-4218 (2003).
149 Kuehner, J. N., Pearson, E. L.,Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews. Molecular cell biology* **12**, 283-294, doi:10.1038/nrm3098 (2011).
150 Kwak, H., Fuda, N. J., Core, L. J.,Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950-953, doi:10.1126/science.1229386 (2013).
151 Lai, D. P., Tan, S., Kang, Y. N., Wu, J., Ooi, H. S., Chen, J., Shen, T. T., Qi, Y., Zhang, X., Guo, Y., Zhu, T. et al. Genome-wide profiling of polyadenylation sites reveals a link between selective polyadenylation and cancer metastasis. *Hum Mol Genet* **24**, 3410-3417, doi:10.1093/hmg/ddv089 (2015).
152 Lai, W. K. M., Pugh, B. F. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nature reviews. Molecular cell biology* **18**, 548-562, doi:10.1038/nrm.2017.47 (2017).
153 Lardenois, A., Liu, Y., Walther, T., Chalmel, F., Evrard, B., Granovskaia, M., Chu, A., Davis, R. W., Steinmetz, L. M.,Primig, M. 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, doi:10.1073/pnas.1016459108 (2011).
154 Lardenois, A., Stuparevic, I., Liu, Y., Law, M. J., Becker, E., Smagulova, F., Waern, K., Guilleux, M. H., Horecka, J., Chu, A., Kervarrec, C. et al. The conserved histone deacetylase Rpd3 and its DNA binding subunit Ume6 control dynamic transcript architecture during mitotic growth and meiotic development. *Nucleic Acids Res* **43**, 115-128, doi:10.1093/nar/gku1185 (2015).
155 Latos, P. A., Pauler, F. M., Koerner, M. V., Senergin, H. B., Hudson, Q. J., Stocsits, R. R., Allhoff, W., Stricker, S. H., Klement, R. M., Warczok, K. E., Aumayr, K. et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* **338**, 1469-1472, doi:10.1126/science.1228110 (2012).
156 Le Hir, H., Nott, A.,Moore, M. J. How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* **28**, 215-220, doi:10.1016/S0968-0004(03)00052-5 (2003).
157 Lemire, A., Lea, K., Batten, D., Jian Gu, S., Whitley, P., Bramlett, K.,Qu, L. Development of ERCC RNA Spike-In Control Mixes. *Journal of Biomolecular Techniques : JBT* **22**, S46-S46 (2011).
158 Lemon, B., Tjian, R. Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev* **14**, 2551-2569 (2000).
159 Lenstra, T. L., Benschop, J. J., Kim, T., Schulze, J. M., Brabers, N. A., Margaritis, T., van de Pasch, L. A., van Heesch, S. A., Brok, M. O., Groot Koerkamp, M. J., Ko, C. W. et al. The specificity and topology of chromatin interaction pathways in yeast. *Molecular cell* **42**, 536-549, doi:10.1016/j.molcel.2011.03.026 (2011).

160 Lenstra, T. L., Coulon, A., Chow, C. C.,Larson, D. R. Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. *Molecular cell* **60**, 597-610, doi:10.1016/j.molcel.2015.09.028 (2015).
161 Li, B., Gogol, M., Carey, M., Pattenden, S. G., Seidel, C.,Workman, J. L. Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. *Genes Dev* **21**, 1422-1430, doi:10.1101/gad.1539307 (2007).
162 Li, B., Jackson, J., Simon, M. D., Fleharty, B., Gogol, M., Seidel, C., Workman, J. L.,Shilatifard, A. Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. *J Biol Chem* **284**, 7970-7976, doi:10.1074/jbc.M808220200 (2009).
163 Lin, D., Hiron, T. K.,O'Callaghan, C. A. Intragenic transcriptional interference regulates the human immune ligand MICA. *The EMBO journal* **37**, doi:10.15252/embj.201797138 (2018).
164 Ling, J., Ainol, L., Zhang, L., Yu, X., Pi, W.,Tuan, D. HS2 enhancer function is blocked by a transcriptional terminator inserted between the enhancer and the promoter. *J Biol Chem* **279**, 51704-51713, doi:10.1074/jbc.M404039200 (2004).
165 Ling, J., Baibakov, B., Pi, W., Emerson, B. M.,Tuan, D. The HS2 enhancer of the beta-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a cis-linked globin promoter. *J Mol Biol* **350**, 883-896, doi:10.1016/j.jmb.2005.05.039 (2005).
166 Liu, Y., Beyer, A.,Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535-550, doi:10.1016/j.cell.2016.03.014 (2016).
167 Liu, Y., Stuparevic, I., Xie, B., Becker, E., Law, M. J.,Primig, M. The conserved histone deacetylase Rpd3 and the DNA binding regulator Ume6 repress BOI1's meiotic transcript isoform during vegetative growth in Saccharomyces cerevisiae. *Molecular microbiology* **96**, 861-874, doi:10.1111/mmi.12976 (2015).
168 Lo, H. C., Wan, L., Rosebrock, A., Futcher, B.,Hollingsworth, N. M. Cdc7-Dbf4 regulates NDT80 transcription as well as reductional segregation during budding yeast meiosis. *Molecular biology of the cell* **19**, 4956-4967, doi:10.1091/mbc.E08-07-0755 (2008).
169 Longtine, M. S., McKenzie, A., 3rd, Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P.,Pringle, J. R. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14**, 953-961, doi:10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U (1998).
170 Lopez-Maury, L., Marguerat, S.,Bahler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* **9**, 583-593, doi:10.1038/nrg2398 (2008).
171 Love, M. I., Huber, W.,Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
172 Loven, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., Levens, D. L., Lee, T. I.,Young, R. A. Revisiting global gene expression analysis. *Cell* **151**, 476-482, doi:10.1016/j.cell.2012.10.012 (2012).
173 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F.,Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).
174 Mahat, D. B., Kwak, H., Booth, G. T., Jonkers, I. H., Danko, C. G., Patel, R. K., Waters, C. T., Munson, K., Core, L. J.,Lis, J. T. Base-pair-resolution genome-

wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**, 1455-1476, doi:10.1038/nprot.2016.086 (2016).
175 Mai, B., Breeden, L. CLN1 and its repression by Xbp1 are important for efficient sporulation in budding yeast. *Mol Cell Biol* **20**, 478-487 (2000).
176 Malabat, C., Feuerbach, F., Ma, L., Saveanu, C.,Jacquier, A. Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *Elife* **4**, doi:10.7554/eLife.06722 (2015).
177 Marcu, K. B., Patel, A. J.,Yang, Y. Differential regulation of the c-MYC P1 and P2 promoters in the absence of functional tumor suppressors: implications for mechanisms of deregulated MYC transcription. *Curr Top Microbiol Immunol* **224**, 47-56 (1997).
178 Marinov, G. K. On the design and prospects of direct RNA sequencing. *Brief Funct Genomics* **16**, 326-335, doi:10.1093/bfgp/elw043 (2017).
179 Marquardt, S., Escalante-Chong, R., Pho, N., Wang, J., Churchman, L. S., Springer, M.,Buratowski, S. A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* **157**, 1712-1723, doi:10.1016/j.cell.2014.04.036 (2014).
180 Marston, A. L., Amon, A. Meiosis: cell-cycle controls shuffle and deal. *Nature reviews. Molecular cell biology* **5**, 983-997, doi:10.1038/nrm1526 (2004).
181 Martens, J. A., Laprade, L.,Winston, F. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature* **429**, 571-574, doi:10.1038/nature02538 (2004).
182 Martianov, I., Ramadass, A., Serra Barros, A., Chow, N.,Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666-670, doi:10.1038/nature05519 (2007).
183 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3, doi:10.14806/ej.17.1.200 (2011).
184 Mason, P. B., Struhl, K. The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Mol Cell Biol* **23**, 8323-8333 (2003).
185 Masters, J. N., Attardi, G. Discrete human dihydrofolate reductase gene transcripts present in polysomal RNA map with their 5' ends several hundred nucleotides upstream of the main mRNA start site. *Mol Cell Biol* **5**, 493-500 (1985).
186 Matsuura, A., Treinin, M., Mitsuzawa, H., Kassir, Y., Uno, I.,Simchen, G. The adenylate cyclase/protein kinase cascade regulates entry into meiosis in Saccharomyces cerevisiae through the gene IME1. *The EMBO journal* **9**, 3225-3232 (1990).
187 Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J. A.,Churchman, L. S. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541-554, doi:10.1016/j.cell.2015.03.010 (2015).
188 McCullough, L., Rawlins, R., Olsen, A., Xin, H., Stillman, D. J.,Formosa, T. Insight into the mechanism of nucleosome reorganization from histone mutants that suppress defects in the FACT histone chaperone. *Genetics* **188**, 835-846, doi:10.1534/genetics.111.128769 (2011).
189 McDaniel, S. L., Hepperla, A. J., Huang, J., Dronamraju, R., Adams, A. T., Kulkarni, V. G., Davis, I. J.,Strahl, B. D. H3K36 Methylation Regulates Nutrient Stress Response in Saccharomyces cerevisiae by Enforcing Transcriptional Fidelity. *Cell Rep* **19**, 2371-2382, doi:10.1016/j.celrep.2017.05.057 (2017).
190 Mellor, J., Woloszczuk, R.,Howe, F. S. The Interleaved Genome. *Trends Genet* **32**, 57-71, doi:10.1016/j.tig.2015.10.006 (2016).

191 Meyer, R. E., Chuong, H. H., Hild, M., Hansen, C. L., Kinter, M.,Dawson, D. S. Ipl1/Aurora-B is necessary for kinetochore restructuring in meiosis I in Saccharomyces cerevisiae. *Molecular biology of the cell* **26**, 2986-3000, doi:10.1091/mbc.E15-01-0032 (2015).
192 Miller, M. P., Unal, E., Brar, G. A.,Amon, A. Meiosis I chromosome segregation is established through regulation of microtubule-kinetochore interactions. *Elife* **1**, e00117, doi:10.7554/eLife.00117 (2012).
193 Mischo, H. E., Chun, Y., Harlen, K. M., Smalec, B. M., Dhir, S., Churchman, L. S.,Buratowski, S. Cell-Cycle Modulation of Transcription Termination Factor Sen1. *Molecular cell* **70**, 312-326 e317, doi:10.1016/j.molcel.2018.03.010 (2018).
194 Mitchell, A. P., Bowdish, K. S. Selection for early meiotic mutants in yeast. *Genetics* **131**, 65-72 (1992).
195 Mizuno, T., Nakazawa, N., Remgsamrarn, P., Kunoh, T., Oshima, Y.,Harashima, S. The Tup1-Ssn6 general repressor is involved in repression of IME1 encoding a transcriptional activator of meiosis in Saccharomyces cerevisiae. *Current genetics* **33**, 239-247 (1998).
196 Mizunuma, M., Tsubakiyama, R., Ogawa, T., Shitamukai, A., Kobayashi, Y., Inai, T., Kume, K.,Hirata, D. Ras/cAMP-dependent protein kinase (PKA) regulates multiple aspects of cellular events by phosphorylating the Whi3 cell cycle regulator in budding yeast. *J Biol Chem* **288**, 10558-10566, doi:10.1074/jbc.M112.402214 (2013).
197 Moreno-Torres, M., Jaquenoud, M.,De Virgilio, C. TORC1 controls G1-S cell cycle transition in yeast via Mpk1 and the greatwall kinase pathway. *Nature communications* **6**, 8256, doi:10.1038/ncomms9256 (2015).
198 Moretto, F., Wood, N. E., Kelly, G., Doncic, A.,van Werven, F. J. A regulatory circuit of two lncRNAs and a master regulator directs cell fate in yeast. *Nature communications* **9**, 780, doi:10.1038/s41467-018-03213-z (2018).
199 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L.,Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).
200 Murray, S. C., Haenni, S., Howe, F. S., Fischl, H., Chocian, K., Nair, A.,Mellor, J. Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Res* **43**, 7823-7837, doi:10.1093/nar/gkv666 (2015).
201 Murray, S. C., Mellor, J. Using both strands: The fundamental nature of antisense transcription. *Bioarchitecture* **6**, 12-21, doi:10.1080/19490992.2015.1130779 (2016).
202 Myers, C. N., Berner, G. B., Holthoff, J. H., Martinez-Fonts, K., Harper, J. A., Alford, S., Taylor, M. N.,Duina, A. A. Mutant versions of the S. cerevisiae transcription elongation factor Spt16 define regions of Spt16 that functionally interact with histone H3. *PloS one* **6**, e20847, doi:10.1371/journal.pone.0020847 (2011).
203 Nachman, I., Regev, A.,Ramanathan, S. Dissecting timing variability in yeast meiosis. *Cell* **131**, 544-556, doi:10.1016/j.cell.2007.09.044 (2007).
204 Nadal-Ribelles, M., Mas, G., Millan-Zambrano, G., Sole, C., Ammerer, G., Chavez, S., Posas, F.,de Nadal, E. H3K4 monomethylation dictates nucleosome dynamics and chromatin remodeling at stress-responsive genes. *Nucleic Acids Res* **43**, 4937-4949, doi:10.1093/nar/gkv220 (2015).
205 Nadal-Ribelles, M., Sole, C., Xu, Z., Steinmetz, L. M., de Nadal, E.,Posas, F. Control of Cdc28 CDK1 by a stress-induced lncRNA. *Molecular cell* **53**, 549-561, doi:10.1016/j.molcel.2014.01.006 (2014).

206 Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M.,Snyder, M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344-1349, doi:10.1126/science.1158441 (2008).
207 Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L. M.,Jacquier, A. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**, 1038-1042, doi:10.1038/nature07747 (2009).
208 Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F.,Oliviero, S. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72-77, doi:10.1038/nature21373 (2017).
209 Ng, H. H., Robert, F., Young, R. A.,Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell* **11**, 709-719 (2003).
210 Ng, P., Wei, C. L., Sung, W. K., Chiu, K. P., Lipovich, L., Ang, C. C., Gupta, S., Shahab, A., Ridwan, A., Wong, C. H., Liu, E. T. et al. Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nature methods* **2**, 105-111, doi:10.1038/nmeth733 (2005).
211 Nguyen, T., Fischl, H., Howe, F. S., Woloszczuk, R., Serra Barros, A., Xu, Z., Brown, D., Murray, S. C., Haenni, S., Halstead, J. M., O'Connor, L. et al. Transcription mediated insulation and interference direct gene cluster expression switches. *Elife* **3**, e03635, doi:10.7554/eLife.03635 (2014).
212 Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T.,Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature methods* **6**, 917-922, doi:10.1038/nmeth.1401 (2009).
213 Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., Carmo-Fonseca, M.,Proudfoot, N. J. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526-540, doi:10.1016/j.cell.2015.03.027 (2015).
214 O'Sullivan, J. M., Tan-Wong, S. M., Morillon, A., Lee, B., Coles, J., Mellor, J.,Proudfoot, N. J. Gene loops juxtapose promoters and terminators in yeast. *Nature genetics* **36**, 1014-1018, doi:10.1038/ng1411 (2004).
215 Ohkura, H. Meiosis: an overview of key differences from mitosis. *Cold Spring Harb Perspect Biol* **7**, doi:10.1101/cshperspect.a015859 (2015).
216 Orphanides, G., LeRoy, G., Chang, C. H., Luse, D. S.,Reinberg, D. FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92**, 105-116 (1998).
217 Ottoz, D. S., Rudolf, F.,Stelling, J. Inducible, tightly regulated and growth condition-independent transcription factor in Saccharomyces cerevisiae. *Nucleic Acids Res* **42**, e130, doi:10.1093/nar/gku616 (2014).
218 Pak, J., Segall, J. Regulation of the premiddle and middle phases of expression of the NDT80 gene during sporulation of Saccharomyces cerevisiae. *Mol Cell Biol* **22**, 6417-6429 (2002a).
219 Pak, J., Segall, J. Role of Ndt80, Sum1, and Swe1 as targets of the meiotic recombination checkpoint that control exit from pachytene and spore formation in Saccharomyces cerevisiae. *Mol Cell Biol* **22**, 6430-6440 (2002b).
220 Pande, A., Brosius, J., Makalowska, I., Makalowski, W.,Raabe, C. A. Transcriptional interference by small transcripts in proximal promoter regions. *Nucleic Acids Res* **46**, 1069-1088, doi:10.1093/nar/gkx1242 (2018).
221 Park, D., Morris, A. R., Battenhouse, A.,Iyer, V. R. Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Res* **42**, 3736-3749, doi:10.1093/nar/gkt1366 (2014).

222 Parker, S., Fraczek, M. G., Wu, J., Shamsah, S., Manousaki, A., Dungrattanalert, K., de Almeida, R. A., Estrada-Rivadeneyra, D., Omara, W., Delneri, D.,O'Keefe, R. T. A resource for functional profiling of noncoding RNA in the yeast Saccharomyces cerevisiae. *RNA* **23**, 1166-1171, doi:10.1261/rna.061564.117 (2017).

223 Parker, S., Fraczek, M. G., Wu, J., Shamsah, S., Manousaki, A., Dungrattanalert, K., de Almeida, R. A., Invernizzi, E., Burgis, T., Omara, W., Griffiths-Jones, S. et al. Large-scale profiling of noncoding RNA function in yeast. *PLoS Genet* **14**, e1007253, doi:10.1371/journal.pgen.1007253 (2018).

224 Pedruzzi, I., Dubouloz, F., Cameroni, E., Wanke, V., Roosen, J., Winderickx, J.,De Virgilio, C. TOR and PKA signaling pathways converge on the protein kinase Rim15 to control entry into G0. *Molecular cell* **12**, 1607-1613 (2003).

225 Pelechano, V., Wei, W., Jakob, P.,Steinmetz, L. M. Genome-wide identification of transcript start and end sites by transcript isoform sequencing. *Nat Protoc* **9**, 1740-1759, doi:10.1038/nprot.2014.121 (2014).

226 Pelechano, V., Wei, W.,Steinmetz, L. M. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**, 127-131, doi:10.1038/nature12121 (2013).

227 Pierce, M., Benjamin, K. R., Montano, S. P., Georgiadis, M. M., Winter, E.,Vershon, A. K. Sum1 and Ndt80 proteins compete for binding to middle sporulation element sequences that control meiotic gene expression. *Mol Cell Biol* **23**, 4814-4825 (2003).

228 Pijnappel, W. W., Schaft, D., Roguev, A., Shevchenko, A., Tekotte, H., Wilm, M., Rigaut, G., Seraphin, B., Aasland, R.,Stewart, A. F. The S. cerevisiae SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. *Genes Dev* **15**, 2991-3004, doi:10.1101/gad.207401 (2001).

229 Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., Zeitlinger, J. et al. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517-527, doi:10.1016/j.cell.2005.06.026 (2005).

230 Prescott, E. M., Proudfoot, N. J. Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A* **99**, 8796-8801, doi:10.1073/pnas.132270899 (2002).

231 Primig, M., Williams, R. M., Winzeler, E. A., Tevzadze, G. G., Conway, A. R., Hwang, S. Y., Davis, R. W.,Esposito, R. E. The core meiotic transcriptome in budding yeasts. *Nature genetics* **26**, 415-423, doi:10.1038/82539 (2000).

232 Proffitt, J. H., Davie, J. R., Swinton, D.,Hattman, S. 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. *Mol Cell Biol* **4**, 985-988 (1984).

233 Qing, T., Yu, Y., Du, T.,Shi, L. mRNA enrichment protocols determine the quantification characteristics of external RNA spike-in controls in RNA-Seq studies. *Sci China Life Sci* **56**, 134-142, doi:10.1007/s11427-013-4437-9 (2013).

234 Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P.,Gu, Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *Bmc Genomics* **13**, 341, doi:10.1186/1471-2164-13-341 (2012).

235 Quinlan, A. R., Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

236 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A.,Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191, doi:10.1093/nar/gku365 (2014).

237 Rando, O. J. Genome-wide mapping of nucleosomes in yeast. *Methods in enzymology* **470**, 105-118, doi:10.1016/S0076-6879(10)70005-7 (2010).
238 Rando, O. J. Genome-wide measurement of histone H3 replacement dynamics in yeast. *Methods Mol Biol* **759**, 41-60, doi:10.1007/978-1-61779-173-4_3 (2011).
239 Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A.,Segal, E. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nature genetics* **44**, 743-750, doi:10.1038/ng.2305 (2012).
240 Ray, D., Ye, P. Characterization of the metabolic requirements in yeast meiosis. *PloS one* **8**, e63707, doi:10.1371/journal.pone.0063707 (2013).
241 Rincon-Arano, H., Halow, J., Delrow, J. J., Parkhurst, S. M.,Groudine, M. UpSET recruits HDAC complexes and restricts chromatin accessibility and acetylation at promoter regions. *Cell* **151**, 1214-1228, doi:10.1016/j.cell.2012.11.009 (2012).
242 Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E.,Chang, H. Y. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323, doi:10.1016/j.cell.2007.05.022 (2007).
243 Risso, D., Ngai, J., Speed, T. P.,Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature biotechnology* **32**, 896-902, doi:10.1038/nbt.2931 (2014).
244 Roberts, S. M., Winston, F. Essential functional interactions of SAGA, a Saccharomyces cerevisiae complex of Spt, Ada, and Gcn5 proteins, with the Snf/Swi and Srb/mediator complexes. *Genetics* **147**, 451-465 (1997).
245 Robinson, P. J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R. G.,Rhodes, D. 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J Mol Biol* **381**, 816-825, doi:10.1016/j.jmb.2008.04.050 (2008).
246 Roca, C. P., Gomes, S. I., Amorim, M. J.,Scott-Fordsmand, J. J. Variation-preserving normalization unveils blind spots in gene expression profiling. *Scientific reports* **7**, 42460, doi:10.1038/srep42460 (2017).
247 Roy, D., Zhang, Z., Lu, Z., Hsieh, C. L.,Lieber, M. R. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol Cell Biol* **30**, 146-159, doi:10.1128/MCB.00897-09 (2010).
248 Rubin-Bejerano, I., Mandel, S., Robzyk, K.,Kassir, Y. Induction of meiosis in Saccharomyces cerevisiae depends on conversion of the transcriptional represssor Ume6 to a positive regulator by its regulated association with the transcriptional activator Ime1. *Mol Cell Biol* **16**, 2518-2526 (1996).
249 Rufiange, A., Jacques, P. E., Bhat, W., Robert, F.,Nourani, A. Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *Molecular cell* **27**, 393-405, doi:10.1016/j.molcel.2007.07.011 (2007).
250 Salama, S. R., Hendricks, K. B.,Thorner, J. G1 cyclin degradation: the PEST motif of yeast Cln2 is necessary, but not sufficient, for rapid protein turnover. *Mol Cell Biol* **14**, 7953-7966 (1994).
251 Samarsky, D. A., Ferbeyre, G., Bertrand, E., Singer, R. H., Cedergren, R.,Fournier, M. J. A small nucleolar RNA:ribozyme hybrid cleaves a nucleolar RNA target in vivo with near-perfect efficiency. *Proc Natl Acad Sci U S A* **96**, 6609-6614 (1999).

252 Santangelo, G. M. Glucose signaling in Saccharomyces cerevisiae. *Microbiology and molecular biology reviews : MMBR* **70**, 253-282, doi:10.1128/MMBR.70.1.253-282.2006 (2006).
253 Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J.,Kouzarides, T. Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-411, doi:10.1038/nature01080 (2002).
254 Sanz, L. A., Hartono, S. R., Lim, Y. W., Steyaert, S., Rajpurkar, A., Ginno, P. A., Xu, X.,Chedin, F. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Molecular cell* **63**, 167-178, doi:10.1016/j.molcel.2016.05.032 (2016).
255 Sarkar, S., Dalgaard, J. Z., Millar, J. B.,Arumugam, P. The Rim15-endosulfine-PP2ACdc55 signalling module regulates entry into gametogenesis and quiescence via distinct mechanisms in budding yeast. *PLoS Genet* **10**, e1004456, doi:10.1371/journal.pgen.1004456 (2014).
256 Sayou, C., Millan-Zambrano, G., Santos-Rosa, H., Petfalski, E., Robson, S., Houseley, J., Kouzarides, T.,Tollervey, D. RNA Binding by Histone Methyltransferases Set1 and Set2. *Mol Cell Biol* **37**, doi:10.1128/MCB.00165-17 (2017).
257 Schindler, K., Benjamin, K. R., Martin, A., Boglioli, A., Herskowitz, I.,Winter, E. The Cdk-activating kinase Cak1p promotes meiotic S phase through Ime2p. *Mol Cell Biol* **23**, 8718-8728 (2003).
258 Schmitt, S., Prestel, M.,Paro, R. Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev* **19**, 697-708, doi:10.1101/gad.326205 (2005).
259 Schneider, C. A., Rasband, W. S.,Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**, 671-675 (2012).
260 Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J.,Cramer, P. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* **155**, 1075-1087, doi:10.1016/j.cell.2013.10.024 (2013).
261 Schwabish, M. A., Struhl, K. Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol Cell Biol* **24**, 10111-10117, doi:10.1128/MCB.24.23.10111-10117.2004 (2004).
262 Schwartz, S., Agarwala, S. D., Mumbach, M. R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T. S., Satija, R., Ruvkun, G., Carr, S. A. et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**, 1409-1421, doi:10.1016/j.cell.2013.10.047 (2013).
263 Sein, H., Varv, S.,Kristjuhan, A. Distribution and maintenance of histone H3 lysine 36 trimethylation in transcribed locus. *PloS one* **10**, e0120200, doi:10.1371/journal.pone.0120200 (2015).
264 Shah, J. C., Clancy, M. J. IME4, a gene that mediates MAT and nutritional control of meiosis in Saccharomyces cerevisiae. *Mol Cell Biol* **12**, 1078-1086 (1992).
265 Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A.,Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843-852 (2000).
266 Shearwin, K. E., Callen, B. P.,Egan, J. B. Transcriptional interference--a crash course. *Trends Genet* **21**, 339-345, doi:10.1016/j.tig.2005.04.009 (2005).
267 Shepard, P. J., Choi, E. A., Lu, J., Flanagan, L. A., Hertel, K. J.,Shi, Y. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **17**, 761-772, doi:10.1261/rna.2581711 (2011).

268 Sheth, U., Parker, R. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **300**, 805-808, doi:10.1126/science.1082320 (2003).
269 Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem* **81**, 65-95, doi:10.1146/annurev-biochem-051710-134100 (2012).
270 Shin, M. E., Skokotas, A.,Winter, E. The Cdk1 and Ime2 protein kinases trigger exit from meiotic prophase in Saccharomyces cerevisiae by inhibiting the Sum1 transcriptional repressor. *Mol Cell Biol* **30**, 2996-3003, doi:10.1128/MCB.01682-09 (2010).
271 Smith, H. E., Su, S. S., Neigeborn, L., Driscoll, S. E.,Mitchell, A. P. Role of IME1 expression in regulation of meiosis in Saccharomyces cerevisiae. *Mol Cell Biol* **10**, 6103-6113 (1990).
272 Smolle, M., Venkatesh, S., Gogol, M. M., Li, H., Zhang, Y., Florens, L., Washburn, M. P.,Workman, J. L. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol* **19**, 884-892, doi:10.1038/nsmb.2312 (2012).
273 Snijder, B., Pelkmans, L. Origins of regulated cell-to-cell variability. *Nature reviews. Molecular cell biology* **12**, 119-125, doi:10.1038/nrm3044 (2011).
274 Soares, L. M., He, P. C., Chun, Y., Suh, H., Kim, T.,Buratowski, S. Determinants of Histone H3K4 Methylation Patterns. *Molecular cell* **68**, 773-785 e776, doi:10.1016/j.molcel.2017.10.013 (2017).
275 Sommermeyer, V., Beneut, C., Chaplais, E., Serrentino, M. E.,Borde, V. Spp1, a member of the Set1 Complex, promotes meiotic DSB formation in promoters by tethering histone H3K4 methylation sites to chromosome axes. *Molecular cell* **49**, 43-54, doi:10.1016/j.molcel.2012.11.008 (2013).
276 Sopko, R., Raithatha, S.,Stuart, D. Phosphorylation and maximal activity of Saccharomyces cerevisiae meiosis-specific transcription factor Ndt80 is dependent on Ime2. *Mol Cell Biol* **22**, 7024-7040 (2002).
277 Stiller, J. W., Cook, M. S. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryotic cell* **3**, 735-740, doi:10.1128/EC.3.3.735-740.2004 (2004).
278 Stossi, F., Madak-Erdogan, Z.,Katzenellenbogen, B. S. Estrogen receptor alpha represses transcription of early target genes via p300 and CtBP1. *Mol Cell Biol* **29**, 1749-1759, doi:10.1128/MCB.01476-08 (2009).
279 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.,Allis, C. D. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol Cell Biol* **22**, 1298-1306 (2002).
280 Strich, R., Khakhina, S.,Mallory, M. J. Ume6p is required for germination and early colony development of yeast ascospores. *FEMS Yeast Res* **11**, 104-113, doi:10.1111/j.1567-1364.2010.00696.x (2011).
281 Strich, R., Surosky, R. T., Steber, C., Dubois, E., Messenguy, F.,Esposito, R. E. UME6 is a key regulator of nitrogen repression and meiotic development. *Genes Dev* **8**, 796-810 (1994).
282 Struhl, K., Segal, E. Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**, 267-273, doi:10.1038/nsmb.2506 (2013).
283 Stuart, D. The meiotic differentiation program uncouples S-phase from cell size control in Saccharomyces cerevisiae. *Cell cycle* **7**, 777-786, doi:10.4161/cc.7.6.5562 (2008).
284 Stuart, D. T. Selection of G1 Phase Yeast Cells for Synchronous Meiosis and Sporulation. *Methods Mol Biol* **1471**, 123-132, doi:10.1007/978-1-4939-6340-9_5 (2017).

285 Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H.,Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**, 66-71, doi:10.1038/nature13007 (2014).
286 Sun, M., Lariviere, L., Dengl, S., Mayer, A.,Cramer, P. A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). *J Biol Chem* **285**, 41597-41603, doi:10.1074/jbc.M110.144568 (2010).
287 Suzuki, C., Hori, Y.,Kashiwagi, Y. Screening and characterization of transposon-insertion mutants in a pseudohyphal strain of Saccharomyces cerevisiae. *Yeast* **20**, 407-415, doi:10.1002/yea.970 (2003).
288 Takahashi, H., Kato, S., Murata, M.,Carninci, P. CAGE (cap analysis of gene expression): a protocol for the detection of promoter and transcriptional networks. *Methods Mol Biol* **786**, 181-200, doi:10.1007/978-1-61779-292-2_11 (2012).
289 Takemata, N., Oda, A., Yamada, T., Galipon, J., Miyoshi, T., Suzuki, Y., Sugano, S., Hoffman, C. S., Hirota, K.,Ohta, K. Local potentiation of stress-responsive genes by upstream noncoding transcription. *Nucleic Acids Res* **44**, 5174-5189, doi:10.1093/nar/gkw142 (2016).
290 Tan-Wong, S. M., Zaugg, J. B., Camblong, J., Xu, Z., Zhang, D. W., Mischo, H. E., Ansari, A. Z., Luscombe, N. M., Steinmetz, L. M.,Proudfoot, N. J. Gene loops enhance transcriptional directionality. *Science* **338**, 671-675, doi:10.1126/science.1224350 (2012).
291 Teves, S. S., Weber, C. M.,Henikoff, S. Transcribing through the nucleosome. *Trends Biochem Sci* **39**, 577-586, doi:10.1016/j.tibs.2014.10.004 (2014).
292 The Gene Ontology, C. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res* **45**, D331-D338, doi:10.1093/nar/gkw1108 (2017).
293 Thebault, P., Boutin, G., Bhat, W., Rufiange, A., Martens, J.,Nourani, A. 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, doi:10.1128/MCB.01083-10 (2011).
294 Tresenrider, A., Unal, E. One-two punch mechanism of gene repression: a fresh perspective on gene regulation. *Current genetics* **64**, 581-588, doi:10.1007/s00294-017-0793-5 (2018).
295 Tuan, D., Kong, S.,Hu, K. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A* **89**, 11219-11223 (1992).
296 Tufarelli, C., Stanley, J. A., Garrick, D., Sharpe, J. A., Ayyub, H., Wood, W. G.,Higgs, D. R. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nature genetics* **34**, 157-165, doi:10.1038/ng1157 (2003).
297 Uhler, J. P., Hertel, C.,Svejstrup, J. Q. A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc Natl Acad Sci U S A* **104**, 8011-8016, doi:10.1073/pnas.0702431104 (2007).
298 Ushijima, T., Hanada, K., Gotoh, E., Yamori, W., Kodama, Y., Tanaka, H., Kusano, M., Fukushima, A., Tokizawa, M., Yamamoto, Y. Y., Tada, Y. et al. Light Controls Protein Localization through Phytochrome-Mediated Alternative Promoter Selection. *Cell* **171**, 1316-1325 e1312, doi:10.1016/j.cell.2017.10.018 (2017).
299 Van Dalfsen, K. M., Hodapp, S., Keskin, A., Otto, G. M., Berdan, C. A., Higdon, A., Cheunkarndee, T., Nomura, D. K., Jovanovic, M.,Brar, G. A. Global Proteome Remodeling during ER Stress Involves Hac1-Driven Expression of Long Undecoded Transcript Isoforms. *Developmental cell* **46**, 219-235 e218, doi:10.1016/j.devcel.2018.06.016 (2018).

300 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., Nicolas, A. et al. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475**, 114-117, doi:10.1038/nature10118 (2011).
301 van Nuland, R., Schram, A. W., van Schaik, F. M., Jansen, P. W., Vermeulen, M.,Marc Timmers, H. T. Multivalent engagement of TFIID to nucleosomes. *PloS one* **8**, e73495, doi:10.1371/journal.pone.0073495 (2013).
302 van Werven, F. J., Amon, A. Regulation of entry into gametogenesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **366**, 3521-3531, doi:10.1098/rstb.2011.0081 (2011).
303 van Werven, F. J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., Primig, M.,Amon, A. Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* **150**, 1170-1181, doi:10.1016/j.cell.2012.06.049 (2012).
304 Vasiljeva, L., Buratowski, S. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Molecular cell* **21**, 239-248, doi:10.1016/j.molcel.2005.11.028 (2006).
305 Venkatesh, S., Li, H., Gogol, M. M.,Workman, J. L. Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. *Nature communications* **7**, 13610, doi:10.1038/ncomms13610 (2016).
306 Venkatesh, S., Smolle, M., Li, H., Gogol, M. M., Saint, M., Kumar, S., Natarajan, K.,Workman, J. L. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **489**, 452-455, doi:10.1038/nature11326 (2012).
307 Venkatesh, S., Workman, J. L. Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development. *Wiley Interdiscip Rev Dev Biol* **2**, 685-700, doi:10.1002/wdev.109 (2013).
308 Venkatesh, S., Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nature reviews. Molecular cell biology* **16**, 178-189, doi:10.1038/nrm3941 (2015).
309 Vidan, S., Mitchell, A. P. Stimulation of yeast meiotic gene expression by the glucose-repressible protein kinase Rim15p. *Mol Cell Biol* **17**, 2688-2697 (1997).
310 Wagner, E. J., Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nature reviews. Molecular cell biology* **13**, 115-126, doi:10.1038/nrm3274 (2012).
311 Walther, T., Letisse, F., Peyriga, L., Alkim, C., Liu, Y., Lardenois, A., Martin-Yken, H., Portais, J. C., Primig, M.,Francois, J. Developmental stage dependent metabolic regulation during meiotic differentiation in budding yeast. *BMC biology* **12**, 60, doi:10.1186/s12915-014-0060-x (2014).
312 Wan, L., Niu, H., Futcher, B., Zhang, C., Shokat, K. M., Boulton, S. J.,Hollingsworth, N. M. Cdc28-Clb5 (CDK-S) and Cdc7-Dbf4 (DDK) collaborate to initiate meiotic recombination in yeast. *Genes Dev* **22**, 386-397, doi:10.1101/gad.1626408 (2008).
313 Wan, L., Zhang, C., Shokat, K. M.,Hollingsworth, N. M. Chemical inactivation of cdc7 kinase in budding yeast results in a reversible arrest that allows efficient cell synchronization prior to meiotic recombination. *Genetics* **174**, 1767-1774, doi:10.1534/genetics.106.064303 (2006).
314 Wang, A., Kurdistani, S. K.,Grunstein, M. Requirement of Hos2 histone deacetylase for gene activity in yeast. *Science* **298**, 1412-1414, doi:10.1126/science.1077790 (2002).
315 Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P.,Burge, C. B. Alternative isoform regulation in

human tissue transcriptomes. *Nature* **456**, 470-476, doi:10.1038/nature07509 (2008).

316 Wapinski, I., Pfeffer, A., Friedman, N.,Regev, A. Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**, 54-61, doi:10.1038/nature06107 (2007).

317 Watanabe, Y., Yamamoto, M. S. pombe mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell* **78**, 487-498 (1994).

318 Weber, C. M., Ramachandran, S.,Henikoff, S. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Molecular cell* **53**, 819-830, doi:10.1016/j.molcel.2014.02.014 (2014).

319 Weidberg, H., Moretto, F., Spedale, G., Amon, A.,van Werven, F. J. Nutrient Control of Yeast Gametogenesis Is Mediated by TORC1, PKA and Energy Availability. *PLoS Genet* **12**, e1006075, doi:10.1371/journal.pgen.1006075 (2016).

320 Wilkening, S., Pelechano, V.,Steinmetz, L. M. Genome-Wide Identification of Alternative Polyadenylation Events Using 3'T-Fill. *Methods Mol Biol* **1358**, 295-302, doi:10.1007/978-1-4939-3067-8_18 (2016).

321 Williams, R. M., Primig, M., Washburn, B. K., Winzeler, E. A., Bellis, M., Sarrauste de Menthiere, C., Davis, R. W.,Esposito, R. E. The Ume6 regulon coordinates metabolic and meiotic gene expression in yeast. *Proc Natl Acad Sci U S A* **99**, 13431-13436, doi:10.1073/pnas.202495299 (2002).

322 Williams, S. K., Truong, D.,Tyler, J. K. Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc Natl Acad Sci U S A* **105**, 9000-9005, doi:10.1073/pnas.0800057105 (2008).

323 Winter, E. The Sum1/Ndt80 transcriptional switch and commitment to meiosis in Saccharomyces cerevisiae. *Microbiology and molecular biology reviews : MMBR* **76**, 1-15, doi:10.1128/MMBR.05010-11 (2012).

324 Woo, Y. H., Li, W. H. Gene clustering pattern, promoter architecture, and gene expression stability in eukaryotic genomes. *Proc Natl Acad Sci U S A* **108**, 3306-3311, doi:10.1073/pnas.1100210108 (2011).

325 Wyers, F., Rougemaille, M., Badis, G., Rousselle, J. C., Dufour, M. E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., Libri, D. et al. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725-737, doi:10.1016/j.cell.2005.04.030 (2005).

326 Xie, B., Horecka, J., Chu, A., Davis, R. W., Becker, E.,Primig, M. Ndt80 activates the meiotic ORC1 transcript isoform and SMA2 via a bi-directional middle sporulation element in Saccharomyces cerevisiae. *RNA biology* **13**, 772-782, doi:10.1080/15476286.2016.1191738 (2016).

327 Xu, L., Ajimura, M., Padmore, R., Klein, C.,Kleckner, N. NDT80, a meiosis-specific gene required for exit from pachytene in Saccharomyces cerevisiae. *Mol Cell Biol* **15**, 6572-6581 (1995).

328 Xu, Z., Wei, W., Gagneur, J., Clauder-Munster, S., Smolik, M., Huber, W.,Steinmetz, L. M. Antisense expression increases gene expression variability and locus interdependency. *Mol Syst Biol* **7**, 468, doi:10.1038/msb.2011.1 (2011).

329 Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W.,Steinmetz, L. M. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033-1037, doi:10.1038/nature07728 (2009).

330 Yamashita, A., Watanabe, Y., Nukina, N.,Yamamoto, M. RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast. *Cell* **95**, 115-123 (1998).
331 Yang, Y., McBride, K. M., Hensley, S., Lu, Y., Chedin, F.,Bedford, M. T. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Molecular cell* **53**, 484-497, doi:10.1016/j.molcel.2014.01.011 (2014).
332 Yassour, M., Pfiffner, J., Levin, J. Z., Adiconis, X., Gnirke, A., Nusbaum, C., Thompson, D. A., Friedman, N.,Regev, A. Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome biology* **11**, R87, doi:10.1186/gb-2010-11-8-r87 (2010).
333 Yoh, S. M., Cho, H., Pickle, L., Evans, R. M.,Jones, K. A. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev* **21**, 160-174, doi:10.1101/gad.1503107 (2007).
334 Yu, Y., Neiman, A. M.,Sternglanz, R. The JmjC domain of Gis1 is dispensable for transcriptional activation. *FEMS Yeast Res* **10**, 793-801, doi:10.1111/j.1567-1364.2010.00680.x (2010).
335 Zaugg, J. B., Luscombe, N. M. A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast. *Genome research* **22**, 84-94, doi:10.1101/gr.124099.111 (2012).
336 Zentner, G. E., Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* **20**, 259-266, doi:10.1038/nsmb.2470 (2013).
337 Zhang, P., Dimont, E., Ha, T., Swanson, D. J., Consortium, F., Hide, W.,Goldowitz, D. Relatively frequent switching of transcription start sites during cerebellar development. *Bmc Genomics* **18**, 461, doi:10.1186/s12864-017-3834-z (2017).
338 Zheng, D., Tian, B. Sizing up the poly(A) tail: insights from deep sequencing. *Trends Biochem Sci* **39**, 255-257, doi:10.1016/j.tibs.2014.04.002 (2014).
339 Zheng, X. F., Schreiber, S. L. Target of rapamycin proteins and their kinase activities are required for meiosis. *Proc Natl Acad Sci U S A* **94**, 3070-3075 (1997).
340 Zhou, S., Sternglanz, R.,Neiman, A. M. Developmentally regulated internal transcription initiation during meiosis in budding yeast. *PloS one* **12**, e0188001, doi:10.1371/journal.pone.0188001 (2017).