Complex Transcription Units in *Saccharomyces cerevisiae*

Doctorate of Philosophy, University of Oxford



Tania Nguyen

Keble College, Trinity Term 2013

Approximate word count: 47, 000

Complex Transcription Units in *Saccharomyces cerevisiae*

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

Tania Nguyen, Keble College, Trinity Term 2013.

Abstract

Overlapping antisense transcription, once thought to be an exceptional event, has become established as a common feature of transcription units. However, despite its prevalence, a bona fide biological consequence or function of overlapping transcription has yet to be ascribed, with the exception of a few locus-specific examples. In this analysis a multi-dimensional systematic approach was employed to interrogate the antisense-producing *HMS2* locus. As a result, two additional layers of transcription unit complexity were described: temporal separation of sense and antisense transcription states and transcript-end heterogeneity. Additionally, by observing transcription state-switching in near real time, evidence was provided that links the two phenomena in which changes in the dominant antisense isoform correlate with sense/antisense state switching. Furthermore, in the case of *HMS2* regulation, a gene displaying expression periodicity in the metabolic cycle, it was demonstrated that the adoption of sense or antisense states bears consequences at neighbouring loci, showing that modulating antisense transcription can coordinate expression of a cluster of genes. Taken together, this study supports a general model in which sense/antisense state switching at a single locus, modulated by transcriptional regulation of both 5' and 3'-end promoters, as well as adopting alternative transcript isoforms, can have long-range effects at neighbouring genes, ultimately providing cells with the ability to tailor a wider transcriptional landscape, so that they may adjust their genetic responses to their specific environmental or physiological needs.

Acknowledgements

First and foremost I would like to thank my supervisor, Jane Mellor, for taking a risk back in 2009 and allowing me the opportunity to join her lab. I came to Oxford very inexperienced, but with her constant support and guidance I have learned a great deal and look forward to furthering my career in molecular biology research. I would also like to acknowledge the Clarendon Fund, Keble College, and the Natural Sciences and Engineering Council of Canada for their financial support during my studies, without which this work would not have been possible. Next, I would like to thank the members of the Mellor Lab, in particular Anitha Nair for contributing the *URA3 xrn1* Δ strain and for providing overall technical support, as well as Struan Murray for helping with the statistical analysis of the RNA FISH data. I would also like to thank Simon Haenni for the *GAL1* RNA FISH samples (Figure 48), Ana Serra Barros, and Mike Youdell for helping me get started in the lab. Additionally, I would like to express my tremendous gratitude to Françoise Howe, thank you for reading through this work and for your friendship throughout this entire experience. Furthermore, I would like to acknowledge our collaborators from the Lars Steinmetz group at EMBL.

Secondly, to all the friends I've encountered along the way, especially Jayati Jain, Photeini Dimirouli, Ruchi Patel, Sameer Sengupta, the rest of the Keble lot, and the Oxford University Women's Rugby Football Club, though some may argue otherwise, thank you for keeping me sane during these last four years. In between the messy northern blots, failed 3'RACE attempts, and tricky microscopy, I will remember the late night conversations about existential and economic crises, the hysterical laughter, and the joy you have brought to my life.

Last, but definitely not least, I would like to thank my family. Phil, you've always let me talk your ear off about anything and everything. Hermie, thanks for holding down the fort while I was away. But most of all, Mom, I couldn't have done this without you. You always know just what to say. Thank you so much for your love and support. Finally, I would like to make special mention of my father and "little brother" Javi. I keep the two of you close to my heart everyday and are the inspiration for all I do.

Abbreviations

Abbreviation	Term in full
ac	acetylation
AS	antisense
ATP	adenosine triphosphate
bp	base-pair
BSA	bovine serum albumin
ChIP	chromatin immunoprecipitation
CPF	cleavage and polyadenylation factor
CRE	*cis*-acting regulatory element
CTD	C-terminal domain
CUT	cryptic unstable transcript
DNA	deoxyribonucleic acid
FISH	fluorescence *in situ* hybridization
FKBP	FK506 binding protein
FRB	FKPB-Rapamycin binding
Gal	galactose
Glu	glucose
GO	gene ontology
GTF	general transcription factor
H3	histone H3
HAT	histone acetyltransferase
HDAC	histone deacetylase
ICR/ICE	imprint control region/element
IVT	*in vitro* transcription
Kan	kanamycin
lncRNA	long non-coding RNA
me	methylation
me1	monomethylation
me2	dimethylation
me3	trimethylation
MNase	micrococcal nuclease
mRNA	messenger RNA
ncRNA	non-coding RNA
NDR	nucleosome depleted region
NET-seq	native elongating transcript sequencing
NFR	nucleosome free region
NPS	nucleosome positioning sequence
nt	nucleotide
OD	optical density

OX	oxidative phase of the yeast metabolic cycle
PAS	polyadenylation signal
PCR	polymerase chain reaction
PIC	pre-initiation complex
qPCR	quantitative PCR
RACE	rapid amplification of cDNA ends
RB	reductive building phase of the yeast metabolic cycle
RC	reductive charging phase of the yeast metabolic cycle
RNA	ribonucleic acid
RNAi	RNA interference
RNAPII	RNA polymerase II
RNP	ribonucleoprotein
RRM	RNA recognition motif
RT-PCR	Reverse Transcription PCR
SAGA	Spt-Ada-Gcn5 acetyltransferase
SAP	Sense/antisense pair
SSC	saline-sodium citrate
SUT	stable unanotated transcript
TBP	TATA-binding protein
TCS	transcriptional control similarity
TF	transcription factor
TI	transcription interference
tRNA	transfer RNA
TSS	transcription start site
UAS	upstream activating sequence
URS	upstream repressing sequence
WT	wild-type
WT	WT
Xa	active X chromosome
XCI	X chromosome inactivation
Xi	inactive X chromosome
XIC	X-inactivation center
XUT	Xrn1-sensitive unstable transcript
YMC	yeast metabolic cycle
YPD	yeast extract – bactopeptone – glucose
YPGal	yeast extract – bactopeptone – galactose

Table of contents

1. Introduction 2
1.1. Gene expression networks: 2
1.2. Gene Transcription: 3
1.2.1. The Transcription Cycle: 3
1.2.2. Transcription factors interact with the basal machinery to initiate transcription: 6
1.3. Gene expression in a chromatin context: 8
1.3.1. The nucleosome and histone tail modifications: 8
1.3.2. Functions of histone modifications: 9
1.3.3. H3K36 methylation is an important co-transcriptional histone modification that influences histone acetylation: 10
1.3.4. The nucleosome depleted region: 11
1.3.5. The recruitment model and promoter architecture as exemplified by the Gal4 activator system: 14
1.4. A systems approach to understanding gene expression regulation: 16
1.4.1. Individual genes in a genome-wide network: 16
1.4.2. Temporal compartmentalization of gene expression in the metabolic cycle: 17
1.4.3. Spatial regulatory connections and co-expression of neighbouring genes: 22
1.5. Integrating non-coding RNAs into existing models for genome regulation: 24
1.5.1. The genome is pervasively transcribed: 24
1.5.2. Origin and biogenesis of long non-coding RNAs: 26
1.5.3. Overlapping antisense transcription: 30
1.5.4. Consequences of compact transcription: 34
1.6. Perspectives on lncRNAs in humans: 38
1.6.1. X-inactivation: 39
1.6.2. Genomic Imprinting: 40
1.7. Aims: 41

***2.* Methods and Materials 43**
2.1. Media Composition: 43
2.2. Yeast strains and transformation: 43
2.3. Cell culture: 46
2.4. Anchor Away: 47
2.5. RNA extraction: 49
2.6. 3' RACE: 49
2.7. Northern blotting: 51
2.8. In vitro transcription and radiolabelling: 51
2.9. Approximating transcript levels: 55
2.10. Hybridization of RNA to strand-specific tiling arrays: 56
2.11. RNA Fluorescence In Situ Hybridization (FISH): 58
2.12. Chromatin Immunoprecipitation: 68

***3.* Investigating the role of overlapping transcription in transcription unit regulation using the *HMS2* locus as a model system 73**
3.1. Introduction: 73
3.1.1. Unstable Antisense Transcripts: 73

3.1.2. Stable Antisense Transcripts: 77
3.1.3. Aims: studying antisense transcription at the HMS2 locus 78
3.2. Results: *80*
3.2.1. The HMS2 sense and antisense promoters are overlapped by transcription from the opposite strand: 80
3.2.2. HMS2 sense and antisense are substrates of Rrp6p and Xrn1p but are regulated through distinct mechanisms: 84
3.2.3. Regulation of the HMS2 locus: 86
3.2.4. TFIIB regulates HMS2 sense and antisense transcription: 89
3.2.5. Inter-locus dependency at the HMS2 gene cluster: 94
3.2.6. Single cell analysis of HMS2 expression using RNA FISH reveals temporal separation of sense and antisense transcription: 96
3.3. Discussion: *103*
3.3.1. Anti-correlated sense/antisense levels depicted by Northern blot and tiling array analysis is a manifestation of single cells generating either sense or antisense: 103
3.3.2. Establishing a model: 106
4. Disrupting Overlapping Antisense Transcription at the *HMS2* locus 110
4.1. Introduction: *110*
4.1.1. Abolishing Antisense overlap: 110
4.1.2. Gene disruption: 114
4.1.3. Aims: 114
4.2. Results: *116*
4.2.1. The *HMS2.kan* and HMS2.ADH gene constructs produce novel but distinct transcription units: 116
4.2.2. Perturbation of the TEF promoter results in a more complex transcriptional profile: 127
4.2.3. Further dissection of the TEF promoter: 135
4.2.4. Investigating the relationship between 5' and 3' end regulatory elements: 142
4.2.5. Abolishing antisense transcription via gene substitution: 146
4.2.6. Transcription interference represses the *BAT2* promoter: 151
4.2.7. Single-cell analysis of antisense-less transcription units 152
4.3. Discussion: *160*
4.3.1. Creation of a fixed state system may generate noisy expression of HMS2 in the metabolic cycle: 160
4.3.2. Antisense transcription in an interleaved genome: 162
4.3.3. Promoter/terminator strength and functional duality determine transcriptional landscapes: 163
5. Devising a Mechanisms for Sense/Antisense State-switching 167
5.1. Introduction: *167*
5.1.1. Regulatory Transcription: 167
5.1.2. Aims: 171
5.2. Results: *172*
5.2.1. Coordinating a state-switch using a galactose inducible form of HMS2: 172
5.2.2. Activator binding is required for maintaining a state switch: 177
5.2.3. Transcriptional state switching corresponds to changes in antisense transcript-end sites: 181
5.2.4. Antisense shortening is Gal4p-independent in pGAL1.HMS2 and is inherent to the HMS2 AS: 183

5.2.5. Chromatin remodelling and modification at the HMS2 locus: 189
5.2.6. Preliminary investigation of the HMS2 promoter chromatin architecture: 191
5.3. Discussion: 194
5.3.1. Contributors to "plasticity" or high expression variability: 194

6. Discussion 197
6.1. Reviewing the state-switching model: 198
6.2. Sense/antisense state switching exists as a part of a larger network of interlinked overlapping transcription, making antisense-related phenotypes difficult to discern: 200
6.3. Potential roles for overlapping transcription units in establishing relationships between genome spatial- and temporal-structures: 203
6.4. Assessing the role of overlapping transcription in biological oscillatory networks: . 206
6.5. Extending the model- antisense-mediated post-transcriptional gene regulation: 210
6.6. Other future work: 218
6.7. Perspectives: 218
6.8. Concluding remarks: 219

7. Appendix 221

8. Bibliography 223

Table of Figures

Figure 1. The canonical view is that transcription is linear and unidirectional..... 4
Figure 2. The Yeast Metabolic Cycle.....19
Figure 3. Major classes of long non-coding transcripts in *Saccharomyces cerevisiae.* 27
Figure 4. Overlapping transcription in a compact genome.....31
Figure 5. Nucleosome depleted regions support sense and antisense transcription. 32
Figure 6. Mechanisms of transcription interference.....36
Figure 8. Generating Strand-Specific Probes.....54
Figure 9. Testing probe strand-specificity.....55
Figure 10. Single molecule RNA Fluorescence *in situ* hybridization.....59
Figure 12. Levels of H3 at the *TUB2* locus change between glucose and galactose, but not transcript levels.....71
Figure 13. Nascent transcription at the *HMS2* and *PHO84* loci.....76
Figure 14. The *HMS2* gene cluster. 79
Figure 15. Mapping the 3' ends of the *HMS2* sense and antisense and *BAT2* sense transcripts. 81
Figure 16. Overlapping Transcription at the *HMS2* locus.....83
Figure 17. Existing mechanisms of CUT and XUT sense/antisense regulation do not apply to the *HMS2* locus.....86
Figure 18. The *HMS2* sense/antisense pair can be manipulated in galactose. 88
Figure 19. Potential transcriptional regulators of the *HMS2* locus.....90
Figure 20. Disruption of transcription factor binding does not alter sense/antisense transcript levels.....91
Figure 21. Depletion of TFIIB using the Anchor Away technique. 93
Figure 22. Transcription over the *RPS4A*, *BAT2*, and *YJR149W* loci is linked to *HMS2* transcription in glucose- versus galactose-containing media.....95
Figure 23. Single-cell analysis of *HMS2* expression using RNA FISH. 98
Figure 24. The *HMS2* locus displays regulation in the cell cycle.....100
Figure 25. Periodicity of genes in the *HMS2* cluster in the Yeast Metabolic Cycle...102
Figure 26. Knocking out the 3' Gal4p site does not abolish overlapping antisense transcription at the *SUR7* locus.....105
Figure 27. Model of a two-state system at *HMS2.*.....108
Figure 28. Coding and non-coding transcripts at the *S. cerevisiae GAL* locus.....112
Figure 29. Summary of transcripts in the *HMS2.kan* and *HMS2.ADH* disruption strains.....117
Figure 30. Verifying *kanMX6* expression.....119
Figure 31. Sense transcription across the *HMS2.kan* and *HMS2.ADH* gene constructs 120
Figure 32. Antisense transcription across the *HMS2.kan* and *HMS2.ADH* gene constructs.....123
Figure 33. Quantitation of sense transcripts in wild-type and disrupted *HMS2*.....126

Figure 34. Rrp6p sensitivity of wild-type and ^{T}*ADH1*-truncated *HMS2* sense and antisense transcripts.....128
Figure 35. The 5' ^{P}*TEF* Δ-1 strain.....131
Figure 36. Summary map of the *HMS2.kan* and 5' ^{P}*TEF* Δ-1 transcript profiles.....132
Figure 37. Disrupting the *TEF* promoter does not abolish antisense transcription. 134
Figure 38. Dissection of the 5' region of the *TEF* promoter. 137
Figure 39. Effects of dissecting the 5' region of the *TEF* promoter.....138
Figure 40. Mutational dissection of "Region D" in the *TEF* promoter.....140
Figure 41. Northern blot analysis of the dissection of "Region D" in the *TEF* promoter 141
Figure 42. Sequences at the 3' end of a transcription unit can also affect transcript levels. 144
Figure 43. Summary map of wild-type and *URA3* transcript profiles.....148
Figure 44. Substitution of the *HMS2* CDS with that of *URA3* ablates antisense transcription.....149
Figure 45. Rrp6p and Xrn1p sensitivity of *URA3* transcripts 150
Figure 46. Insertion of the *ADH1* terminator results in increased levels of *BAT2* transcript.....153
Figure 47. Single-cell analysis of *HMS2* sense expression in wild-type and *HMS2.ADH* cells. 155
Figure 48 Single-cell analysis of *GAL1* sense expression in wild-type cells. 156
Figure 49. Assessing differences between growth of wild-type and *HMS2.ADH* cells. 157
Figure 50. Summary map of *pGAL1.HMS2*, *pGAL1.URA3* and wild-type *HMS2* transcript profiles in cells grown in glucose- and galactose-containing media. 173
Figure 51. Replacement of the *HMS2* promoter with that of *GAL1* creates an antisense-dominant state in cells grown in glucose-containing medium.....174
Figure 52. Induction profile of galactose-inducible *HMS2*.....176
Figure 53. Induction profiles of the galactose inducible *HMS2* and *URA3* transcripts. 178
Figure 54. Induction profile of *pGAL1.HMS2* after Gal4p was depleted from the nucleus. 179
Figure 55. Induction profile of *pGAL1.HMS2* after TBP was depleted from the nucleus. 180
Figure 56. Mapping the 3' ends of the *pGAL1.HMS2* antisense transcript isoforms.182
Figure 57. Observing *pGAL1.HMS2* antisense transcript isoforms during glucose repression. 184
Figure 58. Observing *pGAL1.HMS2* antisense transcript isoforms in the absence of the Gal4p activator. 186
Figure 59. Antisense transcript shortening in galactose-grown cells is inherent to the *HMS2* antisense transcript.....188
Figure 60. Inactivation of Spt6p does not result in derepression of the *HMS2* sense promoter. 190
Figure 61. Distribution of H3K36 trimethylation across the *HMS2* promoter in glucose- and galactose-grown cells. 193

Figure 62. Revising the state-switching model..... 199
Figure 63 . Synteny of HMS2 and *BAT2* in three related species of budding yeast.. 204
Figure 64. The relationships described at *HMS2* are not unique to the gene cluster. 208
Figure 65. A potential role for perinuclear antisense RNA dots in mediating elasticity in sense/antisense state switching. 216
Figure 66. Evidence for enrichment of *HMS2* sense but not antisense transcripts in extracts enriched for polyadenylated RNA. 222

Chapter 1.

Introduction to complex transcription units in *Saccharomyces cerevisiae*.

1. Introduction

1.1. Gene expression networks:

The central dogma of molecular biology describes the sequential transfer of information from an initial DNA template, to an RNA molecule (transcription), which then serves as either a functional gene product or an intermediary messenger RNA (mRNA) template used for protein synthesis (translation). This progression from stored genetic information in "coding" DNA sequences (CDS) to functional gene product, protein or otherwise, is called gene expression. Cellular structure and function in development, metabolism, and response to stress and other environmental cues are all orchestrated through tightly coordinated regulation of gene expression programmes likely involving thousands of genes that are turned either "off" or "on", depending on the circumstance and necessity. While gene expression is subject to regulation at multiple stages, much of this regulation occurs at the level of transcription. Indeed, in higher eukaryotes failure to correctly manage transcription can manifest in different disease states, including genomic imprinting disorders in development and cancer, emphasizing the requirement for multiple, potentially redundant devices of control. As a result, much work has been devoted to understanding these mechanisms and deconstructing the transcriptional regulatory circuitry; however, with the advent of large-scale, high-throughput genome-wide studies, the complexity of regulation quickly surpassed expectations. In particular, the discovery of widespread transcription that did not generate "functional" RNA prompted the reassessment of long held principles in the central tenets.

1.2. Gene Transcription:

1.2.1. The Transcription Cycle:

As the entry point, transcription is poised as a critical rate-limiting step in controlling the gene expression pathway. In eukaryotic systems, nuclear DNA can be transcribed by three related but functionally distinct multi-subunit RNA polymerase (Pol) complexes, RNA polymerase I, II, and III, with each complex subject to particular mechanisms of control. These polymerase complexes synthesize RNAs complementary to the DNA template by repeatedly catalyzing the addition of nucleotide (nt) building blocks to the 3' ends of growing RNA chains. While transcription by RNA polymerase I (RNAPI) and III (RNAPIII) is mainly responsible for generating ribosomal (rRNA), transfer (tRNA), and other structural RNAs, constituting the greater part of a cell's transcriptional activity, transcription by RNA polymerase II (RNAPII) synthesizes mRNA from protein-coding genes, and thus has been the focus of the bulk of studies on transcription regulation (reviewed in Kaplan, 2013). The transcription cycle is itself divided into three phases: initiation, elongation, and termination, with each step also associated with specific aspects of gene anatomy and presenting various substrates for regulation (Hahn & Young, 2011; Venters & Pugh, 2009a) (Figure 1). For instance, initiation traditionally occurs at the 5' ends of genes, where gene-specific transcriptional activators help to recruit RNA polymerase II alongside a set of general transcription

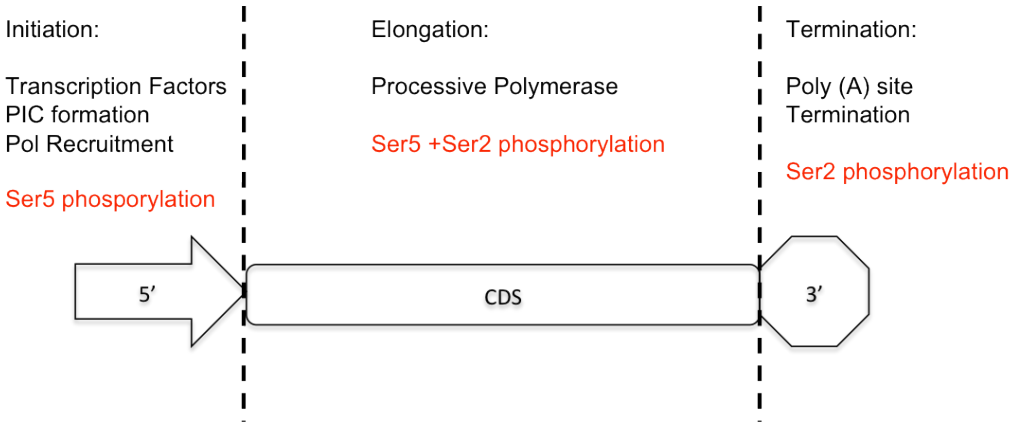


Figure 1. The canonical view is that transcription is linear and unidirectional.

The phases of the transcription cycle are associated with distinct aspects of gene anatomy, such that a single region of DNA possesses a unique function in either initiation, coding potential, and termination. The transcription cycle is also associated with different phosphorylation states of the serine residues in the RNAPII CTD, displayed in red font.

factors (GTF) to assemble a pre-initiation complex (PIC) at the core regulatory promoter, which houses the transcription start site (TSS). This assembly prompts the "opening" of the promoter DNA such that the RNA polymerase is poised to begin transcribing the information-containing segment of the gene or open reading frame (ORF). The progression from a poised state to a translocating processive polymerase marks the beginning of the elongation phase. Finally, towards the 3' end of most genes, the transcribing polymerase encounters an AAUAAA polyadenylation (poly (A)) signal (PAS), which recruits the polyadenylation factors that mediate the addition of a 3' end adenosine tract called the poly (A) tail to the nascent pre-mRNA after 3' end cleavage, a process collectively termed 3' end formation (N. J. Proudfoot, 2011). Shortly after transcribing through the PAS, the polymerase is released from the DNA template through still largely elusive mechanisms, resulting in termination of transcription. Co-transcriptional modification of the pre-mRNA's 5' and 3' ends results in a mature capped

and polyadenylated mRNA species competent for export to the cytoplasm and translation.

The phases of the transcription cycle can also be characterized by differential phosphorylation states of the RNAPII C-terminal domain (CTD) (reviewed in Buratowski, 2009). This CTD "tail" of RNAPII's Rbp1p subunit consists of the repeating heptapeptide sequence, Tyr-Ser-Pro-Thr-Ser-Pro-Ser (YSPTSPS). There are 26 repeats in *S. cerevisiae* and 52 in humans. Serines 2 and 5 (Ser2 and Ser5) of the tail are major sites for phosphorylation; however phosphorylation of serine 7 has also been identified in yeast and mammals, but is less well described (Akhtar et al., 2009; Egloff et al., 2007; Kim, Suh, Cho, & Buratowski, 2009). In brief, an unphosphorylated RNAPII is recruited to the promoter where it is phosphorylated at Ser5 by TFIIH. A drop in the levels of Ser5-phosphorylation after the first few hundred base pairs of the coding region indicates an elongation-competent form of RNAPII, while towards the 3' of the gene, Ser2 phosphorylation, by kinase Ctk1, accumulates and is associated with 3' end formation (Figure 1). Importantly, these different phosphorylation patterns link the transcription cycle with the recruitment of the RNA surveillance and pre-mRNA-processing machinery, as well as chromatin modifiers and remodelers, providing an additional avenue for transcriptional and post-transcriptional regulation.

1.2.2. Transcription factors interact with the basal machinery to initiate transcription:

While each step in the transcription cycle can be modulated to alter the final transcriptional output, by and large, transcription initiation is suspected to tender the most substantial dimensions for transcriptional regulation, by virtue of the complex interplay between the combinatorial nature of *cis*-acting regulatory elements (CRE), their associated sequence-specific *trans*-acting transcription factors, and the chromatin architecture found at the promoter regions of genes (reviewed in Venters & Pugh, 2009). Collectively, this process is known as transcriptional activation by recruitment (reviewed in Mark Ptashne & Alexander Gann, 1997). Overall, it is the interaction between these protein and DNA constituents that is responsible for dictating the spatial and temporal gene expression patterns in response to different biological requirements.

In general, transcription factors bind to CREs, either upstream activating (UAS) or repressing (URS) sequences and work to stimulate or discourage the recruitment of the basal transcription machinery to the core promoter, respectively. The basal machinery, whose assembly at core promoter sequences is essential for catalyzing correctly positioned transcription initiation, consists of RNAPII, and the GTFs including TFIIA, -B, -D, -E, -F, and -H (reviewed in Sikorski & Buratowski, 2009). TFIID is a multi-subunit complex, containing TBP (TATA-binding protein), which binds to the TATA DNA motif, the most highly conserved core-promoter element in eukaryotes (reviewed in Hernandez, 1993). While the basal machinery can produce transcription *in vitro* from

isolated core promoters, *in vivo* the interaction between PIC components and sequence-specific activators is a necessary step in initiating robust gene-specific transcription and to overcome nucleosomal barriers to transcription (Knezetic & Luse, 1986; Lorch, LaPointe, & Kornberg, 1987, 1992; Svaren & Horz, 1997). In light of these findings, decades of research have been dedicated to characterizing transcription factors, their interacting partners, structure, relationships to chromatin and functions in gene regulation.

Currently, there are 169 genes designated as transcription factors in the Yeastract database (Teixeira et al., 2006). The number of genes regulated by any given transcription factor is highly variable; for instance, the galactose-dependent transcriptional activator, Gal4p, is thought to target 10 genes in *S. cerevisiae* (Ren et al., 2000). By contrast, the regulator Rap1p (Repressor Activator Protein 1), which has both an activating and repressing domain, is estimated to bind to over 300 promoters (Buck & Lieb, 2006). Adding to the complexity and customization of transcriptional regulation are a host of general co-regulators, which act to bridge the regulatory signals conveyed by the sequence-specific factors and the basal machinery. These co-regulators come in many different forms, including multi-subunit complexes such as the Mediator co-activator, as well as chromatin modifying and chromatin remodelling machinery, to facilitate the interaction between RNAPII and promoter DNA (Lemon & Robert, 2000).

1.3. Gene expression in a chromatin context:

1.3.1. The nucleosome and histone tail modifications:

In eukaryotes, DNA exists in a highly plastic protein-DNA assembly called chromatin. The basic unit of chromatin, the nucleosome, is comprised of 147 base pairs (bp) of DNA wrapped 1.65 times around a heterooctamer of core histone proteins. Repeating nucleosome units make up the chromatin fibre, which serves as the basic template on which regulatory forces can act (reviewed in Jansen & Verstrepen, 2011).

The octameric assembly of histones consists of two units each of histone H2A (H2A), histone H2B (H2B), histone H3 (H3), and histone H4 (H4). Core histone proteins consist of a globular core domain and protruding amino-terminal tail that are subject to covalent post-translational modifications. Some of the best-characterized modifications include methylation (mono-, di-, and trimethylation), acetylation, and ubiquitination of lysine (K) residues; however, there are many more including sumoylation of lysines; methylation of arginines; and phosphorylation of serines and threonines. Briefly, the addition or erasure of these modifications are catalyzed by histone modifying enzymes, including histone methyltransferases (HMTs), histone acetyltransferases (HATs), histone deacetylases (HDACs), and histone demethylases (reviewed in Kouzarides, 2007). Histone modifying enzymes often exist in the context of protein complexes; for instance, the SAGA (Spt-Ada-Gcn Acetyltransferase) complex is a general transcriptional co-activator with acetyltransferase activity modulated through its Gcn5p subunit (Koutelou, Hirsch, & Dent, 2010) (reviewed in Koutelou et al. 2010).

1.3.2. Functions of histone modifications:

The influences exerted by these individual histone modifications on transcription are, at best, superficially understood. In general, histone modifications are thought to function in two ways. Firstly, some of these marks, in particular acetylation of H3 and H4 lysine residues, have been proposed to directly alter the electrostatic affinity between the positively charged histone lysines and the negatively charged phosphate backbone of nucleosomal DNA, resulting in loosened DNA-histone interactions. This is consistent with studies that have demonstrated *in vivo* that acetylated histones are easier to displace (Reinke & Hörz, 2003; Zhao, Herrera-diaz, & Gross, 2005). Secondly, histone tail modifications may provide binding platforms for effector proteins bearing specific domains such as bromodomains or chromodomains, which can recognize acetylated and methylated histone tails, respectively. These effector proteins come in many forms, including other histone modifying enzymes, consistent with the popular notion of "histone crosstalk", and/or chromatin remodelers. For example, the RSC (Remodels Chromatin Structure) complex is a multi-bromodomain-containing ATP-dependent chromatin remodeler (Cairns et al., 1996). In terms of executing a downstream function, RSC can recognize SAGA (H3 HAT) and NuA4 (H4 HAT) acetylated histones through its bromodomains and can facilitate the passage of RNAPII through nucleosomes by translocation of nucleosome-bound DNA through its ATPase catalytic subunit, Sth1p (Carey, Li, & Workman, 2006; Saha, Wittmeyer, & Cairns, 2002).

1.3.3. H3K36 methylation is an important co-transcriptional histone modification that influences histone acetylation:

Recent advances using high-throughput genome-wide techniques have demonstrated that histone modifications can be associated with specific regions of gene anatomy. For instance, at the average gene, H3K36 methylation is found predominantly in the coding regions of genes. The H3K36 methyltransferase, Set2 (Strahl et al., 2002), travels with elongating polymerases phosphorylated on Ser2/Ser5, resulting in co-transcriptional deposition of methyl groups on H3K36 (Keogh et al., 2005). The function of trimethylated H3K36 (H3K36me3) is still unclear, but dimethylated H3K36 (H3K36me2) marks "active" genes (Rao, Shibata, Strahl, Lieb, & Lieb, 2005). Additionally, at least H3K36me2 is required to recruit the histone deactylase complex, Rpd3S, which deacetylates histones H3 and H4 in the wake of transcription in order to reset the chromatin to a transcription impervious state, preventing potentially deleterious cryptic initiation (Carrozza et al., 2005; Joshi & Struhl, 2005; Keogh et al., 2005; Youdell et al., 2008). Thus, it seems as though H3K36me2 indentifies with two conflicting roles, marking genes as "on" and suppressing spurious intragenic transcription.

While the process of Set2/Rpd3S-mediated histone deacetylation has been well described, the understanding of histone acetylation in coding regions is less well characterized. However, a study by Rufiange and colleagues (2007) demonstrated that replication-independent histone H3/H4 exchange in the coding region was dependent on high levels of transcription and that newly assembled histones were marked by

H3K56 acetylation. Acetylation of histone H3 K56, a modification to the globular core that disrupts electrostatic interactions with the nucleosomal DNA, is carried out by histone acetyltransferase Rtt109 on newly synthesized soluble H3 (Ozdemir et al., 2005; Tsubota et al., 2007). By exploiting H3K56ac as an indicator of histone exchange, a recent study identified novel links between co-transcriptional H3K36 methylation and histone H3/H4 acetylation (Venkatesh et al., 2012). The study showed that the accumulation of acetylated histones generated by transcription-dependent histone exchange over the coding region of genes, is generally blocked by H3K36 methylation, which in addition to its role in recruiting the Rpd3S deacetylase, suppresses the interaction of H3 with histone chaperones, thus reducing histone exchange. While Venkatesh and colleagues (2012) demonstrated this relationship in the coding region of genes, similar co-transcriptional mechanisms might work to regulate gene promoter regions. For instance, promoter nucleosomes are dynamic and participate in histone exchange even in repressive conditions, potentially poising them for activation; however, what targets these promoters, in particular repressed promoters, for histone exchange remains ambiguous (Dion et al., 2007; Rufiange, Jacques, Bhat, Robert, & Nourani, 2007).

1.3.4. The nucleosome depleted region:

Regardless of modification state, by limiting accessibility, packaging of DNA into nucleosomes is thought to be generally repressive to gene transcription; a notion that has been supported by a number of pioneer studies demonstrating that nucleosome

depletion causes derepression of genes (Durrin, Mann, & Grunstein, 1992; M Han & Grunstein, 1988; Min Han, Kim, & Kayne, 1988). Thus, eukaryotes have devised a host of mechanisms to cope with these nucleosomal barriers. For instance, the FACT complex (facilitates chromatin transcription) allows polymerase to translocate on nucleosome arrays by disassembling nucleosome hurdles (reviewed in Reinberg & Sims, 2006).

As previously discussed, transcription initiation is particularly sensitive to a chromatinized template, as *in vitro* studies demonstrated that polymerase and the GTFs were insufficient to initiate transcription on nucleosome arrays (Knezetic & Luse, 1986; Lorch et al., 1987, 1992; Svaren & Horz, 1997). To facilitate access to promoter DNA, most eukaryotic promoters bear a distinct chromatin signature characterized by nucleosome depletion, or a nucleosome depleted region (NDR). The NDR, otherwise referred to as the nucleosome free region (NFR), is characterized by a ~150 bp region depleted of nucleosome occupation, and flanked by an upstream "-1" and downstream "+1" nucleosome (Yuan et al., 2005). Both the -1 and +1 nucleosomes contain the H2A histone variant, H2A.Z, which supposedly acts to enhance nucleosome eviction and thus relieve any general inhibitory effects of nucleosome occupation (Raisner et al., 2005; Zhang, Roberts, & Cairns, 2005). The positioning of these nucleosomes is determined by a multitude of different factors including nucleosome-deterring poly (dA/dT) tracts flanked by bendable AA/TT/TA dinucleotides spaced at 10 bp intervals, otherwise known as nucleosome-positioning sequences (NPS). While sequence determinants act *in cis*,

protein factors such as the general regulatory factors (GRF) Reb1p and Rap1p can exert influence on nucleosome positioning *in trans* (reviewed in Struhl & Segal, 2013).

The promoter architecture described above is widespread; a genome-wide analysis of nucleosomal DNA performed by Mavrich and colleagues (2008) found that 95% of all genes in *S. cerevisiae* contained a 5' nucleosome depleted region (Mavrich et al., 2008). Likewise, similar organization has been demonstrated in a variety of metazoans including worms, flies, and humans (Bai & Morozov, 2010; Radman-Livaja & Rando, 2010). In relation to *cis*-acting regulatory sequences, binding motifs, in particular UAS/URS sequences, generally reside within areas of nucleosome depletion, which likely facilitates their accessibility to their target regulator and resultant PIC formation (reviewed in Bai & Morozov, 2010; Jansen & Verstrepen, 2011). Consequently, core promoter elements, especially TATA boxes dwelling in the vicinity of NDRs can be subject to repressive nucleosome occlusion and might require the activity of activators and subsequent chromatin remodelers in order for activation to take place (Svaren & Horz, 1997).

As one might expect, different classes of genes are distinguishable by their promoter nucleosomal architecture (Bai & Morozov, 2010; Bradley R Cairns, 2009). For instance, constitutively expressed genes such as housekeeping/growth genes generally bear the canonical -1/+1 nucleosome arrangement. These "open" promoters are characterized DNA sequences that possess the propensity for NDR formation: a central poly (dA:dT)

tract, flanked by nucleosome positioning sequences. Furthermore, the *cis*-regulatory sequences, such as binding sites for transcriptional activators, of these constitutively expressed genes typically reside within the exposed NDR. However, TATA boxes are noticeably scarce from constitutive promoters (Basehoar, Zanton, & Pugh, 2004; Huisinga & Pugh, 2004).

In contrast, "covered" promoters, which bear multiple binding sites for transcriptional activators, and are enriched for TATA boxes (Basehoar et al., 2004; Huisinga & Pugh, 2004), are generally occupied by nucleosomes. This promoter architecture is associated with highly regulated inducible genes or genes displaying high degrees of expression variability, such as stress response genes. Promoters with a "covered" nucleosome architecture, limit DNA accessibility to transcriptional activators and other regulatory factors, exhibiting bi-modal expression patterns in which a gene is either fully activated or switched off, such that single cells in a genetically homogenous population grown under constant conditions display high expression "noise" (Field et al., 2008).

1.3.5. The recruitment model and promoter architecture as exemplified by the Gal4 activator system:

One of the best-characterized examples of transcriptional activation is the galactose-mediated control of the *GAL* genes in *S. cerevisiae* by the Gal4p activator. These genes are repressed under growth in glucose-containing media but are rapidly induced upon transfer to galactose-containing media. Through its activation domain, Gal4p switches on its target genes through sequential recruitment of coactivators and general

transcription factors. More specifically, under permissible conditions, the UAS_{GAL}-bound Gal4p interacts directly with the SAGA and Mediator co-activator complexes through its activation domain, which then act as scaffolds for the binding of the GTFs, including TBP which is directly recruited by SAGA (reviewed in Traven, Jelicic, & Sopta, 2006). Accordingly, SAGA is necessary for full activation, whereas Mediator only makes a moderate contribution (Reeves & Hahn, 2005). Moreover, the Gal4p activator has been shown to stimulate transcription in all eukaryotes examined, ranging from yeast to humans, as well as drive transcription from heterologous cassettes, demonstrating that the recruitment pathway displayed by the Gal4p activator is well conserved (Longtine et al., 1998).

In terms of promoter architecture, it was originally thought that the *GAL* genes exhibited standard inducible promoter nucleosomes, with the UAS_{GAL} motifs found in nucleosome-depleted regions, but with their corresponding TATA boxes experiencing nucleosome occlusion (reviewed Lohr, 1997). However, more recent work by Floer and colleagues (2010) demonstrated that at the *GAL10-1* bidirectional promoter, the UAS_{GAL}, is bound by an H2A.Z-containing nucleosome, partially unwrapped via an interaction with the RSC chromatin-remodelling complex (Floer et al., 2010). Additionally, in *S. cerevisiae*, RSC has a DNA binding subunit that is thought to function in positioning this RSC-nucleosome complex at the UAS_{GAL}. The RSC-nucleosome complex positively affects galactose induction by creating a more permissive platform for Gal4p binding, as in its absence, non-RSC nucleosomes bind to the UAS_{GAL}, antagonizing Gal4p binding and

impairing induction kinetics. Floer and colleagues (2010) further demonstrated that this RSC-nucleosome complex is sufficient to phase the surrounding promoter nucleosomes, which, in an earlier study, were shown to be evicted by the Gal4p-recruited the SWI/SNF chromatin remodeler, which itself works to slide or eject nucleosomes, facilitating recruitment of downstream coactivators such as SAGA and Mediator to their respective binding domains (Bryant et al., 2008). Additionally, pioneering studies demonstrated that nucleosomes occluding the promoter are disrupted in a Gal4p activation domain-dependent manner, underscoring the critical relationship between activators and promoter chromatin remodelling (Axelrod, Reagan, & Majors, 1993).

1.4. A systems approach to understanding gene expression regulation:

1.4.1. Individual genes in a genome-wide network:

While in-depth analysis of individual gene regulation provides a mechanistic understanding of how genes work as isolated units, in order to fully appreciate the associated biological outcomes of gene regulation, a systems-level perspective is required. Since the development of high-throughput genome-scale tools transformed the molecular toolkit, it has become clear that the investigation of genetic instructions and the subsequent derivation of their encoded products, informs only superficially the connections between genotype and phenotype; alternatively, a systems approach considers the structure, behaviour and dynamics of genetic interactions rather than dissecting the isolated parts (reviewed in Ideker, Galitski, & Hood, 2001; Kitano, 2002). Indeed, modulation of single genes occurs within the context of a broader circuitry, involving multiple interacting inputs and outputs, both intrinsic (e.g. genetic) and

extrinsic (e.g. environmental), and constrained by spatial and temporal dimensions. The resultant networks consist of thousands of genes regulating each other's expression by an array of largely poorly understood interactions and mechanisms. For instance, this wider framework includes co-expression of functionally similar genes, signal transduction cascades in response to environmental cues, as well as overarching expression patterns driven by metabolic and cell division requirements. Given these relationships, there lies great value in exploring how the regulation of individual genes figures into the dynamics of genome-wide expression networks.

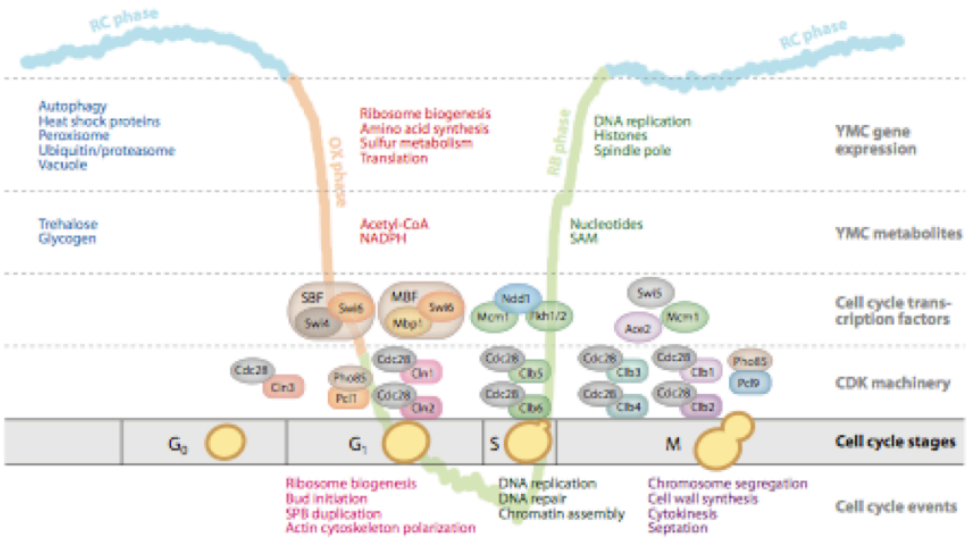
1.4.2. Temporal compartmentalization of gene expression in the metabolic cycle:

A perhaps intuitive, but underappreciated aspect of gene regulation is that in order to cope with a large and diverse genetic load, encoding products with oftentimes competing functions, cells organize gene transcription into biologically relevant temporal windows. Most famously are the circadian rhythms, in which gene expression is organized according to day-night cycles, primarily to suit metabolic needs during distinct temporal periods (Rutter, Reick, & McKnight, 2002). While largely unreceptive to day-night cycles, in glucose-limited continuous batch culture, budding yeast will synchronize their respiratory and glycolytic activities, which persist in a stable oscillatory relationship (reviewed in Richard, 2003). In two different strains of *S. cerevisiae*, independent studies by Klevecz and colleagues (2004) and Tu and colleagues (2005) were able to detect robust 40- and 300-minute period oscillations in oxygen consumption, respectively, both characterized by non-respiratory and respiratory phases and accompanied by phase-specific transcription signatures involving over half

the yeast genome. In spite of the discrepancy in period length, which is primarily a function of glucose concentration, both studies found that genes with common functions displayed coordinate transcription and that these transcription patterns were structured relative to oxygen consumption. While these studies measured RNA levels, additional work has been done using fluorescent protein reporters, which correlates with the RNA expression profiles (Laxman, Sutter, & Tu, 2010).

Based on these findings, Tu and colleagues (2005) conceived a model for a Yeast Metabolic Cycle (YMC) comprised of three phases: the respiratory oxidative (OX) phase, and non-respiratory reductive building (RB) and reductive charging (RC) phases (Figure 2A, B). While based on the "longer" YMC, these annotations by and large reflect the oscillatory dynamics represented in the "shorter" YMC described by Klevecz and colleagues (2004), which they annotate as having oxidative, early, and late reductive phases. In either scenario, the OX phase is mainly associated with biomass accumulation and respiration. Genes expressed during this phase include those encoding cytoplasmic ribosomal proteins and proteins involved in translation. In the RB phase, characterized by minimal oxygen consumption, a new expression programme is instigated, dominated by genes encoding mitochondrial ribosomal proteins, histones, and spindle pole formation components. In accordance with this genetic profile, DNA replication appears to be confined to the RB phase, with approximately half of all cells entering the cell cycle near the beginning of the RB phase (Tu, Kudlicki, Rowicka, & McKnight, 2005). Klevecz and colleagues (2004) demonstrated similar findings and contend that the temporal

A)



B)

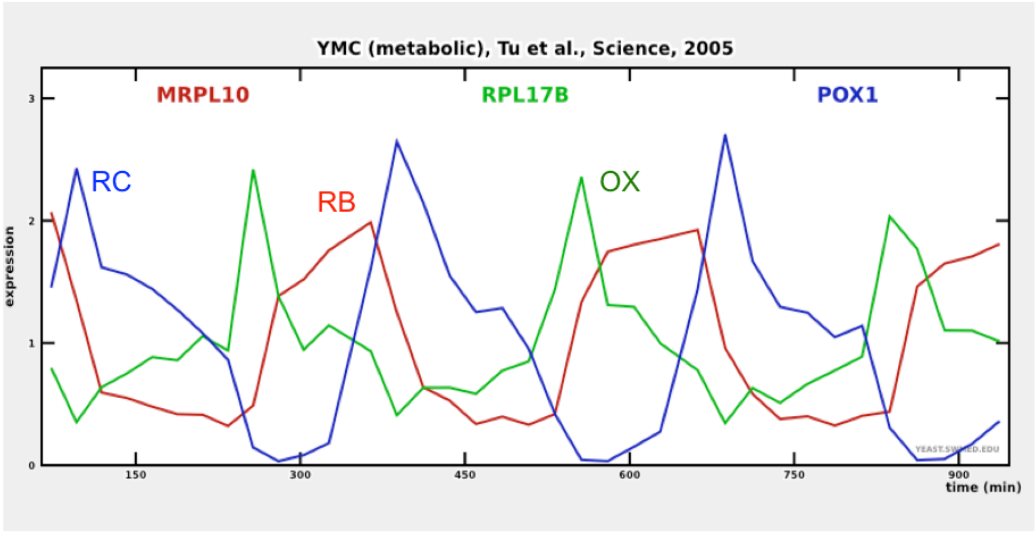


Figure 2. The Yeast Metabolic Cycle.

(A) The YMC is comprised of three phases based on oxygen consumption: the high oxygen consumption (HOC) oxidative phase and low oxygen consumption (LOC) reductive building and reductive charging phases. In a population of metabolically synchronous cells, ~ 50% of the population will enter the cell cycle during the reductive building phase. Adapted from Cai & Tu (2012). **(B)** *MRPL10, RPL17B,* and *POX1* are sentinel genes for the RB (red), OX (green), and RC (blue) phases, respectively, as described by Tu et al. (2005).

partitioning of respiratory metabolism and DNA synthesis might represent an evolutionary adaptation for reducing oxidative damage to nascent DNA via reactive oxygen species (ROS) primarily generated during the OX phase. Accordingly, a later study by Chen and colleagues (2007) found that decoupling cell division from the metabolic cycle resulted in increased spontaneous mutation rates (Z. Chen, Odstrcil, Tu, & McKnight, 2007). Additionally, between 70-87% of mRNAs are maximally expressed in the non-respiratory phases of the YMC, potentially because of the same selective pressures (Klevecz, Bolen, Forrest, & Murray, 2004). Finally, the RC phase, which is also associated with low oxygen consumption, has an expression programme enriched for genes involved in autophagy, protein degradation, and non-respiratory forms of metabolism, in particular, fatty acid oxidation. As a result, the RC phase is marked by the production of acetyl-coA, presumably in preparation for the impending transition to the respiratory oxidative phase, during which it would be used for the production of adenosine triphosphate (ATP) (Tu et al., 2005).

The robustness with which these transcription super-clusters oscillate necessitates tight control of the timing of RNA synthesis and degradation. Temporal kinetics of RNA synthesis of functionally-related genes could be managed to some extent through shared regulatory motifs in the promoters. In accordance with this theory, Tu and colleagues (2005) demonstrated that ~60% of transcription factors in yeast displayed periodicity in the YMC. Post-transcriptionally, the half-life of an mRNA could also affect its availability during the YMC. A study by Soranzo and colleagues (2009) found that

genes with shorter half-lives have very sharp oscillations, indicating short-lived availability, whereas transcripts with longer half-lives displayed broader, more moderate amplitude oscillations, indicating prolonged availability (Soranzo, Zampieri, Farina, & Altafini, 2009). While the interplay between transcription factor-mediated activation bursts and transcript-specific half-lives might specify the temporal context in which a gene is expressed, given the stability of these oscillations in general, it is worth exploring potential auxiliary forms of regulation employed by cells to ensure the fidelity of the YMC.

Both the Klevecz et al (2004) and Tu et al (2005) investigations revealed metabolic periodicity using yeast cultures synchronized in nutrient-restricted continuous growth conditions; however, further studies have independently corroborated the existence of a YMC in individual and asynchronous cells using fluorescent reporters or RNA fluorescence in situ hybridization (FISH) to observe periodic expression (Laxman et al., 2010; Silverman et al., 2010). Together these works indicate metabolic cycling as an intrinsic property of yeast metabolism, which prompts further investigation into why metabolic time structures exist; how they integrate with other biological phenomena; and how they influence or are influenced by changes in the environment. Likely, the biological rationale underlying this oscillatory behaviour is to generate biochemical reactions within optimized conditions, by enhancing the coordination of functionally-related genes with their diffusible regulators, minimizing futile reactions, but also to ensure these reactions occur during low-risk windows of time; for instance, the

partitioning of respiration and DNA synthesis (Klevecz et al., 2004; Tu et al., 2005). Considering these incentives, it is reasonable to believe that these metabolic rhythms are not unique to yeast, but may also manifest in other eukaryotes and influence other rhythmicities found in nature such as circadian rhythms, sleep-wake cycles, and hibernation (reviewed in Tu & McKnight, 2006). Indeed, a study by Slavov and colleagues (2012) provided evidence for a metabolic cycle in *Schizosaccharomyces pombe*, as well human epithelial and primary fibroblast cells (Slavov, Airoldi, Van Oudenaarden, & Botstein, 2012). Additionally, considering the breadth of genes regulated through this network in *S. cerevisiae*, valuable information might be ascertained about individual genes by understanding how both their regulators and their products fit into this scheme.

1.4.3. Spatial regulatory connections and co-expression of neighbouring genes:

While metabolic requirements create a basis for higher order expression programmes, genomic spatial arrangements have also been shown to contribute to regulatory relationships between genes, albeit on a smaller scale. Indeed, it has become increasingly evident that gene order in eukaryotes is non-random and may bear some influence on gene regulation (reviewed in Hurst, Pál, & Lercher, 2004). Several studies have reached the consensus that proximity between genes, in particular adjacent and proximal non-adjacent genes, imparts a strong driving force for positively correlated patterns of expression (Batada, Urrutia, & Zhang, 2007; Cohen, Mitra, Hughes, & Church, 2000; J. M. Lee & Sonnhammer, 2003). While genes displaying functional relatedness, such as the *GAL* genes, might be regulated similarly by possessing the same

cis-acting sequences (i.e. the UAS_{GAL}), co-expression of clustered genes in *S. cerevisiae* is largely independent of transcriptional control similarity (TCS), gene orientation, and/or shared regulatory sequences, (Batada et al., 2007; Cohen et al., 2000). However, members of adjacent gene pairs or clusters are more likely to belong to the same functional pathway than expected by chance (Cohen et al., 2000; J. M. Lee & Sonnhammer, 2003). Additionally, genes belonging to the same stable protein complexes favour co-expression (R. Jansen, Greenbaum, & Gerstein, 2002) and tend to co-localize within 10-30 kb of each other (Teichmann & Veitia, 2004). Further evidence for functionally-related gene clustering was provided in an earlier report by Cho and colleagues (1998), who demonstrated that 25% of genes displaying periodicity in the cell cycle were found adjacent to genes expressed in the same period, linking spatial gene arrangement with physiologically-relevant rhythmicities in gene expression (Cho et al., 1998).

While the mechanism by which these clustered genes are co-regulated remains largely elusive, work by Cohen and colleagues (2000) indicated that distance between genes is the main factor contributing to positively correlated expression patterns. Compatible with this finding, work by Batada and colleagues (2007) strongly supported a model in which genes are organized together in open chromatin domains to facilitate regional co-expression in the face of repressive nucleosome occupancy. However, while proximity increases the likelihood of co-expression, the fact that localization in these domains is not deterministic of positive correlations, such that within these clusters reside genes

that do not display co-expression with their compatriots, implies that there exist some, yet to be determined, fine-tuning mechanisms in addition to these position effects.

1.5. Integrating non-coding RNAs into existing models for genome regulation:

The structure and dynamics of relationships between genes either through time-dependent physiological programmes or spatial domains have been to a great extent accessed by large-scale genome-wide studies. The same technology unveiling this circuitry has also uncovered an unanticipated level of non-coding transcription, adding more complexity to an already complicated system and challenging the framework for gene regulation conveyed by the central dogma. While great strides have been made in characterizing a diverse body of non-coding transcripts, fewer successes have been realized in determining if and how non-coding transcription contributes to gene regulation, let alone how it might integrate into a greater transcriptional network. A few examples of regulatory non-coding RNAs have been described in *S. cerevisiae*, once again exemplifying its benchmark status as a genetic model organism; however, whether these examples represent functionality for ncRNAs in general or exceptional cases in, what otherwise are spurious transcription events, has yet to be fully resolved (Brosius, 2005).

1.5.1. The genome is pervasively transcribed:

Initial unbiased interrogation of the transcriptome revealed that rather than being exclusive to protein-coding regions and infrastructural RNAs, transcription is pervasive and widespread, with estimates ranging from 85% to 98% of nuclear DNA being transcribed into RNA in yeast and humans, respectively (Bertone et al., 2004; Birney et

al., 2007; Carninci et al., 2005; David et al., 2006). In yeast, this is echoed by genome-wide RNAPII chromatin immunoprecipitation (ChIP) data, which detected polymerase in non-genic regions (Steinmetz et al., 2006). This volume of non-coding transcription is in contrast to the protein-coding portion of most genomes; for instance, transcription of protein-coding DNA is estimated at approximately 1-2% in humans (Cheng et al., 2005). So, while proteins might comprise the most obvious structural and functional units of cells, they make up a proportionally minor fraction of the genetic load. It is possible that the majority of this non-coding transcription is non-functional and is simply fortuitous (Brosius, 2005). However, with transcriptional coverage verging on full capacity, the explanation that the generation of these RNAs is just biological noise seems inadequate.

One indication for the functionality of widespread non-coding transcription is the existing positive correlation between organism complexity and the proportion of non-coding DNA, which ranges from ~12% in prokaryotes, ~30% in *S. cerevisiae,* ~70% in plants, to ~98% in humans (Taft & Mattick, 2003). Thus, a plausible role for non-coding transcription is to contribute fine-tuned regulation of the genome's protein-coding fraction, such that the greater the proportion of non-coding DNA, the greater the capacity for nuanced phenotypes. How this complexity is implemented remains the central question, though it likely exerted through several different mechanisms, given the lack of consensus in non-coding RNA function. Accordingly, across eukaryotic taxa, non-coding RNAs vary in length, origins, biogenesis, transcriptional and post-transcriptional regulation, and have been shown to act in both *cis* and *trans*.

1.5.2. Origin and biogenesis of long non-coding RNAs:

Ostensibly, in eukaryotes there are two arbitrary classes of non-coding RNAs based on size: small non-coding RNAs that are shorter than 200 nt and long non-coding RNAs that are greater than 200 nt in length. In recent years, short RNAs have dominated the literature, in particular with the discovery of small interfering RNAs (siRNA) and RNA interference (RNAi) in *Caenorhabditis elegans*, a process by which generation of siRNAs from double-stranded RNA (dsRNA) results in the in *trans* silencing of genes possessing complementary sequences (Fire et al., 1998). Since this discovery, similar RNA interference-mediated processes have been discovered in other genetic model organisms such as fission yeast, plants, *Drosophila*, zebrafish, frogs, mice and humans (reviewed in Brantl, 2002). Curiously, *S. cerevisiae* is lacking the genes for the molecular machinery of this pathway (Aravind, Watanabe, Lipman, & Koonin, 2000). Since it is devoid of the components necessary for RNAi, *S. cerevisiae* serendipitously provides a simplified model organism to study long non-coding RNAs.

In *S. cerevisiae*, long non-coding RNAs can be generated from intra- and intergenic regions and can partially or fully overlap annotated regions on either strand, with some transcripts displaying polycistronic architecture (David et al., 2006). Three non-mutually exclusive subclasses, distinguished according to their sensitivity to different RNA decay pathways, are typically used to describe these RNAs (Figure 3A, B, C). The most widely characterized are the cryptic unstable transcripts (CUTs), mainly small (<1Kb) transcripts

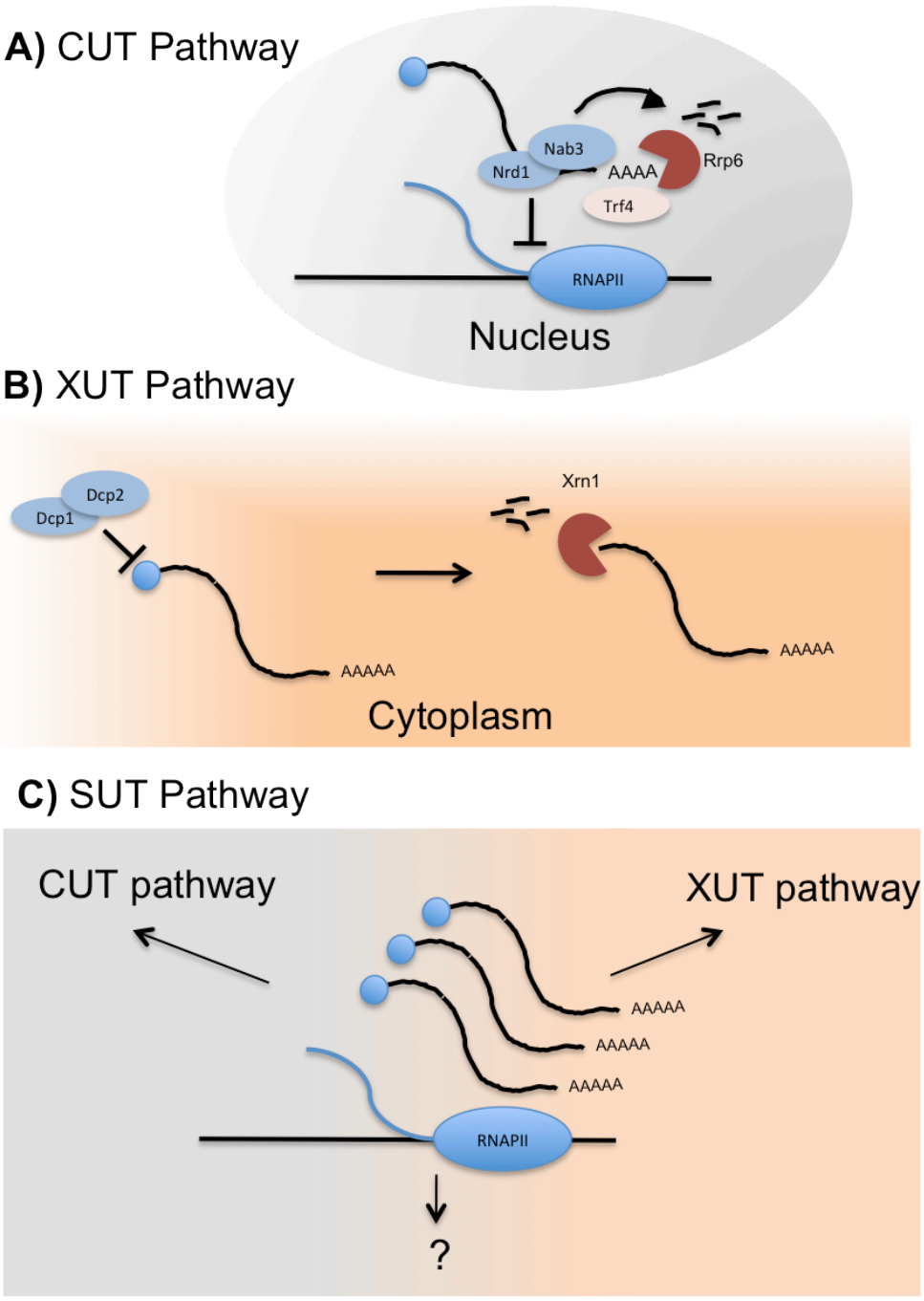


Figure 3. Major classes of long non-coding transcripts in *Saccharomyces cerevisiae*.

In *S. cerevisiae* there are 3 major classes of long non-coding RNAs based on sensitivity to decay pathways: **(A)** Cryptic unstable transcripts (CUTs) are degraded in the nucleus by nuclear exosome component Rrp6p; **(B)** XUTs (Xrn1 unstable transcripts) are sensitive to the Xrn1p-mediated cytoplasmic decay pathway; and **(C)** SUTs (Stable untranslated transcripts) are relatively stable non-coding RNAs that can accumulate before being degraded by different pathways.

antisense to annotated genes, which are subject to degradation by the nuclear exosome, the nuclear form of a large complex of 3'-5' exonucleases (Davis & Ares, 2006; Wyers et al., 2005). CUTs, by means of RNA recognition motifs (RRM), are specifically targeted by the alternative Nrd1-Nab3-Sen1 termination pathway (Arigo, Eyler, Carroll, & Corden, 2006; Thiebaut, Kisseleva-Romanova, Rougemaille, Boulay, & Libri, 2006), which in turn recruits the TRAMP (Trf4p-Air1p-Air2p-Mtr4p) complex, resulting in Trf4p-mediated transcript oligo-adenylation, essentially marking the transcripts for degradation by the nuclear RNA decay pathway (LaCava et al., 2005; Vanácová et al., 2005; Vasiljeva & Buratowski, 2006; Vasiljeva, Kim, Mutschler, Buratowski, & Meinhart, 2008; Wyers et al., 2005). This is in contrast to stable coding RNAs, whose termination/cleavage and polyadenylation is mediated by the Polyadenylation Cleavage Factors (CPF) (Birse, 1998). Since CUTs are degraded almost immediately after synthesis, only when the nuclear exosome is perturbed by deletion of its catalytic subunit *RRP6*, can these transcripts accumulate to detectable levels (Burkard & Butler, 2000; Davis & Ares, 2006; Xu et al., 2009).Moreover, native elongating transcript sequencing (NET-seq), a novel technique that measures nascent transcription in a strand-specific manner, corroborated the generation of these unstable transcripts in wild-type cells (Churchman & Weissman, 2011).

A second class of unstable ncRNAs displays sensitivity to the cytosolic mRNA decay pathway mediated by the de-capping enzyme complex Dcp1p-Dcp2p and the 5'-3' exonuclease Xrn1p (Geisler, Lojek, Khalil, Baker, & Coller, 2012; Tourrière, Chebli, & Tazi,

2002; van Dijk et al., 2011; Wilusz, Wormington, & Peltz, 2001). Transcripts belonging to this class are called Xrn1 unstable transcripts (XUT) (Van Dijk et al., 2011). Unlike the CUTs, which must be stabilized by perturbing the nuclear exosome, some XUTs can be detected at low levels without deletion of *XRN1* (Van Dijk et al., 2011). These quasi-stable transcripts partially represent a third class of low-abundance stable non-coding RNAs, or stable unannotated transcripts (SUTs), which can be detected in wild-type cells (Xu et al., 2009).

Regardless of post-transcriptional fate, most of these non-coding RNAs are transcribed from 5' NDRs that typically overlap with the 5' NDRs of other annotated or unannotated transcripts (Xu et al., 2009). Strikingly, 63% of all divergent transcript pairs containing at least one non-coding transcript, and 66% of all divergent protein-coding transcript pairs manifest as a result of divergent transcription from a shared 5' NDR, strongly indicating that promoter regions in yeast are inherently bidirectional (Xu et al., 2009). Moreover, bidirectionality is not unique to the yeast genome as studies in humans predict that more than 10% of promoters are bidirectional, likely an underestimate since this analysis only considered coding gene pairs (Trinklein et al., 2004). In both yeast and humans, most bidirectional transcript pairs display a modest propensity for co-expression, with some pairs exhibiting anti-regulation (Churchman & Weissman, 2011; Trinklein et al., 2004; Xu et al., 2009). Whatever the relationship between divergent gene pairs, the use of bidirectional promoters and consequent non-coding transcription

comprises an additional aspect to bear in mind in the organization of transcription regulatory networks.

1.5.3. Overlapping antisense transcription:

As a corollary of bidirectional promoters generating pervasive transcription and the compact nature of the yeast genome, one might expect to observe high degrees of transcriptional overlap, such that a single segment of DNA might occupy multiple transcriptional contexts (Figure 4). While overlap can occur to varying degrees on either the same or opposite strand of the DNA duplex, one of the more striking examples is the production of non-coding antisense RNAs from the opposite strands of coding templates. Generally, these transcripts initiate from the 3' ends of an ORF, are transcribed antisense to the coding RNA, and terminate at the 5' end of the gene, constituting a conjugate sense/antisense pair (SAP). Like other classes of ncRNAs, antisense transcripts are subject to different modes of post-transcriptional regulation including CUT and XUT pathways, which might also bear consequences on their functionality (Neil et al., 2009; van Dijk et al., 2011).

Implicitly, the architecture of antisense-producing transcription units requires that regulatory sequences at the 5' and 3' ends of the coding gene be capable of directing both transcription initiation and termination. Indeed, parallels between the 5' and 3' ends of genes are apparent, particularly in terms of chromatin architecture. For instance, similar to the 5' end NDR, a ~173 bp 3' NDR can be found at the 3' ends of 95% of genes in *S. cerevisiae* (Mavrich et al., 2008) (Figure 5). While these 3' NDRs might function in transcription termination, it was also postulated that they might help

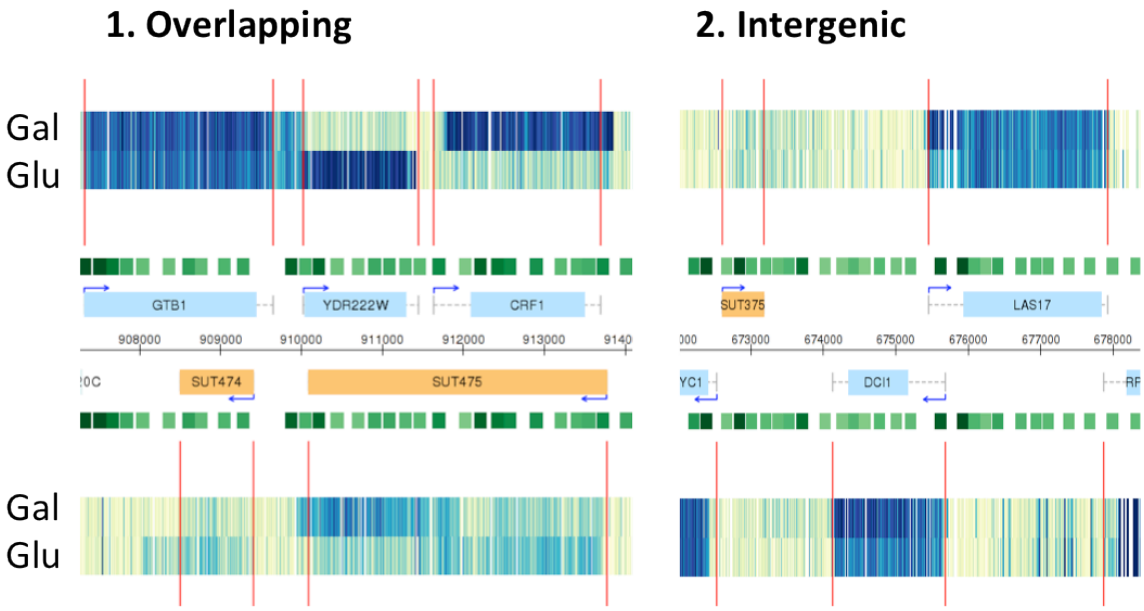


Figure 4. Overlapping transcription in a compact genome

Unbiased interrogation of the transcriptome demonstrated that transcription is not limited to genes, but is overlapping on both strands in regions of both coding and non-coding potential, such as **(1)** strands opposite to coding genes or **(2)** intergenic regions. No hybridization is indicated in yellow, hybridization is indicated in blue. Polyadenylated RNA (purified by Ana Serra Barros) from cells grown in glucose- (Glu) and galactose-containing media (Gal) was hybridized to strand-specific high resolution tiling arrays in collaboration with Lars Steinmetz at the European Molecular Biology Laboratories.

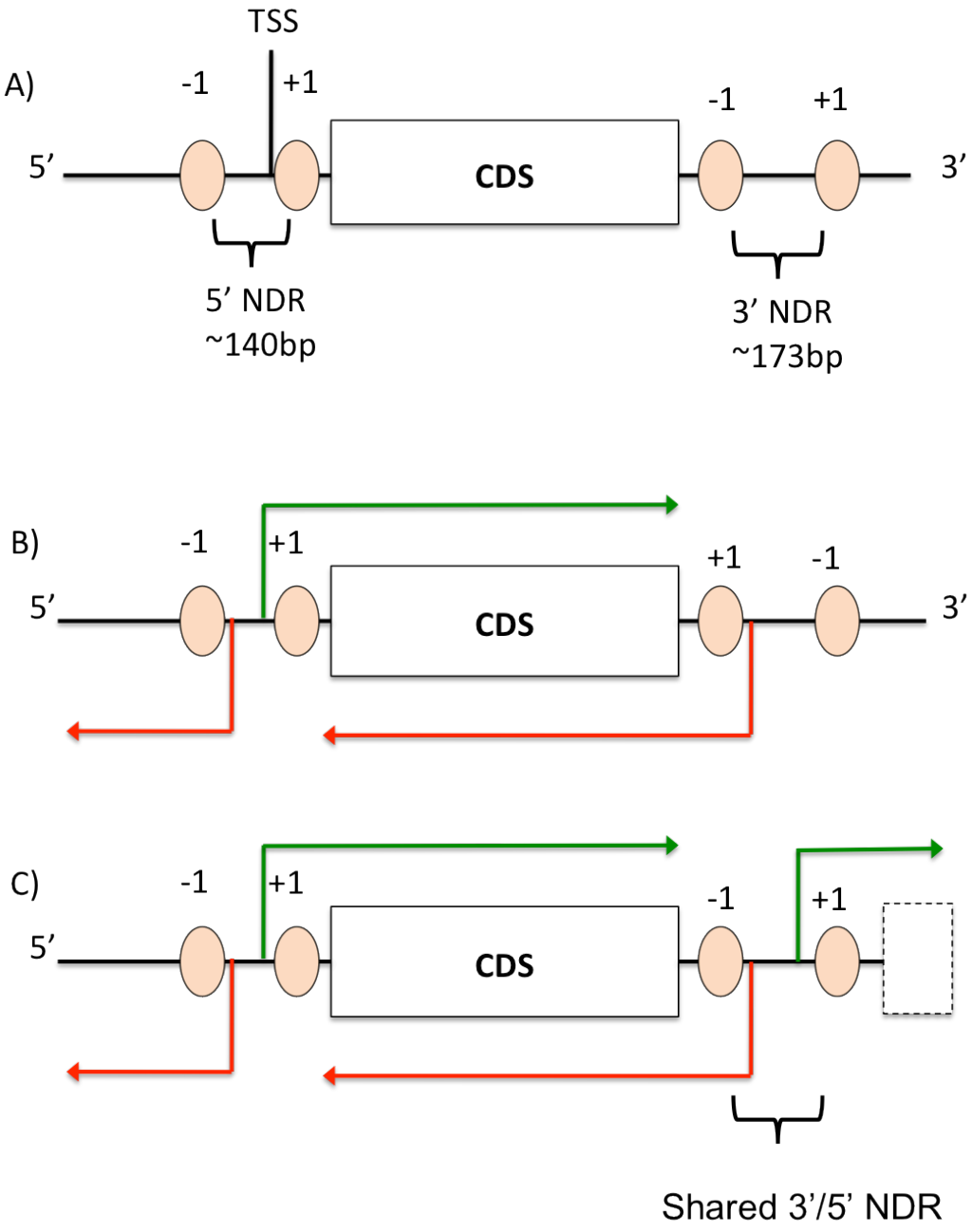


Figure 5. Nucleosome depleted regions support sense and antisense transcription.

(A) Promoter and terminator associated NDRs are flanked by an upstream -1, and an immediately downstream +1, H2A.Z containing nucleosome. **(B)** 3' NFRs are capable of producing antisense transcripts. **(C)** The region of a 3' NFR of a gene is sometimes the 5' NFR of the neighbouring, tandemly oriented gene.

accommodate antisense initiation by acting as antisense promoters (Mavrich et al., 2008). Moreover, work by Venters & Pugh (2009) demonstrated that these antisense promoters adopted the same canonical "-1, NFR, +1" H2A.Z-containing 5'NFR arrangement and were also characterized by the presence of PICs or partial RNAPII-deficient PICs, further highlighting similarities between the 5' and 3' ends of genes (Venters & Pugh, 2009b). In line with these observations, it was found that most antisense RNAs arise from within the 3' NDRs of their coding gene pairs (Mavrich et al., 2008; Xu et al., 2009).

Some studies have challenged the prevalence of NDRs at the 3' ends of transcription units, indicating that chromatin digestion using micrococcal nuclease (MNase), a technique employed by most of the early nucleosome mapping studies, creates biases for A-T rich sequences, such as terminator regions (Fan et al., 2010). Using sonicated chromatin to avoid these sequence biases, Fan and colleagues (2010) contend that 3' end nucleosome depletion exists, but is much less pronounced and not a ubiquitous feature of the 3' ends of genes. Rather, the authors found that 3' NDRs are highly correlated with the presence of downstream genes oriented in tandem. This finding reflects another common arrangement supporting antisense transcription, in which antisense producing 3' NDRs overlap the 5' NDR of an adjacent gene (Xu et al., 2009). These reports suggest that antisense transcription and its associated promoter architecture are artefacts of the bidirectional nature of gene promoters. However, a later genome-wide study by Murray and colleagues (2012) used the Fan et al. (2010)

dataset to demonstrate that at antisense-producing units in particular, the 3' end of the gene was characterized by marked nucleosome depletion that was independent of the 5' NDRs of downstream genes. Furthermore, the study demonstrated that these antisense-producing 3' NDRs supported PIC formation at levels corresponding to detectable levels of nascent antisense transcription (Murray et al., 2012). What this evidence suggests, is that antisense promoters may exist as discrete regulatory entities that are susceptible to similar forms of regulation as mRNAs discussed earlier, perhaps favouring a functional role for their generation.

1.5.4. Consequences of compact transcription:

By virtue of their shared physical space and sequence complementarity, it is widely thought that antisense transcription might regulate its sense pair, potentially representing a subclass of a functional non-coding transcription. A few studies in yeast have characterized some unique examples of regulatory non-coding antisense RNAs, which will be discussed in greater detail in the following chapters. Briefly, these functions ranged from roles in galactose and phosphate metabolism, to controlling meiotic entry (Camblong, Iglesias, Fickentscher, Dieppois, & Stutz, 2007; Hongay, Grisafi, Galitski, & Fink, 2006; Houseley, Rubbi, Grunstein, Tollervey, & Vogelauer, 2008; Uhler, Hertel, & Svejstrup, 2007). The accompanying molecular mechanisms, if available, were just as disparate, varying from histone deacetylation of the sense promoter through the Hda1 (histone deacetylase 1) or Rpd3S deacetylase complexes to nucleosome eviction and sense activation. A general function for antisense transcription in generating greater expression variability for the sense gene has been demonstrated; however, by

undefined mechanisms (Xu et al., 2011). A communication by Brosius (2005) contends these individual examples of regulatory non-coding RNAs are more likely to reflect exceptional cases of natural selection, rather than general functionality.

Alternatively, rather than assessing the function of non-coding transcription in terms of its end product (i.e. non-functional RNA, CUT/XUT/SUT/antisense), it is possible that the very process of transcribing these RNAs manifests in some general regulatory consequences, in effect broadening the capacity of non-coding transcription to affect gene regulation. For instance, transcription interference (TI), is the process by which transcription over one region of DNA can directly antagonize transcription from a neighbouring region in *cis* and is applicable to transcription generated from the same or opposite DNA strands (Callen, Shearwin, & Egan, 2004; Martens, Laprade, & Winston, 2004). While most studies will agree on the outcome of transcription interference, there is still much debate surrounding how the repression is exerted (reviewed in Shearwin, Callen, & Egan, 2005). The most widely accepted model for repressive transcription interference is “promoter occlusion”, or when an “aggressor” promoter induces transcription so that the subsequent travelling RNA polymerase traverses a second “sensitive” promoter, which then suffers impaired initiation (Figure 6A). Still, there exist several competing views regarding the exact molecular mechanism behind promoter occlusion.

The conventional and earliest understanding of promoter occlusion was described in a study looking at the effects of an upstream prophage lambda promoter on a

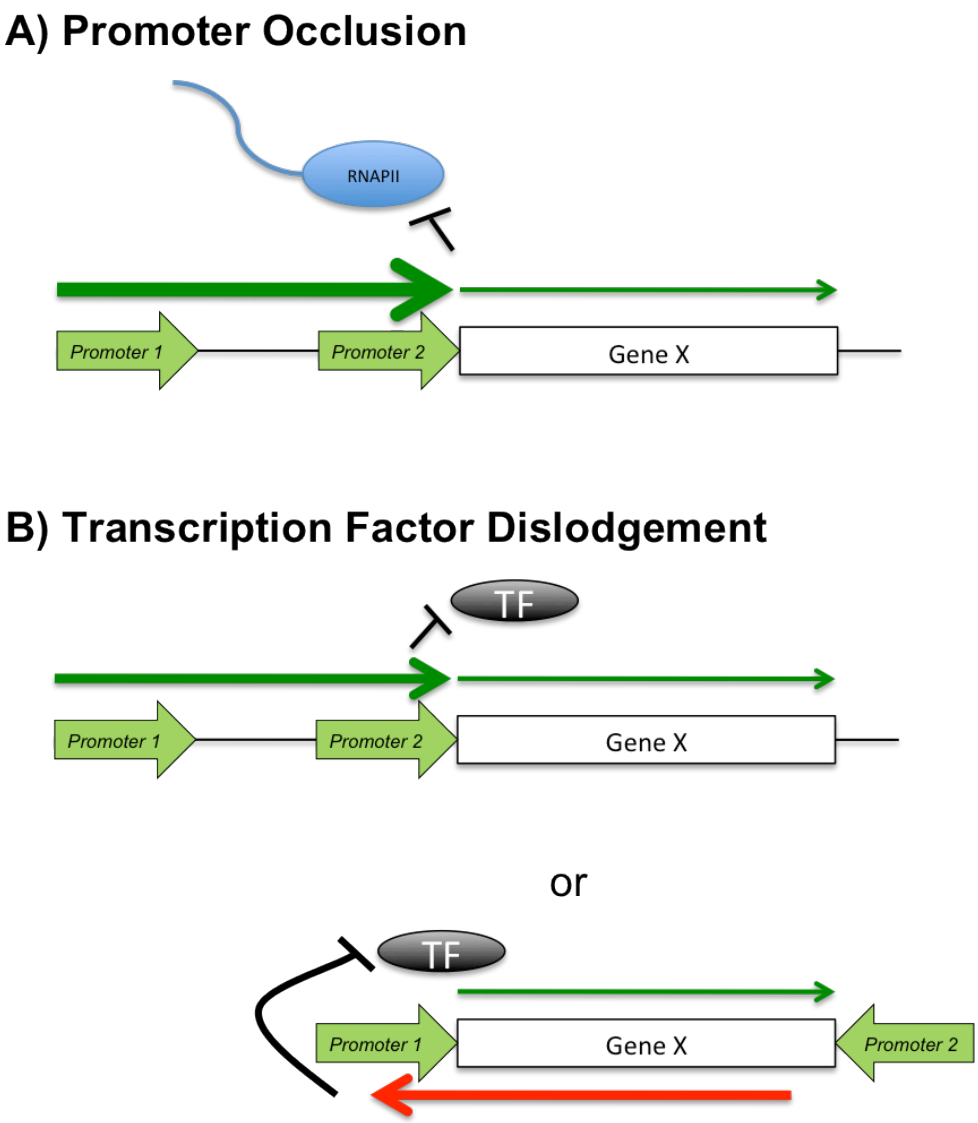


Figure 6. Mechanisms of transcription interference.

Schematics of two widely accepted models for transcription interference. **(A)** Promoter occlusion occurs when a stronger upstream promoter initiates transcription at a rate that greatly exceeds the initiation rate of a downstream promoter, resulting in occupation of the promoter and prevention of further polymerase binding. **(B)** Alternatively, transcription from a stronger promoter can antagonize transcription factor binding of a downstream promoter, though a consensus mechanism has not yet been defined. Transcription factor dislodgement is thought to occur on the same or opposite strands.

downstream *E. coli gal* operon promoter. Using this model system, the authors indicated that promoter occlusion occurs when an RNA polymerase complex, driven by an upstream promoter, transiently occupies a downstream promoter, physically blocking the binding of other RNA polymerase molecules by either steric hindrance or altering the structure of the promoter DNA (Adhya & Gottesman, 1982). The study concedes that this type of control is limited by the strength of the upstream promoter, with the aggressor promoter having to initiate transcription much more frequently than the sensitive promoter, otherwise the time of polymerase occupation is too short and unlikely to manifest in extensive repression.

Later work in *S. cerevisiae* by Greger and coworkers (2000) examined promoter occlusion of the *GAL7* promoter as a result of read-through transcription from upstream gene, *GAL10* (Greger, Aranda, & Proudfoot, 2000). In this instance, the authors suggested that the occlusion and the resultant drop in transcription from the *GAL7* promoter was a function of impaired interactions between the Gal4p activator and the *GAL7* promoter DNA, or promoter occlusion by transcription factor displacement (Figure 6B). A later communication by Palmer and colleagues (2011) distinguishes this type of interaction at the *GAL7* promoter as “transcription factor dislodgement”, a separate, but not isolated, consequence of interfering transcription distinct from the promoter occlusion described by Adyha & Gottesman (1982) (Palmer, Egan, & Shearwin, 2011). By using mathematical modelling, Palmer and colleagues (2011) further assert that transcription factor dislodgement in concert with promoter occlusion in its canonical

sense, but enhanced by polymerase pausing to create lengthier promoter occupation times, could generate sustained interference at promoters. Alternatively, given what is known about co-transcriptional remodelling/modification of chromatin architecture, it is also possible that interfering transcription creates chromatin domains unfavourable to initiating transcription. Evidence for this type of chromatin-mediated transcription interference is discussed in much greater detail in Chapter 5.

With these mechanisms in mind, a highly compact genome, as exemplified by *S. cerevisiae,* would create a favourable arena for transcriptional interference to impose a basic level of regulation on genes, made more likely given the extent of non-coding transcription. In essence, through transcription interference, a pervasively transcribed genome creates physical connections between transcription units, both coding and non-coding, with sense/antisense pairs potentially representing the most intimate association existing between two transcription units. Furthermore, knowing the relationships between protein-coding genes in either temporally-regulated physiological programmes or chromatin-mediated spatial domains, it is likely that transcription of non-coding units participates in these highly integrated networks of gene transcription.

1.6. Perspectives on lncRNAs in humans:

Like *S. cerevisiae*, the genomes of higher eukaryotes such as mice and humans exemplify an interleaved architecture, which gives rise to pervasive non-coding transcription (reviewed in Kapranov, Willingham, & Gingeras, 2007). There are a few classical examples of non-coding RNA-mediated gene regulation in mammalian and human

development, with the best-studied examples bearing pivotal roles in managing dosage compensation. For instance, X chromosome inactivation and genomic imprinting are two developmentally crucial processes, which, through still somewhat controversial mechanisms, involve non-coding RNAs. In both cases, failure to properly manage dosage produces some of the most common human congenital disorders including Rett, fragile X, Prader-Willi/Angelman, and Beckwith-Wiedemann syndromes (reviewed in J. T. Lee & Bartolomei, 2013).

1.6.1. X-inactivation:

X chromosome inactivation (XCI) is the process by which one copy of the X chromosome is silenced in female XX cells, manifesting as monoallelic expression from the active X (a), and strikingly, a condensed inactive X chromosome (Xi) known as the Barr body. Mechanistically, XCI is managed through a single 100-500 kb *cis*-acting X inactivation center (*Xic*). A host of lncRNAs is expressed in the vicinity of the *Xic*, the most prominent of them being *Xist* and its antisense pair *Tsix*, and these lncRNAs control the establishment of repressive chromatin (reviewed in Pauli, Rinn, & Schier, 2011). In brief, expressed *Xist* RNA binds to the Polycomb repressive complex 2 (PRC2), directing PRC2 to the X chromosome destined for inactivation, where it deposits the repressive H3K27me3 mark. Interestingly, *Tsix* transcription marks the active X by negatively controlling the effects of *Xist* expression through multiple mechanisms. For instance, *Tsix* RNA expression acts to silence expression from the *Xist* promoter in *cis* through recruitment of the DNA methyltransferase, Dnmt3a. Additionally, *Tsix* can antagonize *Xist* interaction with PRC2, potentially by itself binding and sequestering PRC2,

stoichiometrically preventing *Xist*/PRC2 binding (reviewed in J. T. Lee & Bartolomei, 2013). Thus, XCI demonstrates that a long non-coding RNA can direct silencing of an entire chromosome, while its antisense prevents this silencing.

1.6.2. Genomic Imprinting:

Like XCI, genomic imprinting manifests as monoallelic expression, such that regulation of genes localized within a discrete domain is coordinated in a parent-of-origin dependent manner. There are four well-characterized imprinting clusters in humans: the *H19/Igf2*, *Kcnq1*, *Igf2r*, and *Snprn* domains, with each of them harbouring lncRNAs that are also imprinted (reviewed in J. T. Lee & Bartolomei, 2013). Imprinted clusters are controlled by an imprint control region or element (ICR/ICE) that is differentially epigenetically modified (e.g. histone modifications/covalent DNA methylation) in a parent-of-origin dependent fashion. Like the XIC, these lncRNAs originate in the vicinity of these ICRs . In each case, it appears as though the lncRNA acts in *cis* to help establish an imprinted state. For instance, at the *Igf2r* cluster, the lncRNA *Airn* is expressed in *cis* from the paternal chromosome and is required to silence all three coding genes situated in the cluster (Pauler, Barlow, & Hudson, 2012). The prevailing hypothesis is that these non-coding RNAs create sense/antisense overlap with imprinted mRNA genes causing transcriptional interference at a regulatory region, which then directs the establishment of repressive chromatin domains that induces silencing. Indeed, truncating *Airn* by using early polyadenylation signals, such that the *Igf2r* promoter is not overlapped, results in the loss of *Igf2r* repression (Sleutels, Zwart, & Barlow, 2002). A study by Latos et al. (2012) later demonstrated that *Airn*-mediated imprinting of *Igf2r* is dependent on

overlap of the *Igf2r* promoter, which interferes with RNAPII recruitment (Latos et al., 2012).

Taken together, XCI and genomic imprinting in humans demonstrate that lncRNAs play an important role in developmentally important gene regulation. Furthermore, non-coding antisense RNAs have been implicated in epigenetic silencing of tumour suppressor genes (TSG) in tumourigenesis (Yu et al., 2008). It is apparent that, as in *S. cerevisiae*, non-coding transcription in humans is rampant across the genome and linked to multiple facets of biology. Additionally, in both organisms, the interplay between non-coding transcription and chromatin architecture is at the heart of regulating protein-coding genes. Thus, perhaps the mechanisms and non-coding phenomena described in this study will bear relevance to models of human disease involving non-coding transcription.

1.7. Aims:

The aim of this work is to evaluate transcription unit regulation in consideration of the genetic complexities described in recent years. By implementing a systematic approach, this report endeavours to integrate different aspects of gene regulation, including regulation in *trans* by gene-specific transcription factors, remodelling and modification of chromatin, spatial influences on co-expression, and temporal compartmentalization of gene expression, but most importantly, in the context of a pervasively transcribed, non-coding genome. Detailed aims are found at the beginning of each results section.

Chapter 2.

Methods and Materials.

2. Methods and Materials

A multidirectional approach was employed in order to deduce the relationship between sense and antisense transcription. Techniques included strand-specific Northern blot analysis, 3' RACE, strand-specific high-resolution tiling array analysis, and single molecule RNA FISH. Unless otherwise stated, all experiments were performed at least in duplicate to ensure that the trends observed were reproducible.

2.1. Media Composition:

Unless otherwise stated, cells were grown to exponential phase in rich medium (YP) (1% bactpopeptone, 1% Difco yeast extract) supplemented with 2% glucose (YPD) or 2% galactose (YPGal) for galactose inducing conditions (described below).

2.2. Yeast strains and transformation:

The strains of *S. cerevisiae* used for this study are shown in Table 1. Genetic manipulation of strains was carried out as described in Longtine et al. 1998. Polymerase chain reaction products engineered for deletion strains were designed to contain either a *kanMX6 or HISMX6* expression cassette, which consisted of either the *kan*^{r} or *HIS5* selectable marker gene flanked by the *TEF* promoter and terminator sequences from *Ashbya gossypii,* as well as 40 bp termini that were homologous to the flanking intergenic regions of the target gene (Longtine et al., 1998). Generally, deletion strains encompassed the entire coding region of the target gene, from start to stop codon. *HMS2* disruption constructs were generated using either the *pFA6a-GFP(S65T)-kanMX6*

or *pFA6a-kanMX6* plasmids, in order to create disruption strains with and without incorporating a 5' *ADH1* terminator, respectively. The resultant strains featured a 10bp deletion (651-660 bp) in the *HMS2* coding region, substituted by *kanMX6*-containing constructs. To generate disruption strains in the Δ5'^{P}*TEF* mutation series, PCR amplification was carried out using forward primers that excluded various segments of the *TEF* promoter. Four-primer site-directed mutagenesis was performed in order to generate the *TEF* terminator TATA box mutant disruption strains. The *URA3* substitution strain was engineered using the pRS306 plasmid to generate a complete *HMS2*::*URA3* substitution strain, without altering the 5' or 3' intergenic regions or introducing any exogenous regulatory sequences to the *HMS2* locus. *pGAL1* driven strains were engineered using the pFA6a-*HISMX6-pGAL1* plasmid.

Competent cells were prepared overnight in 0.1 M lithium acetate (LiAc) and transformed in 0.1M LiAc/40% polyethylene glycol (PEG) using 10µl of the appropriate PCR product, purified using the QIAquick PCR purification columns (Qiagen), as specified by the manufacturer. Transformed cells were subjected to two rounds of positive selection for kanamycin resistance, *HIS* or *URA* competence. Over-night cultures were prepared from positive clones in 1 ml of YPD. DNA from these cultures was extracted using phenol:chloroform:isoamyl alchohol and subjected to diagnostic PCR. PCR products were run on 1% agarose gels in 1X TBE running buffer at 100 V for 20 minutes. Based on size discrimination, positive clones containing the desired manipulation were prepared in glycerol (Anitha Nair) and stored at -80°C. To ensure the correct deletion,

PCR products displaying the correct sizes for the Δ5'^{P}*TEF* deletions were purified and sent for sequencing (Source Bioscience, Oxford).

Strain Name	Parent Strain	Genotype	Origins
BY4741 (WT)	S288C	*MATa; his3Δ 1; leu2Δ 0; met15Δ 0; ura3Δ*	Euroscarf
pGAL1.HMS2	BY4741	*pHMS2::HisMX6-pGAL1*	This Study
rrp6 Δ	BY4741	*rrp6::KanMX*	Euroscarf
xrn1 Δ	BY4741	*xrn1::KanMX4*	Open Biosystems
hda1 Δ	BY4741	*hda1::KanMX6*	David Clynes
cat8 Δ	BY4741	*cat8::KanMX4*	Open Biosystems
adr1 Δ	BY4741	*adr1::KanMX4*	Open Biosystems
cat8 Δ adr1 Δ	*cat8 Δ*	*adr1::HisMX6*	This Study
HHY168	W303	HHY168 *MATα tor1-1 fpr1::NAT RPL13A-2×FKBP12::TRP1*	Euroscarf
TFIIB-FRB	HHY168	*Sua7-FRB::kanMX6*	This Study
hms2 Δ	BY4741	*hms2::KanMX4*	Open Biosystems
HMS2.kan	BY4741	*HMS2 Δ(651-660)::kanMX6*	David Brown
HMS2.ADH	BY4741	*HMS2 Δ(651-660)::^{T}ADH1.kanMX6*	David Brown
HMS2.ADH rrp6 Δ	*HMS2.ADH*	*rrp6 Δ::HISMX6*	This Study
Δ5'^{P}*TEF*-1, 1a	BY4741	*HMS2 Δ(651-660)::kanMX6 Δ(1-217)*	This Study
Δ5'^{P}*TEF*-2, 2a	BY4741	*HMS2 Δ(651-660)::kanMX6 Δ(1-120)*	This Study
Δ5'^{P}*TEF*-3, 3a	BY4741	*HMS2 Δ(651-660)::kanMX6 Δ(1-73)*	This Study
Δ5'^{P}*TEF*-1'	BY4741	*HMS2 Δ(651-660)::kanMX6 Δ(1-167)*	This Study
Δ5'^{P}*TEF*-1''	BY4741	*HMS2 Δ(651-660)::kanMX6 Δ(1-181)*	This Study
xTATACAAA	BY4741	*HMS2 Δ(651-660)::kanMX6* (TATACAAA::TTTCAAAA)	This Study
xTATAAAAA	BY4741	*HMS2 Δ(651-660)::kanMX6* (TATAAAAA::AAAAAATT)	This Study
URA3 xrn1 Δ	*URA3*	*xrn1 Δ::kanMX6*	Anitha Nair
URA3 rrp6 Δ	*URA3*	*rrp6 Δ::KanMX6*	This Study
pGAL1.URA3	*URA3*	*pHMS2::HisMX6-pGAL1*	This Study
HHY183	HHY168	*GAL4-FRB::KanMX6*	Euroscarf
HHY154	HHY168	*TBP-FRB::KanMX6*	Euroscarf
pGAL1.HMS2 Gal4-FRB	HHY183	*pHMS2::HisMX6-pGAL1*	This Study

pGAL1.HMS2 TBP-FRB	HHY154	*pHMS2::HisMX6-pGAL1*	This Study
gal4 Δ	BY4741	*gal4 Δ::kanMX*	Euroscarf
pGAL1.HMS2 gal4 Δ	*gal4 Δ*	*pHMS2::HisMX6-pGAL1*	This Study
FY2181 (WT Spt6)		*MATa leu2D1 his4-912d lys2-128d FLAG-Spt6*	F. Winston
FY2180 (*spt6-1004*)	FY2181	*FLAG-spt6-1004*	F. Winston

Table 1. Genotypes of yeast strains used in this study.*

*Additional strains of interest not mentioned in this thesis are indicated in Table 7 in the Appendix. These strains could also be used to manipulate levels of *HMS2* sense and antisense.

2.3. Cell culture:

Yeast strains stored in glycerol at -80°C were streaked onto 2% agar, YPD plates and incubated at 30°C for 48 hours (h). Liquid cultures were then prepared in YPD and incubated overnight at 30°C, rotating at 200 rpm. These overnight cultures were used to inoculate fresh YPD so that cultures started at an OD_{600} (optical density) of 0.2. These cells were grown to exponential phase at 30°C, 200rpm or an approximate OD_{600} of 0.5. Exponential phase cells were harvested and stored at -20°C or immediately used for RNA extraction or other downstream application. Unless otherwise indicated, galactose induction was carried out using exponentially growing cell cultures (OD_{600}= 0.5), which were pelleted, removed of YPD, and then transferred to yeast extract/peptone containing 2% galactose (YPGal). Cells were grown in galactose and harvested after 5 minutes (min), 30 min, 1 h, 2 h, 3h, or 4 h, as specified by the experiment. In the case of cell cycle arrest experiments, cells were grown for 3 h in YPD, after which they were treated with either α-factor (600 ng/ml final concentration) or nocodazole (15 ng/ml

final concentration) for 2.5 h. The α-factor pheromone arrests Mat a cells in late G1 while nocodazole arrests cells in the G_2/M transition of the cell cycle.

2.4. Anchor Away:

The anchor away technique (Figure 7) exploits the high affinity interaction between the FK506 Binding Protein (*FKBP12*) and *FKBP12* Rapamycin binding (FRB) domain of human mTOR protein in order to deplete target proteins from the nucleus in a time-dependent manner (Haruki, Nishikawa, & Laemmli, 2008). Rapamycin binds to the *FKBP12* domain, providing a platform for interaction with FRB, which subsequently binds to form a ternary complex of nanomolar dissociation constant (J. Chen, Zheng, Brown, & Schreiber, 1995). Since rapamycin, a potent antifungal, is toxic to wild-type yeast, anchor-away strains are mutant for endogenous *TOR1* (*tor1-1*) and have *FPR1* deleted, the yeast homolog to human *FKBP12*. The *FPR1* deletion also minimizes competition for *FRB* binding in the presence of rapamycin. To deplete TFIIB from the nucleus, TFIIB was *FRB*-tagged by homologous recombination as described earlier using the *pFA6α-FRB-kanMX6* plasmid (Euroscarf). To deplete Gal4 and TBP from the nucleus in strains containing the *pGAL1.HMS2* construct, the *pGAL1.HMS2* construct was transformed into the HHY183 (Euroscarf) or HHY148 (Euroscarf) strain backgrounds, respectively. The HHY183 and HHY154 strains feature the human FRB domain C-terminally fused to either Gal4p or TBP, respectively, and the human *FKBP12* C-terminally fused to *RPL13A* in a W303 background carrying the *tor1-1* mutation and *fpr1* deletion. Transformations were carried out as described previously. Rapamycin dissolved in DMSO was added (1 µg/ml final concentration) to *pGAL1.HMS2 GAL4*-FRB cells induced in galactose-containing

medium for 3 h (Rap^+). A control (Rap^-) group was treated with 15 µl of DMSO as a control. Both the Rap^-- control and Rap^+- treated cells were harvested after 1 h of rapamycin exposure. Cell pellets were either stored at -20°C or immediately used for RNA extraction.

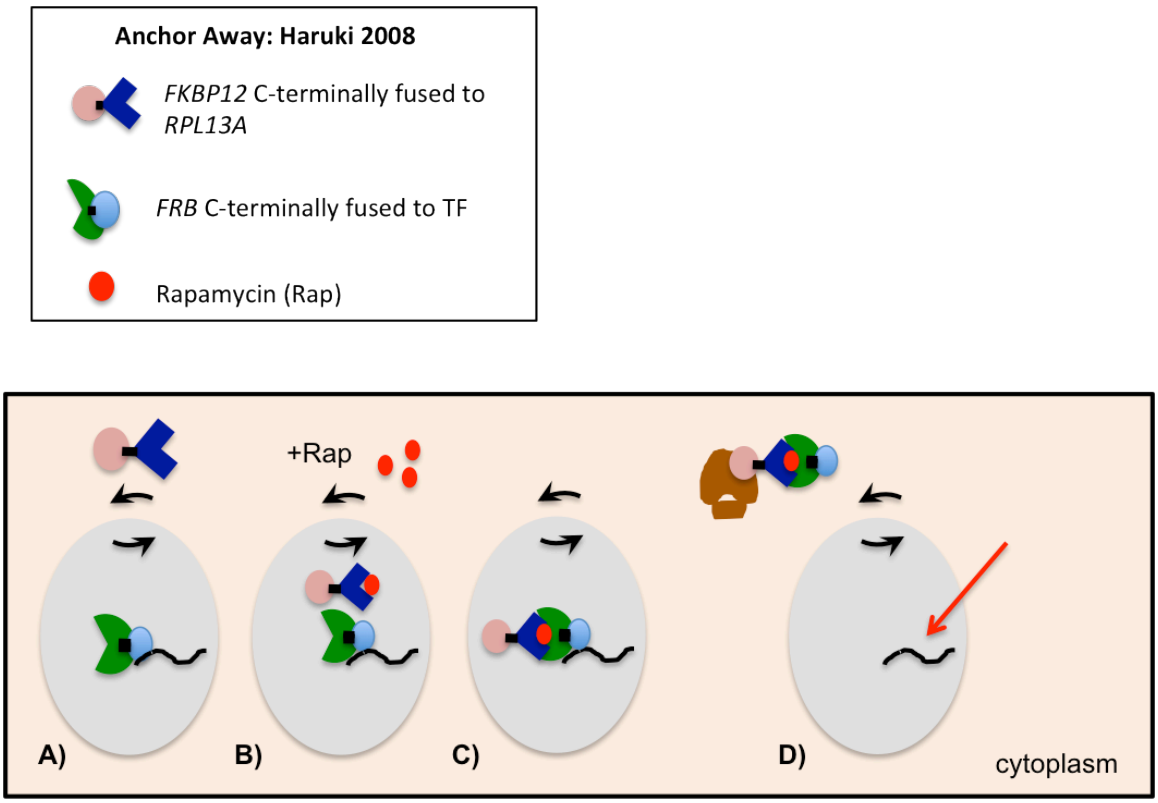


Figure 7. The Anchor Away Technique

Schematic diagram of the mobile Anchor Away scheme as described by Haruki et al (2008). **(A)** The anchor, RPL13A, is C-terminally fused to the human FKBP12, while the target protein (in this case Gal4p) is C-terminally fused to the FRB domain of human mTOR. **(B)** RPL13A shuttles from the cytoplasm to the nucleus (grey circle) during assembly of the 40S and 60S ribosomal particles. **(C)** In the presence of rapamycin, FKBP12-bound rapamycin is shuttled to the nucleus where it is brought into contact with the FRB-fused target protein, with which it forms a stable ternary complex. **(D)** This complex is shuttled out to the cytoplasm, where RPL13A is incorporated into ribosomal particles, thus depleting the nucleus of the target protein.

2.5. RNA extraction:

Total RNA extracts were acquired by hot phenol extraction (Collart & Oliviero, 2001). In some cases, RNA extracts were enriched for polyadenylated transcripts using Oligotex-dT as directed by the manufacturer (QIAGEN). The concentration of RNA extract was measured and standardized to 1000 ng/µl using a nanodrop spectrophotometer.

2.6. 3' RACE:

The protocol for rapid amplification of cDNA 3' ends (3' RACE) used in this study was modified from the methods provided by the 3´ RACE System for Rapid Amplification of cDNA Ends kit (Invitrogen). RNA was extracted and standardised to 1000 ng/µl as described previously. Following treatment with DNase-I, 100 µg RNA was reprecipitated using phenol:chloroform. The upper phase was treated in chloroform, spun down (13,000 rpm, 5 min) at 4°C, after which the supernatant was transferred to a 1.5 ml Eppendorf tube containing 1 ml 100% ethanol/40 µl sodium acetate. Samples were stored at -80°C for 20 minutes before being pelleted (13,000 rpm, 15 min) at 4°C. RNA pellets were washed with 70% ethanol and resuspended in 20 µl water. The concentration of the reprecipitated RNA was measured using a nanodrop spectrophotometer.

Reverse transcription PCR reactions were prepared using ~0.5 µg RNA (~0.4 µl/reaction), 5 µl Adapter primer (AP primer) (10 µM), 1 µl dNTPs (10 mM each), and 7.6 µl water.

Samples were heated in a Thermocycler to 70°C for 10 min. After being cooled and held at 4°C, 4 µl 5x First Strand buffer, 1 µl DTT (1 mM) and 0.25 µl RNase OUT was added to the reactions. Samples were heated to 55°C for 3 min in the Thermocycler, after which 1 µl Superscript III (Invitrogen) was added. Reverse transcription was carried out at 55°C for 1 h. After stopping the reaction by heating samples to 70°C for 15 min, 0.4 µl RNase H was added and samples were kept in the Thermocycler for 30 min at 37°C. cDNA was either stored at -20°C or used immediately for PCR amplification.

Amplification of target cDNAs was carried out in 50 µl reactions, using 2.5 µl of either a forward (to target "sense" cDNAs) or reverse (to target "antisense" cDNAs) gene-specific primer at a 20 µM concentration, and 1 µl of the universal amplification primer (UAP) primer at a 50 µM concentration. Primers were added in excess concentration to increase the specificity of the reaction. Primer sequences for 3'RACE used in this study are displayed in Table 2. The remainder of the amplification buffer was comprised of 5 µl dNTPs (2.5 µM), 5 µl 10x Taq buffer (Sigma), 34 µl water, and 0.5 µl Taq polymerase (Sigma). Amplification was performed in a Thermocycler using 30 cycles, an annealing temperature of 55°C and an elongation time of 30 seconds.

PCR products were loaded onto a 1% agarose gel and visualised under a blue light transilluminator. Bands were cut from the gel and DNA was extracted using the QIAquick Gel Extraction kit (Qiagen). Using the same forward (to sequence sense

transcript 3' ends) or reverse (to sequence antisense transcript 3' ends) primers as in the amplification step, samples were sent for sequencing (Source Bioscience, Oxford).

Primer	**Sequence 5'-3'**
Adapter Primer (AP)	GCCCACGCGTCGACTAGTACTTTTTTTTTTTTTTT
Universal amplification primer (UAP)	GCCCACGCGTCGACTAGTAC
HMS2 FWD (sense transcript 3' end)	CTCCTTTCCAAACATCTTCG
HMS2 REV (antisense transcript 3' end)	AGAGACACACTTGGCGACTG
pGAL1.HMS2 REV (antisense transcript 3' end)	CAAAGACATCAGGCTGCTCC
BAT2 (sense transcript 3' end)	CTGATCCAGCAAGATAAGTG

Table 2. Primers used for 3'RACE.

2.7. Northern blotting:

Total RNA samples (~10 µg) were separated by electrophoresis on formaldehyde 1% agarose gels containing ethidium bromide (EtBr) for 3 or 6 hours at 100V. Formaldehyde running buffers were adapted from Maniatis et al. (1989). To check for RNA degradation, the 25S and 18S ribosomal RNA (rRNA) bands were photographed under UV light using a 0.5 s or 0.1 s exposure. These images also served as loading controls. RNA was transferred onto nitrocellulose membranes through capillary action using a 20X saline-sodium citrate (SSC) buffer as described in Maniatis et al. (1989). Following the transfer, membranes were placed at 80°C for at least 2 h in order to fix the transferred RNA to the membrane. The 25S and 18S ribosomal bands were visualized and marked under UV light and served as an indicator of size.

2.8. In vitro transcription and radiolabelling:

DNA templates incorporating a unidirectional T7 phage promoter sequence on one strand were generated by PCR (Table 3). For each targeted region of *HMS2,* two ~200-300bp templates, a "sense" and an "antisense", were created, differing only in the

location of the T7 sequence. With these templates, single-stranded, radiolabelled RNA probes were synthesized through *in vitro* transcription using a T7-promoter-dependent RNA polymerase (Ambion) in the presence of $^{32}P_{\alpha}$-UTP (Figure 8). The location of the T7 promoter sequence in the template DNA imparted directionality to the reaction therefore generating strand-specific probes. To increase hybridization specificity and reduce background signal, RNA probes were treated with DNase-I to degrade residual dsDNA, and filtered through a G-350 sepharose and glass filter column to remove excess nucleotide triphosphates. Membranes were incubated with the purified probes overnight as in Maniatis et al. (1989). X-ray film was then exposed to the radiolabelled membranes in lightproof film cassettes. Images of sense and antisense Northern blots represent equal exposure times, unless otherwise stated in the figure legends.

Primer	Name	Sequence
H1 Sense	FWD 5'UTR *HMS2*	GACCCGTCCTTCTTTTACCC
	REV 5'UTR *HMS2* (T7)	(TAATACGACTCACTATAGGG)* AGAGACACACTTGGCGACTG
H1 Antisense	FWD 5'UTR *HMS2* (T7)	(TAATACGACTCACTATAGGG) GACCCGTCCTTCTTTTACCC
	REV 5'UTR *HMS2*	AGAGACACACTTGGCGACTG
H2 Sense	FWD *HMS2* +244	GGCGGGCCGCCCGGACTGCG
	REV *HMS2* +555 (T7)	(TAATACGACTCACTATAGGG) CGGTTTGGTAGGCGTGAAAC
H2 Antisense	FWD *HMS2* +244 (T7)	(TAATACGACTCACTATAGGG) GGCGGGCCGCCCGGACTGCG
	REV *HMS2* +792	CGGGTGCCACTGAGTGGCCG
H3 Sense	FWD *HMS2* +694	CCAAACGCATCCGAATTGTC
	REV *HMS2* +1007 (T7)	(TAATACGACTCACTATAGGG) GGGGCTAACTAGGATGTCAC
H3 Antisense	FWD *kan* +246 (T7)	(TAATACGACTCACTATAGGG) TTATGCCTCTTCCGACCATC
	REV *HMS2* +1007	GGGGCTAACTAGGATGTCAC
K1 Sense	FWD *kan* +246	TTATGCCTCTTCCGACCATC
	REV *kan* +430 (T7)	(TAATACGACTCACTATAGGG) GCCTGAGCGAGACGAAATAC
B1 Sense	FWD *BAT2*	AGGCTTACAGAACGGTGGAC
	REV *BAT2* (T7)	(TAATACGACTCACTATAGGG) GCTCTATCAGGCGTGGAAAC
B1 Antisense	FWD *BAT2* (T7)	(TAATACGACTCACTATAGGG) AGGCTTACAGAACGGTGGAC
	REV *BAT2*	GCTCTATCAGGCGTGGAAAC
I1 Sense	FWD *HMS2* STOP +3	CCTGCCTCTGAGGGTCATTC
	REV *HMS2* STOP +200 (T7)	(TAATACGACTCACTATAGGG) GATTAGATAACAAGAGGTGTCAGAGAG
I1 Antisense	FWD *HMS2* STOP +3 (T7)	(TAATACGACTCACTATAGGG) CCTGCCTCTGAGGGTCATTC
	REV *HMS2* STOP +200	GATTAGATAACAAGAGGTGTCAGAGAG

Table 3. Primers used to generate strand-specific probes for Northern blot analysis.
*The T7 promoter sequence is shown in parentheses.

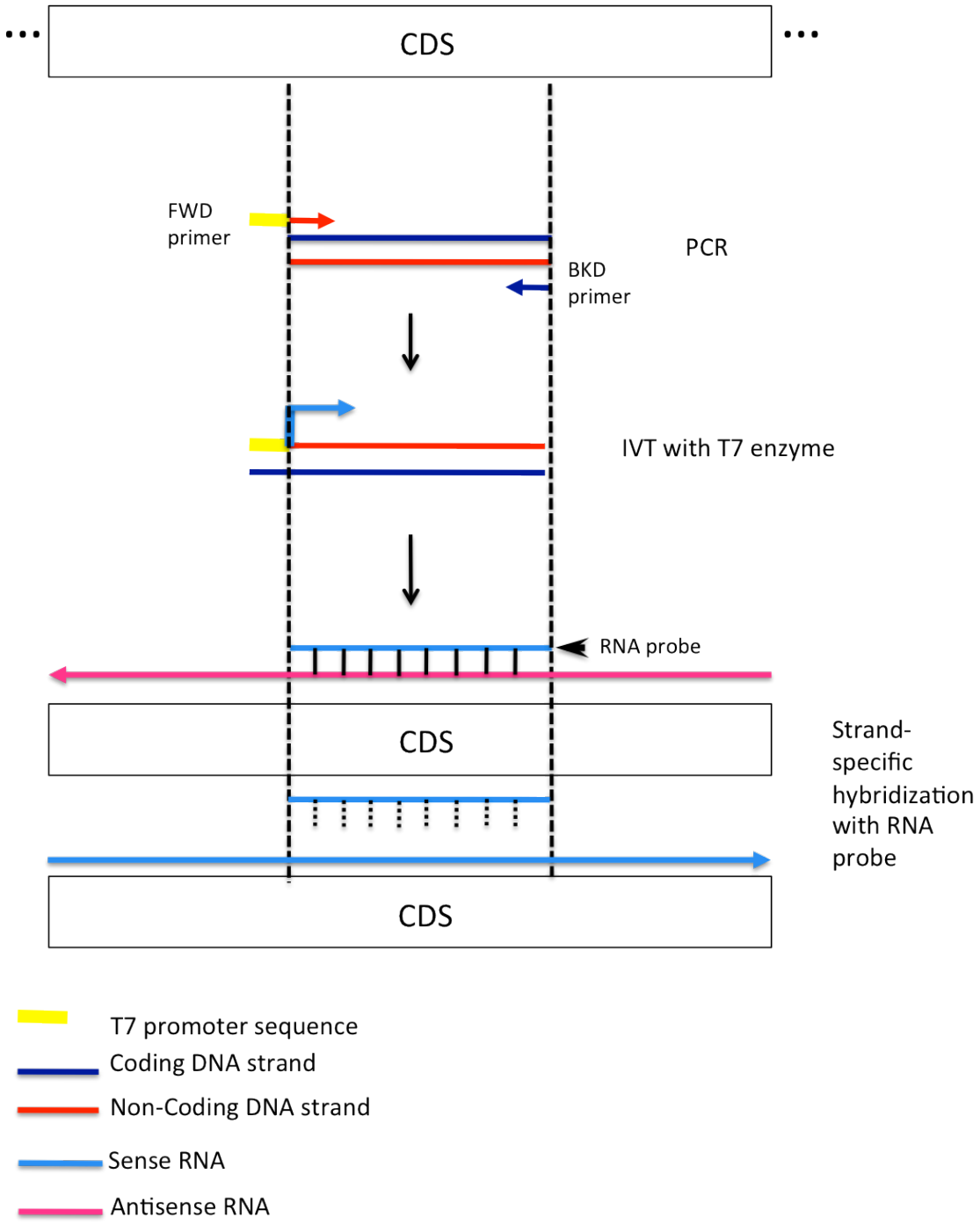


Figure 8. Generating Strand-Specific Probes.

In vitro transcription using $^{32}P_{\alpha}$-UTP results in radiolabelled, strand-specific probes. For example, placing the T7 promoter sequence (yellow) at the 5' region of the forward primer results in a probe that is specific to the antisense transcript (pink).

Though most of the Northern blots produced resulted in asymmetric transcript profiles, indicating that the T7-generated probes were strand-specific, the strand-specificity of the H1 and K1 probes was tested. Sense and antisense T7-transcribed probes were generated using the appropriate primer sets in the absence of radiolabel and dot blots were prepared on nitrocellulose membranes and placed at 80°C as described previously. Dot blots were exposed to either radiolabelled sense or antisense probes. If probes were indeed strand-specific, then the radiolabelled sense probe should hybridize to the cold antisense probes on the dot blots, and vice versa. The H1 and K1 probes demonstrated strand-specificity; however, a K1 antisense probe was not (Figure 9).

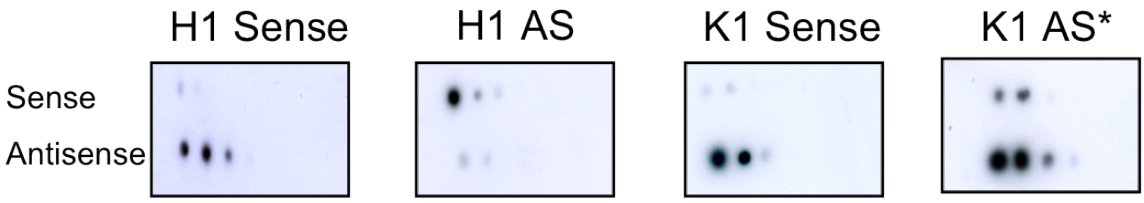


Figure 9. Testing probe strand-specificity.

Dot blots were prepared using cold T7-generated sense or antisense probes and then incubated with the appropriate radiolabelled probe. The dilution series consisted of unlabelled probe at neat, 1:10, 1:100, 1:1000, and 1:10000 dilutions. Dot blots depicted here represent a 5-minute exposure of membranes to X-ray film. The 1:10000 "dot" was only visible after 3 days of exposure to X-ray film. If probes were strand-specific it was expected that they were hybridize to the opposite probe (i.e. a radiolabelled sense probe would hybridize to a cold antisense probe, and vice versa). The H1 sense and antisense probes as well as the K1 sense probe display strand-specificity. *However, a K1 antisense probe displayed cross hybridization, and was thus not used in this study.

2.9. Approximating transcript levels:

To approximate transcript levels detected via Northern blot analysis, H1 probe PCR products were generated without the T7 promoter. A dilution series (50, 5, 0.5, 0.05, and 0.005 ng/µl) was prepared using this T7-less DNA template. After denaturing the

DNA, 1µl of each dilution was dispensed, in descending order, onto nitrocellulose membranes post-transfer. RNA and DNA were fixed onto the membranes at 80°C for 2h. These membranes were hybridized with the H1 sense and antisense probes. Transcript concentrations were estimated by comparing band intensity to the intensity displayed by the dilution series.

2.10. Hybridization of RNA to strand-specific tiling arrays:

Hybridization of RNA to strand-specific, high resolution tiling arrays was performed as described in Perocchi et al. (2007) (Perocchi, Xu, Clauder-Münster, & Steinmetz, 2007). In brief, 100 ml cultures of wild-type cells were grown to an OD_{600}= 0.6 in 2% YPD. Cultures prepared for galactose induction were grown to an OD_{600}= 0.6 in 2% YPD, shifted to 2% YPGal for 3 h and then harvested. Total RNA was extracted as described previously. Poly (A)+ RNA was isolated by Ana Serra Barros. RNA was treated with RNase-free DNase-I using the Turbo DNA-free kit (Ambion). To increase yield, first-strand cDNA synthesis was carried out in duplicate using 20 µg of total RNA, incubated with 1.72 µg random hexamers (RH6, Invitrogen) and 0.034 µg Oligo(dT) (0.5 µl/µg stock diluted to 0.05 µg, Invitrogen), and placed in a Thermocycler (70°C, 10 min; 25°C, 10 min; 4°C, hold). Keeping samples on ice, 20 µl of 5X first strand buffer, 10 µl 0.1M DTT, 5 µl dNTP + dUTP mix, 1.68 µl Actinomycin D, and 10 µl Superscript II Reverse Transcriptase were added to each reaction, and placed in a Thermocycler (25°C, 10 min; 37°C, 30 min; 42°C, 30 min; 70°C, 10 min; 4°C, hold). To remove any residual RNA, samples were treated with RNase H (Epicentre) and RNase cocktail (Ambion). Duplicate cDNA samples were then combined and purified using the Affymetrix cDNA clean up

columns and eluted in 30 µl diethylpyrocarbonate (DEPC) water. After measuring the concentration of cDNA, 5.5 µg of cDNA was fragmented and labelled using the Affymetrix GeneChip® WT terminal labelling kit, as specified by the manufacturer. Briefly, samples were treated with 10U/µl Uracil-DNA Glycosylase (UDG) and 1000U/µl apurinic/apyrimidinic endonuclease (APE) in a Thermocyler (37°C, 60 min; 93°C, 2 min; 4°C, hold for at least 2 min) in order to fragment the cDNA. Fragmented cDNA was compared to undigested DNA on a 1% agarose gel, with the resultant fragments appearing approximately 50-100 bp in size. Samples were then end-labelled using Terminal Transferase (30U/µl) in the presence of a biotinylated compound (GeneChip® DNA Labelling Reagent) at 37°C for 60 min.

As specified by the manufacturer, labelled cDNA fragments were hybridized to an Affymetrix *S. cerevisiae* tiling array (PN 520055) using the GeneChip® Hybridization, Wash and Stain kit (Affymetrix) after being denatured in the presence of Oligo B2 (Affymetrix) at 99°C, which aids in alignment during analysis. Hybridization was carried out at 45°C at 60 rpm in a rotisserie for 16 h. After overnight hybridization, arrays were washed and then stained using the provided Stain Cocktails (Affymetrix) and then placed in an Affymetrix fluidics station 450. Hybridized and stained arrays were scanned using an Affymetrix 3000 7G scanner. Data analysis was performed by members of the Lars Steinmetz group at the European Molecular Biology Labs in Heidelberg, Germany.

2.11. RNA Fluorescence In Situ Hybridization (FISH):

The single molecule RNA Fluorescence *in situ* hybridization (FISH) protocol was adapted from Zenklusen and Singer (2010) (Figure 10). By hybridizing a set of fluorescently-labelled DNA probes to complementary target RNA sequences, the use of this technique allows the detection of single RNAs in single yeast cells without necessitating genetic modification (Zenklusen & Singer, 2010) . In brief, four DNA probes of approximately 50 nt in length and 50% GC content were designed (Table 4). These probes contained four amino-allyl dT modified bases, which were then coupled to fluorescent cyanine dyes. Modified bases were evenly spaced by approximately 10 nt in order to prevent quenching of the conjugated dyes. In order to capture both *HMS2* sense and antisense RNAs, probes were designed complementary to either the 5' or 3' of *HMS2*, respectively, so to avoid cross-hybridization between sense and antisense probe sets.

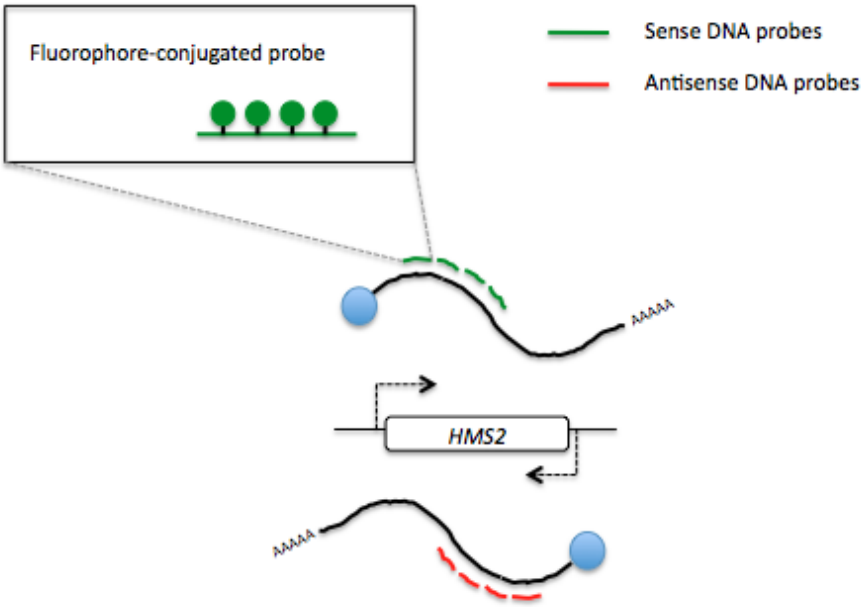


Figure 10. Single molecule RNA Fluorescence *in situ* hybridization

A combination of four 50 nt DNA oligonucleotides, each containing 4 evenly spaced amino-allyl dTs conjugated to 4 fluorescent dyes, was used to target single RNA molecules of either sense (Cy5-labelled probes in green) or antisense (Cy3-labelled probes in red) transcript in paraformaldehyde fixed cells. Probes are homologous to an approximately 200 nt region at the 5' end of the respective transcripts, have an ~50% G/C content and have low probability for forming stable secondary structures (~$\Delta G \leq -2.5$ kCal/mol). Probe sequences were ran in BLAST (SGD^{TM}) to monitor for any potential cross-hybridization to other genes. (Procedure adapted from Zenklusen and Singer, 2010).

Probe	Sequence	G/C content
*HMS2*_S_1	CAGG**Z**GGTAGAGTTTGC**Z**TACAAAGACA**Z**CAGGCTGCTCCATCC**Z**CGATG*	50
*HMS2*_S_2	GAC**Z**TTGGTGAACTGG**Z**CCGGATTCCAAA**Z**GACGAGCTTGCTGCCA**Z**CAG	52
*HMS2*_S_3	GCGGCCCGCC**Z**TCTGGAAGTTA**Z**ATTTGCTCAAT**Z**GCTTAACGAA**Z**GCTG	48
*HMS2*_S_4	ATTGGGAG**Z**GGCGGCAGAC**Z**GATGAGCTT**Z**AACCAGTGAGAGGC**Z**ATTAA	48
*HMS2*_AS_1	CCATTAT**Z**AACAGACCCG**Z**CTCCCCTCAAGA**Z**ACCAGCATTGC**Z**CCCTCC	52
*HMS2*_AS_2	TTTTCGTCACCG**Z**GTCGCCCAACA**Z**GCAACAAGAACA**Z**CTCATCTACT**Z**C	46
*HMS2*_AS_3	TTCGCCGGCG**Z**GTTCGAATTTA**Z**GAACTGCTCTT**Z**GGATACGGCCAC**Z**CA	50
*HMS2*_AS_4	CT**Z**ACAGACAATCTCC**Z**ACCACCAGTTTCACGCC**Z**ACCAAACCGC**Z**AACT	40

Table 4. RNA FISH probes.
Amino-allyl dT-modified probes are indicated by a "**Z**".

HMS2 sense and antisense probe sets were labelled with either Cy5 or Cy3 cyanine dye (CyDye^{TM}, GE Healthcare), respectively. Probe labelling was carried out as specified by the manufacturer (GE, Healthcare), with some minor alterations. First, for each labelling reaction, all four unlabeled probes (belonging to either the *HMS2* sense or antisense set) were combined to make a total 5 µg probe mix (i.e. 1.25 µg of each probe). Prior to labelling, 500 µl of buffer PN was added to the probe mixes, which were then purified using the QIAquick Nucleotide Removal Column (Qiagen) system. Purification was carried out according to the manufacturer; however, to increase yield, the samples were loaded onto the columns twice and unlabeled probe mixes were then eluted in 40 µl Milli-Q purified water (referred to as Milli-Q water), not the elution buffer provided in

the kit. Probe mixes were then lyophilized in a SpeedVac, after which the dried DNA pellets were each resuspended in 10 µl of labelling buffer and added to tubes containing the respective dyes. To facilitate resuspension, samples were vortexed vigorously and then subjected to a quick pulse centrifugation in order to collect the labelling reaction at the bottom of the tube.

After reactions were incubated in the dark overnight, the labelled probe mixes were then purified using the QIAquick columns. Purification was performed as before except columns were washed twice with buffer PE to fully remove the unincorporated dye, and then eluted using 100 µl of the elution buffer provided in the kit. The concentration and labelling efficiencies of the *HMS2* sense and antisense probe mixes were measured using a spectrophotometer before probes were stored in the dark at -20°C. Labelling efficiency was calculated as in Zenklusen and Singer (2010). Efficiencies for both Cy5-conjugated sense probes and Cy3-conjugated antisense probes were comparable (Sense =0.96, Antisense = 1.00).

Labelled DNA probe mixes were hybridized to paraformaldehyde-fixed spheroplasted cells, immobilized on glass slides. All buffers and coverslips used in the following steps were pre-prepared in order to reduce time lags that might introduce inconsistencies between experiments. To prepare cells, cultures were grown in glucose and induced (if applicable) in galactose as described previously. For each growth condition, 50 ml of cell culture was prepared. Furthermore, as a control, for each growth condition/strain used,

an *hms2* Δ cell culture was also prepared. Cells pellets were harvested in 50 ml falcons by centrifugation (5 min/3000 rpm), and then fixed in 4 % (v/v) paraformaldehyde (EM grade, Electron Microscopy Science) in 1X phosphate buffered saline (PBS) for 45 minutes on top of a table top shaker (75 rpm). Fixed cells were collected by centrifugation (3500 rpm, 5 min) at 4°C and washed 3 times with 10 ml pre-prepared and filter sterilized cold buffer B (1.2 M sorbitol, 100 mM KHPO4, pH 7.5). Buffer B is an isotonic buffer that is used to prevent cells from bursting after their cell walls are digested. After each wash cells are harvested by centrifugation (3500 rpm, 2:30 min) at 4°C. Following the final wash, cells are resuspended in 1 ml of buffer B in a 1.5 ml Eppendorf tube, and pelleted by centrifugation (4,000 rpm, 1 min) at 4°C.

In order to allow probes to access the nucleus and hybridize to their target RNAs, cell walls are enzymatically digested to produce spheroplasts. After centrifugation cell pellets were removed of buffer B and resuspended in 1 ml of spheroplast buffer (1.2 M sorbitol, 100 mM KHPO4, pH 7.5, 20 mM ribonucleoside–vanadyl complex (VRC), 20 mM β-mercaptoethanol). Cell walls were digested with the addition of 30 µl of lyticase (25U/µl) (Sigma-Aldrich), which digests the β-glucan polysaccharides comprising the yeast cell wall. Cells grown in YPD were treated with lyticase for 20 minutes, while cells grown in galactose were treated with the enzyme for 25 minutes, as growth in galactose seemed to impede the digestion reaction. To confirm digestion of approximately 80% of cells, 5 µl of sample was pipetted onto a glass slide and checked under a light microscope. Spheroplasts were then collected by centrifugation (3500 rpm, 3 min) at

4°C and then washed in 1 ml cold resuspension buffer (1.2 M sorbitol, 100 mM $KHPO_4$, pH 7.5, 20 mM ribonucleoside–vanadyl complex). After centrifugation (3500 rpm, 3 min) at 4°C, spheroplasts were gently resuspended in 1 ml cold resuspension buffer by pipetting and then kept on ice.

To immobilize cells on glass coverslips, round 18 mm in diameter coverslips were coated with poly-L-lysine (Sigma-Aldrich), which provides a substratum for cellular attachment. To prepare coverslips for poly-L-lysine treatment, coverslips were boiled in 500 ml 0.1 N HCl for 10 minutes, rinsed thoroughly with Milli-Q water, and then dipped in 100% ethanol (EtOH). Coverslips were then dried on transparent acetate sheets at room temperature, covered to avoid dust accumulation. Following drying, 100 µl of 0.01% poly-L-lysine solution was added to cover each coverslip. After a 5 minute, room temperature incubation, the poly-L-lysine solution was removed using a vacuum pump and the coverslips were air-dried at room temperature. Using forceps, coverslips were individually dipped in Milli-Q water twice to wash off excess poly-L-lysine solution. Coverslips were then air-dried and then transferred, face-up, to 12-well plates, where they could be stored for several months.

Using these poly-L-lysine coated coverslips, 150 µl of spheroplasts were pipetted onto the centre of the coverslips. After allowing the spheroplasts to settle onto the coverslips for 30 min at 4°C, 2 ml of cold buffer B was added to each well, then carefully removed using a vacuum pump. For storage, 2ml of 70% EtOH was added to each well.

Cells were stored for at least 3h at -20°C before being used for hybridization; however, they could be stored at -20°C for up to 6 months.

To prepare probe mixes for hybridization, 2 ng each of sense and antisense probe mix for double RNA FISH, or 2ng of sense probe mix for single RNA FISH, was added to 10 µg *Escherichia coli* tRNA and 10 µg of salmon sperm DNA in a single Eppendorf tube. Probe mixes were then lyophilized using a SpeedVac. While probe mixes were lyophilizing, spheroplasts adhered to coverslips were rehydrated by removing the EtOH from the 12-well plate using a vacuum pump, after which they were washed by adding 2ml 2X SSC at room temperature for 5 minutes. This wash was repeated a second time before cells were washed once in a 40% formamide/2X SSC solution at room temperature for 5 minutes. Additionally, for each coverslip to be used, 12 µl of "solution F" (40% formamide, 2x SSC, 10 mM $NaHPO_4$, pH 7.5), and 12 µl of "solution H" (2x SSC, 2 mg/ml bovine serum albumin, 10 mM VRC) were prepared. Solution F, pre-warmed to 95°C, was added to the DNA probe pellet, which was then heated at 95°C for 3 minutes. The probe mix was then removed from the heat, and allowed to cool for 1 minute, before solution H was added to create the final "hybridization mix".

Using a flat glass plate, parafilm and slips of cardboard, a "hybridization chamber" was prepared as in Figure 11. A 22 µl drop of hybridization mix was placed on the parafilm-covered glass plate, onto which a coverslip was inverted, face down, carefully in order to avoid generating air bubbles. Generally, two coverslips per condition/strain were

prepared per experiment. Hybridization chambers were sealed with an additional layer of parafilm, taped down around the edges. Cardboard spacers within the hybridization chamber acted to prevent the top layer of parafilm from contacting the bottom later and/or disrupting contact between the coverslips and hybridization mix. Hybridization chambers were placed into cardboard boxes to prevent light from bleaching the dyes and incubated overnight at 37°C. A 40% formamide/2X SSC solution was also placed at 37°C overnight.

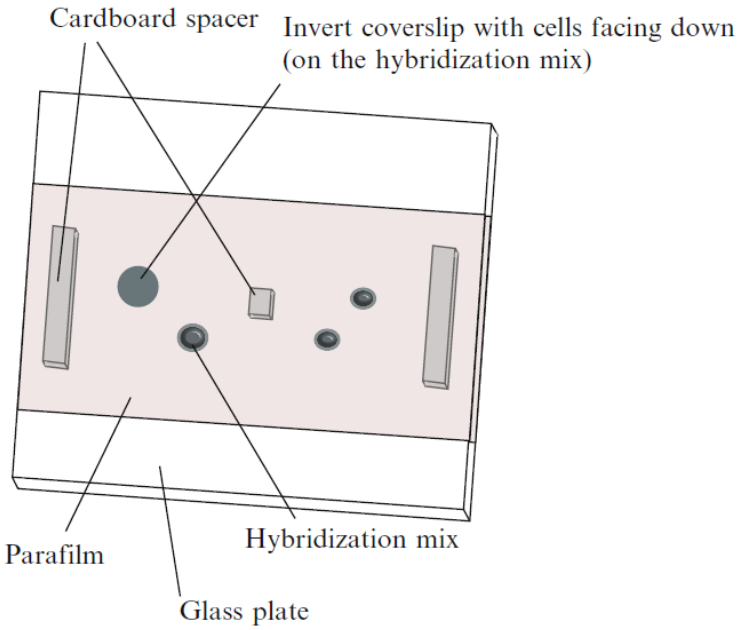


Figure 11. The hybridization chamber.

The hybridization chamber is constructed using a glass plate, parafilm, and cardboard spacers. Cells attached to poly-L-lysine coated coverslips are inverted onto drops of hybridization solution situated on parafilm stretched over a glass plate. A second layer of parafilm is used to seal the hybridization chamber, with cardboard spacers in between the two layers of parafilm to prevent contact. Image taken from Zenklusen and Singer (2010).

Using forceps, coverslips were removed from the hybridization chamber and placed into fresh 12-well plates, face up. Coverslips were washed and incubated twice using 2ml/coverslip of the pre-warmed 40% formamide/2X SSC solution for 15 min at 37°C. Coverslips were then washed once with 2ml 2X SSC/0.1% Triton X-100 at room temperature for 15 minutes and then again in 2ml 1X SSC at room temperature for 15 minutes. Coverslips were then washed in 2ml 1X PBS and DAPI (final concentration, 0.5 μg/ml) for 2 minutes. After DAPI exposure, coverslips were washed once with 2ml 1X PBS for 2 minutes. Using forceps, coverslips were then dipped into 100% EtOH and allowed to air-dry on parafilm, and covered with a light-proof shield. Coverslips were inverted (cells facing downward) onto a drop of pre-warmed Pro-long Gold mounting solution on clean glass slides and allowed to polymerize at room temperature overnight in a light-proof box. Coverslips were then sealed with clear nail polish.

Image acquisition was performed in collaboration with the Ilan Davis group, with a DeltaVision CORE: Wide-field fluorescence deconvolution imaging microscope. To obtain accurate expression profiles and information about RNA spatial distribution, images were collected in 3D. Using a 100x objective lens, 31 “z-stacks” were collected. For each stack, an exposure time of 0.5 seconds for DAPI and 1 s for Cy3 (TRITC) and Cy5 filters was applied. Images were captured using a CoolSNAP HQ camera (Photometrics, Tucson, AZ)

To facilitate image analysis, 3D data sets were compressed into 2D images by using a maximum projection, a display of the maximum value of all fluorescence signals in a z-stack for a particular pixel location. Reduction of 3D datasets to accurate 2D representations is possible, in particular for lowly transcribed genes, since the probably that two RNAs found in the same x-y position, but a different z coordinate is assumed to be low. To determine whether or not a cell was expressing *HMS2* sense and/or antisense, a threshold distinguishing genuine signal from expressed RNAs from background/non-specific signal was applied to the datasets. To determine this threshold, coverslips containing the *hms2* Δ strain (grown in equivalent growth conditions to *HMS2* expressing cells), were prepared as a negative control. The maximum intensity maximum projection was measured for each z-stack of individual *hms2* Δ cells for the *HMS2* sense and, if applicable, antisense probes. These maximum intensity maximum projections were averaged across the 31 stacks to produce an average "max" signal for a single cell. These max values were calculated for each intact *hms2* Δ cell surveyed and an average max value for the population was determined. To take into account the spread in the dataset, the value two standard deviations above the calculated population max average was used to mark the background threshold. Then, the maximum intensity maximum projection signals from cells containing the complementary *HMS2* sequences were compared to this threshold. Cells displaying average max signals above this threshold were considered to represent cells expressing *HMS2* sense and/or antisense, while cells with average max signals below the threshold were considered to be non-expressing. Images displayed in this study represent 2D

compressions of z-stacks 12-22 (11 stacks), which included most of the nucleus and cytoplasm. Error bars represent standard deviations, which were determined by producing 1000 bootstrap pseudoreplicates of both data sets (glucose vs. galactose or antisense overlap vs. no antisense overlap) and comparing the resultant distributions of percentages using a Wilcoxon rank-sum test (Statistical analysis performed by Struan Murray).

2.12. Chromatin Immunoprecipitation:

Chromatin immunoprecipitation (ChIP) was performed as described by Morillion et al. 2003 (Morillon, O'Sullivan, Azad, Proudfoot, & Mellor, 2003). In short, 50 ml of cells were grown to OD600=~0.5 in YPD and then induced in galactose-containing media for 1h as previously described. Cells were fixed in 50 ml 1% formaldehyde/PBS for 30 min. The cross-linking reaction was stopped with the addition of 2.5 ml of 2.5 M glycine. Fixed cells were harvested and then physically lysed by transferring cells into tubes filled with acid washed glass beads, which were then shaken vigorously using a MagnaLyser (Roche). Fixed chromatin was then sheared by sonication for 30 min, 1 min on/20s off on the medium setting using a bioruptor (Diagenode). Cell lysis and sonication steps were both carried out at 4°C. Cellular debris was removed from chromatin extracts by centrifugation (10, 000 rpm, 15 min) at 4°C. Sonicated chromatin samples were stored at -80°C.

One in four dilutions of fixed chromatin were incubated with 5 µl of H3K36me3 antibody (ab9050) or 5 µl of H3 antibody in siliconised 1.5ml Eppendorf tubes. These tubes were

placed on a rotating wheel overnight at 4°C. Negative controls ("no antibody") were carried out and did not contain the antibody. Additionally, for the analysis, a one in four dilutions of fixed chromatin were also prepared and immediately stored at -20°C ("totals"), serving as internal controls for chromatin extraction efficiencies between samples. Antibody-bound chromatin was pulled down by incubating samples with 40 µl protein A-sepharose beads for 90 min on a rotating wheel at room temperature. Beads were washed one time each with TSE-150/500 (1 % Triton-X-100, 0.1% SDS, 2 mM EDTA, 20 mM Tris-Cl pH 8.0, 150/500 mM NaCl), lithium chloride (LiCl) wash (0.25 M LiCl, 1 % NP-40, 1 % sodium dioxycholate, 1 mM EDTA, 10 mM Tris-Cl pH 8.0), and finally twice in 1X TE. Excess TE was removed using a vacuum pump. Bound chromatin was eluted by incubating the beads in 100 µl of pre-warmed elution buffer (1 % SDS, 0.1 M $NaHCO_3$) at 65°C in a shaking incubator. Beads were pelleted by centrifugation (13,000 rpm, 1 min) and the supernatant was transferred to fresh 1.5 ml Eppendorf tubes. To reverse crosslinking, 7 µl of 5 M NaCl was added to the immunoprecipitate, the negative controls, and the total controls. Samples were then incubated for 3h at 65°C, followed by a 1 h treatment with RNaseA at 37°C. After an overnight incubation with 1 µl proteinase K (Roche), samples were purified using the QIAquick PCR purification columns (Qiagen) and eluted in 200 µl 1 mM Tris-Cl pH 8.0. Total DNA was diluted accordingly in 300 µl Tris-Cl. The proportion of immunoprecipitated DNA (IP) to total DNA was measured by performing real-time quantitative PCR (qPCR) on all of the samples using a Corbett Rotorgene machine and Sybr green mix (Bioline). Primers used for RT qPCR are displayed in Table 5.

Primer	Sequence
*HMS2*_-809 FWD	AAGAACGCAGGAAGAAAAAGG
*HMS2*_-648 REV	TGTTTTCAATATTCAGTTCAGCAG
*HMS2*_-367 FWD	CGTGTCTGGTGTGCAAAAAC
*HMS2*_ -222 REV	TAAAAGAAGGACGGGTCCAG
*HMS2*_-83 FWD	GTCGCTACACGAAAGAGACAC
*HMS2*_-11 REV	AAGAGTAAGTGGACGCATCAG
HMS2 _-40 FWD	CGCCAAGTGTGTCTCTTTCG
*HMS2*_+129 REV	ATTCCAAATGACGAGCTTGC
TUB2 CDS FWD +484	CGTATGATGGCCACCTTCTC
TUB2 CDS BKD +577	CCAATTGGTGCACAGAGAAC

Table 5. Primers used for real-time PCR.

Data were expressed as the proportion of IP (adjusted for signal accounted for by the no antibody controls) to total DNA (% Input), normalized to proportions (% input) of H3 immunoprecipitate. Additionally, it was found that between growth in YPD and YPGal, proportions of H3 changed dramatically at loci whose transcription was seemingly unaffected by the change in carbon source. For instance, at the *TUB2* locus, which encodes β-tubulin, levels of H3 in galactose were approximately five times of those detected in glucose, despite the gene's transcript levels remaining relatively constant via tiling array analysis (Figure 12 A, B). Thus, to ensure that the trends observed took into account potential global shifts in H3 occupancy, *TUB2* was chosen as a standard gene. Thus, data in the graphs represented are expressed as proportions H3-normalised levels of H3K36me3 for regions of *HMS2*, normalised to the H3-normalised proportions of H3K36me3 in the coding region *TUB2*. Error bars represent the compounded standard error of three real-time PCR reactions.

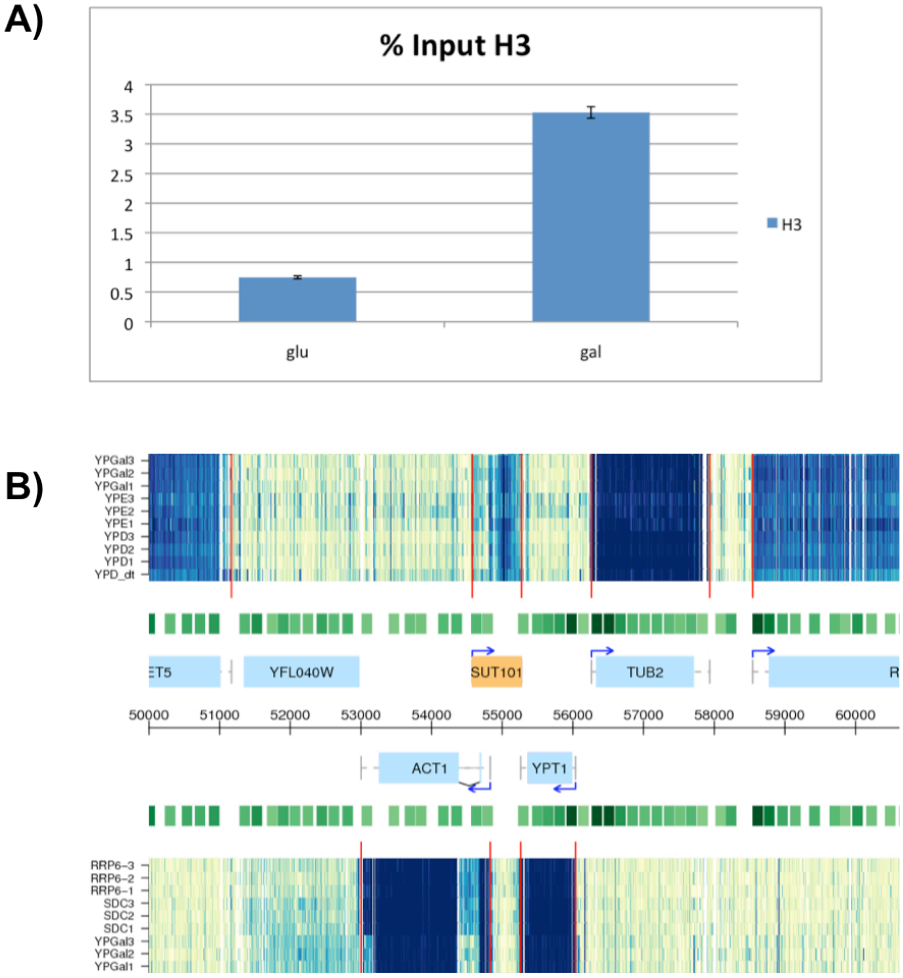


Figure 12. Levels of H3 at the *TUB2* locus change between glucose and galactose, but not transcript levels.

(A) Levels of H3 increased at the *TUB2* locus between growth in glucose- and galactose-containing media. The y-axis represents the percent input of H3 in relation to total input. **(B)** Levels of *TUB2* transcript do not change after a change in carbon source. Tiling array data was taken from: http://steinmetzlab.embl.de/cgi-bin/viewNFRsharing.pl?showSamples=NFRsharing&gene=EGT2 Blue boxes= coding regions, orange boxes= non-coding transcription units, green boxes= nucleosome positions. The error bars on the bar graph represent compounded standard error of three RT-PCR reactions.

Chapter 3.

Investigating the role of overlapping transcription in transcription unit regulation using the *HMS2* locus as a model system.

3. Investigating the role of overlapping transcription in transcription unit regulation using the *HMS2* locus as a model system

3.1. Introduction:

3.1.1. Unstable Antisense Transcripts:

The biological role of overlapping, antisense transcription has remained a contentious issue despite numerous locus-specific and genome-wide studies devoted to uncovering what, if any, function these non-coding transcripts or, more specifically, their transcription plays in genetic regulation. One of the pioneering and most widely cited examples of ncRNA-mediated gene regulation in *S. cerevisiae* was described at the *PHO84* locus, where an antisense transcript is produced but rapidly degraded by the nuclear exosome (Camblong et al., 2007). The study found that stabilisation and subsequent accumulation of the antisense transcript in aged cells or through deletion of the crucial nuclear exosome component, Rrp6p, resulted in silencing of *PHO84* sense transcription. Additionally, the authors demonstrated that this RNA-mediated gene silencing acts through histone deacetylation of the *PHO84* promoter via histone deacetylase, Hda1p. In a later publication, the authors also demonstrated that the *PHO84* antisense transcript could silence sense transcription *in trans*; however, through an Hda1p-independent mechanism (Camblong et al., 2009). Subsequent genome-wide studies identified 925 non-coding Rrp6p-sensitive transcripts (approximately 12.7% of the total transcript pool), or “cryptic unstable transcripts” (CUTs) (Neil et al., 2009; Xu et

al., 2009). While many RNA molecules, both coding and non-coding, display stabilisation in the absence of Rrp6p, as a mechanism, directed deacetylation through Rrp6-mediated antisense stabilisation appears to be unique to the *PHO84* locus.

In addition to degradation by the nuclear exosome, non-coding RNAs can also be subject to rapid degradation by the 5'-3' cytoplasmic exonuclease, Xrn1p (Kem1p), constituting a novel class of transcript named Xrn1-sensitive unstable transcripts (XUTs) (Van Dijk et al., 2011). The XUTs comprise a large faction of transcripts (1,658 transcripts in total), 66% of which are transcribed antisense to open reading frames. Of these annotated genes, when their overlapping antisense XUTs were stabilized in the absence of Xrn1p, 25% displayed decreased RNAPII occupancy, suggesting a reciprocal relationship between sense transcription and antisense stabilisation, similar to that of *PHO84*. Further investigation by van Dijk et al. (2011) demonstrated that the repressive activity of the antisense XUTs was dependent on the activity of the Set1 complex but, curiously, necessitated the absence of H3K4me3 at the sense promoter. While the authors postulated that the presence of H3K4me1/2 might be required for XUT-mediated silencing, they stopped short of providing an exact mechanism.

Immediate degradation of a large body of non-coding RNAs by Rrp6p and Xrn1p might appear like a wasteful process in terms of energy and resources; however, it is possible that the act of simply transcribing these unstable RNAs might bear regulatory consequences, perhaps through co-transcriptional remodeling or modification of

chromatin. Native elongating transcript sequencing (NET-seq), a technique developed by Churchman and Weissman (2011), captures nascent transcription, and thus is well suited to the purpose of analysing highly unstable transcripts. Using NET-seq, Churchman and Weissman measured nascent transcription from divergent promoters, specifically those producing an upstream antisense CUT and a downstream sense messenger RNA. The analysis found that 80% of unstable antisense was produced at a level three-fold lower than their sense counterparts. Moreover, the study demonstrated that directionality of divergent, bidirectional promoters was imparted via Rpd3S-mediated suppression of antisense initiation. While the NET-seq analysis did not elucidate a biological function for unstable antisense transcription, it indicated that not only are antisense CUTs tightly monitored post-transcriptionally by RNA surveillance pathways, but that their generation is subject to regulation at the level of initiation, perhaps to modulate the effects of their transcription on upstream promoters.

PHO84 has a peculiar NET-seq profile that does not necessarily coincide with predictions given the model proposed by Camblong and colleagues (2007) (Figure 13A). In wild-type, glucose-grown cells, a drop in nascent *PHO84* sense transcription, approximately 200 nt from the translation stop codon, coincides with the detection of a lowly transcribed antisense transcript, not detected in the Camblong et al. (2007) study. What implications this profile has on the regulation of the locus are unclear; however, the additional intricacies unveiled by the nascent transcription profile of *PHO84* suggest that the Rrp6p/Hda1p model for antisense-mediated regulation might not fully reflect the

sense/antisense relationship at the locus and underscores the need for a more comprehensive evaluation of overlapping transcription.

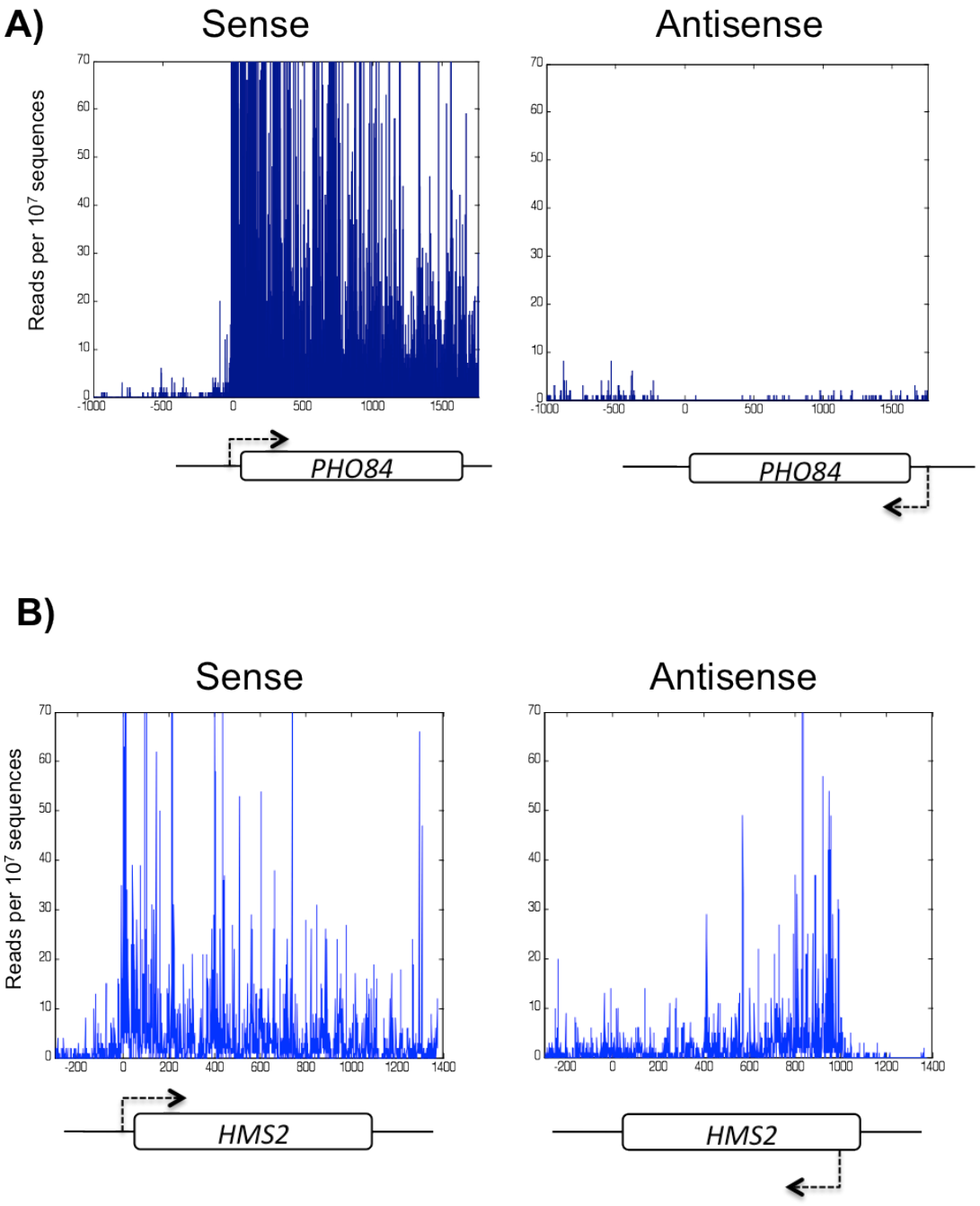


Figure 13. Nascent transcription at the *HMS2* and *PHO84* loci.

Nascent transcription at the **(A)** *PHO84* and **(B)** *HMS2* loci as measured by NET-Seq (Churchman and Weissman, 2011). "0" on the x-axis marks the sense transcript TSS, also marked by the broken-line arrows. The boxes represent ORFs.

3.1.2. Stable Antisense Transcripts:

Whereas several studies have focused on the functionality of highly labile RNAs, a more general model for stable antisense-mediated gene regulation has been largely neglected. A study by Xu et al. (2011) demonstrated, genome-wide, that in the presence of both Rrp6p and Xrn1p, antisense transcription has a negative effect on lowly transcribed genes and may act to fine tune gene expression at the lower end of the expression spectrum. This relationship was contingent on antisense transcription overlapping the transcription start site (TSS) of the sense promoter. While this study was able to demonstrate a reciprocal relationship between sense and stable antisense transcript levels, it also fell short of describing a viable mechanism through which antisense transcription might influence that of its sense gene.

Despite the lack of an exact and general molecular mechanism through which it might act, antisense transcription has been implicated in various facets of *S. cerevisiae* biology, ranging from being associated with genes controlling the genetic response to nutrient availability; for instance *PHO84* (Camblong et al., 2007)*, GAL10* (Houseley et al., 2008), and *PHO5* (Uhler et al., 2007); to controlling meiotic entry (Hongay et al., 2006; van Werven et al., 2012); and more broadly, displaying periodicity in the cell cycle (Granovskaia et al., 2010). Given the seemingly ubiquitous presence of antisense transcripts, both stable and processed, in all aspects of biological function, it is anticipated that their generation plays an important, but perhaps varied, role in gene expression regulation.

3.1.3. Aims: studying antisense transcription at the HMS2 locus

While genome-wide studies have been invaluable in establishing the wide reach of antisense transcription, many of the characterized mechanisms were first established using single-gene studies, such as the analysis of *PHO84*. In this study, the *HMS2* (High-copy Mep Suppressor 2) locus in *S. cerevisiae* was chosen as the preferred model system. The HMS2 protein is largely uncharacterized, bearing similarity to heat shock transcription factors, but likely to function in growth pathways, as over-expression of *HMS2* rescues pseudohyphal growth defects in diploid *MEP1* and *MEP2* mutants (Lorenz & Heitman, 1998). Pseudohyphal growth is a type of elongated growth pattern exhibited by diploid and haploid cells found in nitrogen-starved conditions and is intimately related to pathways affecting polarized growth, in particular progression through the G2/M transition of the cell cycle (reviewed in Song & Kumar, 2012). Interestingly, the yeast strain S288C and its derivative strains, such as BY4741 used in this study, cannot form pseudohyphae because of a naturally occurring nonsense mutation in the *FLO8* (FLOcculation 8) gene (Liu, Styles, & Fink, 1996).

Unlike its elusive protein function, multiple genome-wide studies (Granovskaia et al., 2010; Xu et al., 2009) and previous work done in our laboratory (James Halstead, 2007; David Brown, 2009) have produced a relatively well-characterized *HMS2* transcript profile, which, most notably, bears a stably transcribed sense/antisense pair. Additionally, unlike *PHO84*, the NET-seq profile of *HMS2* is more representative of a typical gene, perhaps indicating its suitability as a model for more general

sense/antisense relationships (Figure 13B) (Churchman & Weissman, 2011). Moreover, the locus also provides a convenient setting in which to study the role of overlapping transcription in an interleaved genome as it lies in the middle of a gene cluster consisting of upstream gene *RPS4A* and downstream genes *BAT2*, and *YJR149W* (Figure 14).

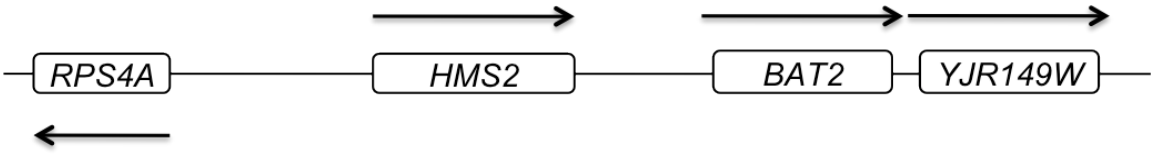


Figure 14. The *HMS2* gene cluster.

The *HMS2* gene cluster consists of upstream gene *RPS4A* and downstream genes *BAT2* and *YJR149W*. White boxes represent ORFs. Black arrows represent the directionality of the sense gene. Images are not drawn to scale. *RPS4A* contains an intron not depicted in the schematic.

To better understand the role of overlapping sense and antisense transcription in the regulation of transcription units in general, the *HMS2* locus was systematically dissected using a multi-dimensional approach: 3'RACE and Northern blot analysis to detect overlapping transcripts in both the sense and antisense directions; high resolution tiling array to describe the transcript profile across multiple neighbouring loci; and RNA FISH to discern RNA localization and cell-to-cell variability in transcription. By addressing *HMS2* regulation from multiple angles, the work in this chapter aims to converge on a comprehensive model for antisense-mediated gene expression regulation, both in terms of *HMS2* sense/antisense pair as an isolated unit, but also as a component of a complex transcriptional regulatory network potentially involving many clustered genes.

3.2. Results:

3.2.1. The HMS2 sense and antisense promoters are overlapped by transcription from the opposite strand:

Published tiling array data by Granovskaia et al (2010) provided the chromosomal coordinates for the *HMS2* sense (Chromosome 10: 704065, 705297) and antisense transcripts (Chromosome 10: 703349, 705149). According to this analysis, the sense transcript initiates 121 nt upstream of the *HMS2* start codon, while the antisense transcript initiates within the *HMS2* CDS, approximately 113 nt upstream of the *HMS2* stop codon. Though tiling arrays are better suited for annotating approximate transcript start and end sites in comparison to Northern blot analysis, overlapping transcription and alternative start/end sites can confound these results. 3'RACE was performed on RNA extracts from galactose grown cells with either wild-type or galactose inducible *HMS2*, in order to improve the 3'end annotation of the antisense and sense transcripts, respectively (Figure 15A). These strains and growth conditions were used since they displayed highest sense or antisense transcript levels via Northern analysis (Figure 52 and Figure 18, respectively). Using this method the sense end was reliably and reproducibly traced to between 44 and 45 nucleotides (nt) upstream of the *BAT2* start codon, and 27nt downstream of a consensus poly (A) site (Figure 15B). This indicates that the *HMS2* sense transcript has an exceptionally long 3' UTR, which may bear consequences on its post-transcriptional regulation (discussed further in Chapter 3).

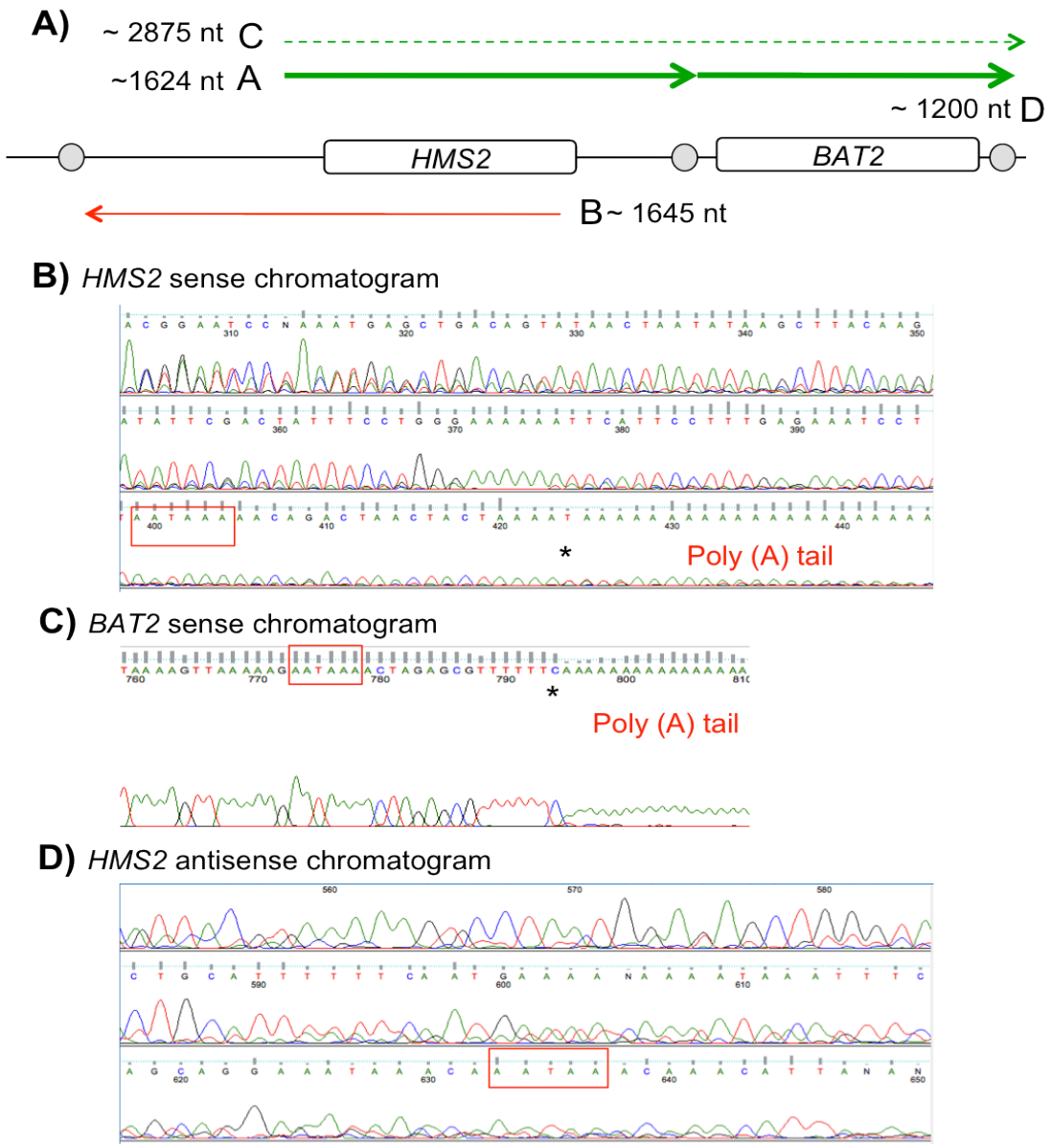


Figure 15. Mapping the 3' ends of the *HMS2* sense and antisense and *BAT2* sense transcripts.

(A) Schematic of the *HMS2* transcription unit. Grey circles represent the approximate locations of consensus Poly (A) sites. Green arrows represent transcripts generated from the same template strand as the coding *HMS2* sense RNA, while red arrows represent transcripts generated from the opposite strand. Arrow lengths represent transcript lengths from estimated transcription start sites to end sites as detected by 3' RACE, while arrow thickness represents approximated steady-state levels of transcript as detected by Northern Blot analysis. White boxes represent gene ORFs. Images are not drawn to scale. **(B)** Results from the 3'RACE performed on a galactose-inducible *HMS2* strain maps the 3' end of the *HMS2* sense transcript (transcript A) to 44 nt upstream of the *BAT2* start codon and 27nt downstream of a consensus poly (A) site (marked by red

box in the chromatogram). **(C)** Using the same strain, the 3' end of the *BAT2* sense transcript (transcript D) was mapped to 63 nt downstream of its stop codon, and 16 nt downstream of a consensus poly (A) site. **(D)** 3' RACE performed on wild-type cells grown in galactose-containing media was also able to approximate a region where the *HMS2* antisense transcript (transcript B) ends, roughly 648 nucleotides upstream of the *HMS2* start codon. However, no distinct poly (A) tail was discernable.

Additionally, the *BAT2* transcript 3' end was precisely mapped in the galactose inducible *HMS2* strain to approximately 63 nt downstream from the *BAT2* stop codon, and 16 nt downstream from a consensus poly (A) site (Figure 15C). While the results of the 3'RACE for the antisense transcript were not as precise, the antisense 3' end was mapped in the vicinity of 644 nt upstream of the *HMS2* start codon, and 28nt downstream of a consensus, antisense poly (A) site (Figure 15D). While this result could not be repeated due to technical failure, the end site detected in this study is consistent with the Granovskaia et al (2010) study, which projected the AS end site to be 835 nt upstream of the *HMS2* start codon. Together start sites defined by previous tiling array analyses and end sites detected in this study confirm that both the *HMS2* sense and antisense promoters are overlapped by transcription from the opposite strand.

Strand-specific Northern blot analysis of the *HMS2* locus was also able to detect the stable sense (transcript A) and antisense (transcript B) transcript pair in YPD rich media (Figure 16A). These transcripts are polyadenylated and, as indicated by their stability, most likely capped (Figure 16B). When ribosomal RNA was depleted from the extract by selecting for poly (A)-containing RNAs, resolution was improved and multiple sense and antisense species were discernible, indicating the potential use of alternative start

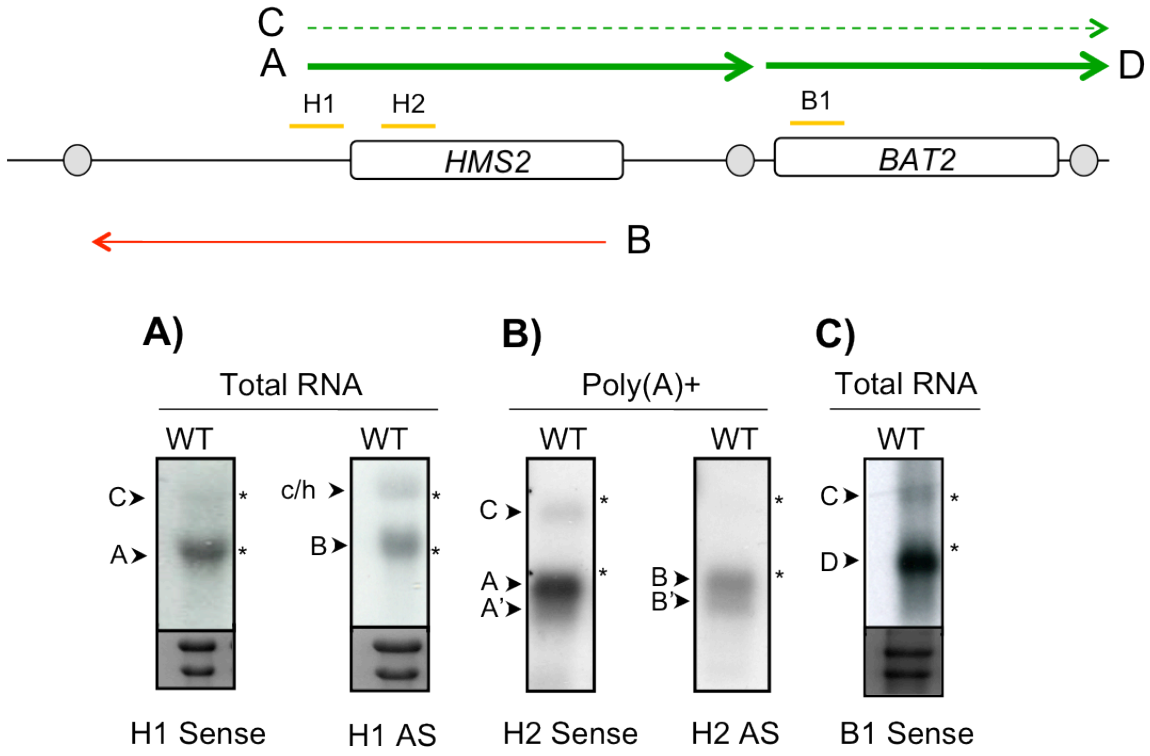


Figure 16. Overlapping Transcription at the *HMS2* locus.

(A) Northern blot analysis of the total RNA using probe H1 detected sense (transcript A) and antisense (transcript B) transcripts. Cross-hybridization (c/h) to the 25S rRNA is especially apparent using the H1 antisense probe. **(B)** The *HMS2* SAP is also detectable in RNA samples enriched for polyadenylated transcripts using probe H2, as well as the presence of a read-through transcript (transcript C). Sense and antisense transcript size heterogeneity is detected in the poly (A)+ extracts (A' and B'). **(C)** Using probe B1, sense transcription over the 5' region of the downstream *BAT2* coding region was also observed, and the *BAT2* sense transcript (transcript D) was detected in addition to the read-through transcript C. Using formaldehyde 1% agarose gels, 10 µg of RNA was separated in each lane for 3h at 100V. The asterisks represent the positions of the 25S (3392nt) and 18S (1789nt) RNA bands. The positions of transcripts A, B, C, D are consistent with their transcript sizes, estimated by collating start site data available in published data (Granovskaia et al. 2010) and end sites determined by 3'RACE in this study. For blots containing total RNA, ethidium bromide-stained rRNA loading controls are displayed below. The asterisks represent the positions of the 25S (3392nt) and 18S (1789nt) RNAs. RNA depleted of ribosomal fractions tended to run further on the gel, attributing for small size differentials between the transcripts in the total and Poly (A)+ RNA samples. Major transcripts and probe locations (yellow bars) are indicated in the above schematic.

and/or end sites not identified by 3'RACE. Therefore, it is possible that the discrepancy between the AS 3' end mapping in this study and the end site identified by Granovskaia et al. (2010) might reflect the transcript isoforms observed via Northern blot analysis. Additionally, by removing non-polyadenylated RNA from the extract, non-specific hybridization to the ribosomal bands was largely avoided, and a substantially longer RNA species was also detected from the sense strand. A sense probe for the coding region of the downstream gene, *BAT2*, identified this longer species as an *HMS2-BAT2* read-through transcript (Figure 16C, transcript C). Given its molecular weight, it is likely that transcript C initiates at the *HMS2* promoter and terminates in the vicinity of the *BAT2* poly (A) site. Furthermore, detection of the *BAT2* sense transcript (transcript D) on a Northern blot is consistent with its predicted size based on 3' RACE analysis. Overall, by estimating transcript sizes relative to the 18S (~1789 nt) and 25S (~3392 nt) ribosomal RNA markers, Northern blot analysis of transcripts occurring across the *HMS2* locus was largely consistent with the mapping data. However, it should be noted that this method of estimating transcript sizes becomes less reliable in terms of accuracy when considering transcripts of higher molecular weight.

3.2.2. HMS2 sense and antisense are substrates of Rrp6p and Xrn1p but are regulated through distinct mechanisms:

While the Northern blot analysis of the *HMS2* locus offered a glimpse of the transcripts present in a population of cells at a particular point in time, it did not give an indication of the nature of the transcripts produced i.e. the rate at which they were transcribed or subjected to decay, properties which might have major consequences on gene

expression regulation, as described at *PHO84*. Thus, potential regulation through an Rrp6p/Hda1p- or Xrn1p-mediated mechanism was tested. Both the *HMS2* sense and antisense transcripts displayed equal sensitivity to degradation via the nuclear exosome (Rrp6p) and cytosolic processing by Xrn1p (Figure 17A). Most notably, in the absence of Rrp6p and Xrn1p, the *HMS2* SAP did not display evidence of a reciprocal relationship. This is in contrast to the Camblong et al. (2007) study on the *PHO84* antisense Rrp6p-sensitivity and the van Dijk et al. (2011) study, which demonstrated Xrn1p-influcenced sense/antisense reciprocity for 273 antisense-producing units, genome-wide. The *HMS2* SAP was also observed in the absence of Hda1p, but no considerable changes in transcript levels were evident via Northern blot analysis (Figure 17B). Together, the evidence from analyzing the *HMS2* SAP in these mutant strains points to a novel, independent regulatory mechanism acting on the *HMS2* SAP, distinct from those described in previous studies by Camblong et al. (2007) and van Dijk et al. (2011).

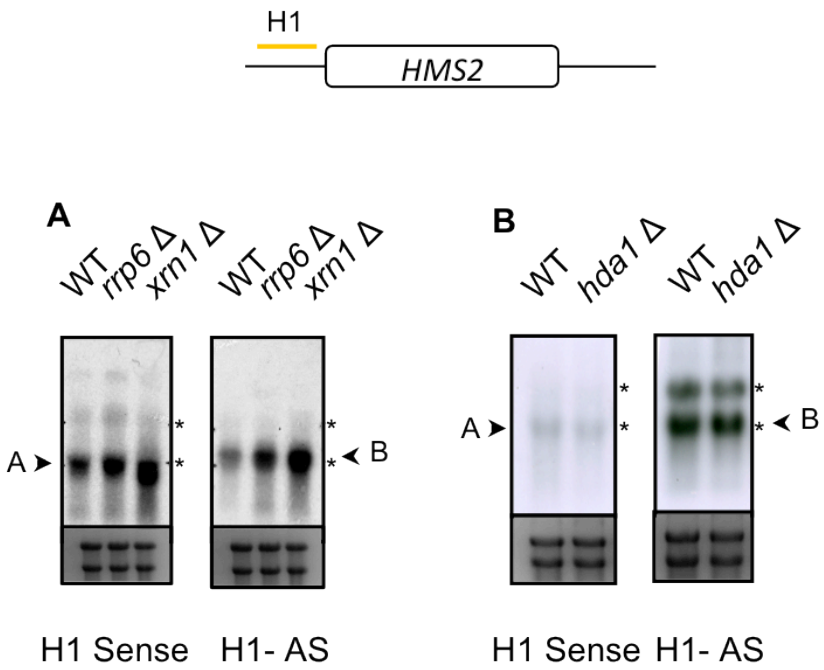


Figure 17. Existing mechanisms of CUT and XUT sense/antisense regulation do not apply to the *HMS2* locus.

Northern blot analysis showing levels of the *HMS2* sense/antisense (transcript A and B, respectively) pair in the presence or absence of the **(A)** RNA processing machinery components Rrp6p and Xrn1p and **(B)** the catalytic subunit of the histone deacetylase complex, Hda1p. Probe location is represented by the yellow bar on the schematic above. Using formaldehyde 1% agarose gels, 10 µg of Total RNA was separated in each lane for 2.5h at 100V. Major transcripts are marked by black arrows. The asterisks represent the positions of the 25S (3392nt) and 18S (1789nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below the blots.

3.2.3. Regulation of the HMS2 locus:

In order to better understand the contribution of overlapping transcription to *HMS2* regulation, techniques to manipulate both sense and antisense expression levels had to be established. Unlike *PHO84*, which is regulated in a switch-like manner by a specific nutrient sensitive transcriptional activator, Pho4p, regulation of *HMS2* transcription remains largely uncharacterized.

According to a study by Lorenz and Heitman (1998), overexpression of *HMS2* is involved in initiating pseudohyphal growth in diploid PHG-deficient *mep1*/*mep2* mutants. While this effect may be through indirect genetic interactions, it was possible that *HMS2* transcript levels could be manipulated by altering growth conditions. For instance, by changing the carbon source from glucose to galactose, we found that *HMS2* sense levels decreased whereas antisense levels increased, indicating a potential reciprocal relationship between the transcript pair (Figure 18). Changes in transcript levels were unlikely to be a direct function of galactose metabolism since there are no Gal4p binding sites in the vicinity of either sense or antisense *HMS2* promoters. Thus, transcriptional regulators that could have elicited these observed changes in sense/antisense levels were investigated.

Based on available ChIP-chip data, MacIssac et al (2006) designed an algorithm to predict the potential binding site motifs for 203 different transcription factors, genome-

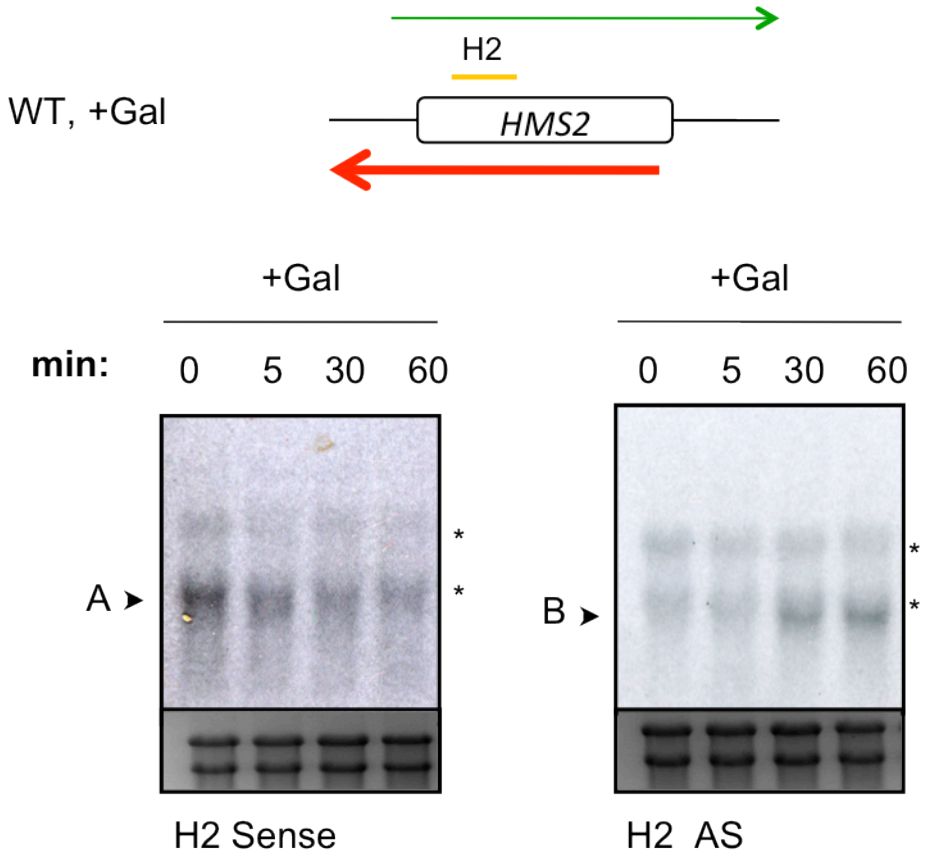


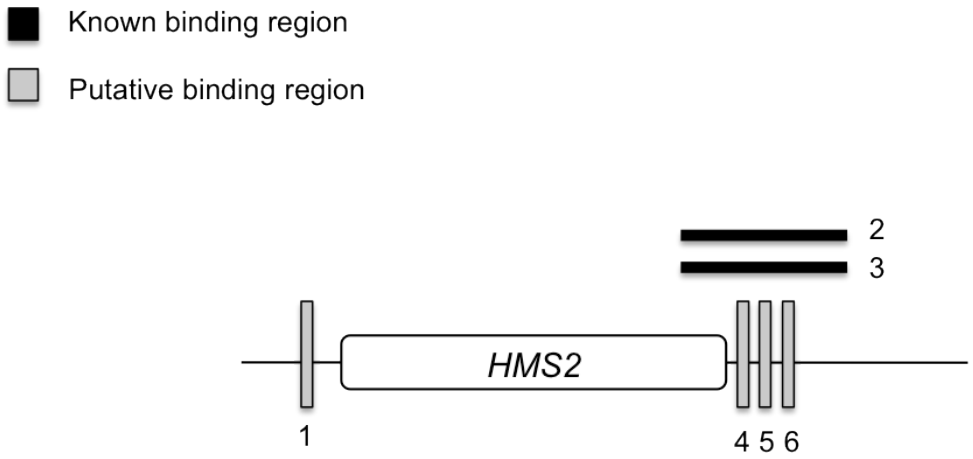
Figure 18. The *HMS2* sense/antisense pair can be manipulated in galactose.

Northern blot analysis showing levels of the *HMS2* sense/antisense pair when cells were grown in glucose-containing medium until OD_{600}= 0.5 (t= 0 min) and after growth in galactose-containing medium for 5, 30, and 60 minutes. During induction *HMS2* sense (transcript A) levels decrease, while antisense levels (transcript B) increase. The probe location and transcripts of interest are depicted in the above schematic. Using formaldehyde 1% agarose gels, 10 µg of total RNA was separated in each lane for 2.5h at 100V. Major transcripts are marked by black arrows. The asterisks represent the positions of the 25S (3392nt) and 18S (1789nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below the blots.

wide (Macisaac et al., 2006). Earlier work by Tachibana et al (2005) had mapped the genomic binding locations of the transcriptional regulators Cat8p and Adr1p using ChIP-chip (Tachibana et al., 2005). Combining these datasets, sequence ±500 base pairs from either the *HMS2* start codon (defined as "5' *HMS2*") or stop codon (defined as "3'*HMS2*") was interrogated for putative transcription factor motifs and bound Cat8p/Adr1p. Multiple putative binding sites were found at the 3' end of *HMS2* including motifs for Rtg3p, Skn7p, and Gln3p as well as a single Cbf1p binding-site motif at the 5' end of *HMS2* (Figure 19). Interestingly, Cat8p and Adr1p were both found to bind to the 3'end of *HMS2*. Since Cat8p and Adr1p are described as carbon source-sensitive transcriptional regulators, the role they might play in altering *HMS2* sense/antisense ratios in the shift from glucose to galactose was examined (Young, Dombek, Tachibana, & Ideker, 2003). However, in the absence of Cat8p, Adr1p, or both regulators, no considerable changes in sense or antisense transcript levels were observed in cells grown in either glucose- or galactose-containing medium in comparison to wild-type levels (Figure 20). Considering that Cat8p and Adr1p were the best candidate regulators of the *HMS2* locus and since redundancy could confound the results of a screen for regulators of *HMS2* expression, alternative methods of manipulating sense/antisense levels were explored and are discussed in greater detail in the next chapter.

3.2.4. TFIIB regulates HMS2 sense and antisense transcription:

While the specific transcription factors regulating the *HMS2* sense/antisense pair remained elusive, it is likely that the same GTFs directing transcription at the 5' end would also operate at the 3'end in order to precisely initiate antisense transcription. For



Transcription Factor	Coordinates	Description	Source
5' end *HMS2*			
1. Cbf1	703798, 703805	Algorithm generated putative binding site	MacIssac et al. 2006
3' end *HMS2*			
2. Adr1p	705188, 705679	ChIP-chip	Tachibana et al. 2005
3. Cat8p	705188, 705679	ChIP-chip	Tachibana et al. 2005
4. Rtg3p	705287, 705293	Algorithm generated putative binding site	MacIssac et al. 2006
5. Skn7p	705374, 705381	Algorithm generated putative binding site	MacIssac et al. 2006
6. Gln3p	705459, 705468	Algorithm generated putative binding site	MacIssac et al. 2006

Figure 19. Potential transcriptional regulators of the *HMS2* locus.

A schematic of computationally predicted and known transcription factor binding sites (via ChIP-chip) found +/- 500 base pairs from either the *HMS2* start (5' end) or stop (3' end) codon. Grey bars represent putative transcription factor binding sites while black bars represent locations of known bound sites.

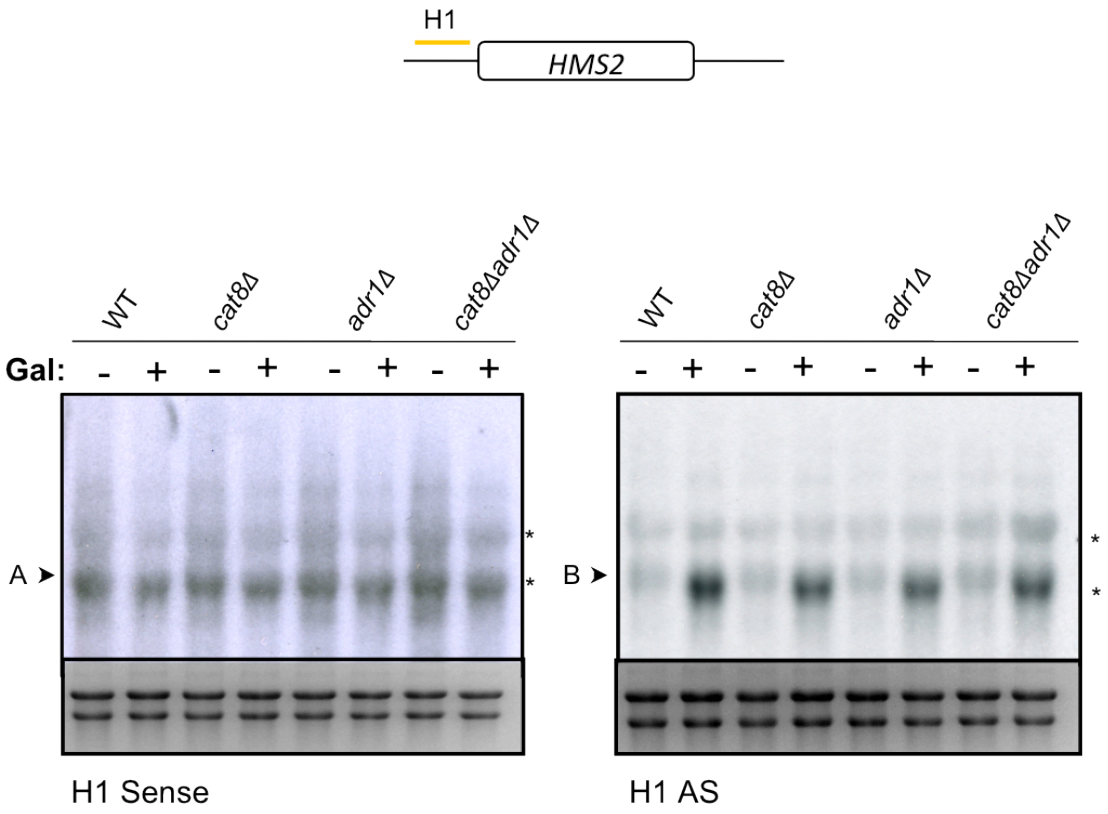


Figure 20. Disruption of transcription factor binding does not alter sense/antisense transcript levels.

Northern blot analysis of *HMS2* sense and antisense (transcripts A and B, respectively) levels in the presence or absence of transcription factors Cat8p and Adr1p. Probe location (H1) is represented by the yellow bar in the above schematic. Cells were grown in glucose-containing medium (-) until OD_{600}= 0.5 and shifted to galactose-containing (+) medium for 1h. Using formaldehyde 1% agarose gels, 10 µg of Total RNA was separated in each lane for 2.5h at 100V. Major transcripts are marked by black arrows. The asterisks represent the positions of the 25S (3392nt) and 18S (1789nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below.

instance, a recent study by Murray and colleagues (2012) demonstrated that at antisense producing transcription units, average levels of PIC components such as TBP and TFIIB are ~60% of those at the 5' end of the sense gene. To assess whether the *HMS2* antisense transcription unit adhered to this archetypal promoter architecture, *HMS2* antisense levels were observed when TFIIB was depleted from the nucleus using the Anchor Away technique (Haruki et al., 2008; see Methods and Materials). Indeed, after 1 h of rapamycin treatment, *HMS2* sense (transcript A), antisense (transcript B), and *BAT2* (transcript D), all demonstrated a decrease in transcript levels (Figure 21A, B, C). This is in contrast to the reciprocal behaviour previously exhibited by the *HMS2* SAP, perhaps indicating that like coding genes *HMS2* and *BAT2*, antisense transcription at the *HMS2* locus is subject to regulation by general transcription factors such as TFIIB. Furthermore, while transcript A demonstrated a decrease in levels, read-through transcription in the sense direction appeared to be enhanced, with a longer read-through transcript, transcript C', being detectable when TFIIB was depleted. Transcript C' is consistent in size with a transcript initiating at the *HMS2* promoter and terminating in the vicinity of a consensus poly (A) signal found mid-way through the gene downstream of *BAT2*, *YJR149W* (data not shown), resulting in a transcript approximately 1 kb larger than read-through transcript C, which reads from *HMS2* through the *BAT2* coding region. Indeed, transcript C' was detectable using the *BAT2* sense probe; however the subtle increases in levels of transcript C' when TFIIB was depleted were not apparent (Figure 21C). This could be due to differences in hybridization efficiencies or threshold effects between probe H2 and B1. The shift to producing sense read-through

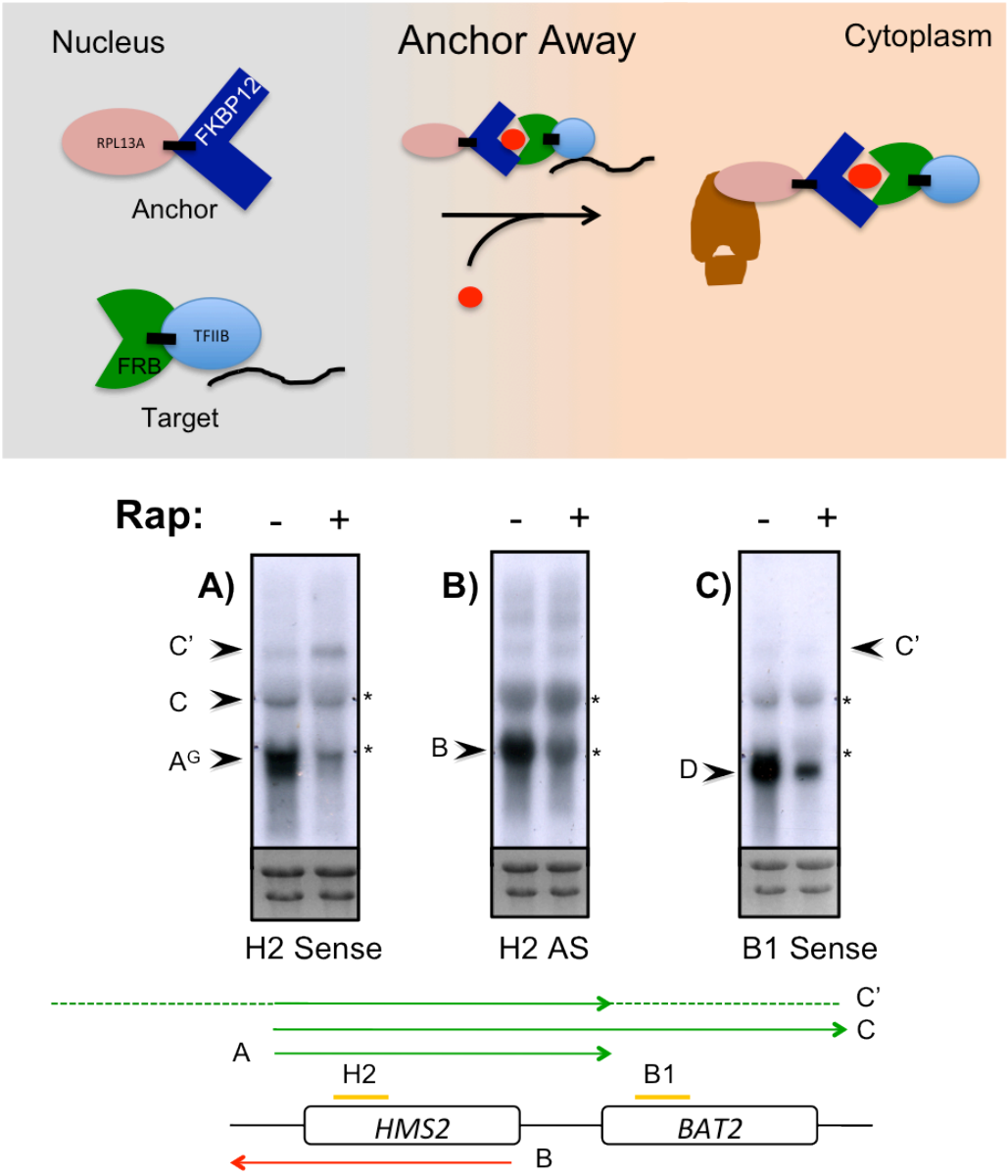


Figure 21. Depletion of TFIIB using the Anchor Away technique.

Northern blot analysis of the *HMS2* transcript profile after TFIIB was depleted from the nucleus using the Anchor Away technique (See Methods and Materials; Haruki et al. 2008), represented by the schematic above. TFIIB was FRB tagged using the pFA6a-*kanMX6*-FRB plasmid (Euroscarf) by homologous recombination (Longtine et al. 1998) in the *RPL13A-FKBP12*-tagged background (Euroscarf). Probe locations (H2, B1) are represented by the yellow bars in the schematic of the locus blow the blots. As in BY4741, a sense/antisense pair is detectable in the Anchor Away, W303 derived, genetic background. Rapamycin was added (Rap+) to exponential phase cells growing in YPD at OD_{600}= ~ 0.5. DMSO was added to a control group of cells (Rap-). Cells were harvested after 1h treatment. **(A)** After 1h treatment with rapamycin, levels of sense transcript (transcript A) were reduced, while levels of the sense read-through (transcript C)

remained relatively constant. A newly-identified longer read-through transcript (transcript C') was detected at increased levels in comparison to the control group. **(B)** In the antisense direction, levels of antisense (transcript B) were also reduced when TFIIB was depleted. **(C)** At the *BAT2* locus, levels of *BAT2* sense transcript (transcript D) decreased in comparison to the control group, while levels of read-through transcript C remained constant. Transcripts of interest are marked by black arrows and are depicted in the schematic below the blots. Using formaldehyde 1% agarose gels, 10 µg total RNA was separated in each lane for 2.5h at 100V. The asterisks represent the approximate positions of the 25S (3392nt) and 18S (1789nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below.

transcripts when antisense transcription was compromised perhaps indicates a striking role for antisense transcription in defining transcription unit boundaries; although, more work must be done in order to substantiate this hypothesis. However in terms of antisense regulation, insofar as it can be ascertained, given its regulation by GTFs, responsiveness to environmental changes, and post-transcriptional modification (e.g. polyadenylation), antisense transcription at the *HMS2* locus is ostensibly regulated in the same manner as transcription of protein-coding genes. Importantly, the purported regulation of antisense through TFIIB, suggests, in support of the work by Murray and colleagues (2012) that transcription units producing sense and antisense transcripts are flanked and regulated by two promoters.

3.2.5. Inter-locus dependency at the HMS2 gene cluster:

Given the detection of the *HMS2-BAT2* read-through transcript (transcript C), it was especially important to gain further insight into the transcriptional events surrounding *HMS2* in order to gauge any potential effects overlapping transcription might have on inter-locus relationships. In collaboration with the Steinmetz group at the European Molecular Biology Labs (EMBL), both total (Figure 22A) and poly (A)+ RNA extracts

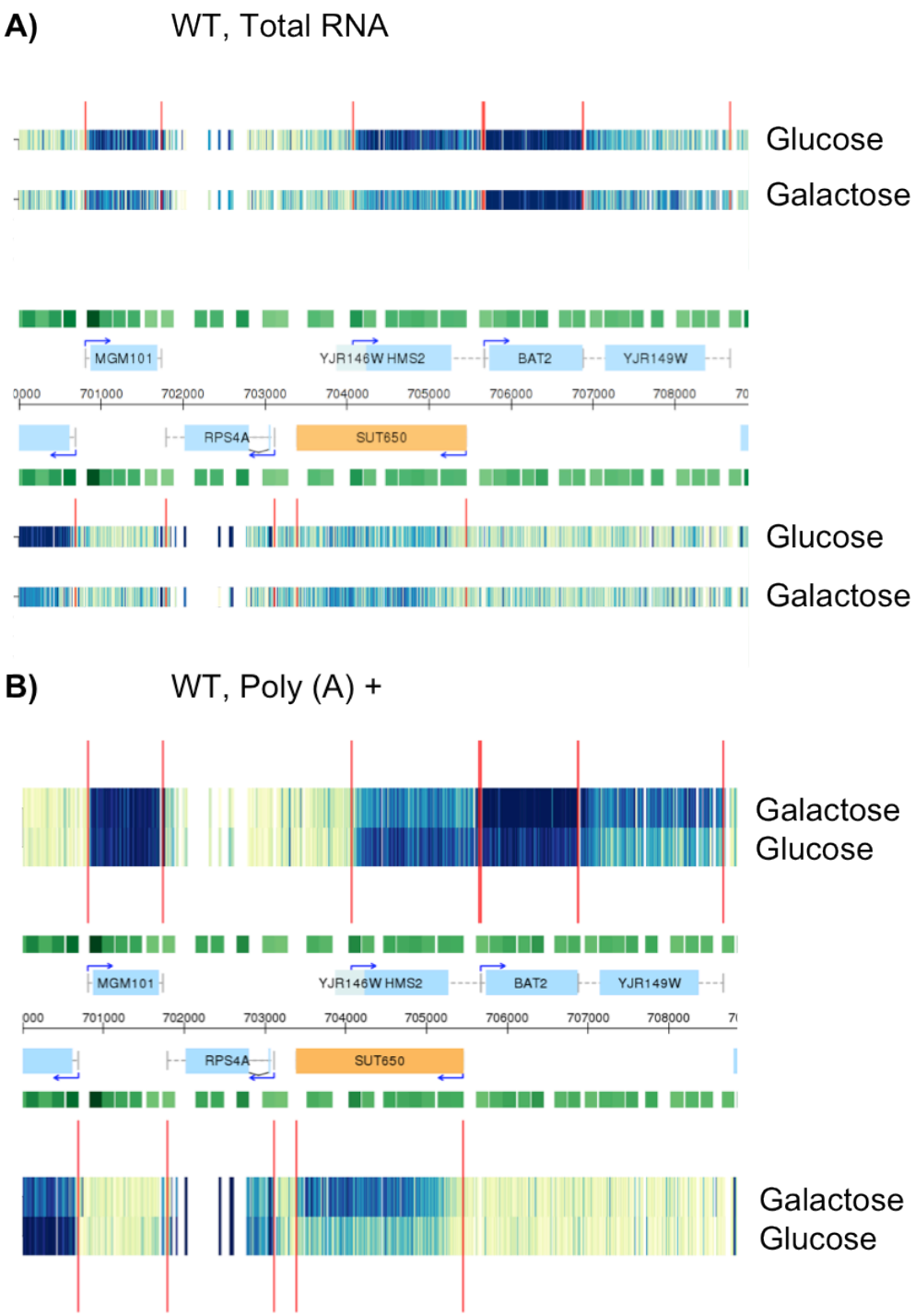


Figure 22. Transcription over the *RPS4A*, *BAT2*, and *YJR149W* loci is linked to *HMS2* transcription in glucose- versus galactose-containing media.

Strand specific high-resolution tiling arrays were hybridized to **(A)** total and **(B)** poly (A)+ RNA (polyadenylated RNA purified by Ana Serra Barros) in collaboration with Lars Steinmetz at the European Molecular Biology Laboratories. Wild-type cells were grown in glucose-containing medium for 3h before being shifted and grown in galactose-

containing medium for an additional 3h. Yellow signifies no hybridization, while blue signifies hybridization. Blue boxes= coding regions, orange boxes= non-coding transcription units, green boxes= nucleosome positions.

(Figure 22) were hybridized onto strand-specific, high-resolution tiling arrays and the transcription profiles of *HMS2*, the upstream gene *RPS4A* and the downstream gene *BAT2* were analysed after cells were grown in glucose or galactose-containing medium. As seen via Northern blot analysis, when cells were shifted from YPD to YPGal, levels of detectable *HMS2* antisense transcript increased, while sense levels decreased. Furthermore, *BAT2* transcript levels reflected the changes in *HMS2* antisense expression, and were also found at increased levels in RNA from cells grown in galactose-containing media, demonstrating potential co-regulation between the *BAT2* sense and *HMS2* antisense promoters. Interestingly, in both total and poly (A)+ extracts, levels of *RPS4A* mirrored those of *HMS2* and demonstrated a modest decrease in the glucose to galactose shift.

3.2.6. Single cell analysis of HMS2 expression using RNA FISH reveals temporal separation of sense and antisense transcription:

While preliminary observations through Northern blot analysis and high-resolution tiling array suggested a reciprocal relationship between the *HMS2* sense/antisense pair, it is important to bear in mind that these methods described the transcript pool of a population of asynchronous cells. Any reliable interpretation of this data was contingent on how the *HMS2* SAP behaved on a single cell level; thus, to complement our Northern blot and tiling array data, cell-to-cell variability in *HMS2* sense/antisense expression was observed under growth in glucose- and galactose-containing media using strand-specific

single molecule RNA FISH (see Methods and Materials). As in Zenklusen et al. (2008), three-dimensional data sets were depicted as two-dimensional images in order to simplify image analysis (Zenklusen, Larson, & Singer, 2008). Using this approach, both sense and antisense transcripts could be detected in glucose (Figure 23 A, B), albeit not in all cells (Figure 23B-ii). Single transcript detection manifested as a small fluorescent spot. Also as in Zenklusen et al. (2008), sites of active transcription were likely represented by higher intensity spots that co-localized with the nuclear DAPI signal, represented in blue (Figure 23A-iv). Most cells contained only one or two spots of sense or antisense. Moreover, closer examination of single cells grown in glucose-containing medium revealed that sense and antisense transcripts were rarely detected in the same cells and if they were, FISH signals were not overlapping, suggesting a degree of temporal separation in their generation. This was particularly apparent in cells with detectable active sites of transcription (Figure 23A-iv, B-v). Thus, what appeared as sense and antisense transcription via tiling array and Northern blot analysis was actually a product of a mixed population of cells producing either the sense or antisense transcript, but not both, at least not at the same time. Likewise, when transcripts from the *HMS2* locus were observed for single cells shifted to galactose-containing medium for 1 h, a greater and smaller proportion of cells were detected producing antisense and sense transcripts, respectively, in comparison to growth in glucose-containing medium (Figure 23C, D). Furthermore, in the representative images, no sites of active sense transcription were discernible in galactose-grown cells, whereas in two fields of view of

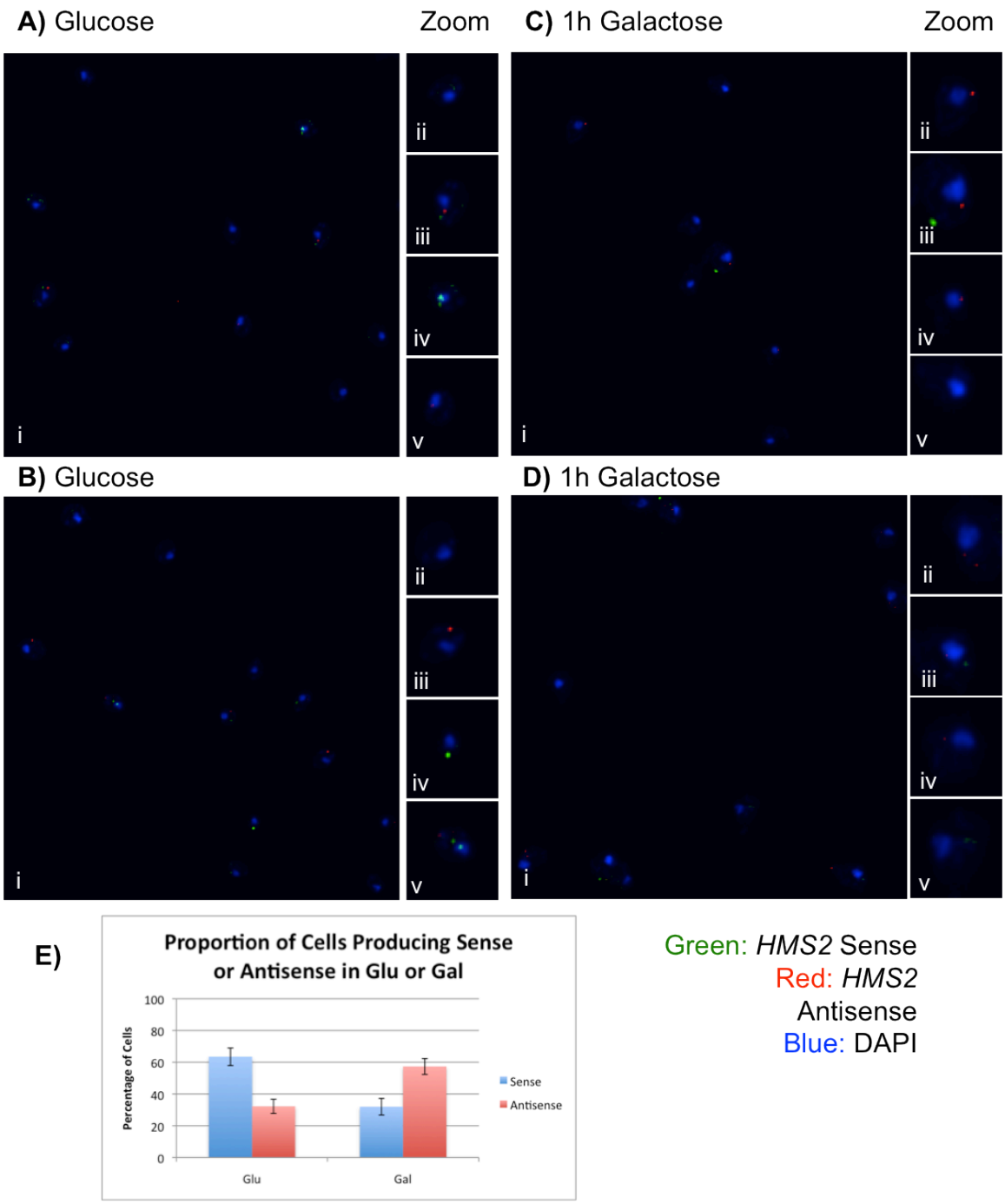


Figure 23. Single-cell analysis of *HMS2* expression using RNA FISH.

HMS2 expression was visualized in single cells grown to OD_{600}= 0.500 in glucose-containing media **(A, B)** then shifted to galactose-containing media for 1h **(C, D)**. In order to capture single *HMS2* sense (green) and antisense (red) transcripts, a combination of four, 50 nt DNA probes labeled with four fluorophores, either Cy5 (sense) or Cy3 (antisense), were hybridized to paraformaldehyde fixed yeast cells. DAPI nuclear stains are displayed in blue. Smaller boxes (ii-v) to the right of the main pictures

are zoomed images of select cells in the field of view (i). The images presented here are part of a larger data set. The graph **(E)** represents the proportion of sense and antisense-expressing cells between growth conditions across the entire sample. A total of 75 cells were assessed in each growth condition. Error bars represent standard deviations, which were determined by producing 1000 bootstrap pseudoreplicates of both datasets (glucose vs. galactose) and comparing the resultant distributions of percentages using a Wilcoxon rank-sum test (statistical analysis performed by Struan Murray). $p= 7.2x10^{-18}$ (sense comparison) and $p= 3.4 \times 10^{-34}$ (antisense comparison).

glucose-grown cells, two cells displayed evidence of actively transcribing *HMS2* sense (Figure 23A-iv, B-v). Thus, while higher rates of *HMS2* antisense transcription in galactose-grown cells cannot be ruled out as an additional possibility, the increases in *HMS2* antisense transcript observed in galactose via Northern analysis can be unambiguously attributed to more cells producing antisense versus sense transcripts.

Systematic phasing or partitioning of transcriptional events through time has been described in earlier reports, specifically in cellular programmes such as the cell and yeast metabolic cycles (Klevecz et al., 2004; Tu et al., 2005). The Granovskaia et al. (2010) study identified the *HMS2* SAP as cell cycle regulated, with sense and antisense having transcript levels peaking during late G1/early S and late M/early G1, respectively. In this study, cell cycle regulation of *HMS2* was confirmed using Northern blot analysis. While no differences in sense/antisense levels were apparent when cells were treated with the α -factor pheromone, which arrests Mat a cells (e.g. BY4741) in G1 (Figure 24A), when cells were treated with nocodazole arresting them in the G2/M phase of the cell cycle,

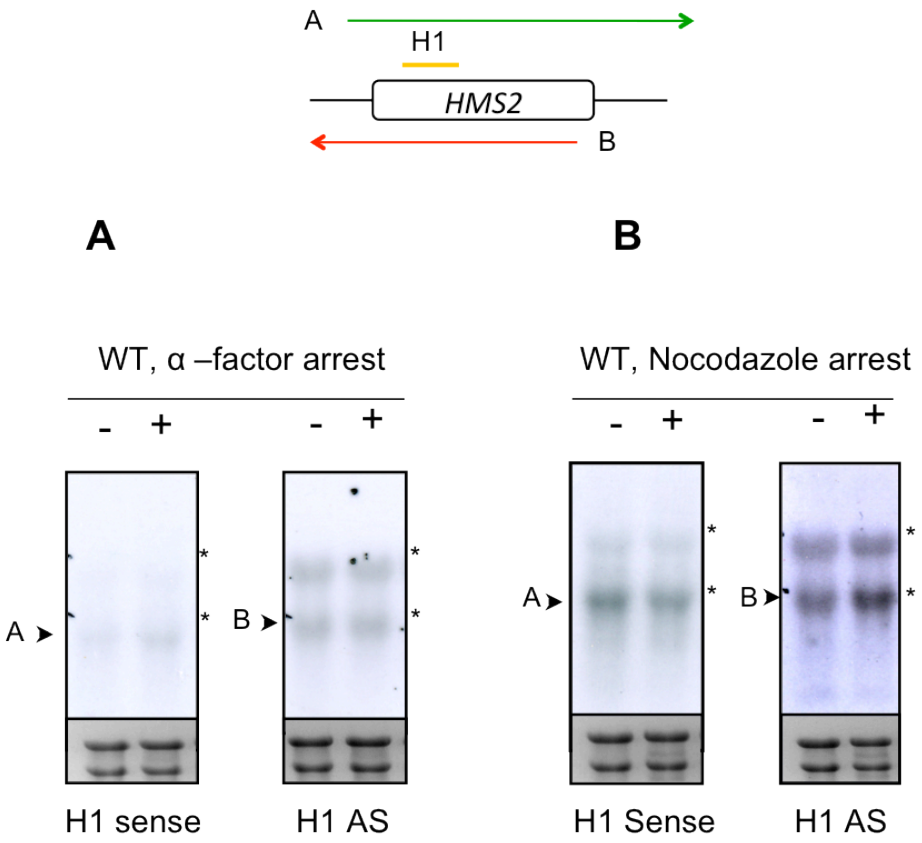


Figure 24. The *HMS2* locus displays regulation in the cell cycle.

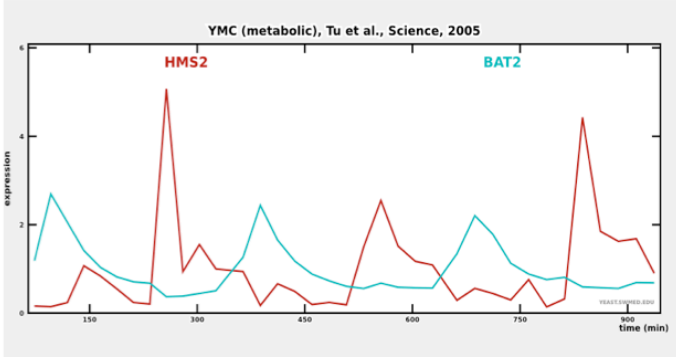
Northern blot analysis of sense/antisense levels in asynchronous (-) and **(A)** alpha-factor/G1 or **(B)** nocodazole/G2/M arrested (+) wild-type cells. The sense and antisense blots in (B) represent a 6 day and 18h exposure, respectively. Using formaldehyde 1% agarose gels, 10ug of Total RNA was separated in each lane for 2.5h at 100V. Major transcripts are marked by black arrows. The asterisks represent the positions of the 25S (3392nt) and 18S (1789nt) RNAs. Ribosomal RNA loading controls are displayed below.

increased levels of antisense and decreased levels of sense transcripts were observed, providing further support for a time-dependent reciprocal relationship (Figure 24B).

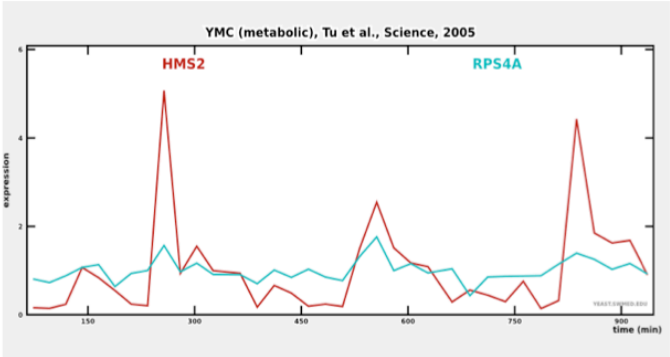
Furthermore, temporal separation of *HMS2* and *BAT2* transcription was previously revealed in a study describing gene expression periodicity in the long phase yeast metabolic cycle (YMC) (Tu et al., 2005). As reported in the study, *HMS2* is most highly

expressed in the growth-associated, high oxygen consumption, oxidative phase of the metabolic cycle along with upstream gene *RPS4A*, while *BAT2* is most highly expressed during the reductive charging (RC) phase, which is largely associated with quiescence and non-respiratory, low oxygen consumption forms of metabolism (Figure 25A, 25B, 25C). Together the inter-locus relationships found in the YMC are consistent with those described in this study when cells were shifted to galactose-containing medium. By extension, since *BAT2* displays coordinate expression with the *HMS2* antisense transcript in galactose, it is possible that the *HMS2* antisense is also transcribed during the reductive charging phase, providing another biological explanation for the cell-to-cell variation in sense/antisense expression observed with RNA FISH. While this study does not provide direct evidence of *HMS2* antisense expression in the reductive charging phase, the conjecture is consistent with Northern analyses that showed increased *HMS2* AS levels and concomitant decreases in sense levels when cells were arrested in the G2/M transition of the cell cycle, as previous studies have indicated that that mitosis and the RC phase of the YMC can occur contemporaneously (Klevecz et al., 2004; Tu et al., 2005). Furthermore, Tu et al. (2005) also identified cycling transcription factors, two of which, Adr1p and Gln3p, are found at the 3'end of *HMS2* and have peak expression levels during the reductive charging phase (Figure 25D). Thus, all evidence is consistent with *HMS2* sense/antisense transcription being controlled within the biological and temporal framework of the Yeast Metabolic Cycle.

A



B



C

Gene	YMC Phase
HMS2	OX
BAT2	RC
RPS4A	OX

D

Transcription Factor	YMC Phase
ADR1	RC
RTG3	RB
GLN3	RC

Figure 25. Periodicity of genes in the *HMS2* cluster in the Yeast Metabolic Cycle.

HMS2 peak transcript levels during three rounds of the Yeast Metabolic Cycle (Tu et al., 2005) in relation to **(A)** *BAT2* and **(B)** *RPS4A*. Taken from: (http://moment.utmb.edu/cgi-bin/sceptrans.cgi). **(C)** Respective phases of the metabolic cycle (OX, oxidative phase; RB, reductive building phase; RB, reductive charging phase) for *HMS2*, *BAT2*, and *RPS4A* as well as **(D)** a list of potential cycling transcription factors regulating the *HMS2* locus.

3.3. Discussion:

It is evident from this analysis that the *HMS2* locus exemplifies several departures from the conventional, linear model of transcription unit architecture. The presence of non-coding, overlapping transcription in both the sense and antisense orientation embodies the characteristics of a pervasively-transcribed, interleaved genome, providing a unique model system in which to study the regulatory consequences of this type of genomic configuration.

3.3.1. Anti-correlated sense/antisense levels depicted by Northern blot and tiling array analysis is a manifestation of single cells generating either sense or antisense:

While previous studies have predominantly reported an anti-correlated relationship between overlapping sense/antisense transcription, few have been able to establish a comprehensive model for why this might be. From the evidence provided in this analysis, *HMS2* does not fit the two chief categories of *PHO84* (CUT) or XUT regulation, but might represent a more general example of mediation through stable antisense expression. Rather, in terms of its organization, the *HMS2* locus more closely resembles the arrangement described by the Xu et al. (2011) study, which demonstrated that the sense/antisense anti-correlation was contingent on transcriptional overlap of either the sense or antisense transcription start sites. *HMS2* exhibits this overlap at both ends of the transcription unit, signifying that at this locus, sense and antisense transcription might share a mutually inhibitory relationship. Indeed, Northern blot and tiling array experiments confirmed that in the shift from glucose- to galactose-containing medium

and while arrested in the G2/M transition, the *HMS2* antisense displays anti-correlated levels in relation to its sense pair.

The Xu et al. (2011) report also contends that promoter overlap allows antisense transcription to shut off sense transcription, specifically when the sense gene is expressed at low levels. They cite *SUR7,* which displays low levels of sense transcription when treated with α-factor, as an example of this phenomenon. *SUR7*'s antisense pair, *SUT719*, is expressed at low levels in glucose-grown cells and high levels in galactose-grown cells. When cells are treated with α-factor and grown in galactose-containing media, high levels of antisense transcript are detected and sense levels appear stifled, as opposed to when cells are not exposed to α -factor and grown in galactose-containing media, in which case high levels of both sense and antisense transcripts are detected. The study also claims that 'deletion' of the *SUT719* antisense alleviates the inhibitory effect under these same growth conditions. While seemingly convincing, this model system suffers from two major drawbacks. The first is that the convoluted changes in growth conditions exerted to elicit the observed effects might also affect the turnover or stability of the sense and antisense transcripts, so that the subtle changes observed in the analysis might not reflect transcriptional events *per se*. Secondly, and more importantly, upon closer inspection of the tiling array data from their antisense deletion strains, evidence at the crux of their assertions, a moderately expressed, read-through antisense RNA can be detected initiating from the middle of downstream gene *GAL80* and overlapping the *SUR7* promoter (Figure 26). Since *SUR7* levels remained unchanged

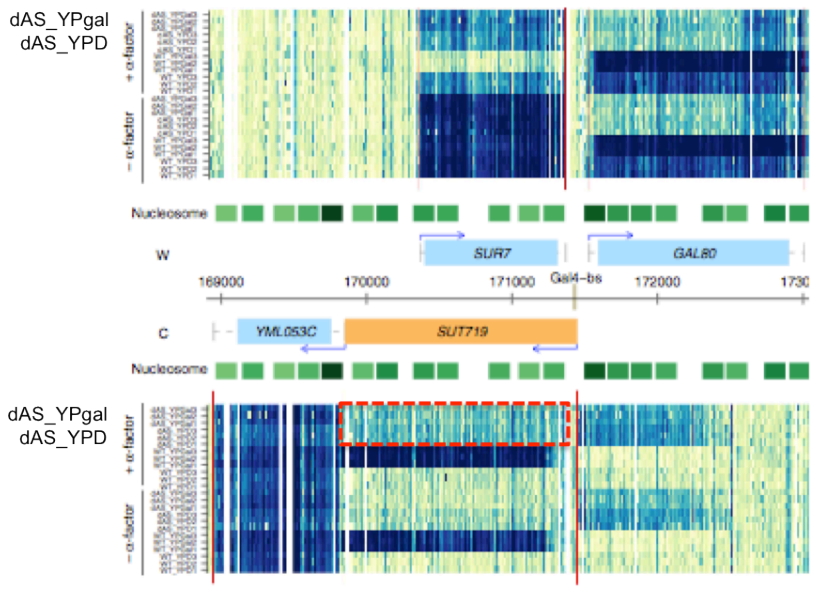


Figure 26. Knocking out the 3' Gal4p site does not abolish overlapping antisense transcription at the *SUR7* locus.

Xu and colleagues (2011) attempted to disrupt antisense transcription at the *SUR7* locus by knocking out the Gal4p binding sites (dAS_YPD, dAS_YPgal) in the shared promoter between *SUT719* and *GAL80*. While the study claims this approach was successful in "disrupting" antisense expression, antisense levels were still detectable over the *SUR7* in the Gal4p binding site deletion strain (boxed in red dotted line). Figure adapted from Xu et al. (2011). Blue boxes= coding regions, orange boxes= non-coding transcription units, green boxes= nucleosome positions.

in this scenario, this demonstrates that moderate levels of sense and antisense can co-exist in a manner similar to that of *HMS2*. While one might conclude from this study that antisense transcription might only exert inhibitory effects on low-activity sense promoters conditionally (i.e. when antisense is transcribed at high levels), a simpler, alternative explanation was elucidated at *HMS2* when sense/antisense expression was observed at a single-cell level.

Single-cell analysis of sense/antisense transcripts in this study demonstrated that sense and antisense were rarely detected in the same cells or cellular compartments, suggesting a degree of temporal separation in their generation. This evidence can reconcile the seemingly conflicting observations in the Xu et al. (2011) study, informing a more likely model in which sense/antisense transcription behaves in a bimodal switch-like manner, through which antisense transcription excludes sense transcription from occurring and vice-versa, as opposed to acting as a level-dependent buffer of transcription. Further evidence from this study revealed that the changes in *HMS2* transcript levels observed via Northern blot and tiling array analysis in glucose- and galactose-grown cells actually represented changes in the proportion of cells in the population expressing sense or antisense transcripts rather than changes in expression range capacity of the sense or antisense promoters. This result is consistent with the co-existence of moderate levels of sense and antisense observed at *HMS2* and *SUR7*, with changes in transcript levels reflecting shifts in the sense/antisense equilibrium on a single cell basis and not inhibition or shut down of the sense promoter in the classical sense. Given these new insights, Northern blot and tiling array data were reinterpreted in the context of the RNA FISH data, and a new model for how overlapping transcription contributes to gene regulation was proposed.

3.3.2. Establishing a model:

Integrating all of the evidence from our analysis of the *HMS2* locus, including a reciprocally-related, temporally-segregated sense/antisense pair; co-regulation of neighbouring genes *RPS4A* and *BAT2;* as well as read-through transcription, a two-state

model for overlapping transcription that combines the effects of transcriptional exclusion (i.e. temporal separation) and the resultant production of long interfering transcripts was proposed (Figure 27). According to this model, as a result of the architecture required for overlapping transcription, a gene can be regulated by integrating environmental signals at the promoters at both its 5' and 3' ends. These signals determine the dominant transcription unit (for instance, *HMS2* antisense in response to growth in galactose-containing media), creating two states, either sense- or antisense-dominant. These states are likely to occur within the context of time-dependent physiological programmes like the cell or metabolic cycles, through which transcription factor availability is determined. During these phases, it is also likely that the production of one transcript excludes the production of its overlapping pair because of polymerase collisions; thus, making temporal separation of the two states favourable (Hobson, Wei, Steinmetz, & Svejstrup, 2012; Prescott & Proudfoot, 2002). Furthermore, transcripts produced in these distinct states, either sense, antisense or additional read-through transcripts, might then transmit the state-defining regulatory signals to neighbouring genes through traditional mechanisms of *cis*-acting transcription interference. Bearing in mind these parameters, in the following chapters the predictive power of this model will be tested (Chapter 4) and the mechanisms and functional consequences of state-switching will be evaluated (Chapter 5).

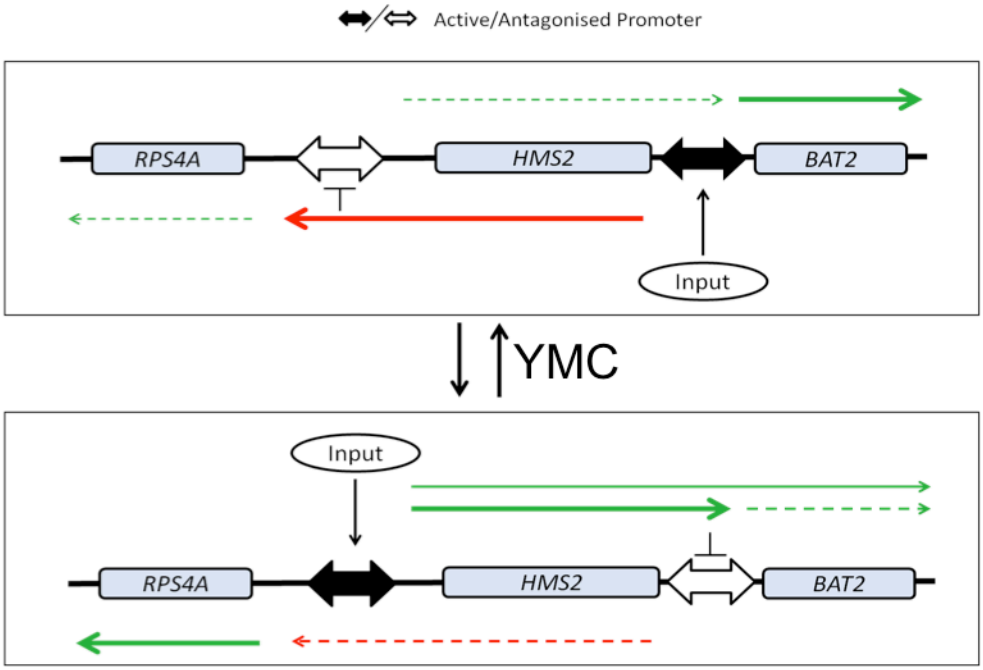


Figure 27. Model of a two-state system at *HMS2.*

Overlapping transcription results in two transcriptional states due to transcriptional exclusion. Transcription from either the sense or antisense promoter can drive the production of interfering transcripts that can act as "long range" regulators. These states are likely to be associated with responses to changes in the environment or physiological programmes such as the yeast metabolic cycle.

Chapter 4.

Disrupting Overlapping Antisense Transcription at the *HMS2* locus.

4. Disrupting Overlapping Antisense Transcription at the *HMS2* locus

4.1. Introduction:

4.1.1. Abolishing Antisense overlap:

The objective of this study was to characterize, in depth, the relationship between the *HMS2* sense/antisense pair in the hope of discerning a general model for antisense-mediated gene regulation. In the previous chapter, the start and end sites, sensitivity to different environmental conditions, stability, and cell-to-cell expression variability were surveyed for the *HMS2* SAP. While these observations were essential to building the proposed state-switching model, further experiments were needed to test the model's robustness. The obvious strategy for determining the importance of antisense transcription in regulating transcription units would be to create either a gain-of-function or loss-of-function scenario; such that antisense transcription is produced or prevented under conditions in which it is normally absent or present, respectively.

While this approach is seemingly straightforward, few studies have been able to perform these gain-of-function/loss-of-function experiments when studying antisense transcription to produce meaningful results in terms of informing a general model. For instance, Houseley et al. (2008) reported a successful loss-of-function experiment, abolishing overlapping antisense transcription from the *GAL10* transcription unit. The *GAL10* antisense transcript, which is initiated from a 3' end internal promoter in the

GAL10 coding region, reads antisense to *GAL10*, through into the *GAL1/GAL10* bidirectional promoter, and extends all the way into the *GAL1* coding region, where it is transcribed in parallel to the sense *GAL1* coding transcript (Figure 28). It is expressed when the *GAL1/GAL10* genes are repressed and is found at reduced levels upon *GAL1*/*GAL10* induction in galactose, indicating a reciprocal relationship between sense and antisense transcription. When point mutations were introduced into putative Reb1p binding sites in the *GAL10* internal promoter, production of the *GAL10*-ncRNA was ostensibly prevented. Using this approach, Houseley and colleagues (2008) found that in the absence of the *GAL10*-ncRNA, the normal repression of the *GAL1/GAL10* cluster in glucose conditions was indeed alleviated, but only when cells were grown in low concentrations of galactose (0.1 g/L) and repressed in low concentrations of glucose (0-0.2 g/L). Additionally, no discernable effects of antisense deletion on *GAL1/GAL10* induction or steady state transcript levels were reported under carbon-rich conditions. This condition-specific sensitivity, similar to that later described by Xu et al. (2011) at the *SUR7* locus, suggests that this antisense transcript might play a role in regulating transcription, but only under conditions in which choosing to either repress or activate gene expression is precarious. In the case of the *GAL10* antisense, it appears as though its production tips the balance in favour of repression. While these findings by Houseley et al. (2008) are appealing and seemingly consistent with the repressive effects of antisense transcripts described in other studies, the interpretation of the results is confounded by the presence of other non-coding RNAs produced in the vicinity. For instance, the report describes a smaller 2.3 kb *GAL10* antisense transcript, also driven by

the *GAL10* 3' internal promoter in parallel to the 4 kb *GAL10*-ncRNA, but does not address how this shorter transcript might also affect the regulation of the *GAL10/GAL1* cluster. Furthermore, unpublished work by Ana Serra Barros provides evidence for many other overlapping non-coding RNAs in the *GAL10/GAL1* cluster, which might influence the locus (Figure 28). Ignoring the effects these additional transcripts might have on the regulation of the gene cluster underestimates the very complexities and nuances in transcriptional regulation the Houseley et al. (2008) study claims to support.

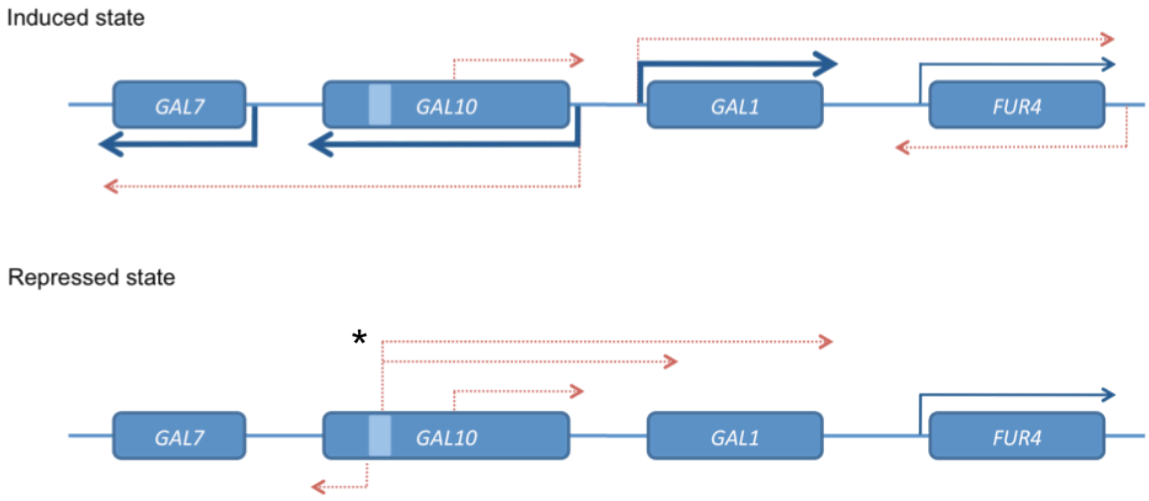


Figure 28. Coding and non-coding transcripts at the *S. cerevisiae GAL* locus.

A schematic of coding (blue solid arrows) and non-coding transcripts (red dotted-line arrows) present at the *S. cerevisiae GAL* locus in inducing (galactose) and repressive (glucose) growth conditions. The *GAL10*-ncRNA described by Houseley et al. (2008) is marked with an asterisk. Transcript map provided by Ana Serra Barros.

An earlier study conducted by Hongay et al. (2006) on the *IME4* (Inducer of Meiosis) locus demonstrated the effects of maintaining antisense transcription in normally repressive conditions. Unsurprisingly, *IME4* sense expression is repressed in haploid cells; however, an overlapping *IME4* antisense transcript is expressed. According to this study, in MAT a/α diploid cells, *IME4* antisense transcription is repressed by the a1/α2

heterodimer, which binds to a consensus a1/α2 binding site at the 3' end of the gene. Under conditions when antisense transcription is repressed, sense *IME4* is transcribed and meiosis is initiated. By creating homozygous mutants of the *IME4* 3' end a1/α2 binding sites, the authors of the study were able to prevent repression of antisense transcription in diploid cells. Consequently, when antisense transcription was permitted in diploid cells, sense transcripts were not detected and induction of meiosis was severely impaired. Conversely, by expressing the a1/α2 heterodimer in haploid cells, Hongay and colleagues showed the consequence of preventing antisense transcription under normally permissible conditions. In this case, the sense *IME4* transcript was detected, providing further evidence of an anti-correlated relationship between sense and antisense transcription. Furthermore, a recent study by van Werven and colleagues (2012) demonstrated that the generation of two non-coding RNAs, the *IME4* antisense transcript and an intergenic transcript, *IRT1* (discussed further in Chapter 5), control meiosis. While the *IME4* study provides the most in-depth dissection of a sense/antisense relationship to date, the scope of its findings in terms of regulation and associated biology are limited to the *IME4* locus, as controlling meiotic entry and antisense regulation via the a1/α2 heterodimer are seemingly unique to the *IME4* antisense. Furthermore, no specific molecular mechanisms for how antisense transcription might repress sense were proposed.

Both the *GAL10* and *IME4* studies relied on manipulating the binding of specific transcriptional regulators, Reb1p and the a1/α2 heterodimer, respectively, in order to

tamper with antisense expression. In the previous chapter, attempts to characterize transcriptional regulators of both *HMS2* sense and antisense expression were unsuccessful. Thus, alternative strategies for abolishing antisense transcription were pursued. Specifically, methods were sought to prevent the overlap of antisense transcription with the sense promoter, where antisense transcription was speculated to exert its function.

4.1.2. Gene disruption:

At the *PHO84* locus, Camblong et al. (2007) inserted a *HIS5* expression cassette to disrupt the sense/antisense overlap, such that the *HIS5* terminator acted bidirectionally to terminate the sense and antisense transcripts. Similar gene disruption methods were used to prevent antisense overlap at *SUR7* (Xu et al., 2011). Both studies demonstrated gene disruption as a useful tool in perturbing native sense/antisense relationships; thus, transcription dynamics at *HMS2* were manipulated by inserting the heterologous *kan*^{r} expression cassette (*kanMX6*), in the presence or absence of an *ADH1* terminator, positioned 5' to the *kanMX6* cassette, into the middle of the *HMS2* coding region using homologous recombination (Longtine et al., 1998).

4.1.3. Aims:

In this chapter, various methods for disrupting antisense overlap at the sense promoter were explored, including gene disruption as in the *PHO84* and *SUR7* studies (Camblong et al., 2007; Xu et al., 2011) and gene substitution. While the experiments pursued were not always successful in preventing antisense overlap at the sense promoter, pairing the *HMS2* promoter and terminator with promoters and terminators from the heterologous

cassettes, generated convenient systems through which the organizational/architectural requirements for high degrees of transcriptional overlap could be scrutinized. Altogether, through Northern blot analysis of these novel transcription units and additional investigation using RNA FISH, the robustness of the state-switching model proposed in Chapter 3 was substantiated.

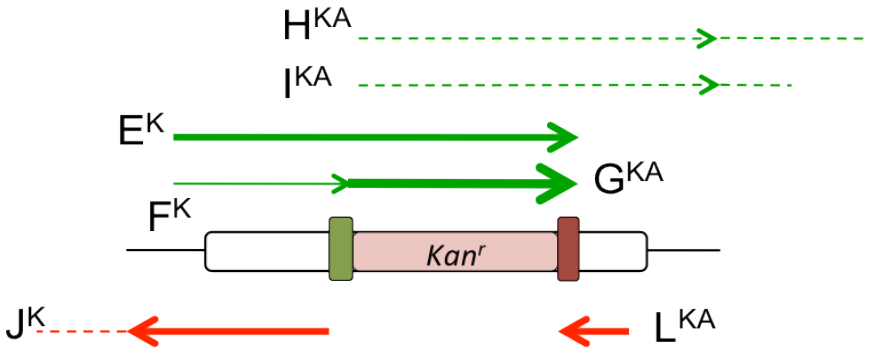
4.2. Results:

In an attempt to gain a better understanding of the *HMS2* SAP, the *HMS2* sense and antisense transcripts were prevented from overlapping by introducing, into the CDS of *HMS2,* a new transcription unit (*kanMX6*) containing a heterologous gene (*kan*^{r}) flanked by a strong promoter (^{P}*TEF*) and terminator (^{T}*TEF*) from *Ashbya gossypii* (Longtine et al 1998). As purported by Camblong et al (2007), this method was successful in disrupting transcription such that the sense and antisense transcripts were terminated by the inserted *HIS5* terminator. In this study of the *HMS2* locus, gene disruption was performed to create strains with (*HMS2.ADH*) and without (*HMS2.kan*) the addition of an upstream *ADH1* terminator (^{T}*ADH1*). Using probes up- and downstream of the inserts, detailed maps (Figure 29A, B) of transcripts of various sizes (Figure 29C) were determined, outlining novel transcription units in both strains. Transcripts detected in only the *HMS2.kan* strain are annotated with a superscript "K", transcripts detected in only the *HMS2.ADH* strain are annotated with a superscript "A", while transcripts found in both strains are annotated with a superscript "KA".

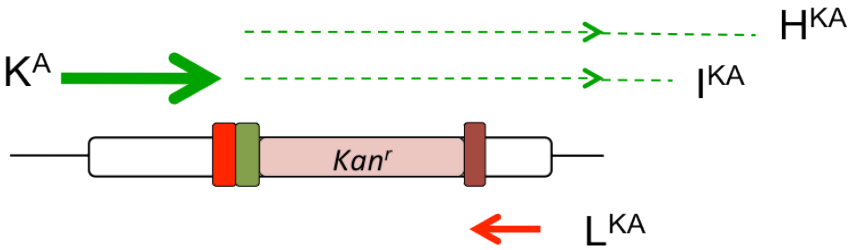
4.2.1. The HMS2.kan and HMS2.ADH gene constructs produce novel but distinct transcription units:

Northern blot analysis of the *HMS2.kan* and *HMS2.ADH* strains using a probe to detect sense transcripts in the *kan*^{R} coding region (K1, Sense) was performed to verify that transcription was occurring across the expression cassettes. As expected, since both strains were able to grow on kanamycin selection media, a strong band consistent in size

A)



B)



C)

Transcript	Estimated Size (nt)
E^K	2300
F^K	1200
G^{KA}	1350
H^{KA}	4450
I^{KA}	3450
J^K	1400
K^A	850
L^{KA}	550
M^{K*}	2800
N^{K*}	3150

Figure 29. Summary of transcripts in the *HMS2.kan* and *HMS2.ADH* disruption strains.

The transcript profiles of **(A)** *HMS2.kan* and **(B)** *HMS2.ADH* derived from Northern blot analysis (see Figure 30 to Figure 32). Transcripts unique to the *HMS2.kan* strain are annotated with a superscript letter "K", while those unique to the *HMS2.ADH* strain are annotated with a superscript letter "A". Transcripts found in both disruption strains are labeled with a superscript "KA". Green boxes= ^{P}*TEF;* Maroon boxes = ^{T}*TEF*; Red boxes =^{T}*ADH1* **(C)** Predicted sizes are displayed in the table. Transcripts marked with an "*" are found in the *HMS2.kan* mutant, 5'^{P}*TEF* Δ-1, see Figure 36 & Figure 37.

(~1350 nt) with a transcript initiating at the *TEF* promoter, reading through the *kan*^{R} CDS, and terminating at the *TEF* terminator, was detected in both *HMS2.kan* and *HMS2.ADH* RNA extracts (Figure 30A, lanes 2 & 5, transcript G^{KA}). Read-through transcripts containing the *kan*^{R} sequences were also detected in both *HMS2.kan* (Figure 30A, lane 2, E^{K}, H^{KA}, I^{KA}) and *HMS2.ADH* strains (Figure 30A, lanes 5, transcript H^{KA}, I^{KA}). Since transcript I^{KA} and H^{KA} were present in both *HMS2.kan* and *HMS2.ADH* extracts, but did not display any apparent changes in size between the two strains, it is likely that these transcripts initiate at the *TEF* promoter and read-through the *TEF* terminator to downstream poly (A) signals. Using the ribosomal RNA markers as an indicator and estimates based on the size of the cassette insertions, the size of transcript I^{KA} (~3450 nt) is consistent with a transcript initiating at the *TEF* promoter and terminating in the vicinity of the *BAT2* poly (A) signal, while transcript H^{KA} is consistent in size (~4450 nt) with a transcript initiating at the *TEF* promoter and terminating in the vicinity of the *YJR149W* intragenic poly (A) signal. On the other hand transcript E^{K} was not detected in the *HMS2.ADH* strain, so was postulated to initiate upstream of the *kanMX6* cassette.

When Northern blots were prepared using a probe to detect *HMS2* sense transcripts upstream of the *kan*^{R} insert (probe H2, sense), very different transcript profiles in the *HMS2.kan* and *HMS2.ADH* strains were apparent. In the *HMS2.kan* strain, an approximately 2300 nt sense transcript consistent with one originating at the *HMS2* promoter (^{p}*HMS2*), reading through the *kan*^{R} CDS, and terminating at the *TEF* terminator was detected (Figure 31, lane 2, transcript E^{K}). The estimated mapping of this transcript

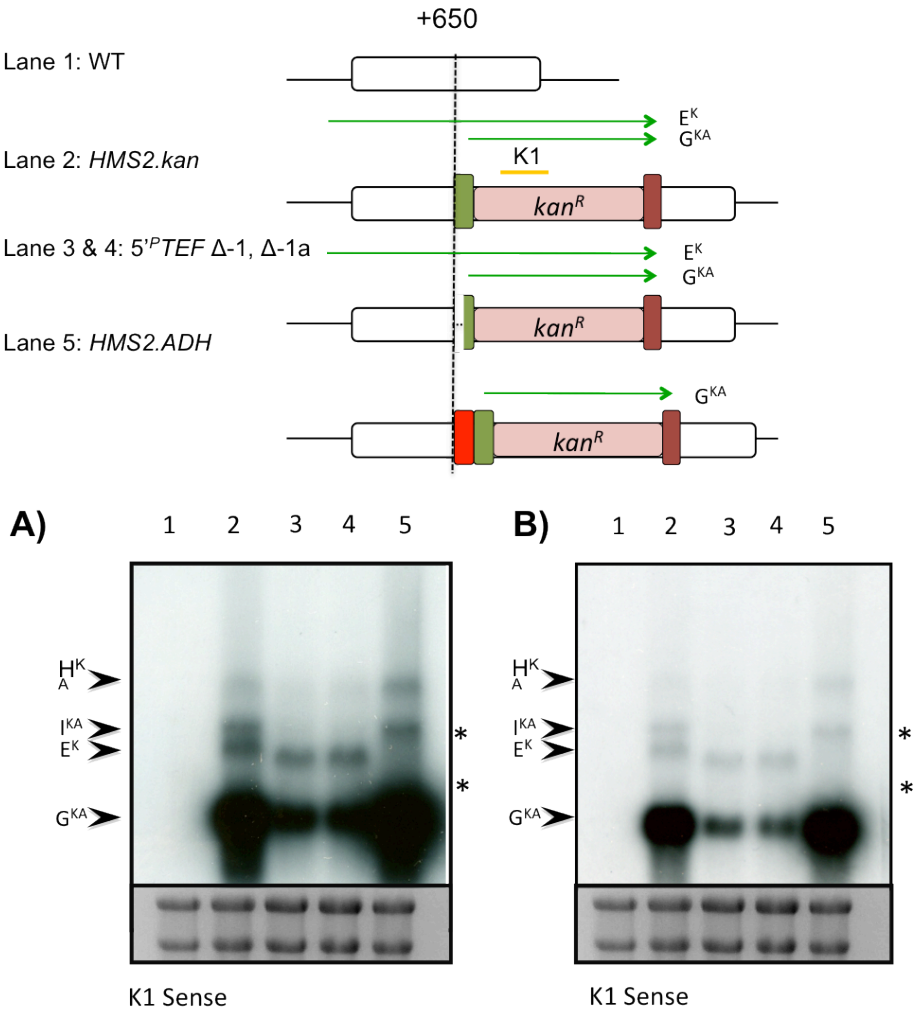


Figure 30. Verifying *kanMX6* expression.

Northern blot analysis of *kanMX6* expression in the disruption strains. Probe location (represented by the yellow bar), strains and major transcripts of interest are depicted in the above schematic; in this depiction, arrow thickness is not indicative of relative steady-state transcript levels. As expected, no kan^R transcripts were detectable in the wild-type strain (lane 1). Expression from *kanMX6* is detectable in the *HMS2.kan* (lane 2) and *HMS2.ADH* (lane 5) strains (transcript G^{KA}). Deletion of the first 217 bp of the *TEF* promoter sequence in the 5'^{P}*TEF* Δ-1 strains (lane 3 & 4) greatly reduces expression from the cassette (transcript G^{KA}). 5'^{P}*TEF* Δ-1 and 5'^{P}*TEF* Δ-1a represent two clones of the 5'^{P}*TEF* Δ-1 deletion strain. Read-through transcripts of varying lengths containing the kan^R sequences are also detectable in all disruption strains (transcripts E^K, I^{KA}, and H^{KA}). RNA was separated for 3 h at 100 V on a 1% agarose/formaldehyde gel. X-ray films were exposed to the radiolabeled membranes for **(A)** 5 days or **(B)** 18h. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNA. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. For full transcript maps please refer to Figure 29.

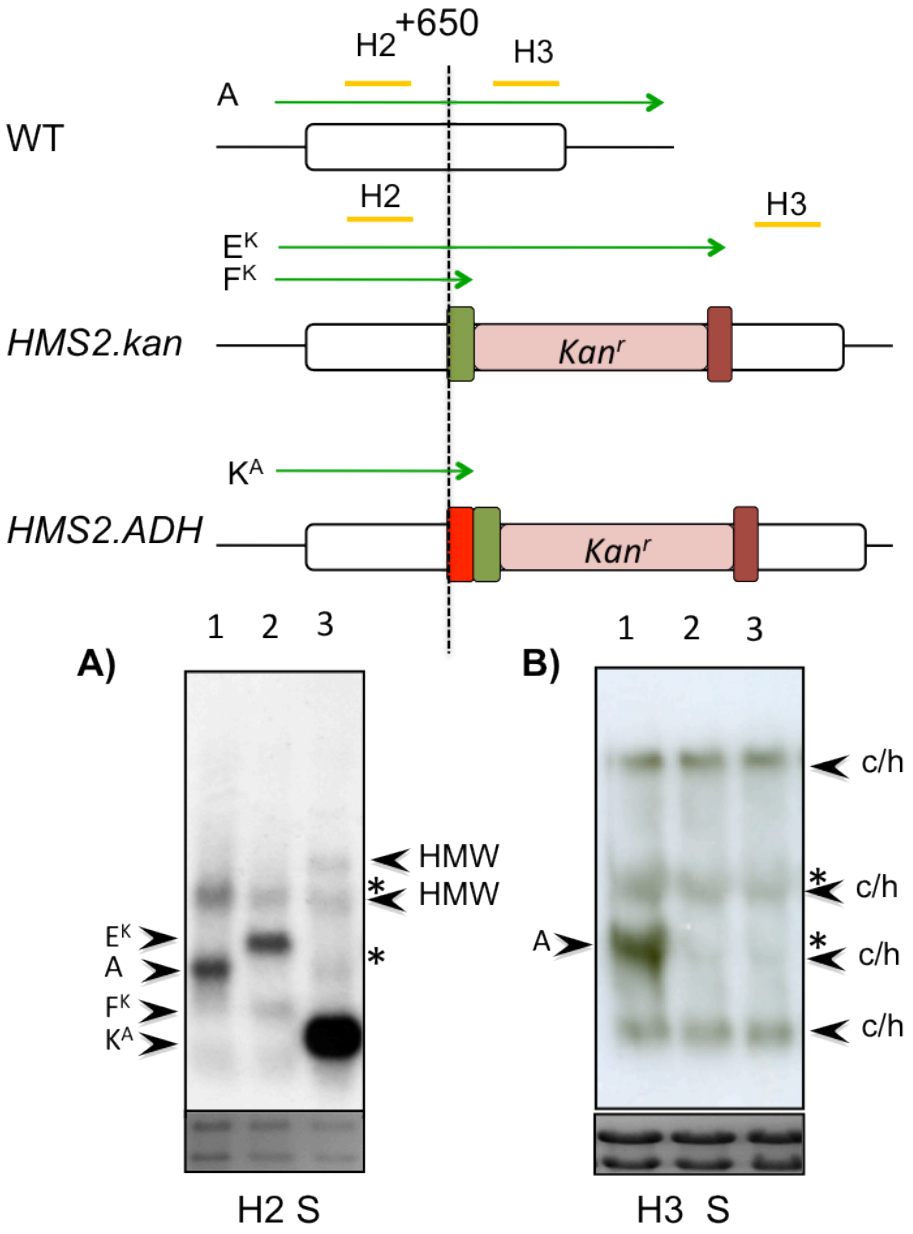


Figure 31. Sense transcription across the *HMS2.kan* and *HMS2.ADH* gene constructs

Northern blot analysis of sense transcription up- (H2) and downstream (H3) of the *kanMX6* cassette in the disruption strains. Probe location (represented by the yellow bar), strains and major transcripts of interest are depicted in the above schematic; in this depiction, arrow thickness is not indicative of relative steady-state transcript levels. **(A)** In the *HMS2.kan* strain (lane 2), the H2 sense probe upstream of the *kanMX6* insert detected a longer sense transcript, ^{P}*HMS2*-^{T}*TEF* (transcript E^{K}), similarly abundant to the endogenous sense transcript (lane 1, transcript A). A shorter sense transcript, ^{P}*HMS2*-^{P}*TEF* (transcript F^{K}), was also detected at lower levels. In *HMS2.ADH* (lane 3), a truncated

sense transcript, ^{P}*HMS2*-^{T}*ADH1* (transcript K^{A}), was detected at increased levels in comparison to wild-type as well as some low-level read-through transcripts (transcript H^{KA} and I^{KA}). **(B)** The H3 sense probe downstream of the insert detected the endogenous antisense transcript (transcript A) in the wild-type strain, but not in the disruption strains. Cross hybridization to ribosomal RNA (c/h) is especially prominent using the H3 sense probe. Using formaldehyde 1% agarose gels, 10 µg of RNA was separated in each lane for 6 h at 100 V (H2 sense) or 3 h at 100 V (H3 sense). The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNA bands. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. For full transcript maps and estimated sizes please refer to Figure 29. Original Northern blots performed by David Brown, repeated by Tania Nguyen.

was corroborated by the read-through transcript previously detected using the K1 sense probe (Figure 30A, lane 2, transcript E^{K}). Transcript E^{K} constituted a novel, dominant transcription unit and signified that the *TEF* terminator was capable of terminating transcription before the endogenous *HMS2* transcription stop site. Additionally, the detection of a fainter band, consistent with a ~1200 nt transcript beginning at the ^{P}*HMS2* and ending at the ^{P}*TEF* constituted a secondary, minor transcription unit (Figure 31A, lane 2, transcript F^{K}). Furthermore, the presence of this minor transcript demonstrated the capacity of the *TEF* promoter to act as a terminator in the "sense" direction.

With the addition of the ^{T}*ADH1*, the H2 sense probe detected a major band, its molecular weight (~850 nt) consistent with a transcript initiating at the *HMS2* promoter and terminating at the *ADH1* terminator (Figure 31A, lane 3, transcript K^{A}). Using this probe, fainter high molecular weight bands were also detected in the *HMS2.ADH* strain. These transcripts differed in size from the *kan*^{R}-containing transcripts I^{KA} and H^{KA} and

could represent transcripts initiating far upstream of *HMS2* and terminating at the *ADH1* terminator; however exact mapping remains unclear (Figure 31A, lane 3, HMW).

A sense probe for the 3' end of the *HMS2* coding region (probe H3, sense) was used to check for transcription occurring downstream of the expression cassettes. As predicted, the 3' *HMS2* sense probe detected the wild-type *HMS2* sense transcript (Figure 31B, transcript A), but did not detect any major sense transcription units in *HMS2.kan* or *HMS2.ADH*, signifying that most transcription was terminated upstream of the *TEF* terminator. Detection of any minor read-through transcription units downstream of the ^{T}*TEF* was confounded by the presence of high levels of non-specific hybridization associated with this probe.

The ability of the *TEF* terminator to terminate bidirectionally and truncate the endogenous antisense transcript arising from the *HMS2* terminator was demonstrated using an antisense probe for the 3' end of *HMS2* (probe H3, AS). Using this 3' antisense probe, the wild-type antisense transcript (Figure 32, transcript B) could no longer be detected in the *HMS2.kan* and *HMS2.ADH* extracts; however, a ~550 nt truncated transcript, consistent with one beginning at the *HMS2* terminator (^{T}*HMS2*) and ending at the ^{T}*TEF* was detected (Figure 32, transcript L^{KA}) in both strains, confirming the ability of the *TEF* terminator to terminate transcripts bidirectionally.

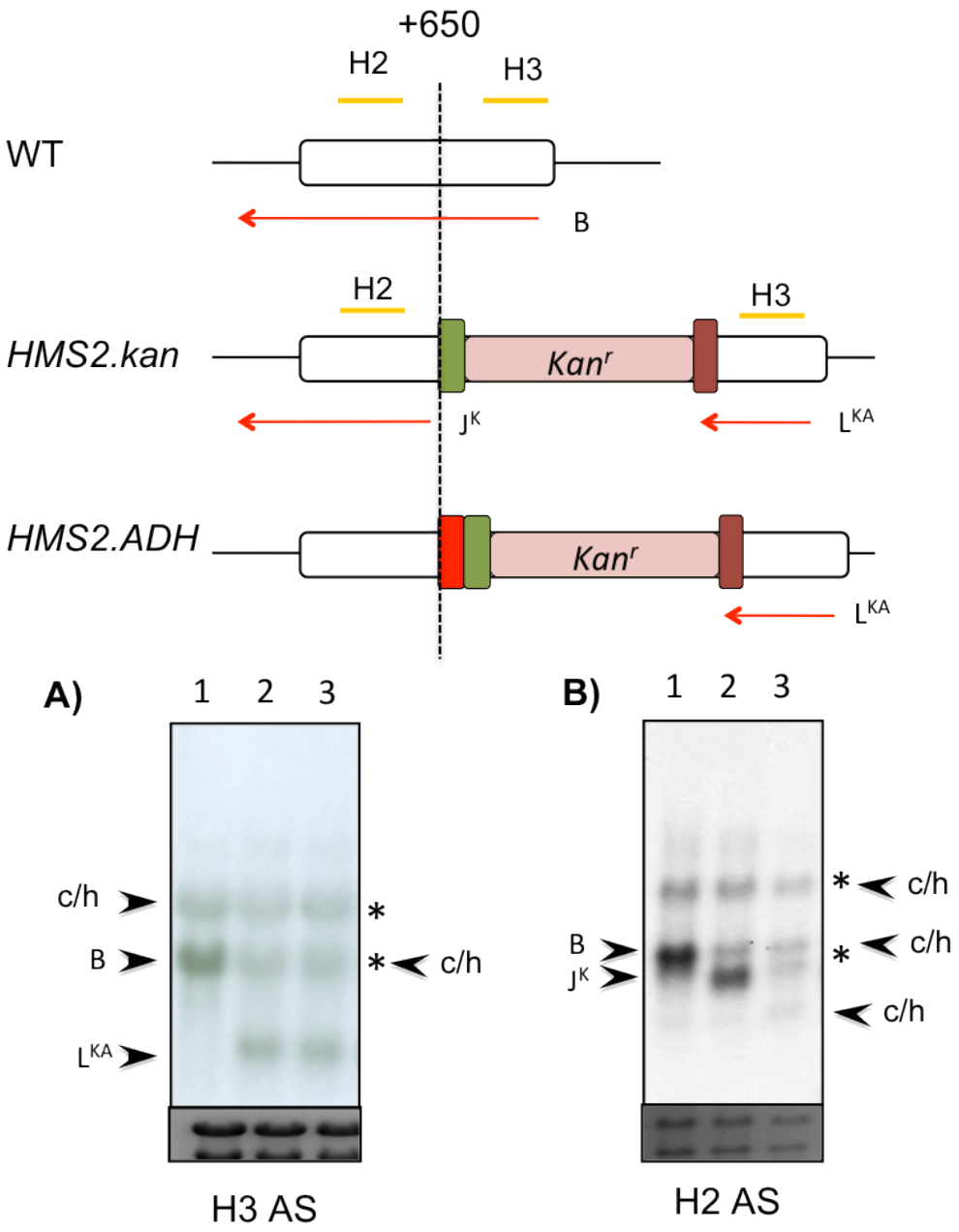


Figure 32. Antisense transcription across the *HMS2.kan* and *HMS2.ADH* gene constructs

Northern blot analysis of antisense transcription down- (H3) and upstream (H2) of the *kanMX6* cassette in the disruption strains. Probe location (represented by the yellow bar), strains and major transcripts of interest are depicted in the above schematic; in this depiction, arrow thickness is not indicative of relative steady-state transcript levels. **(A)** The H3 antisense probe downstream of the insert detected the wild-type (lane 1) antisense transcript (transcript B), which was truncated at the *TEF* terminator in the *HMS2.kan* (lane 2) and *HMS2.ADH* (lane 3) strains resulting in transcript ^{T}*HMS2*-^{T}*TEF*

(transcript L^{KA}). **(B)** The H2 antisense probe upstream of the insert detected the wild-type antisense transcript (transcript B) as well as an antisense transcript in *HMS2.kan* (lane 2), consistent with a transcript initiating near the *TEF* promoter and terminating at the *HMS2* promoter, ^{P}*TEF*-^{P}*HMS2* (transcript J^{K}). No antisense was detected upstream of the *ADH1* terminator in *HMS2.ADH*. Probe cross hybridization is marked by by "c/h". Using formaldehyde 1% agarose gels, 10 µg of RNA was separated in each lane for 6 h at 100 V (H2 sense) or 3 h at 100 V (H3 sense). The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNA bands. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. For full transcript maps and estimated sizes please refer to Figure 29. Original Northern blots performed by David Brown, repeated by Tania Nguyen.

Apart from some lower-level read-through transcription occurring in the sense direction, thus far, the results described were in agreement with the gene disruption described by Camblong et al (2007). However, to ensure that the primary goal of preventing the *HMS2* antisense transcript from overlapping the sense promoter was achieved, an antisense probe upstream of the *kan*^{R} insert was also used (probe H2, AS). Surprisingly in *HMS2.kan*, a prominent band, consistent with a ~1400 nt transcript beginning at the *TEF* promoter and terminating at the *HMS2* promoter was detected (Figure 32B, transcript J^{K}). This result indicated that the *TEF* promoter is able to initiate transcription bi-directionally (transcripts J^{K} and G^{KA}). This bidirectional initiation from the ^{P}*TEF* effectively reconstituted an overlapping SAP at *HMS2,* counteracting our original objectives. While detection of this novel antisense transcript in *HMS2.kan* hampered efforts to describe the effects of abolishing antisense overlap of the sense promoter, similar overlapping transcripts upstream of the *kanMX6* cassette were not detected in the *HMS2.ADH* strain, indicating that a cassette flanked by two strong terminators, the ^{T}*ADH1* and the ^{T}*TEF*, was required to prevent overlapping sense/antisense transcription at the *HMS2* locus (Figure 32B).

Therefore, in *HMS2.ADH* cells, the absence of antisense transcription overlapping the sense promoter is correlated with increased levels of the *HMS2* sense transcript, a result consistent with an anti-correlated relationship between sense and antisense transcription. A dilution series was prepared on Northern blots using denatured dsDNA complementary to the *HMS2* 5' UTR probes (probe H1) in order to estimate the magnitude of this increase in *HMS2* sense transcript levels between wild-type and *HMS2.ADH* cells. Transcript levels were approximated by comparing the band intensities of the sense transcripts to the intensity of the dilution series hybridized with the H1 sense probe. According to this analysis, levels of sense transcript increased 10-fold in the *HMS2.ADH* strain with respect to wild-type levels (Figure 33, transcript A & K^A).

While promising, the absence of antisense detection in *HMS2.ADH* might have reflected a highly unstable transcript, such as the *PHO84* antisense, and not a loss of antisense transcription *per se*. Additionally, alternative 3' end processing by virtue of replacing the *HMS2* terminator with that of *ADH1* could have contributed to increased levels of the sense transcript if the resultant transcript was more stable. Consequently, to rule out stability as the main cause for the increases in sense transcript levels and to ensure that antisense transcription was indeed abolished and not just highly labile, sense and antisense transcription upstream of the *ADH1* terminator was interrogated when *RRP6* was deleted.

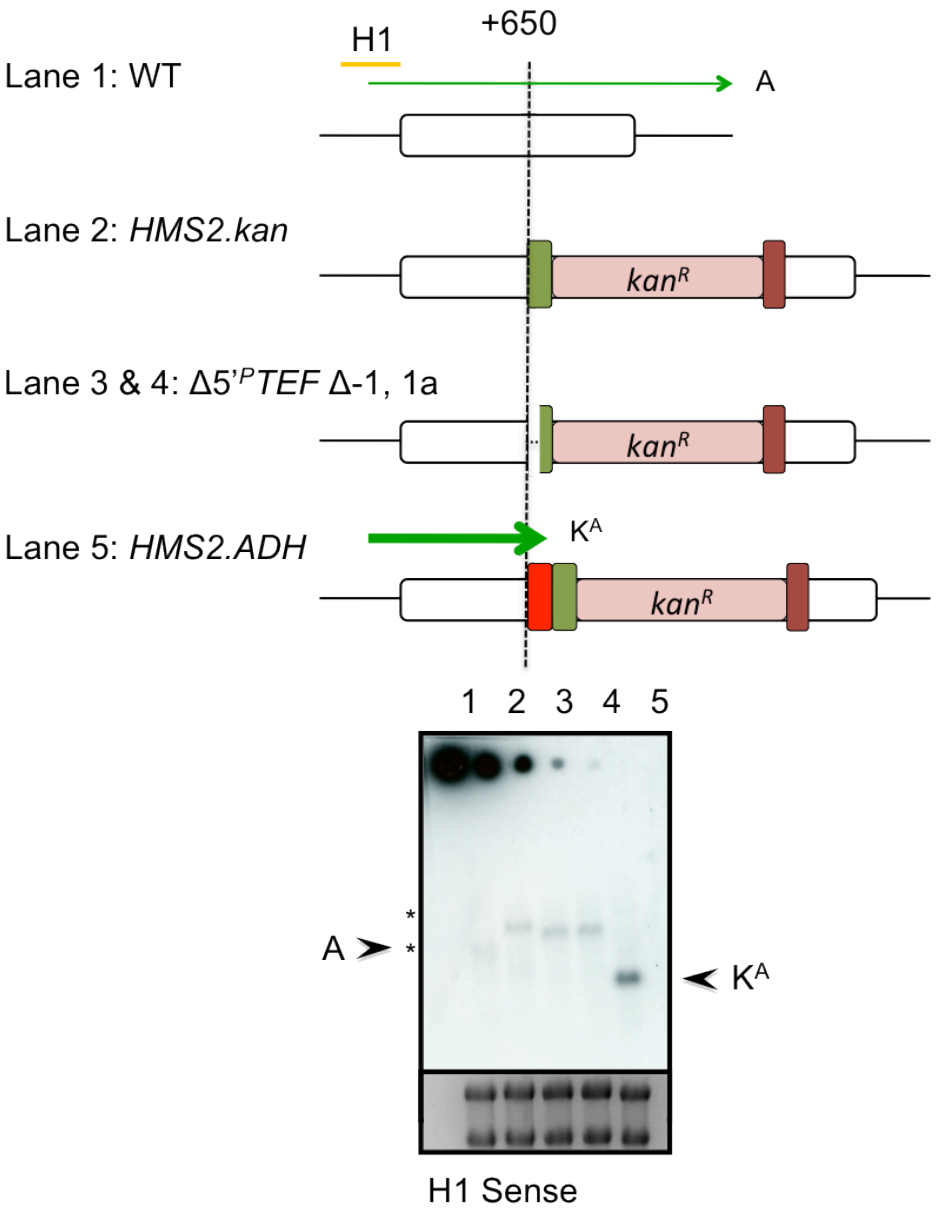


Figure 33. Quantitation of sense transcripts in wild-type and disrupted *HMS2*.

Northern blot analysis of sense transcription upstream (H1 sense) of the *kanMX6* cassette. Probe location (represented by the yellow bar), strains and major transcripts of interest are depicted in the above schematic; in this depiction, arrow thickness is indicative of relative steady-state transcript levels. A dilution series (50 ng/µl, 5 ng/µl, 0.5 ng/µl, 0.05 ng/µl, 0.005 ng/µl) comprised of denatured dsDNA corresponding to the probed region is located at the top of the blot. In WT cells (lane 1) sense transcript levels are <0.005 ng/µl (transcript A). In *HMS2.kan* and Δ5'^{P}*TEF 1, 2* the dominant sense unit is found at comparable levels (lane 2, 3, 4). In *HMS2.ADH* (lane 5) sense levels are found at ~0.05 ng/µl (transcript K^{A}). The blot represents X-ray film exposed for 5 days. Using formaldehyde 1% agarose gels, 10 µg of RNA was separated in each lane for 2.5 h at 100

V (H1 sense). The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNA bands. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. For full transcript maps and estimated sizes please refer to Figure 29.

In the absence of Rrp6p, both the full-length and ^{T}*ADH1*-truncated transcripts displayed equal degrees of stabilization; thus, ruling out decreased Rrp6p sensitivity as the sole reason for the observed increase in sense levels in *HMS2.ADH* (Figure 34A, transcript K^{A}). Most importantly, in the antisense direction, no new transcripts were detected when *RRP6* was deleted, providing further evidence that antisense transcription was, in fact, prevented from traversing the *HMS2* sense promoter in the *HMS2.ADH* strain (Figure 34B).

4.2.2. Perturbation of the TEF promoter results in a more complex transcriptional profile:

While antisense overlap of the sense promoter was indeed alleviated in the *HMS2.ADH* strain, it was still difficult to distinguish to what degree *HMS2* sense transcript levels increased because of antisense ablation, and what was attributable to other transcriptional processes influenced by the insertion of a strong terminator. For instance, it has been documented that promoters and terminators can be brought together in gene loops, consequently affecting transcriptional activation (O'Sullivan et al., 2004). If the addition of the *ADH1* terminator positively affected loop formation, then levels of transcription could be enhanced, regardless of the presence/absence of

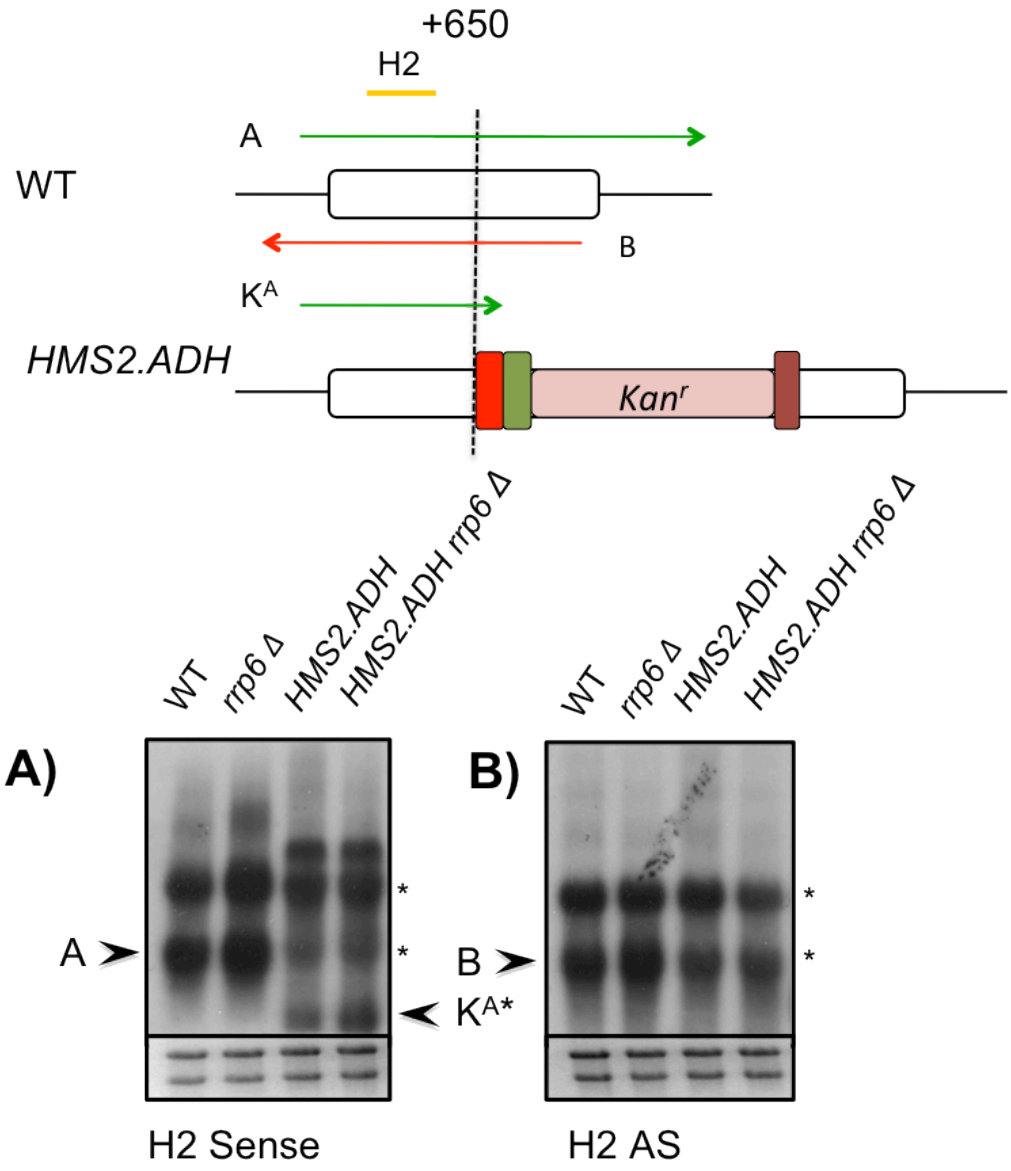


Figure 34. Rrp6p sensitivity of wild-type and ^{T}*ADH1*-truncated *HMS2* sense and antisense transcripts.

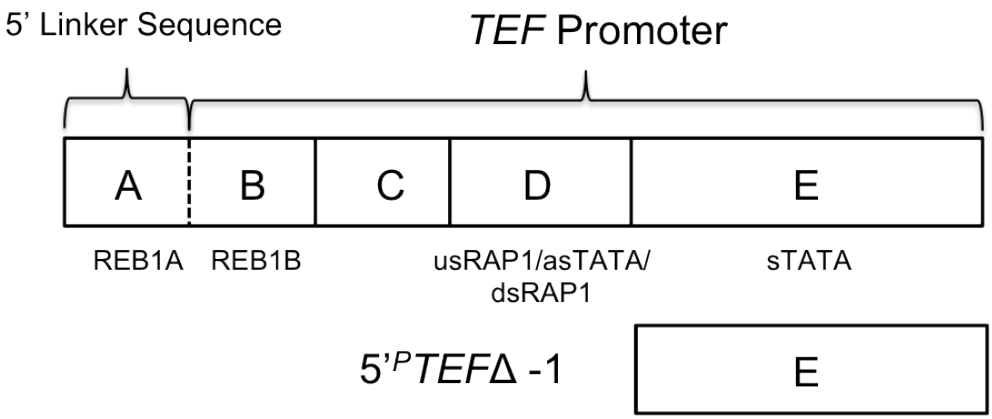
Northern blot analysis depicting levels of **(A)** sense and **(B)** antisense transcripts for the wild-type and *HMS2.ADH* strains in the presence and absence of Rrp6p. Strains and major transcripts of interest are depicted in the above schematic; in this depiction, arrow thickness is not indicative of relative steady-state transcript levels. When the nuclear exosome is impaired in the *rrp6Δ* strain, both the wild-type (transcript A) and *HMS2.ADH* sense transcripts (transcript K^A), and wild-type *HMS2* antisense transcript (transcript B) experienced increases in levels. In *HMS2.ADH* strain, no novel antisense transcripts were detected when the nuclear exosome was impaired. Blots were

hybridized with probe H2, indicated by the yellow bar in the above schematic. Probe cross hybridization is marked by "c/h". Using formaldehyde 1% agarose gels, 10 µg of Total RNA was separated in each lane for 2.5 h at 100 V. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNA. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. Major transcripts are marked by black arrows. For full transcript maps and estimated sizes please refer to Figure 29. *In this representation, the bands corresponding to the K^A transcripts were partially run off the gel; however, changes in levels are still apparent.

antisense transcription. Additionally, 3' end formation, as directed by polyadenylation signals in the terminator, is inter-related with other aspects of the transcription cycle (N. Proudfoot, 2004). Switching terminator elements could also affect transcript levels post-transcriptionally; for example, alterations to 3' UTR length have been shown to influence sensitivity to the Nonsense Mediated Decay (NMD) RNA surveillance pathway, with shorter 3' UTRs displaying immunity and longer ones displaying sensitivity to degradation (Kebaara & Atkin, 2009). *HMS2* has a markedly long 3'UTR of ~427 nt, well above the 144 nt average length of 3' UTRs in *S. cerevisiae*, indicating that it is likely a substrate for NMD (Graber, Cantor, Mohr, & Smith, 1999). Shortening the *HMS2* 3' UTR by utilizing the poly (A) signal of the *ADH1* terminator ($Py(A)_n$: AAAA), which lies approximately 90 nt into the terminator sequences, might result in enhanced post-transcriptional stabilization (Heidmann, Obermaier, Vogel, & Domdey, 1992). Thus, differences at any of these steps described above, provoked by the change in terminator element, could produce an increase in sense transcript levels via Northern blot analysis. These uncertainties prompted the search for another method to disrupt antisense transcription over the *HMS2* promoter in order to corroborate our loss-of-function results obtained in the *HMS2.ADH* strain.

In an attempt to specifically target antisense production from the *TEF* promoter in the *HMS2.kan* strain, a 217 bp sequence at the 5' region of the *TEF* promoter was deleted, creating the *5'^{P}TEF* Δ-1 mutant (later referred to as Δ-1). For the purposes of this mutational analysis, the *TEF* promoter sequences were classed into regions A-E (Figure 35). Section A, which consisted of plasmid linker sequences, contained a putative Reb1p binding site (REB1A). Section B, which marked the beginning of the *TEF* promoter sequence, contained two overlapping putative Reb1p binding sites, one in the sense and another in the antisense orientation (REB1B). Section C did not contain any apparent sequences of interest, while section D contained a putative antisense 7/8 TATA box (asTATA), flanked by an upstream and downstream consensus binding motif for Rap1p (usRAP1 & dsRAP1), both of which are known to be required for high-level expression from the ^{P}*TEF* (Steiner, Philippsen, Gieben, & Str, 1994). Finally, section E contained a consensus TATA box feature in the sense orientation (sTATA). The 217 bp deletion region, which was comprised of sections A-D and thus most of the regulatory motifs, was thought to house the "antisense promoter". Northern blot analysis of the *5'^{P}TEF Δ-1* mutant was performed to monitor any changes in transcription and new transcript maps were defined (Figure 36).

By deleting regions A-D in the *TEF* promoter, *kan^{R}* expression was greatly reduced (Figure 30B, lanes 3 & 4, transcript G^{KA}). The longer hybrid transcript, transcript E^{K}, initiating from the *HMS2* promoter and ending at the *TEF* terminator, was also detected



Region	Size (bp)	Motifs	Sequence
A (linker DNA)	32	REB1A: Putative Reb1 binding site	REB1A: cgggtta
B	41	REB1B: 2 overlapping putative Reb1 binding sites (sense and antisense)	REB1B: cgggcctgcac
C	47	Nothing of interest	n/a
D	97	usRAP1: upstream Rap1 binding site asTATA: 7/8 AS TATA box dsRAP1: downstream Rap1 binding site	usRAP1:gcccatacat asTATA: catgtata dsRAP1: atccatacat
E	200	sTATA: 8/8 sense TATA box	sTATA: tataaaag

Figure 35. The 5' ^{P}*TEF* Δ-1 strain.

A 217 bp deletion was introduced at the 5' region of the *TEF* promoter to create the 5' ^{P}*TEF Δ-1*. The *TEF* promoter can be divided into 5 arbitrary regions based on the presence of putative transcription factor binding sites. Region A consists of linker DNA and contains a Reb1-like binding site (REB1A). Section B comprises the beginning of the *TEF* promoter sequence and contains two overlapping sense and antisense oriented putative Reb1 binding sites (antisense sequence is underlined) (REB1B). Region C does not contain any sequences of interest, while section D contains a putative 7/8 antisense TATA box (asTATA) flanked by two Rap1 binding sites described by Steiner and Philippsen (1994). Remaining after the deletion is section E, which contains a putative consensus TATA box (sTATA). It was thought that by removing the asTATA box, antisense transcription from ^{P}*TEF* could be abolished. The correct deletion was confirmed by DNA sequencing.

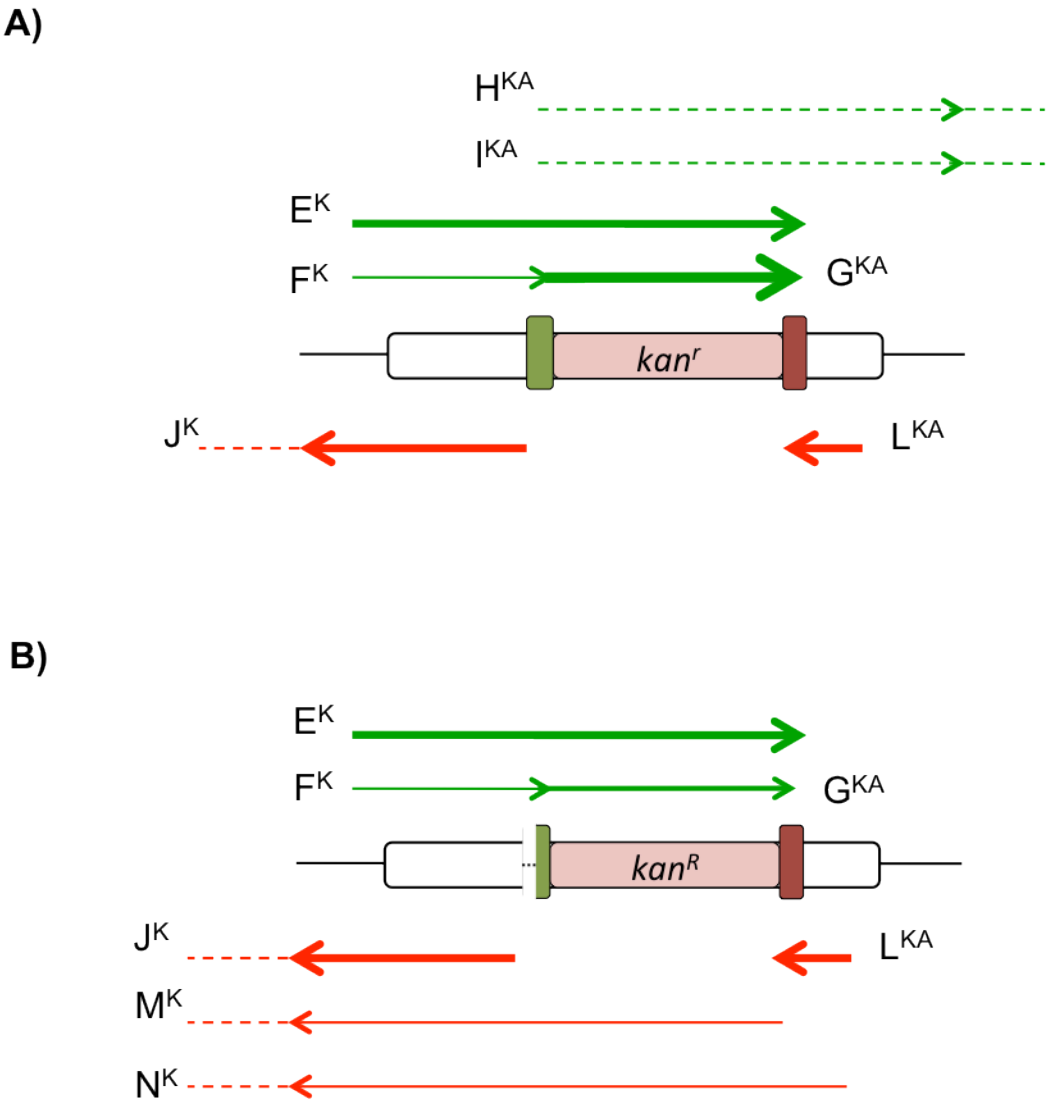


Figure 36. Summary map of the *HMS2.kan* and 5' ^{P}*TEF* Δ-1 transcript profiles

Schematics summarizing the detected transcripts in the *HMS2.kan* (A) and 5' ^{P}*TEF* Δ-1 (B) strains using Northern blot analysis (See Figure 30 and Figure 37). Transcripts detected in both the *HMS2.kan* and 5' ^{P}*TEF* Δ-1 mutants are annotated with a superscript letter "K". Two new longer antisense transcripts (M^K and N^K) are detected in the 5' ^{P}*TEF* Δ-1 mutant (for estimated sizes refer to Figure 29). Additionally, levels of the *kan*^{R} transcript (G^{KA}) are reduced in the 5'^{P}*TEF* Δ-1 mutant in comparison to the *HMS2.kan* strain.

in the *5'^{P}TEF Δ-1* mutant, but shortened in accordance with the size of the deletion (Figure 30A, lanes 3 & 4, transcript E^{K}). The longer read-through transcripts detected in the *HMS2.kan* and *HMS2.ADH* cells (Figure 30A, lane 2 & 5 H^{KA} & I^{KA}) were not found at detectable levels in the ^{P}*TEF* mutants, further suggesting that they might arise from the *TEF* promoter, read through the cassette, and extend to the native *BAT2* or intragenic *YJR149W* poly (A) signals.

When sense transcription upstream of the *kanMX6* cassette was observed in the ^{P}*TEF* mutant, transcript E^{K} was once again detected with its levels unaffected in comparison to the original *HMS2.kan* construct (Figure 37A, lane 3, transcript E^{K}). Though changes to the band corresponding to the shorter sense transcript (transcript F^{K}) do not manifest clearly in the blot hybridized with the H1 probe, in other blots using probe H2 it was shown that transcript F^{K} is found at reduced levels and truncated in length, demonstrating that it terminates downstream of the ^{p}*TEF* deletion (Figure 37B, transcript F^{K}). In the antisense direction, levels of antisense were reduced but surprisingly not abolished, indicating that the remaining sequence of the *TEF* promoter was sufficient to initiate transcription bidirectionally (Figure 37C, lane 3, transcript J^{K}).

Additionally, the ^{p}*TEF* deletion strain provided evidence for the *HMS2* antisense transcript initiating within the *HMS2* coding sequence as no shift in size was observed. Interestingly potentially two other, longer antisense transcripts were detected in the

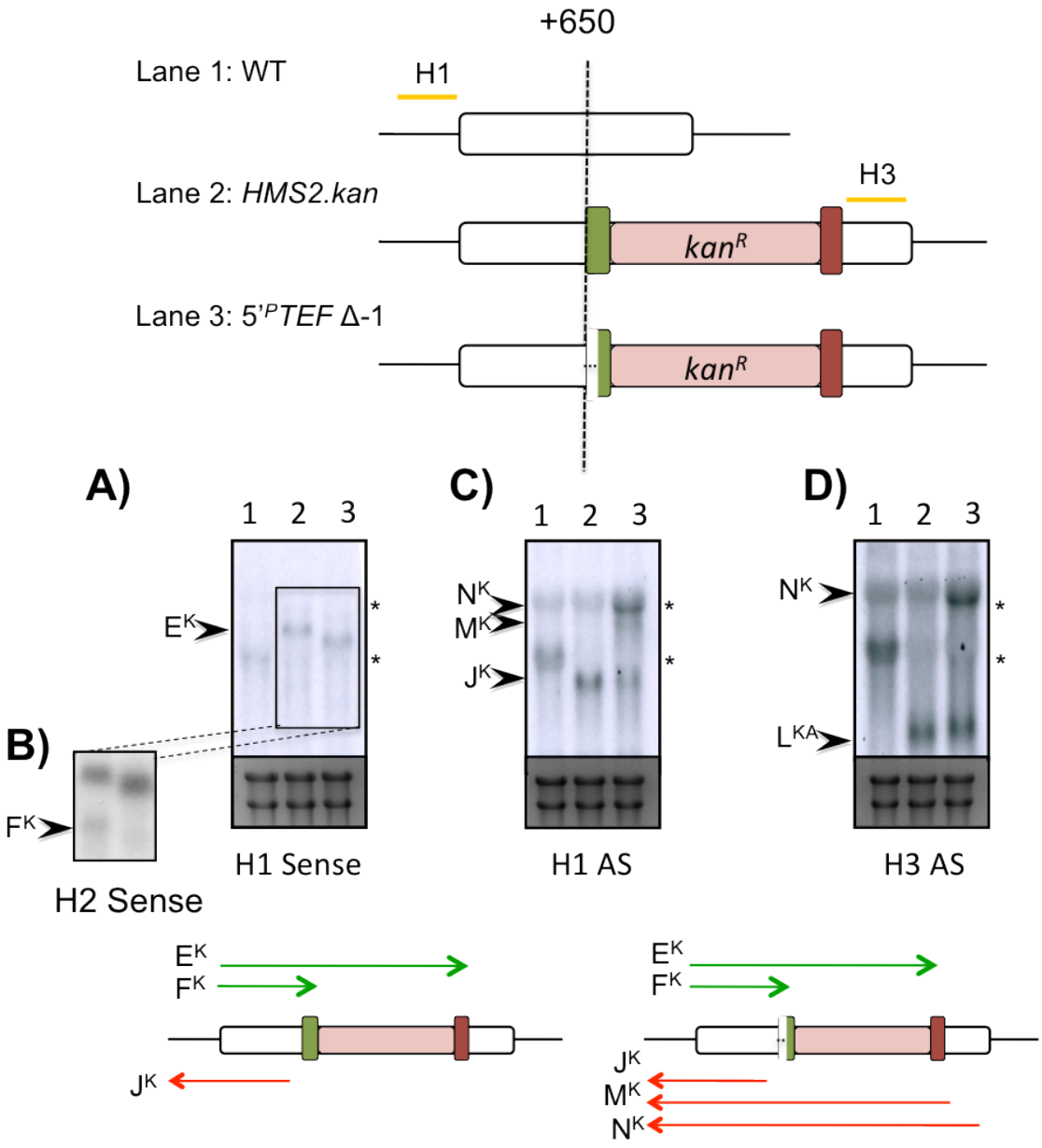


Figure 37. Disrupting the *TEF* promoter does not abolish antisense transcription.

Northern blot analysis of disruption strains upstream (H1 sense/antisense) and downstream (H3 antisense) of the *kanMX6* cassette when the *TEF* promoter was intact (lane 2) or disrupted in the 5' ^{P}*TEF* Δ-1 strain (lane 3) (schematics above). Probe locations are indicated by yellow bars. **(A)** The deletion in the 5' *TEF* strain is reflected by the shift in size of the ^{P}*HMS2-kan-*^{T}*TEF* sense transcript (transcript E^{K}). **(B)** Using the H2 sense probe, short sense transcript F^{K} is detected, shortened in accordance with the size of the deletion. **(C)** In the antisense direction, the deletion results in the reduction of the ^{P}*TEF-*^{P}*HMS2* AS transcript (transcript J^{K}) and the generation of two new transcripts consistent with transcription occurring between ^{T}*TEF-kan-*^{P}*HMS2* (M^{K}) and ^{T}*HMS2-kan-*

^{P}*HMS2* (N^{K}). **(D)** An antisense probe for the 3' end of *HMS2* (H3) detected the truncated antisense transcript initiating from the *HMS2* terminator (transcript J^{KA}) and corroborated the estimated mapping of the ^{T}*HMS2-kan-*^{P}*HMS2* transcript (transcript N^{K}). Using formaldehyde 1% agarose gels, 10 µg of Total RNA was separated in each lane for 2.5 h at 100 V. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNA. Transcripts of interest are marked by black arrows and are also depicted in the miniature transcript maps below the blots. For full transcript maps and estimated sizes please refer to Figure 29 and Figure 36.

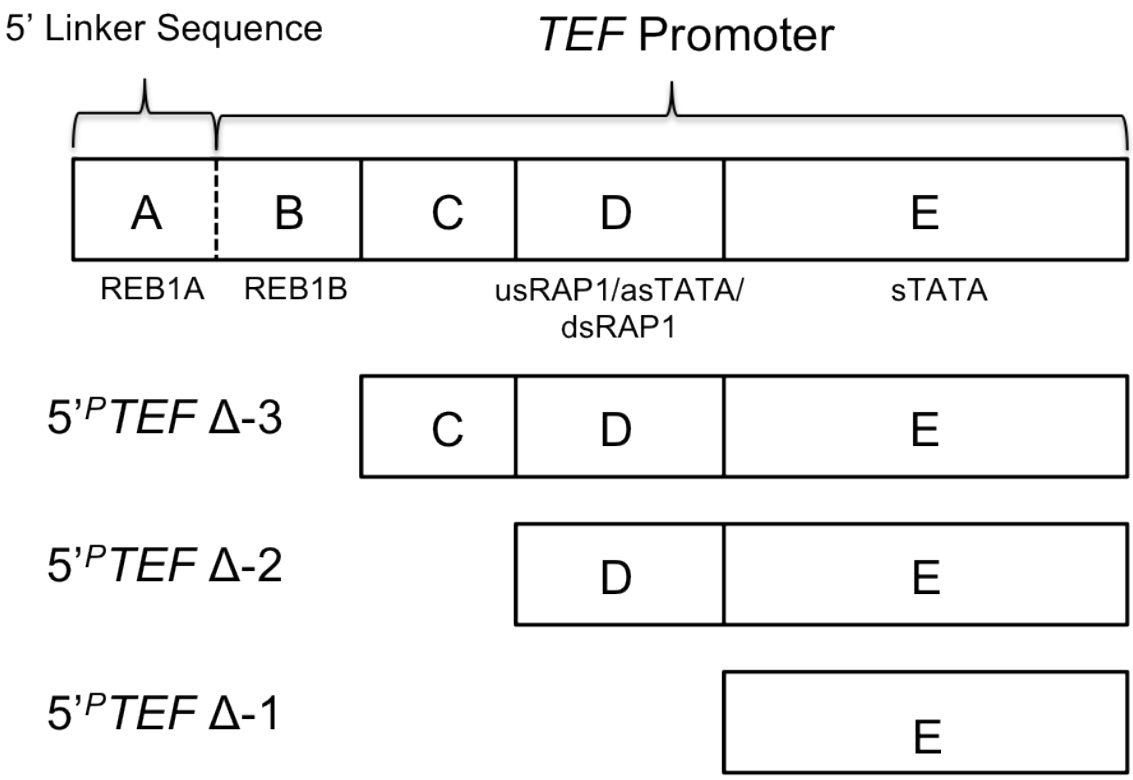
deletion strain (Figure 37C, lane 3, transcripts M^{K} & N^{K}). When an antisense probe for the 3' end of *HMS2* was used (probe H3), we found that one of these transcripts originated from the 3' end of *HMS2* (presumably the native antisense promoter) and given its size, likely reads through to the endogenous *HMS2* promoter (Figure 37D, lane 3, transcript N^{K}). The second, shorter antisense transcript was suspected to initiate from the *TEF* terminator sequence, as this second band was not detected with the 3' *HMS2* antisense probe. Therefore, while physically unaltered, the ability of the *TEF* terminator to terminate bidirectionally was compromised when its corresponding promoter was weakened, but its ability to initiate transcription was possibly enhanced. The behaviour of the *TEF* terminator in the context of the 5'^{P}*TEF* Δ-1 transcriptional landscape underscores the inherent duality of promoters and terminators and implies that the promoter and terminator functions of a single regulatory element exist in a delicate balance contingent on the surrounding transcriptional context and not solely on sequence composition.

4.2.3. Further dissection of the TEF promoter:

In light of the unexpected changes in sense/antisense dynamics observed in the 5' ^{P}*TEF* Δ-1 strain, a more in-depth mutation series was performed on the *TEF* promoter

sequences in order to elucidate how each of its components influenced the transcriptional landscape. Mutational analysis mainly focused on distinguishing the contributions of the REB1A, REB1B, asTATA, us- and dsRAP1 motifs. For instance, additional deletion strains *5′^{P}TEF* Δ-2 (referred to as Δ-2), *5′^{P}TEF* Δ-3 (referred to as Δ-3), were engineered in order to separate the effects of deleting the asTATA box from the effects of deleting the putative Reb1p binding sites, REB1A and REB1B (Figure 38). The *5′^{P}TEF* Δ-2 strain incorporated the asTATA box sequence and Rap1p features in region D, but excluded the Reb1p motifs in sections A and B, as well as the "inert" sequences in section C. The *5′^{P}TEF* Δ-3 strain excluded the REB1A and REB1B features in sections A and B, respectively, but maintained the intervening sequence found in section C in addition to the regulatory motifs in region D. Interestingly, both deletion strains 2 and 3 displayed similar expression profiles, indicating that, as predicted by its sequence composition, section C was unlikely to make major contributions to the transcriptional landscape (Figure 39).

More specifically, in the sense direction, little change was observed in both strains, with levels of the major sense transcription unit, transcript E^{K}, remaining relatively constant, but shortened in size in accordance with the size of the deletions (Figure 39A, transcript E^{K}). Meanwhile, rather than a stark reduction in antisense levels as in the 5′^{P}*TEF* Δ-1 strain, an increase in antisense levels was observed in the Δ-2 and Δ-3 strains (Figure 39B, transcript J^{K}). Furthermore, a sense probe for the *kan^{r}* sequences (probe K1)



Deletion	Size	Excluded Motifs
5'^{P}*TEF*Δ-3	73	REB1A, REB1B
5'^{P}*TEF*Δ-2	120	REB1A, REB1B, Region C
5'^{P}*TEF*Δ-1	217	REB1A, REB1B, Region C, usRAP1, asTATA, dsREB1

Figure 38. Dissection of the 5' region of the *TEF* promoter.

Schematic of *TEF* promoter dissection depicting the remaining promoter sequence after deletions were introduced. 5'^{P}TEFΔ-2 excluded regions C, which contained no sequence of particular interest, and regions A and B, which contained the putative Reb1 binding sites, REB1A and REB1B. In 5'^{P}TEFΔ-3, regions A and B were excluded, effectively deleting all of the putative Reb1 binding sites, but leaving most of the *TEF* promoter intact. The correct deletions were confirmed by DNA sequencing.

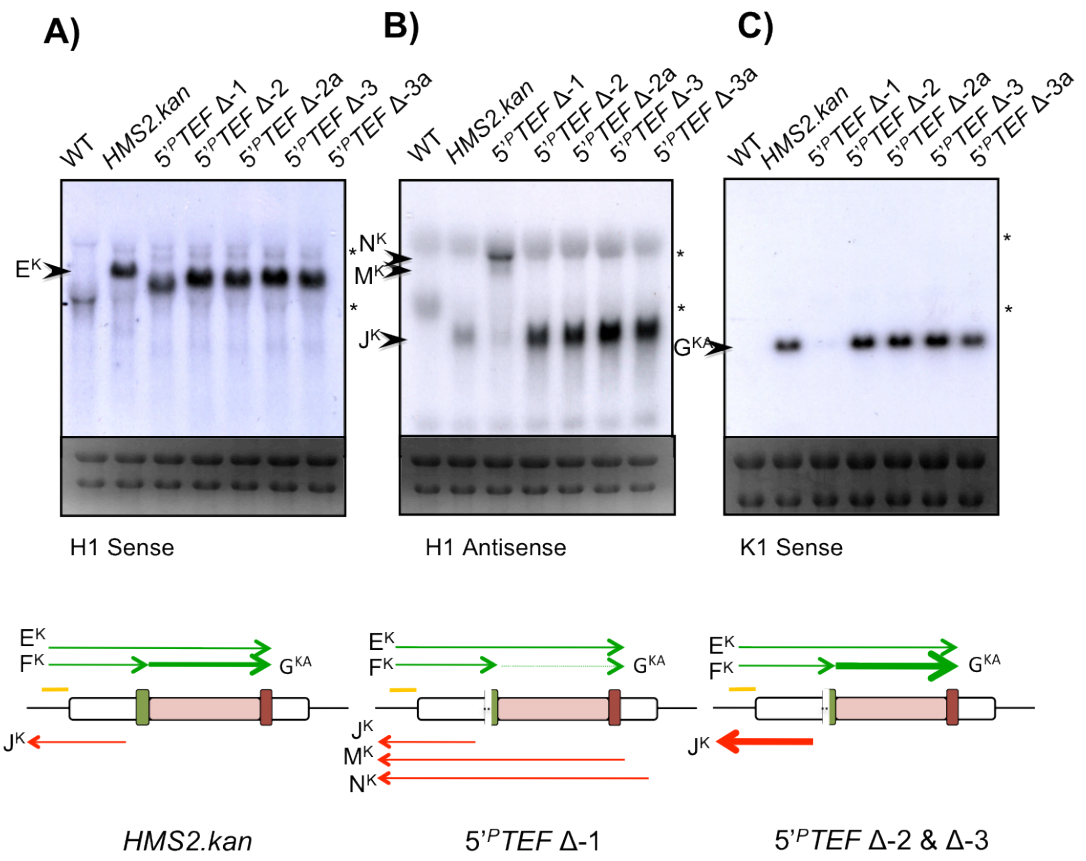
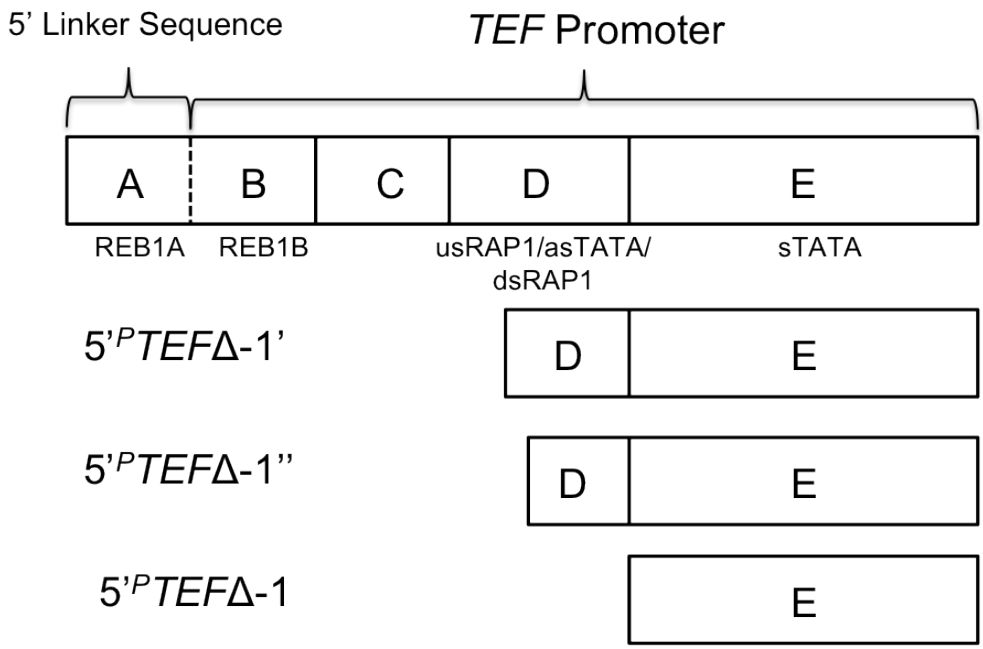


Figure 39. Effects of dissecting the 5' region of the *TEF* promoter

Northern blot analysis of 5'^{P}*TEF* mutant strains (for detailed schematics see (Figure 38) using the H1 sense and antisense. **(A)** The deletions in the 5'*TEF* are reflected by the shift in size of the ^{P}*HMS2-kan-*^{T}*TEF* sense transcript (transcript E^{K}). **(B)** In the antisense direction, only the 5'^{P}*TEF* Δ-1 deletion results in the reduction of the ^{P}*TEF-*^{P}*HMS2* AS transcript (transcript J^{K}) and the generation long antisense transcripts ^{T}*TEF-kan-*^{P}*HMS2* (M^{K}) and ^{T}*HMS2-kan-*^{P}*HMS2* (N^{K}). In contrast, the 5'^{P}*TEF* Δ-2 and 5'^{P}*TEF* Δ-3 mutations resulted in an increase in transcript J^{K} levels. **(C)** Using the K1 sense probe, the increase in transcript J^{K} was reflected by an increase in detectable levels of *kan*^{r} transcript (transcript G^{KA}). Blots A and B represent X-ray films exposed to radiolabeled membranes for 5 days, while blot C represents a 5 min exposure. Using formaldehyde 1% agarose gels, 10 µg of Total RNA was separated in each lane for 2.5 h at 100 V. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. Transcripts of interest are marked by black arrows and are also depicted in the miniature transcript maps below the blots. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNA. For full transcript maps please refer to Figure 36.

indicated that levels of *kan*^{r} transcript (transcript G^{KA}) were also increased in the Δ-2 and Δ-3 strains (Figure 39C, transcript G^{KA}). Thus, in line with observations by Steiner and Philippsen (1994), it was determined that the deletion of region D, which contained the asTATA flanked by the us- and dsRAP1 motifs, was necessary for reducing the level of bidirectional transcription from the *TEF* promoter.

From the analysis of the Δ-2 and Δ-3 strains, it was apparent that the unique transcriptional landscape observed in the 5'^{P}*TEF* Δ-1 mutant was likely attributable to the loss of the motif-rich region D. Thus, to better comprehend the contributions of each of the usRAP1, asTATA, and dsRAP1 motifs in defining the transcription profile, fine-tuned variations of the 5'^{P}*TEF* Δ-1 mutant were constructed (Figure 40). Strain *5''*^{P}*TEF* Δ-1' (referred to as Δ-1') was engineered to exclude the usRAP1 motif, while leaving the asTATA box and dsRAP1 motifs intact. Strain *5''*^{P}*TEF* Δ-1'' (referred to as Δ-1'') was designed to exclude both the usRAP1 motif and asTATA box, leaving only the dsRAP1 motif intact. All the strains Δ-1, Δ-1', and Δ-1'', maintained the sense 8/8 TATA motif in region E. In the sense direction, Northern blot analysis confirmed shortening of major transcription unit E^{K} in accordance with the size of the deletions (Figure 41A, transcript E^{K}). In the antisense direction, it was determined that read-through transcripts M^{K} and N^{K} were specific to the 5' ^{P}*TEF* Δ-1 strain as they were not detectable in Δ-1' or Δ-1'' (Figure 41B). Furthermore, loss of the usRAP1 motif in Δ-1' did not manifest in reduced levels of antisense transcript J^{K}, signifying that usRAP1 was not



Region	Size	Excluded "D" Motifs
5'^{P}*TEF*Δ-1'	167	usRAP1
5'^{P}*TEF*Δ-1''	181	usRAP1, asTATA
5'^{P}*TEF*Δ-1	217	usRAP1, asTATA, dsRAP1

Figure 40. Mutational dissection of "Region D" in the *TEF* promoter

Schematics of deletions 5'^{P}*TEF*Δ-1', 5'^{P}*TEF*Δ-1'' and, 5'^{P}*TEF*Δ-1. All three deletions exclude regions A, B, and C but have varying mutations to region D. 5'^{P}*TEF*Δ-1' only excludes the upstream Rap1 binding site (usRAP1). 5'^{P}*TEF*Δ-1'' excludes both the usRAP1 sequence as well as the neighbouring asTATA box. The original 5'^{P}*TEF*Δ-1 excludes all of region D.

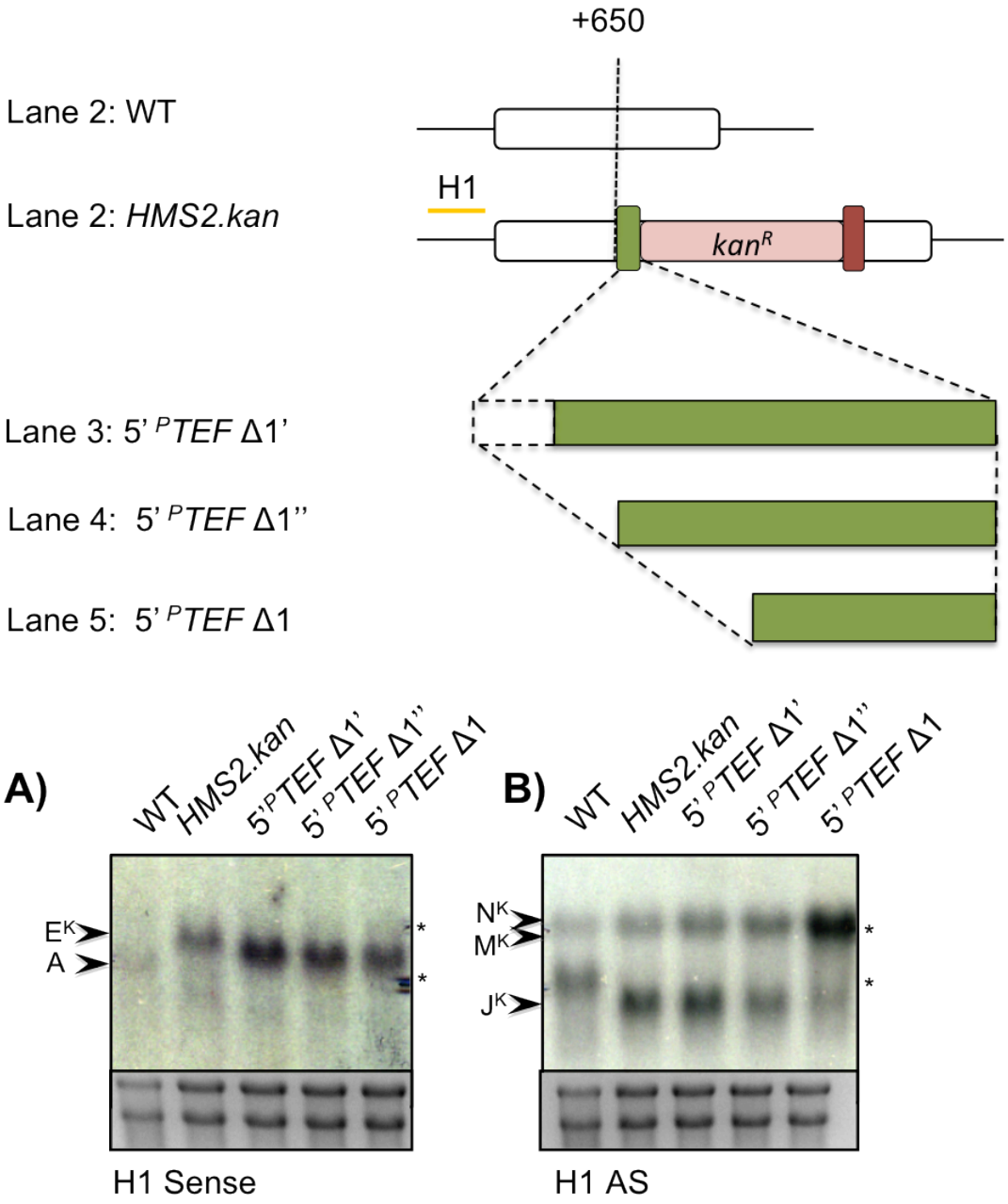


Figure 41. Northern blot analysis of the dissection of "Region D" in the *TEF* promoter

Northern blot analysis of *TEF* promoter mutants 1, 1', and 1'' (For schematics see Figure 40). **(A)** The deletions in the 5' region of the *TEF* promoter are reflected accordingly by the shifts in size of the ^{P}*HMS2-kan-*^{T}*TEF* sense transcript (transcript E^K). **(B)** In the antisense direction, the 5'^{P}*TEF* Δ-1'' mutant, but not the 5'^{P}*TEF* Δ-1' strain, displays decrease in the ^{P}*TEF-*^{P}*HMS2* AS transcript (transcript J^K); however, no read-through transcripts (transcripts M^K or N^K) were detectable as in the original 5'^{P}*TEF* Δ-1 mutant. In this representation of the 5'^{P}*TEF* Δ-1 transcript profile the resolution of transcripts M^K and N^K is relatively low, manifesting as a single diffuse band. Using formaldehyde 1%

agarose gels, 10 µg of Total RNA was separated in each lane for 2.5 h at 100 V. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNA. Transcripts of interest are marked by black arrows and are also depicted in the miniature transcript maps below the blots. For full transcript maps and estimated sizes please refer to Figure 29.

integral to maintaining high levels of antisense transcription from the *TEF* promoter. Only in the Δ-1'' mutant did levels of transcript J^K begin to diminish, indicating that the region containing the asTATA box was important in supporting high levels of antisense transcription from the *TEF* promoter. While levels of *kan*^{r} transcript (transcript G^{KA}) were not measured in these mutants, given the co-regulation displayed in *TEF* mutant strains 1-3, it is likely that levels of transcript J^K also reflect levels of transcript G^{KA} in the Δ-1' and Δ-1'' mutants. In this case, mutants Δ-1, Δ-1', and Δ-1'' illustrate how the regions containing the asTATA box and the dsRAP1 motif are necessary to elicit strong transcription from the *TEF* promoter. Additionally, while weakened, the lone dsRAP1 motif in Δ-1'' mutant was sufficient to support adequate levels of transcription from the *TEF* promoter, such that read-through transcription and/or transcription initiation from the *TEF* terminator was prevented, as indicated by the absence of transcripts M^K and N^K in the Δ-1'' mutant.

4.2.4. Investigating the relationship between 5' and 3' end regulatory elements:

The results from the *TEF* promoter mutational analysis, in particular the original 5' ^{P}*TEF* Δ-1 strain, provided insight into how regulatory elements interact to permit complex, overlapping transcription profiles. Firstly, it became evident that promoters and terminators were capable of performing both 5' and 3' end regulatory functions; for

instance, depending on the context, both the *TEF* promoter and terminator possessed the capacity to initiate and terminate transcription. Moreover, the 5' ^{P}*TEF* Δ-1 mutant illustrated how changing the strength of a promoter element at the 5' end of a transcription unit could affect the activity and function of a distal 3' end terminator, resulting in reshuffling of the transcriptional landscape. However, whether a 3' element merely exists as a passive participant subject to regulation by its promoter pair, or is individually capable of reciprocating the same influence, was largely unclear. Thus, to further elucidate the relationship between 5' and 3' regulatory elements and the their consequence in defining a transcriptional landscape, a mutational analysis of the *TEF* terminator was performed.

Sequence analysis of the *TEF* terminator identified two TATA box-like features in the antisense orientation: an upstream 7/8 variant TATA motif, TATACAAA, and a downstream consensus 8/8 TATA motif, TATAAAAA. Previous unpublished work on the *ADH1* terminator (data not shown), which bears the same two motifs in the sense orientation, identified the TATACAAA sequence, but not the TATAAAAA motif as functional in defining a transcriptional landscape. Thus, it was reasoned that perturbation of this promoter-like element in the *TEF* terminator might influence the activity/function of the *TEF* promoter. To test this hypothesis, the antisense TATACAAA sequence in the *TEF* terminator in the context of *HMS2.kan* was rearranged whilst maintaining the same AT:GC ratio to TTTCAAAA, resulting in the xTATACAAA strain (Figure 42A). As a control, the TATAAAAA sequence was also rearranged to AAAAAATT

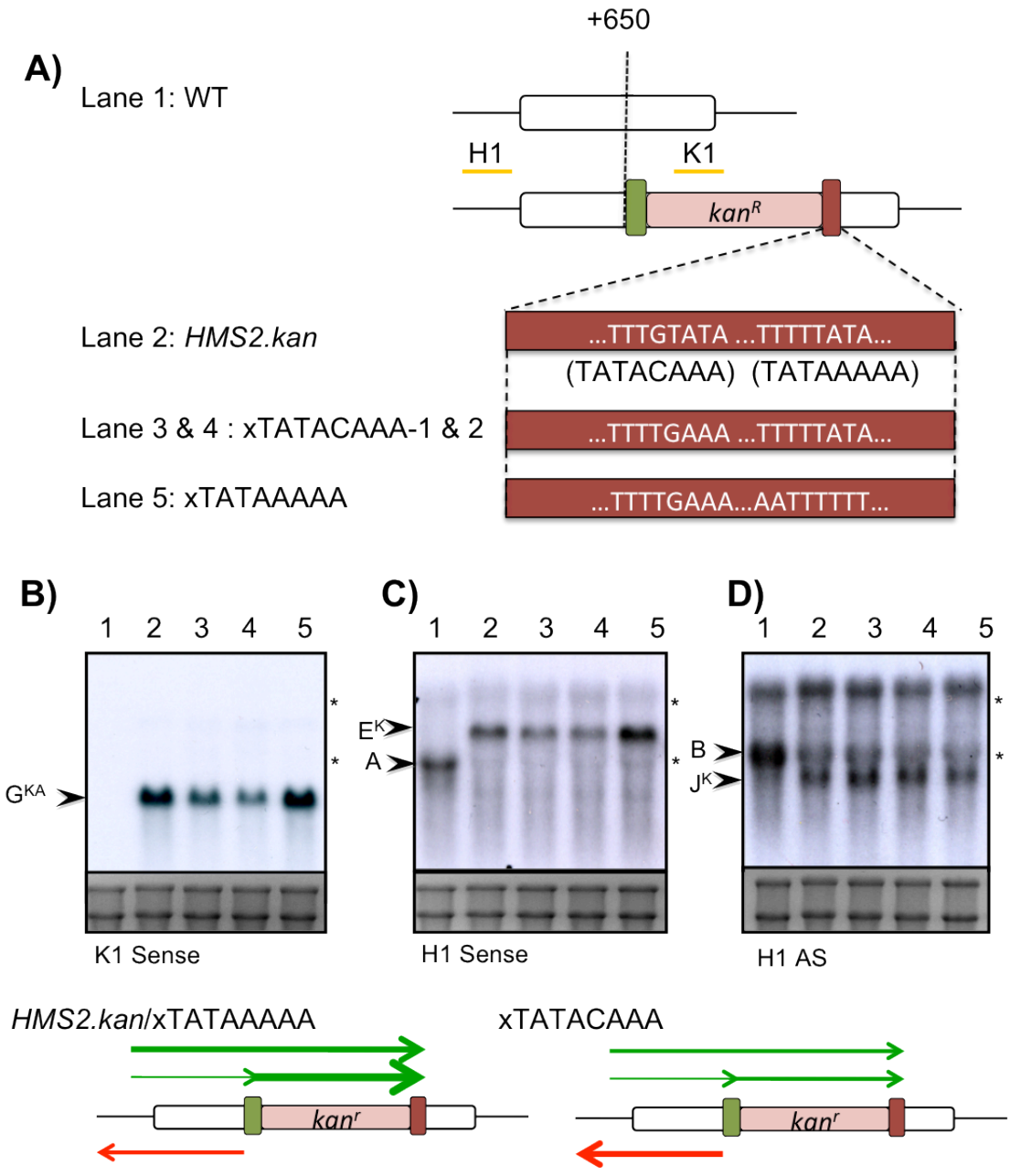


Figure 42. Sequences at the 3' end of a transcription unit can also affect transcript levels.

Northern blot analysis of the **(A)** *TEF* terminator mutants. **(B)** When the 7/8 antisense TATA box-like feature, TATACAAA was scrambled (xTATACAAA-1 & -2) in the ^{T}*TEF*, the ^{P}*TEF-kan-*^{T}*TEF* transcript (transcript G^{KA}) was detected at reduced levels using the K1 sense probe. xTATACAAA-1 & -2 represent two clones of the TATACAAA scramble mutant. When the 8/8 antisense TATA box-like feature, TATAAAAA, was scrambled, no noticeable changes in transcript G^{KA} were apparent. **(C)** Upstream of the *kanMX6* insert (probe H1) in the sense direction, the ^{P}*HMS2-kan-*^{T}*TEF* transcript (transcript E^{K}) was also

detected at reduced levels in the xTATACAAA mutants, while a modest increase was detected in the xTATAAAAA mutant. **(D)** In the antisense direction, a modest increase in levels of ^{P}*TEF*-^{P}*HMS2* (transcript I^{K}) was detected in the xTATACAAA mutants, while no noticeable changes were observed in the xTATAAAAA mutant. Using formaldehyde 1% agarose gels, 10 µg of Total RNA was separated in each lane for 2.5 h at 100 V. Probe locations are shown in yellow in the above schematic. Transcripts of interest are marked by black arrows and are depicted in the schematic below the blots. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below.

to produce the xTATAAAAA strain (Figure 42A). Using a sense probe for the *kan*^{r} coding region (K1), it was revealed that rearrangement of the TATACAAA but not the TATAAAAA sequence resulted in reduced levels of *kan*^{r} transcript (transcript G^{KA}) in comparison to levels in the original *HMS2.kan* strain (Figure 42B). Northern blot analysis performed with the H1 sense probe demonstrated that sense read-through transcript E^{K} also appeared at reduced levels in the xTATACAAA mutant, and at slightly increased levels in the xTATAAAAA mutant (Figure 42C). The decreases in transcript E^{K} and G^{KA} levels in the xTATACAAA mutant might be explained by deficiencies in processing if perturbing the *TEF* terminator sequences resulted in less stable transcripts. Alternatively, it is plausible that mutation of the 3' end TATACAAA motif could negatively affect the activity of the promoters paired with the *TEF* terminator. In support of the latter hypothesis, membranes hybridized with the H1 antisense probe demonstrated increased levels of antisense transcript J^{K} from the *TEF* promoter, a trend unique to the xTATACAAA mutant (Figure 42D). Since both the *TEF* and *HMS2* promoter sequences (where antisense transcript J^{K} is presumed to initiate and terminate, respectively) were unaltered in the xTATACAAA mutant, this observation indicates that altering sequences at the 3' end of a transcription unit can indeed influence the activity

of a 5' element, even in the context of a distinct transcription unit, suggesting communication between the two components. While the nature of this communication is not fully understood, it is an important aspect of gene regulation to consider, especially in the context of an interleaved genome. Additionally, though not further pursued in this study, the TATACAAA element seems to play an important role in transcriptional regulation and merits future investigation.

4.2.5. Abolishing antisense transcription via gene substitution:

While mutational analysis of the *TEF* promoter and terminator provided valuable insight to how relationships between distal regulatory elements affect the transcriptional landscape, the primary goal of finding an alternative method for ablating antisense overlap at the *HMS2* promoter was still unfulfilled. According to the coordinates provided by the Granovskaia et al. (2010) study, the *HMS2* antisense transcription start site lies within the *HMS2* coding region, approximately 100 nt upstream of the translation stop codon. Rather than disrupting overlap by introducing a heterologous cassette, the *HMS2* coding region was substituted for that of *URA3,* effectively deleting the antisense transcription start site without altering the flanking 5' and 3' regulatory sequences; thus avoiding the same concerns raised when the terminator was replaced with that of *ADH1*. Additionally, in light of evidence from the ^{P}*TEF*/^{T}*TEF* mutational analysis, this approach would ostensibly leave 5' and 3' end relationships intact. Northern blot analysis on this *hms2Δ*::*URA3* (referred to as *URA3*) strain was performed using the H1 5' UTR sense and antisense probes, so that transcript levels in the mutant could still be compared to wild-type. *URA3*-containing transcripts, which are 273 nt

shorter than the wild-type *HMS2* transcripts, are annotated with a superscript "U" (Figure 43). Using this method antisense transcription over the *HMS2* locus was successfully abolished (Figure 44A). Furthermore, akin to the changes observed in the *HMS2.ADH* strain, both the sense transcript (transcript A^U) and *URA3-BAT2* read-through transcript (transcript C^U) were found at increased levels, suggesting that the sense and antisense transcripts do exist in an antagonistic relationship (Figure 44B). While the same dot blots were not prepared for the *URA3* strain, it was clear visually by using the same probes (H1) and taking into account the required exposure times of the Northern blot to X-ray film (18 h for *URA3* and 5 day for the *HMS2.ADH* dot blot) that the increase observed in this strain was of at least similar scale to that observed in the *HMS2.ADH* strain.

Sensitivity to Xrn1p and Rrp6p were also investigated to ensure that antisense transcription was truly abolished and not merely destablilzed. Like the wild-type *HMS2* sense transcript, the *URA3* sense transcript (A^U) displayed sensitivity to cytosolic processing via Xrn1p, as a slight increase in sense transcript levels was detected in the *xrn1 Δ* (Figure 45A). Interestingly, in the absence of Rrp6p, the *URA3* sense transcript did not exhibit any notable change in levels. This is in contrast to the native *HMS2* sense transcript, which was detected at increased levels in an Rrp6p deletion strain (Chapter 3). While reduced Rrp6p-sensititvity is possible, an equally plausible explanation could be that the *URA3* sense transcripts are more efficiently exported from the nucleus to the

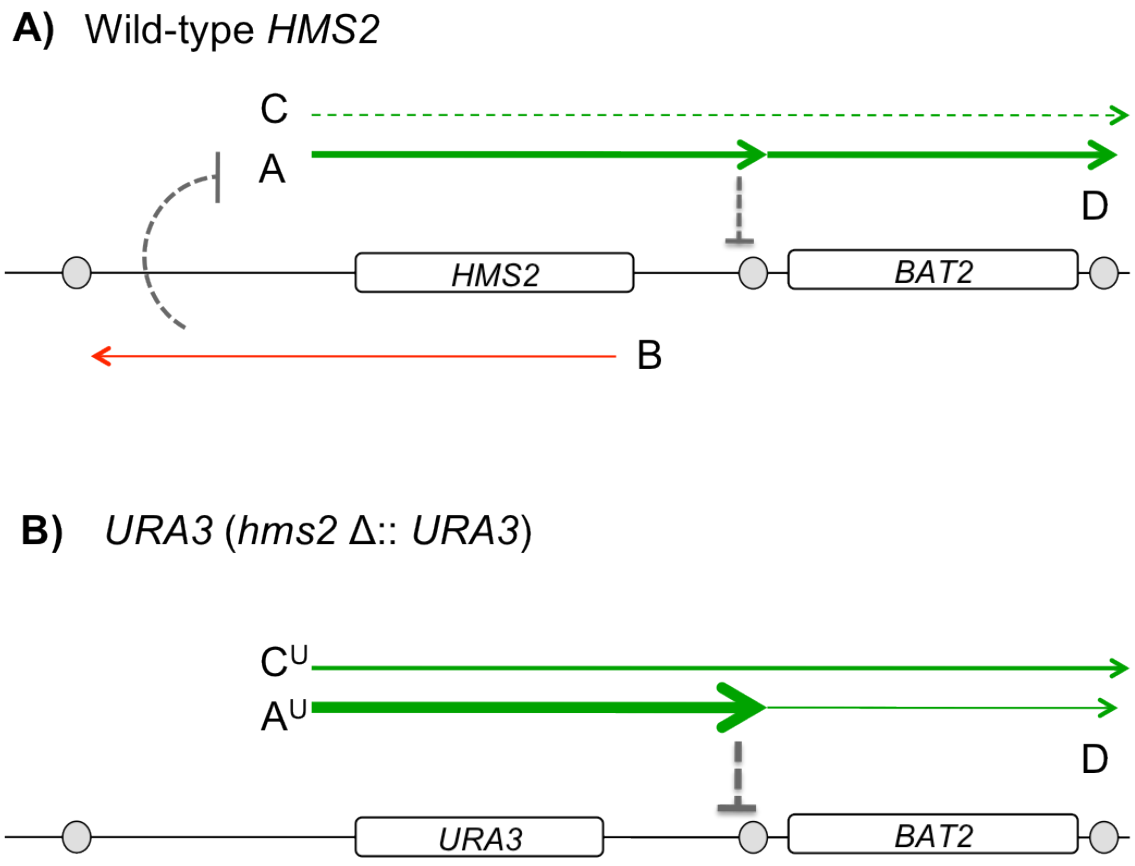


Figure 43. Summary map of wild-type and *URA3* transcript profiles

Schematics summarizing the detected transcripts in (**A)** wild-type and **(B)** *hms2*Δ::*URA3* (*URA3*) strains using Northern blot analysis (See Figure 44). Transcripts detected in the *URA3* strain are annotated with a superscript letter "U". The *URA3* strain does not produce a detectable antisense and displays increased levels of sense (transcript A^U) and sense read-through (transcript C^U) transcripts in comparison to the wild-type equivalents (transcripts A and C, respectively). Additionally, levels of *BAT2* sense transcript (transcript D) are reduced in the URA3 strain relative to wild-type levels. Potential interactions between overlapping transcribed regions are depicted by grey dotted lines.

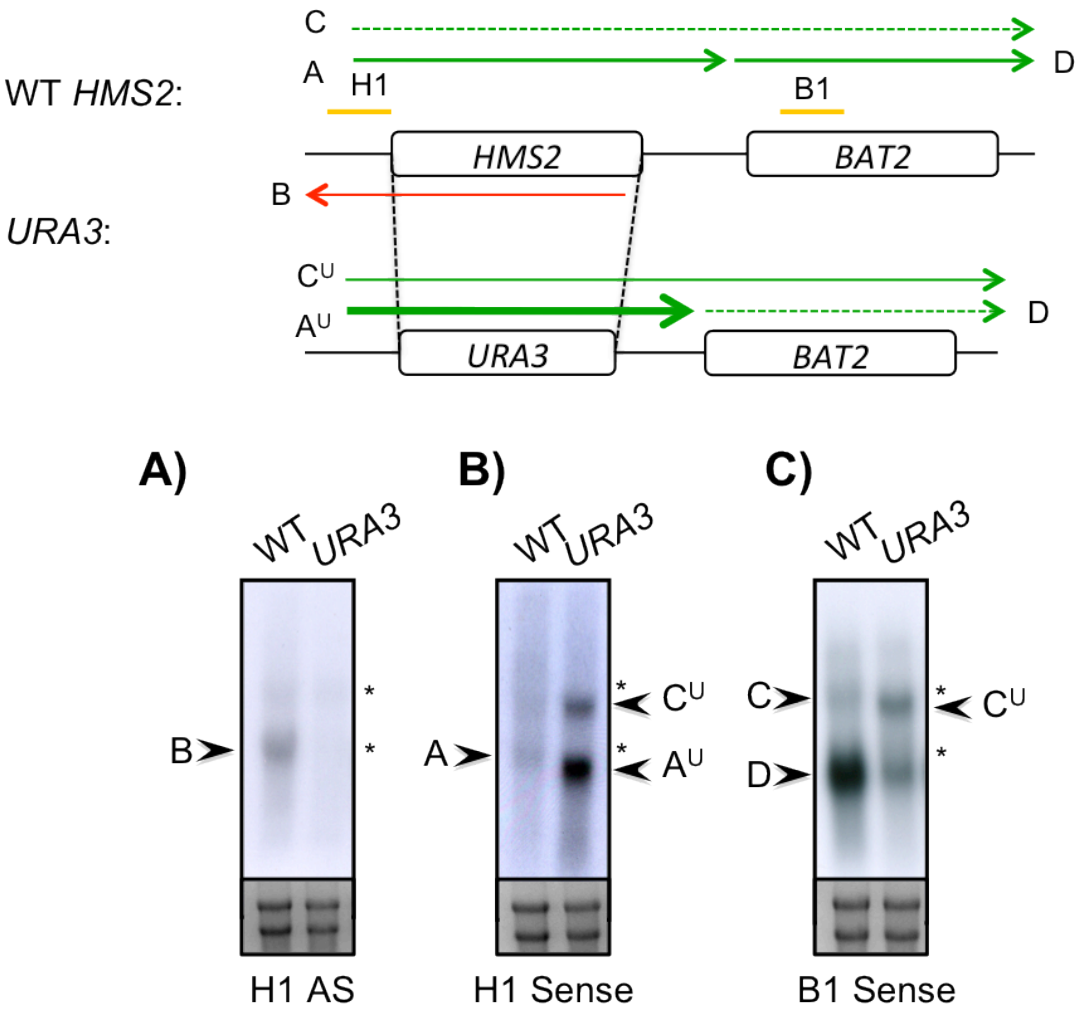


Figure 44. Substitution of the *HMS2* CDS with that of *URA3* ablates antisense transcription

Northern blot analysis of wild-type and *hms2*Δ::*URA3* (*URA3*). Using probe H1, it was demonstrated that substitution of the *HMS2* coding region with that of *URA3* resulted in **(A)** the loss of antisense detection (transcript B), **(B)** a concomitant increase in sense transcript levels (transcript A^U) as well as an increase in the levels of the *HMS2-BAT2* bicistronic transcript (transcript C^U), in comparison to wild-type (transcripts B, A, and C, respectively). **(C)** In the *URA3* substitution strain, a sense probe for the 5' coding region of *BAT2* (B1) detected lower levels of *BAT2* (transcript D) in comparison to wild-type levels. Exposure time =18h. Probe locations (yellow bars) and transcripts of interest are depicted in the schematic above. In this depiction, arrow thickness is indicative of steady-state transcript levels. Using formaldehyde 1% agarose gels, 10 µg of Total RNA was separated in each lane for 2.5 h at 100 V. Transcripts of interest are marked by black arrows. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNA. Ethidium bromide-stained ribosomal RNA loading controls are displayed below.

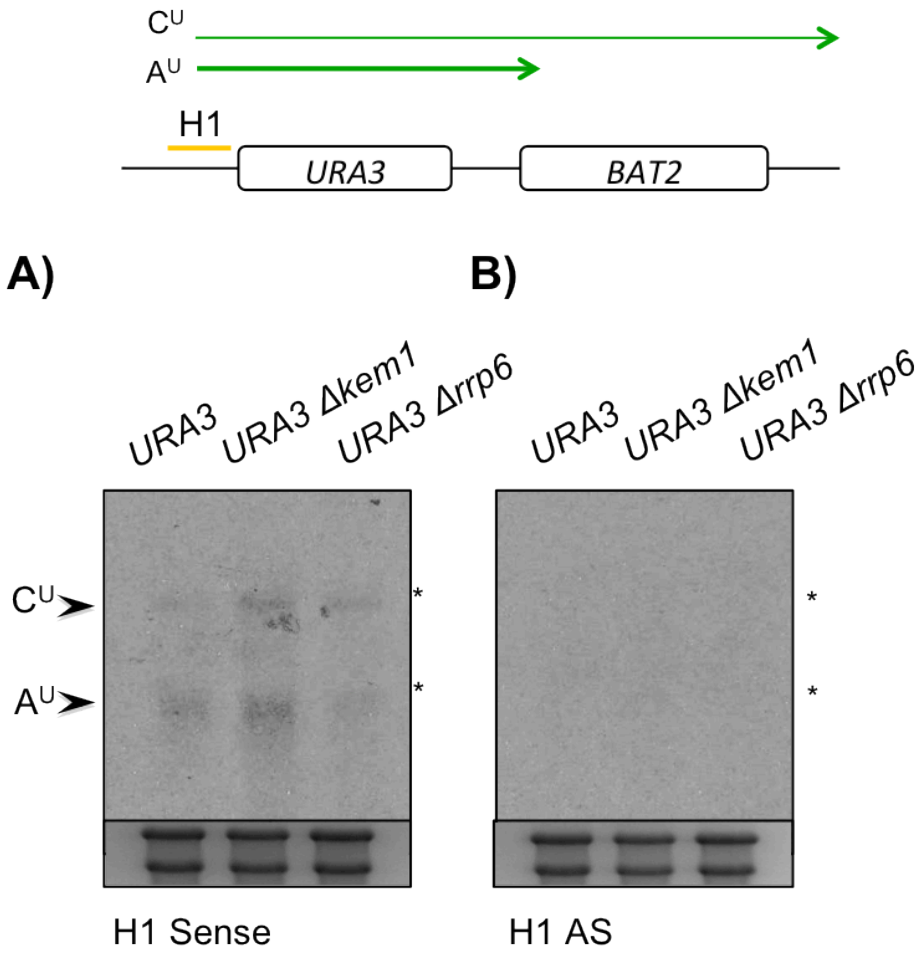


Figure 45. Rrp6p and Xrn1p sensitivity of *URA3* transcripts

Northern blot analysis depicting transcripts in the *URA3* strain in the presence or absence of the RNA processing machinery components Rrp6p and Xrn1p. Blots were hybridized with a probe specific to the *HMS2* 5' UTR (H1, sense/antisense, shown in yellow). No antisense transcripts are stabilized as a result of perturbing the nuclear and cytoplasmic processing machineries. Using formaldehyde 1% agarose gels, 10 µg of Total RNA was separated in each lane for 2.5 h at 100 V. Major transcripts are marked by black arrows and are represented in the above schematic. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below.

cytoplasm, where they are instead subject to Xrn1p-mediated degradation. While determining what features differentiate the post-transcriptional processing of the *HMS2* and *URA3* transcripts would be interesting to pursue, no further work was dedicated to this question in this study. Most importantly, in the antisense direction, no new transcripts were detected in the processing mutants (Figure 45B). Ultimately, with Xrn1p- and Rrp6p-degradation ruled out, substitution of the *HMS2* CDS with that of *URA3* was successful in eliminating antisense transcription at the *HMS2* locus and corroborates our initial findings in the *HMS2.ADH* strain. Collectively, the results from the analysis of the *HMS2.ADH* and *URA3* strains are consistent with the reciprocal relationship observed between the *HMS2* sense/antisense pair in Chapter 3, where increased levels of antisense and decreased levels of sense were detected when cells were shifted from glucose- to galactose-containing media.

4.2.6. Transcription interference represses the BAT2 promoter:

Upon further inspection of the *URA3* transcription profile, the long read-through transcript (transcript C) originally detected in Chapter 3 was also detected at increased levels, shortened in size in accordance with the smaller size of the *URA3* CDS (Figure 44B, transcript C^{U}). Relative increases in this transcript were also detected using the *BAT2* sense probe, as well as concurrent decreases in *BAT2* transcript levels (Figure 44C, transcript C^{U}, transcript D). This observation at *BAT2* provided additional evidence for an anti-correlated relationship between *HMS2* and *BAT2*, previously characterized in the yeast metabolic cycle (Tu et al. 2005) as well as the shift from glucose- to galactose-containing medium (Chapter 3).

In Chapter 3 transcription interference was proposed as the mechanism by which *HMS2* and *BAT2* could be linked, such that increased transcription from *HMS2* would negatively affect the *BAT2* promoter. While the transcription profile of the *URA3* strain supported this proposed relationship, *BAT2* transcript levels were also monitored in the *HMS2.ADH* strain, in which the bulk of *HMS2* sense transcription was terminated before traversing the *BAT2* promoter. In this case, when potential transcription interference at the *BAT2* promoter was prevented, as expected, increased levels of *BAT2* transcript were detected (Figure 46, transcript D). Longer read-through transcripts detected traversing the *BAT2* coding region were likely to initiate from the *TEF* promoter and read through to the *BAT2* poly (A) site or the poly (A) site signal midway through *YJR149W*, mapping consistent with the detection of these transcripts using the K1 sense probe (Figure 30). This additional read-through transcription might explain why the degree of increase experienced by *BAT2* in the *ADH* strain is not very prominent.

4.2.7. Single-cell analysis of antisense-less transcription units

In the previous chapter it was proposed that any sense/antisense reciprocity observed via Northern blot and tiling array analysis was a function of mutually exclusive temporal segregation of sense and antisense transcriptional states in single cells, or more simply put, bi-modal sense/antisense states. According to the proposed model, loss of overlapping antisense transcription would create a one-state system in which single cells could only express the sense transcript. This effective reduction in expression variability in a population of cells would presumably materialize as an increase in sense levels by Northern blot analysis. This conjecture was supported by evidence in the *HMS2.ADH*

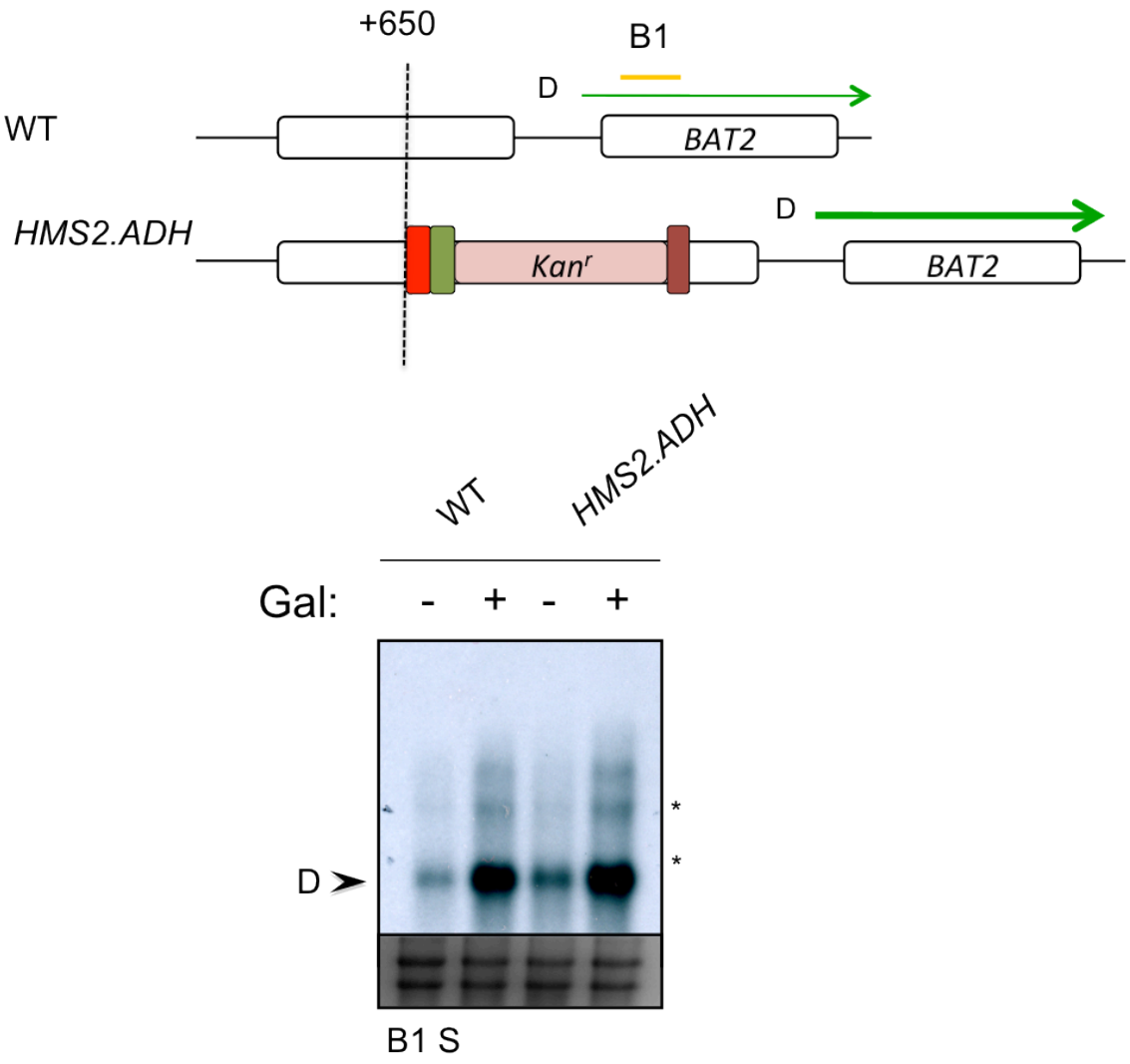


Figure 46. Insertion of the *ADH1* terminator results in increased levels of *BAT2* transcript

Northern blot analysis of *BAT2* transcripts from wild-type and *HMS2.ADH* cells grown to OD_{600}=0.500 in glucose (-) and shifted to galactose-containing medium for 1h (+). The sense probe for the 5' coding region of *BAT2* (B1) detects increased levels of *BAT2* transcript in *HMS2.ADH* cells in comparison to wild-type levels in both growth conditions. Schematics including probe locations (yellow bar) and transcripts of interest are depicted above. In this depiction, arrow thickness is indicative of steady-state transcript levels. Using formaldehyde 1% agarose gels, 10ug of Total RNA was separated in each lane for 2.5h at 100V. Transcripts of interest are marked by black arrows. The asterisks represent the positions of the 25S (3392nt) and 18S (1789nt) RNA. Ethidium bromide-stained ribosomal RNA loading controls are displayed below.

disruption strain and recapitulated in the *URA3* substitution strain. Whether this increase was due to more cells adopting a sense expression state or an increase in the maximum output of the sense promoter was distinguished by performing RNA FISH on the antisense-less *HMS2.ADH* strain. In agreement with our model's prediction, the increase observed in sense transcript levels via Northern blot analysis was a manifestation of a greater proportion of the population expressing *HMS2* sense, with ~62% of the wild-type population and ~81% of the *HMS2.ADH* population expressing the *HMS2* sense transcript, respectively (Figure 47A). In terms of cellular localization, in comparison to wild-type cells (Figure 47B, C) some *HMS2.ADH* cells demonstrated cytoplasmic accumulation of the sense transcript (Figure 47D, E). Though, some *HMS2.ADH* cells still displayed similar accumulation levels to wild-type (Figure 47D-iv and E-v). These observations could be the result of a few possibilities in combination, including more efficient export of *HMS2.ADH* transcripts to the cytoplasm; decreased sensitivity to cytosolic processing; or detection of cells in different phases of the metabolic cycle. Additionally, rough assessment by eye suggests that putative sites of active transcription in *HMS2.ADH* (Figure 47D-ii) do not appear to increase in size/intensity relative to wild-type cells (Figure 47B-iv), possibly indicating that re-initiation rates are not affected in the absence of antisense; a hypothesis consistent with bi-modal state switching. This is made especially clear when the sites of active transcription for *HMS2* are compared to a gene displaying high rates of re-initiation, such as the inducible gene *GAL1* (Figure 48).

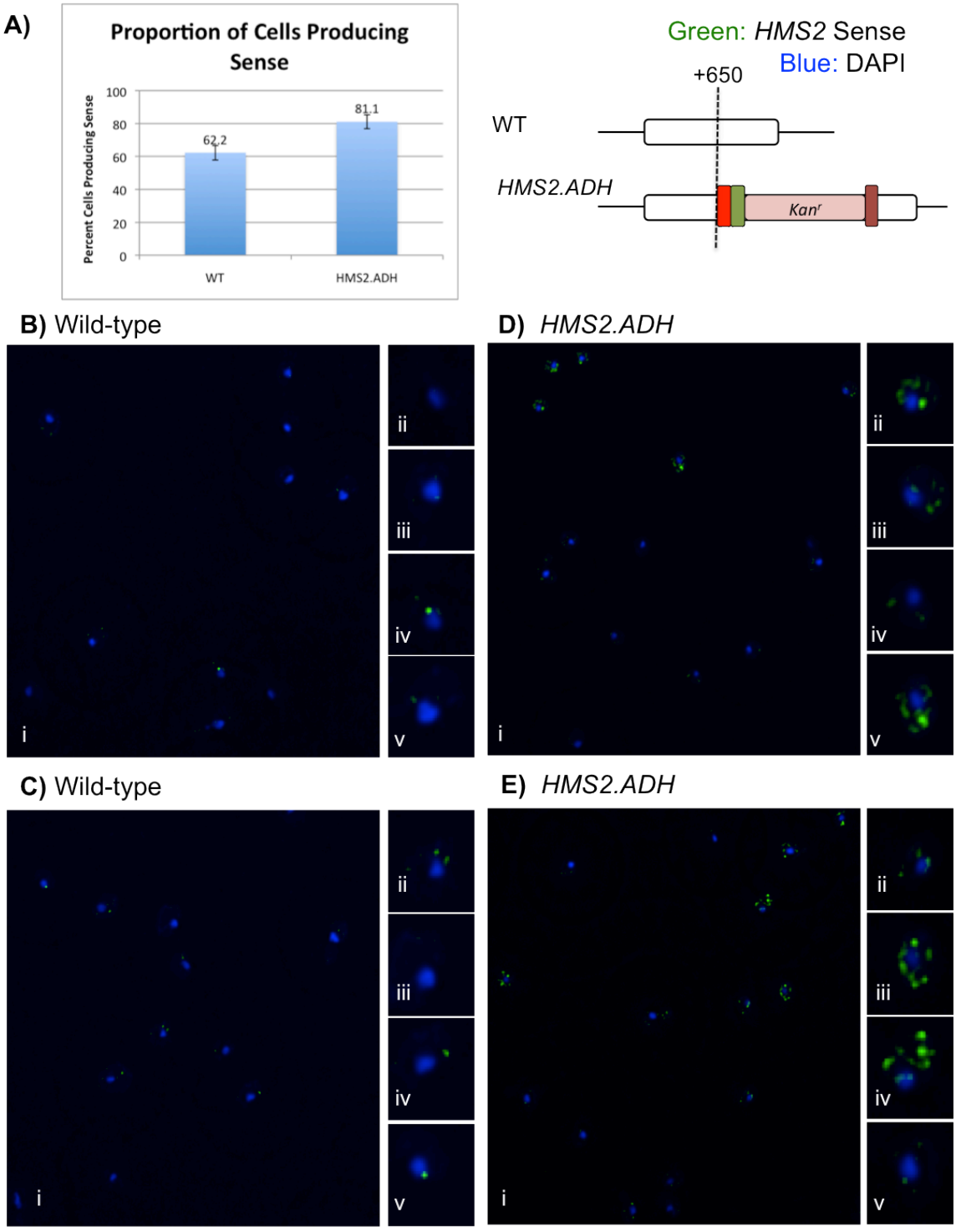


Figure 47. Single-cell analysis of *HMS2* sense expression in wild-type and *HMS2.ADH* cells.

(A) The proportion of sense-expressing cells in the presence or absence of antisense overlap at the *HMS2* sense promoter as determined by RNA FISH. *HMS2* sense

expression was visualized in single cells grown in glucose to OD_{600}= 0.500 in **(B, C)** wild-type (n= 127) and **(D, E)** *HMS2.ADH* cells (n= 95). In order to capture single *HMS2* sense (green) transcripts, a combination of four, 50 nt DNA probes labeled with four Cy3 fluorophores were hybridized to fixed yeast cells. DAPI nuclear stains are displayed in blue. Smaller boxes to the left of the main pictures (i) are zoomed images of select cells (ii-v). Error bars represent standard deviations, which were determined by producing 1000 bootstrap pseudoreplicates of both datasets (glucose vs. galactose) and comparing the resultant distributions of percentages using a Wilcoxon rank-sum test (statistical analysis performed by Struan Murray). p= 0.000168.

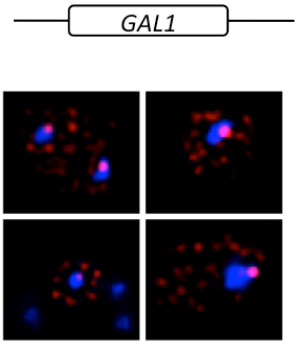


Figure 48 Single-cell analysis of *GAL1* sense expression in wild-type cells.

High rates of reinitiation at the *GAL1* promoter manifest as relatively large, high intensity spots overlapping the nucleus. In order to capture single *GAL1* sense (red) transcripts, a combination of four, 50 nt DNA probes labeled with four Cy3 fluorophores were hybridized to fixed yeast cells (FISH performed by Simon Haenni). DAPI nuclear stains are displayed in blue.

The results from the single cell analysis of the *HMS2.ADH* strain were consistent with the FISH results obtained in wild-type cells shifted from glucose- to galactose-containing media, which demonstrated that after growth in galactose, the increased RNA levels detected via Northern blot analysis equated to more cells in the population producing the antisense transcript. In terms of the biology of this response, previous studies have shown that the proportion of cells in the high oxygen consumption (HOC, oxidative) phases of the metabolic cycles is directly proportional to growth rate (Slavov & Botstein, 2011). Thus, during galactose induction, when growth rate decreases, it is likely that the

number of HOC cells also decreases. Since *HMS2* sense is maximally expressed in the HOC, oxidative phase of the YMC, it is likely that the slower growth, and by extension, more cells in the low oxygen consumption (LOC) phases of the YMC, explains the shift from sense- to antisense-dominant state during the glucose to galactose shift. Accordingly, the increased number of cells expressing the *HMS2* sense transcript in the *HMS2.ADH* strain might be explained by an enhanced growth rate when *HMS2* is disrupted. Thus, to rule out any growth-related effects in the *HMS2.ADH* strain, growth rates of wild-type and *HMS2.ADH* cells were compared on YPD plates; however, no apparent differences were documented (Figure 49).

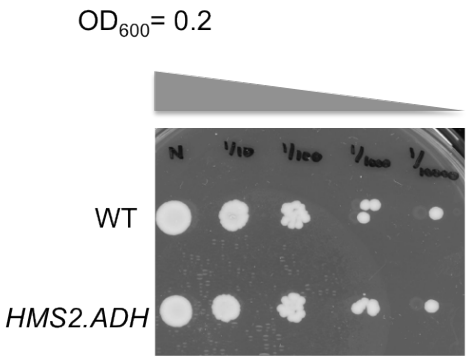


Figure 49. Assessing differences between growth of wild-type and *HMS2.ADH* cells.

Differences in growth rate between wild-type and *HMS2.ADH* cells was determined by drop plate assay. Overnight cultures were normalized to an OD_{600}= 0.200 before 2 µl of "neat" (N) and four serial ten-fold dilutions were pipetted on YPD-agar plates. Plates were examined every 24 h for up to 72 h. The plate in this figure represents a 72 h growth period.

Alternatively, if more cells were not occupying an HOC phase in the antisense-less *HMS2.ADH* strain, then these cells might have been expressing *HMS2* sense out-of-phase, indicating that the overlap of antisense transcription with the sense promoter acts to safeguard robust periodicity by preserving the "off" state of the sense gene,

consistent with the proposed bi-modality in sense/antisense relationships. In a population of metabolically-synchronized cells, this robustness might manifest itself as an increase in the dynamic range of expression. A plausible expectation then, would be that metabolically cycling genes displaying high amplitude peaks in expression, or a high maximum to minimum expression ratio, would be enriched for antisense transcription. For instance, from the Tu et al. (2005) dataset, *HMS2* has a maximum to minimum expression ratio of 36.22. When the protein coding genes with the top ~1% of max/min ratios were observed (36 genes), it appeared as though antisense transcription was a feature of approximately 69.4% of these units, as measured by NET-Seq, or tiling array analysis (Table 6). This is in comparison to the frequency of genes with antisense transcription across the genome as measured by RNA-seq and tiling array analysis, 36.5%, though, the difference between these frequencies may be down to the sensitivity of the techniques used to measure low level and/or unstable transcripts (Murray et al., 2012).

Gene	Min/Max Ratio	Net-Seq (Cut off score ≥5)	Microarray Detection (Xu et al. 2009)	Antisense?
POX1	82.03	✓	✓	✓
GRE1	82.31	✓	✓	✓
HEF3	82.62	×	×	×
CTA1	85.24	✓	✓	✓
SPO11	86.7	✓	✓	✓
UTP5	90.18	×	✓	✓
EGT2	90.22	×	✓	✓
CAT2	90.32	✓	✓	✓
ARX1	92.43	×	×	×
KRS1	96.12	✓	✓	✓
DSE4	98.5	✓	✓	✓
AIR1	99.52	✓	×	✓

SPO16	100.83	✓	✓	✓
CTT1	103.44	✓	✓	✓
BUD27	108.43	×	×	×
DBP3	108.89	✓	✓	✓
FOX2	111.57	✓	✓	✓
AAH1	113.2	✓	✓	✓
SLZ1	120.19	✓	✓	✓
MCD1	120.25	×	×	×
URA7	121.61	×	×	×
CIT3	127.65	✓	✓	✓
BAG7	141.46	✓	✓	✓
ARO9	152.72	×	×	×
YJL216C	163.31	×	×	×
KRR1	164.82	×	×	×
DOG1	178.36	×	×	×
NOP8	187.53	×	×	×
SWI5	198.21	×	×	×
KRE33	198.88	×	✓	✓
NSR1	203.86	×	✓	✓
OPT2	225.71	✓	✓	✓
TOS2	262.76	✓	✓	✓
TKL2	275	✓	✓	✓
CLB6	325.68	✓	✓	✓
PCL1	477.18	✓	✓	✓
			TOTAL	25
			Percentage (%)	69.4

Table 6. Antisense transcription as determined by NET-Seq of metabolically cycling genes with the top 1% max/min ratios.

While just a crude comparison, absence of antisense transcription in high amplitude genes might reflect alternative modes of regulation (e.g. other non-coding RNAs), or limitations in the ability to detect antisense transcripts (e.g. condition specific expression). Interestingly, in several cases antisense transcription was not detected using NET-seq, but antisense transcripts were annotated using tiling array analysis (Xu et al., 2009), underscoring the importance for multilateral analysis of transcription in order to obtain an accurate depiction of the transcriptional landscape.

4.3. Discussion:

In Chapter 3, a bimodal state-switching model was proposed in light of evidence demonstrating that production of overlapping sense/antisense transcripts was temporally segregated. In this chapter, to test the integrity of the model, the outcome of a one-state system was observed by abolishing antisense transcription over the *HMS2* sense promoter. Serendipitously, in the process of creating this antisense-less unit, a better understanding of how a cell might cope with a complex network of multiple overlapping transcription units (both sense and antisense) was attained.

4.3.1. Creation of a fixed state system may generate noisy expression of HMS2 in the metabolic cycle:

Results obtained from the Northern blot and single cell-analysis of the *HMS2.ADH* and *URA3* strains validated the state-switching model proposed in Chapter 3, as removal of the antisense state forced more cells to enter a sense expression state. While sense/antisense bimodality might be an intuitive expression strategy, especially given the steric clashes presented by polymerase collisions (Hobson et al., 2012), interpreting this modularity in terms of a biological function was less clear. In Chapter 3, it was proposed that sense and antisense states could be executed within the context of time-dependent physiological programmes such as the metabolic cycle, such that expression patterns of a gene exist as an ordered, oscillating entity. While availability of cycling transcriptional regulators probably determines whether a gene (sense or antisense) is expressed/repressed in any given phase of the YMC (Tu et al., 2005), a possible role for antisense in oscillating transcriptional networks might be to exert an additional degree

of negative control during the "off" phase of its protein-coding counterpart, resulting in tighter control of periodicity and a reduction in potentially detrimental out-of-phase stochastic expression. For instance, though the effects of antisense elimination on the periodicity of *HMS2* and *BAT2* expression during the metabolic cycle was undetermined, bearing in mind the greater proportion of cells expressing the *HMS2* sense transcript when antisense transcription was ablated, it is plausible that by relieving any negative effects caused by the overlap of antisense transcription into the sense promoter, the locus was partially decoupled from metabolic regulation, such that without repressive antisense, sense transcription was more permissible when its phase-specific activator was not present. In other words, without antisense transcription at *HMS2*, cells experienced an increase in extra-phasic expression "noise" amounting to more cells expressing the *HMS2* sense transcript. Additionally, this antisense-mediated control of robust periodicity is apparently not unique to *HMS2*, but a feature of over two-thirds of the genes displaying high magnitude expression amplitudes. Altogether, this proposed role for antisense as a chaperone of periodicity brings together several initially confounding observations about the *HMS2* locus, including reciprocal and temporally separated sense/antisense expression; metabolic control of *HMS2* and *BAT2*; and the abundance of binding sites for metabolically cycling factors at the 3' end of the *HMS2* locus. To fully test these theories, further examination of *HMS2* sense, antisense and *BAT2* expression states in cells synchronized in the metabolic cycle would be needed.

4.3.2. Antisense transcription in an interleaved genome:

The principal goal of this chapter was to corroborate the relationship between sense and antisense transcription described in Chapter 3. However, in the process of creating antisense-less systems, additional intricacies found in the transcription profiles of the different strains suggested that the role of antisense transcription was more nuanced than to strictly antagonize its sense pair. Rather, the results from the study of conglomerate *kanMX6* constructs indicated that antisense transcription might exist as a component in a more complex array of transcriptional interactions, a likely scenario given the compact and interleaved nature of the yeast genome. Indeed, evidence from the analysis of the antisense-less *URA3* construct hinted that without antisense transcription, the *BAT2* promoter was weakened. The associated increase in *HMS2* sense and *HMS2-BAT2* read-through transcript levels was compatible with a transcription interference model, similar to that described by a study of the SRG1/SER3 locus, which demonstrated that non-coding transcription from a stronger upstream promoter (*SRG1*) reduced expression from the downstream promoter of the coding gene, *SER3* (Martens et al., 2004). This theory was substantiated by the insertion of the *ADH1* terminator, which prevented the bulk of sense transcription from traversing the *HMS2-BAT2* intergenic region and resulted in an increase in detectable *BAT2* levels.

Taken together, the analysis of the *HMS2.ADH* and *URA3* constructs indicated that the function of antisense transcription at the *HMS2* gene cluster is likely twofold: 1) it serves to negatively influence *HMS2* sense transcription and in doing so 2) it prevents the *BAT2*

promoter, which is susceptible to the effects of repressive transcription interference, from experiencing reduced expression levels. In a biological context, this antisense-modulated transcription interference also might explain why peak expression of the *HMS2* and *BAT2* loci are found in separate phases of the metabolic cycle (Tu et al., 2005), which is in agreement with the proposed role of *HMS2* antisense transcription in managing periodicity. Additional repressive antisense effects at the *RPS4A* promoter, consistent with *RPS4A*'s coordinate expression with *HMS2* in the oxidative phase, are also plausible; however, more evidence is needed to confirm this possibility. Given the implications of antisense-mediated transcription interference in inter-locus communication at *HMS2*, the next chapter explores the contribution of transcription interference to state-switching and the mechanisms through which it might act.

4.3.3. Promoter/terminator strength and functional duality determine transcriptional landscapes:

While the addition of the *kanMX6* cassette may be viewed as a highly contrived model system, the interspersion of promoters and terminators between genic regions actually resembles the interleaved gene organization in yeast and mammals (Kapranov et al., 2007). The comparative analysis of the *HMS2.kan* and *5'*Δ^{P}*TEF* strains provided compelling evidence for the conditional and multifunctional nature of promoters and terminators and how an interleaved arrangement of these regulatory elements preserves a state of pervasive transcription. Furthermore, the analysis demonstrated that this type of genomic organization does not create a fixed transcriptional landscape, but can allow for flexibility by exploiting the versatility of regulatory elements. For

instance, the *TEF* promoter, in the context of the *HMS2.kan* construct, drove strong transcription in the sense (transcript G^{KA}) and antisense (transcript J^{K}) directions and terminated transcription from the native *HMS2* promoter (transcript F^{K}), whereas the *TEF* terminator was able to terminate both transcript E^{K} and the natural *HMS2* antisense transcripts (transcript L^{KA}). Unsurprisingly, in the 5' Δ^{P}*TEF* strain, the weakened *TEF* promoter resulted in lower levels of bidirectional initiation. Strikingly, this was also associated with a novel long antisense transcript reading through the *TEF* terminator (transcript N^{K}), indicating that perturbation of the potency of the *TEF* promoter adversely affected the capacity of the intact *TEF* terminator to stop transcription bidirectionally. While these experiments physically modified the sequence of the *TEF* promoter, in a natural setting, promoter/terminator conditionality and strength could be mediated by the presence or absence of transcriptional activators/repressors (e.g. periodic availability in the metabolic cycle) and/or chromatin structure. For instance, previous studies have shown that cells must contend with cryptic promoters by Spt6p-mediated nucleosome occlusion in order to maintain levels of coding transcription (Kaplan, Laprade, & Winston, 2003). The results in the 5'Δ^{P}*TEF* strain might reflect how cells cope with an interleaved organization in general; such that transcription units bound by conditional regulatory elements in "weaker" states permit a more dynamic and complex transcriptional landscape, while promoters and terminators in "stronger" states (for instance, in the presence of an activator) result in a more rigid profile. In relation to the proposed model, the notion that interspersed, multifunctional regulatory elements can display conditionality in both purpose and potency coincides with the

potential for a changeable transcriptional profile and thus reinforces the state-switching theory.

Chapter 5.

Devising a mechanism for sense/antisense state-switching.

5. Devising a Mechanisms for Sense/Antisense State-switching

5.1. Introduction:

5.1.1. Regulatory Transcription:

With the robustness of the state-switching model established (Chapter 4), this chapter explores the potential mechanisms through which transcriptional states are defined and the resultant molecular consequences of adopting and maintaining these states. In particular, as postulated by the model, the importance of transcription factor binding and subsequent generation of interfering transcription was considered.

According to the initial model, transcription factors activating a state will subsequently drive the production of interfering transcripts. Numerous studies in *E. coli*, yeast, and mammals have described transcription interference as a mechanism through which coding and non-coding RNAs could exert repressive functions (reviewed in Mazo, Hodgson, Petruk, Sedkov, & Brock, 2007) and evidence from the previous chapter strongly indicated that this was also likely at the *HMS2* gene cluster. A review by Shearwin et al. (2005) defined transcription interference as "the suppressive influence of one transcriptional process, directly and *in cis*, on a second transcriptional process"; however, there exist several competing views on the exact molecular basis for this suppression. The most comprehensive model for how interfering, non-coding transcription might regulate a gene comes from a series of studies on the *SER3* locus,

mentioned briefly in Chapter 4. Under repressive serine-rich conditions, an upstream promoter drives transcription of a non-coding RNA, *SRG1*, over the *SER3* promoter, causing impaired activator binding (Martens et al., 2004). It was later found by Hainer and colleagues (2011) that abrogation of transcription factor binding was the result of Spt6p/Spt16p-mediated nucleosome assembly, directed by *SRG1* transcription, over the *SER3* promoter. Mutants for these elongation factors displayed reduced nucleosome occupancy at the *SER3* promoter, defective transcription interference of *SER3*, but unaffected levels of *SRG1* transcription (Hainer, Pruneski, Mitchell, Monteverde, & Martens, 2011). Furthermore, the study demonstrated that transcription interference of the *SER3* promoter was independent of histone modifications that suppress cryptic initiation, such as Set2-mediated deposition of H3K36me2. Additionally, in terms of *SRG1* regulation, an earlier study reported that production of the *SRG1* RNA was responsive to environmental levels of serine and was activated by Cha4p, a serine-sensitive transcriptional activator, demonstrating how a cell could directly mediate non-coding interfering transcription and its downstream consequences (Martens, Wu, & Winston, 2005).

A similar regulatory mechanism was described at the *IME1* locus, which is subject to control by *in cis* transcription of an upstream intergenic non-coding RNA, *IRT1* (Van Werven et al., 2012). Similar to the *SRG1/SER3* system, *IRT1* negatively influences *IME1* transcription by antagonizing the binding of its transcriptional activators to the *IME1* promoter. However, rather than Spt6p/Spt16p-mediated nucleosome deposition, van

Werven and colleagues demonstrated that *IRT1* transcription establishes a repressive chromatin state at the *IME1* promoter, modulated by Set2-mediated co-transcriptional deposition of H3K36 methylation and Set3-mediated deacetylation.

The *SRG1/SER3* and *IRT1/IME1* systems are highly attuned with the proposed state-switching model. Like *SRG1* and *IRT1*, *HMS2* antisense transcription is likely regulated by a set of transcriptional activators (Chapter 3), and responsiveness to these regulators might drive transcription with the potential to interfere with the *HMS2* sense promoter during the antisense state. However, whether the mechanism of interference between two genes arranged in tandem, such as *SRG1* and *SER3* or *IRT1* and *IME4*, is applicable to two genes arranged in convergent antisense orientation is largely unknown. For instance, Hongay et al. (2006) proposed that antisense-mediated repression of *IME4* transcription could act through transcription interference, in which transcription from a strong aggressor promoter would diminish transcription from a weaker promoter on the complementary strand; though, unlike the *SER3* and *IME1* studies no molecular basis for the interference was provided.

In an earlier report by Prescott and Proudfoot (2001), the consequence of convergent transcription interference was investigated. When the authors repositioned the *GAL10* and *GAL7* genes, which are normally oriented in tandem, to generate a convergent arrangement, they demonstrated that transcription was unimpaired, so long as the two transcription units did not experience any overlap. However, deletion of the entire

GAL10-GAL7 intergenic region when the genes were arranged convergently, effectively disrupting termination signals and fusing *GAL10* and *GAL7* in a conjugate antisense arrangement, resulted in drastic reductions in steady-state levels of *GAL10* and *GAL7* RNA. The authors attributed this to reduced transcriptional processivity as a result of polymerase collisions, as transcription run-on (TRO) analysis indicated that transcription initiation was not affected.

The dire consequences of colliding polymerases described by Prescott and Proudfoot (2001) indicate a very strong evolutionary pressure against convergent overlapping transcription that is seemingly incompatible with the prevalence of overlapping antisense transcription in yeast. However, the *GAL10-GAL7* example used in the Prescott and Proudfoot study is a highly artificial model system in which the Gal4p activator coordinates concurrent expression of both genes, making polymerase collisions more likely. Accordingly, when the authors prevented simultaneous initiation, interference did not occur. Together, these results underpin the theory that in order for cells to support overlapping transcription over a gene, sense and antisense transcription must be temporally separated, as demonstrated in Chapter 3. With temporal segregation of sense and antisense transcription rendering the significance of polymerase collisions moot, a viable explanation for achieving antisense-mediated transcription interference was still required.

5.1.2. Aims:

The work presented in this chapter endeavours to address the two missing links in the proposed model: 1) how a transcriptional state is determined and maintained and 2) what consequences arise from the resultant state. Thus far, the experiments in chapters 3 and 4 have provoked two distinct transcriptional states at *HMS2*, an antisense-dominant state when cells are grown in galactose-containing media, and a sense-dominant state in cells in which antisense transcription has been artificially ablated. RNA FISH data suggested that these states are defined at a single-cell level, and are separated through time, potentially through oscillating biological programmes like the YMC, a very likely phenomenon given the detrimental effects of polymerase collisions. However, in both of these circumstances, cells were observed that have already achieved a steady state of transcription, in which either a sense or antisense state had been defined. To fully elucidate the molecular mechanism by which a transcriptional state is decided upon, a scenario in which competing sense and antisense promoters could be observed was required. Thus, the *HMS2* antisense-dominant state in galactose-grown cells was exploited and the *HMS2* sense promoter was replaced with the stronger galactose inducible promoter of *GAL1* (*pGAL1)*. It was expected that during galactose induction, the *GAL1* promoter would outcompete the initially strong *HMS2* AS promoter over time, in which case an antisense-to-sense state switch and its imminent consequences could be observed in near real-time.

5.2. Results:

5.2.1. Coordinating a state-switch using a galactose inducible form of HMS2:

Through Northern blot analysis of galactose-induced cells, transcription profiles of the *HMS2* locus in various genetic contexts were compiled in order to capture the details underlying the mechanism of state-switching as well as the contributions of antisense transcription to transcription kinetics (Figure 50). Transcripts detected in the galactose inducible *HMS2* strain are annotated with a superscript "G". Transcripts detected in the galactose inducible *hms2*Δ::*URA3* strain (discussed later in this chapter) are annotated with a superscript "GU".

Upon observation of the hybrid *pGAL1.HMS2* gene construct in glucose-containing media, no band corresponding to the expected size of the *pGAL1.HMS2* sense transcript (~1700 nt) was detected (Figure 51A). Detection of a band below the 18S rRNA marker in the *HMS2* sense blot was determined to be the result of cross-hybridization of the H2 sense probe to the antisense transcript, as use of another probe for the *HMS2-BAT2* intergenic region (probe I1) did not detect this band (Figure 51B). Meanwhile in the antisense direction, levels of the *HMS2* antisense transcript were increased in the *pGAL1.HMS2* strain in comparison to wild-type (Figure 51C). The antisense transcript in *pGAL1.HMS2* (transcript B^{G1}) also demonstrated a shortened length relative to wild-type, likely because its transcription terminates in the *GAL1* promoter sequences rather than extending ~650-800 bp into the *RPS4A-HMS2* intergenic region. Thus, as expected

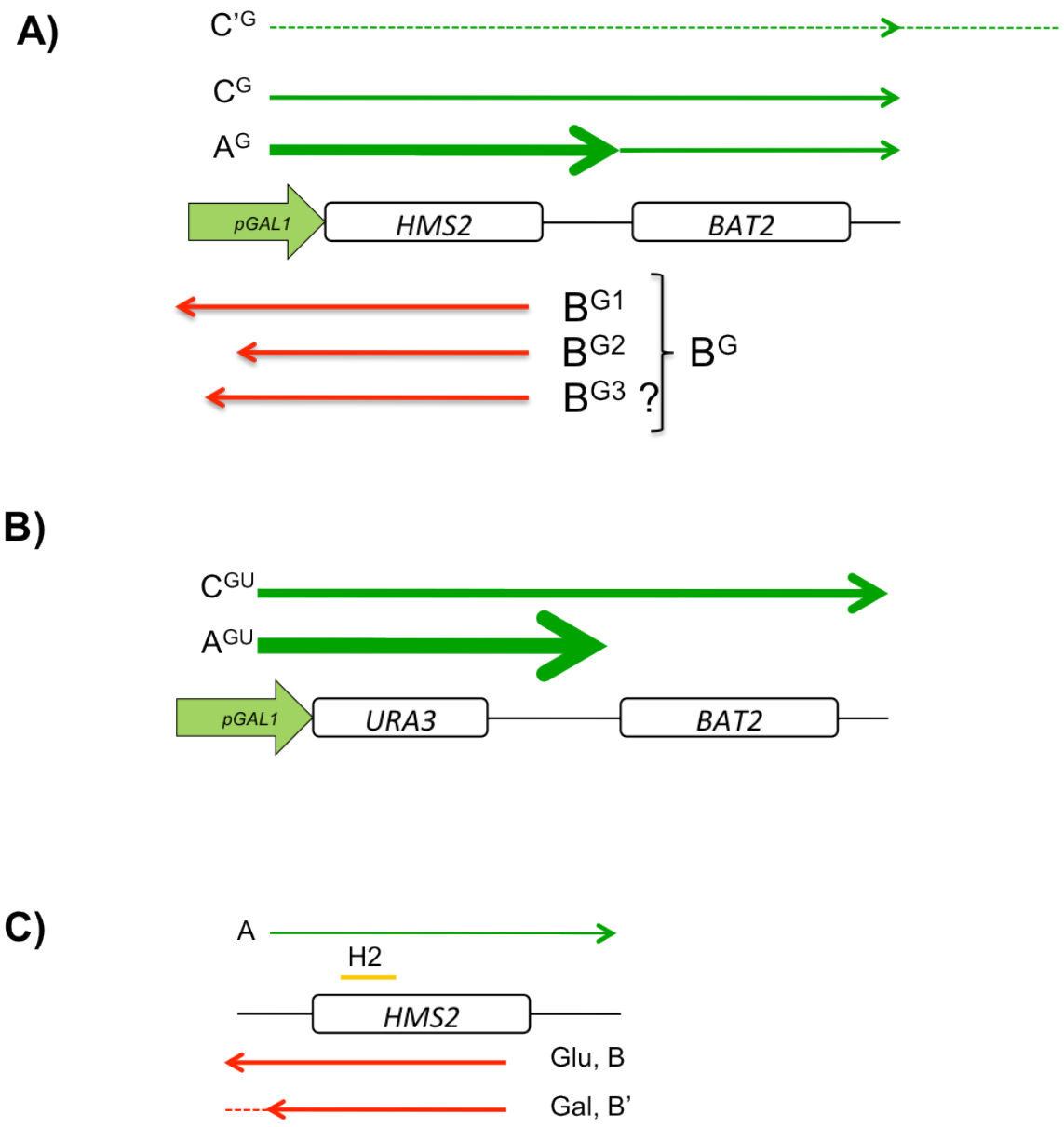


Figure 50. Summary map of *pGAL1.HMS2*, *pGAL1.URA3* and wild-type *HMS2* transcript profiles in cells grown in glucose- and galactose-containing media.

Schematics summarizing the expected transcripts in **(A)** *pGAL1.HMS2* **(B)** *pGAL1.URA3* and **(C)** wild-type *HMS2* cells grown in glucose- and galactose-containing media using Northern blot analysis (See Figure 52, Figure 53, and Figure 59). Transcripts detected in the *pGAL1.HMS2* strain are annotated with a superscript letter "G", while transcripts detected in the *pGAL1.URA3* strain are annotated with a superscript letter "GU". The *pGAL1.HMS2* antisense transcript isoforms are numerated, but are referred to collectively as antisense transcripts "B^{G}". *HMS2* antisense transcript heterogeneity (transcripts B and B') in wild-type cells was demonstrated to be condition-specific, with the shorter form manifesting during growth in galactose-containing medium.

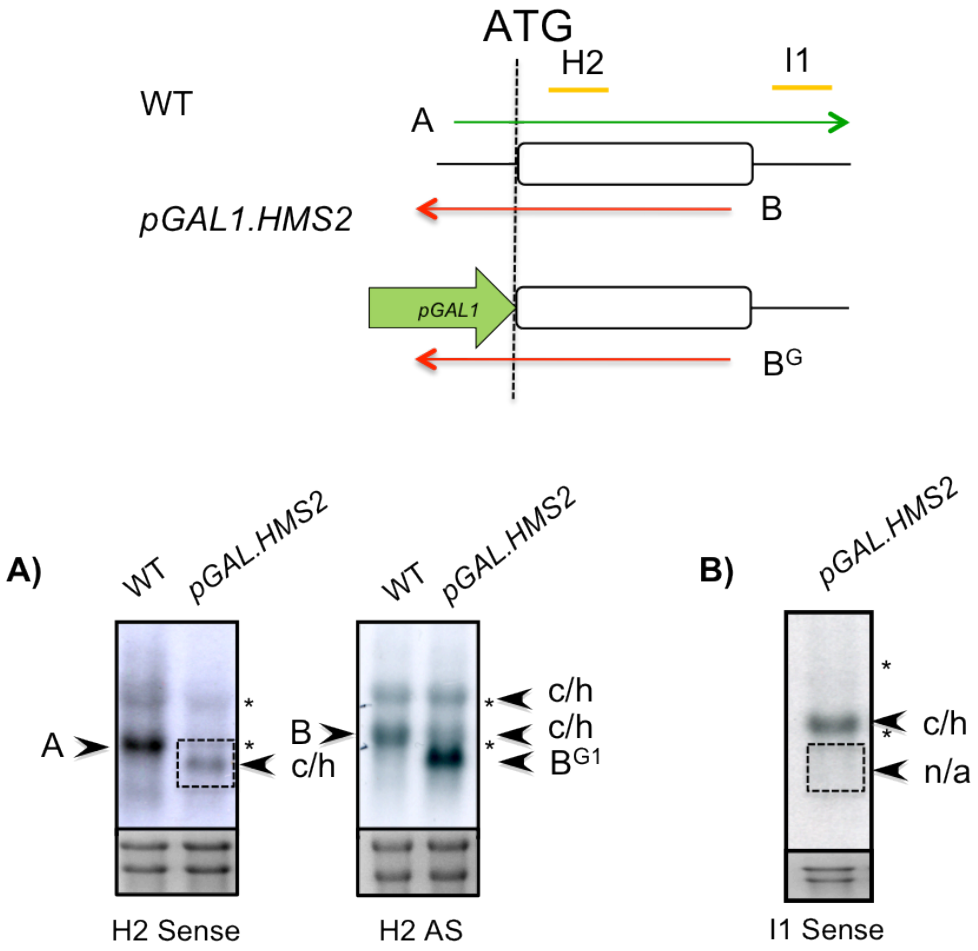


Figure 51. Replacement of the *HMS2* promoter with that of *GAL1* creates an antisense-dominant state in cells grown in glucose-containing medium.

Northern blot analysis of the galactose inducible *HMS2* strain, *pGAL1.HMS2* using **(A)** the H2 sense and antisense and **(B)** I1 sense probes (indicated by the yellow bars in the above schematic). In glucose conditions, insertion of the *GAL1* promoter (*pGAL1*) results in decreased levels of sense and concordant increased levels of antisense (transcript B^{G1}) in comparison to wild-type (transcripts A and B). The band detected in the *pGAL1.HMS2* H2 sense blot represents the sense probe cross hybridizing to the antisense transcript (c/h), as sense detection using a probe for the *HMS2-BAT2* intergenic region (probe I1) does not detect this band (n/a). Also, cross-hybridization to the 18S rRNA using probe I1 is reduced when RNA extracts are enriched for polyadenylated transcripts (See Figure 53). Using formaldehyde 1% agarose gels, 10 µg of total RNA was separated in each lane for 2.5 h at 100 V. Transcripts of interest are marked by black arrows and are depicted in the above schematic; in this representation, arrow thickness is not indicative of steady-state transcript levels. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below.

in *pGAL1.HMS2*, under repressing glucose conditions an *HMS2* antisense-dominant state was effectively recapitulated.

After shifting to galactose-containing medium, an antisense-dominant state was maintained until 1 h (Figure 52B, transcript B^{G1}/B^{G2}), with a concomitant increase in transcript levels of the downstream gene, *BAT2* (Figure 52C, transcript D). After 3 h growth in galactose-containing medium, the *GAL1* promoter was able to initiate a switch to a sense-dominant state (Figure 52A, transcript A^{G}). Additionally, the long *HMS2-BAT2* bicistronic transcript was also detected after 3 h growth in galactose-containing medium (Figure 52A, transcript C^{G}). Meanwhile, antisense transcript levels were almost completely abolished (Figure 52B, transcript B^{G1}/B^{G2}). Furthermore, as predicted by the model, *BAT2* transcript levels also exhibited a reduction in the *HMS2* sense-dominant state (Figure 52C, transcript D).

To determine whether antisense transcription had any effect on *pGAL1.HMS2* sense transcript induction rates, the antisense-less *hms2Δ::URA3* construct was placed under the control of the *GAL1* promoter. The sense transcript profile was interrogated after 1, 2, 2.5, and 3 h growth in galactose-containing medium using a probe for the *HMS2-BAT2* intergenic region (probe I1). For this experiment, RNA extracts enriched for polyadenylated RNA were used to reduce cross-hybridization and noisy background signal. No changes in induction rate were apparent, with the sense transcript only detectable after 2 h in both strains, signifying that antisense transcription does not

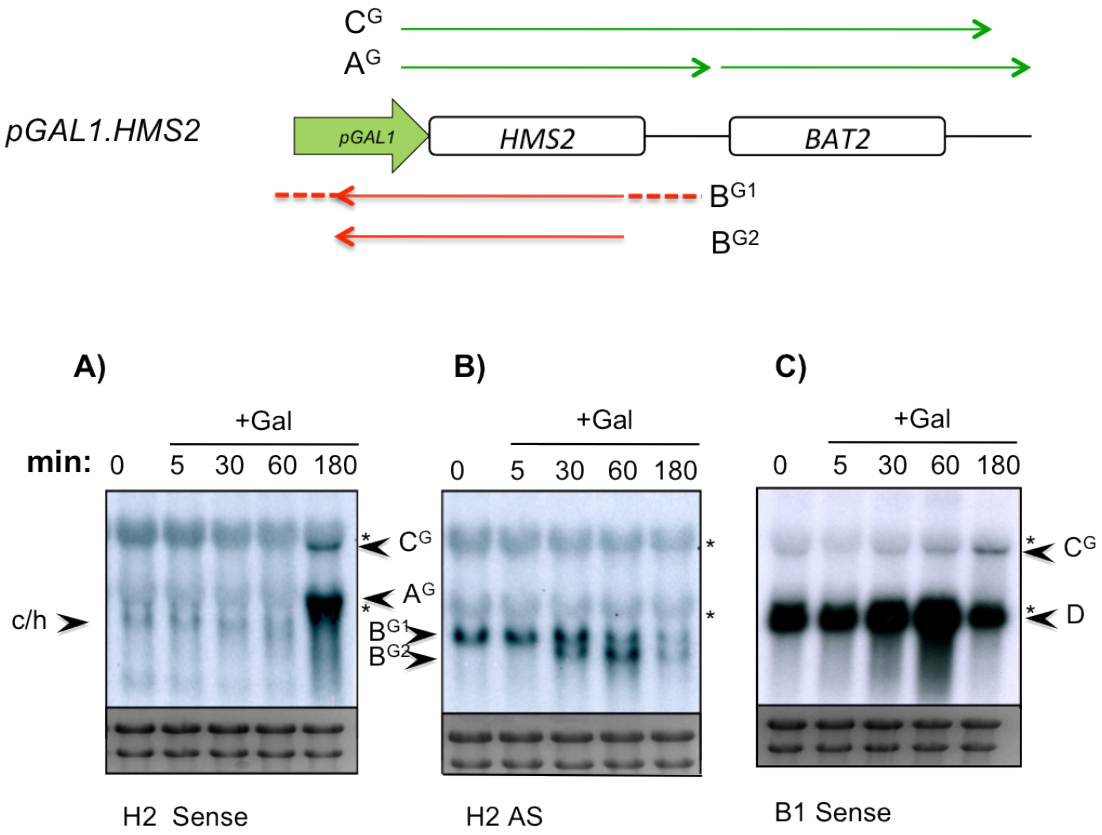


Figure 52. Induction profile of galactose-inducible *HMS2*.

Northern blot analysis of the *pGAL1.HMS2* **(A)** sense, **(B)** antisense and **(C)** *BAT2* transcript profile between 0 min (Glu) and 3 h (180 min) growth in galactose-containing medium. Upon galactose induction, an antisense-dominant state is maintained, with two antisense species being detected between 30 min and 1 h (60 min) (transcripts B^{G1} and B^{G2}), but with no detectable levels of sense after 1 h. Signal detected between 0 min and 1 h time points using the H2 sense probe are the result of cross hybridization with the antisense transcripts (c/h). After 3 h the *GAL1* promoter creates a sense-dominant state (transcript A^{G}), with antisense transcript levels almost completely abolished. Note the appearance of the *HMS2-BAT2* bicistronic transcript at 3 h (transcript C^{G}) and concomitant decrease in *BAT2* transcript levels (transcript D). Using formaldehyde 1% agarose gels, 10 µg of Total RNA was separated in each lane for 2.5 h at 100 V. Transcripts of interest are marked by black arrows and are represented in the schematic above; in this depiction, arrow thickness is not indicative of relative steady-state transcript levels. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. For full transcript map please refer to Figure 50.

impair sense activation (Figure 53, transcript A^{GU}). Rather, this finding is compatible with the proposed bi-modality of sense/antisense expression, such that without an antisense-dominant state, more cells would express the sense transcript. With transcription in the *pGAL1.HMS2* strain adhering to the model's predictions, the intermediate steps required for state-switching were investigated using this system.

5.2.2. Activator binding is required for maintaining a state switch:

According to the proposed model, signals integrated at either the 5' or 3' end of a gene decide a transcriptional state by driving either a sense or antisense transcript, respectively. Likewise, it was expected that the reverse would hold true: under permissive conditions, should communication between a state-determining signal and its target promoter be severed, a state-switch should ensue. Thus, the inducible *pGAL1.HMS2* system in conjunction with the Anchor Away technique (Haruki et al., 2008) was employed to deplete the nucleus of the Gal4p activator during the antisense-to-sense state switch. In keeping with the model, removal of Gal4p after activation of the *GAL1* promoter at 3 h abolished *HMS2* sense transcription (Figure 54A, transcript A^{G}) and an antisense-dominant state was recuperated (Figure 54B, transcript B^{G}). Levels of the long read-through transcripts (transcript C^{G} and C'^{G}) were also abolished with the addition of rapamycin, confirming that these transcripts initiate from the *GAL1* promoter (Figure 54A). These effects were specific to the removal of Gal4p, as depletion of TBP from the nucleus in order to disrupt PIC formation did not completely abolish sense transcription but resulted in a partial reduction in levels of transcripts A^{G}, C^{G}, and C'^{G} (Figure 55A). TBP depletion also did not result in recuperation of antisense transcript

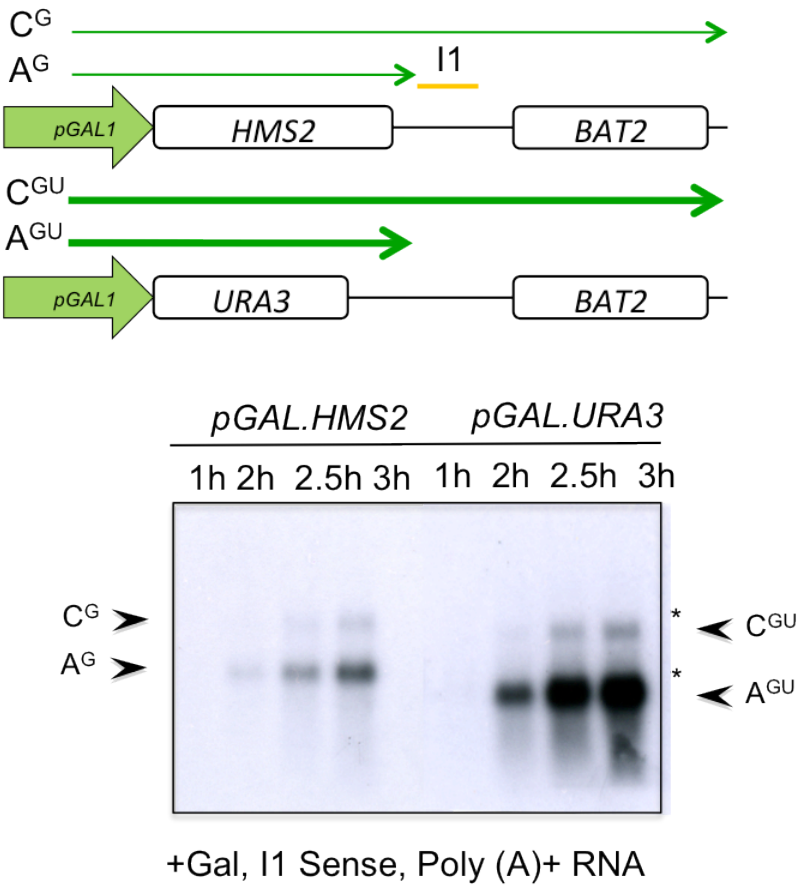


Figure 53. Induction profiles of the galactose inducible *HMS2* and *URA3* transcripts.

Northern blot analysis of the *pGAL1.HMS2* and *pGAL1.URA3* induction profiles between 1 h and 3 h growth in galactose-containing medium using a sense probe (I1) for the *HMS2-BAT2* intergenic region (marked in yellow). *pGAL1.HMS2* transcripts (A^{G}) are found at lower levels than *pGAL1.URA3* (A^{GU}) transcripts; however, activation for both strains is only detected after approximately 2 h of growth in galactose-containing medium. Using formaldehyde 1% agarose gels, Poly (A) + RNA was separated in each lane for 2.5 h at 100 V. Transcripts of interest are marked by black arrows and are represented in the schematic above; in this depiction, arrow thickness is indicative of relative steady-state transcript levels. The asterisks represent the approximate positions of the 25S (3392 nt) and 18S (1789 nt) RNAs. For full transcript maps please refer to Figure 50.

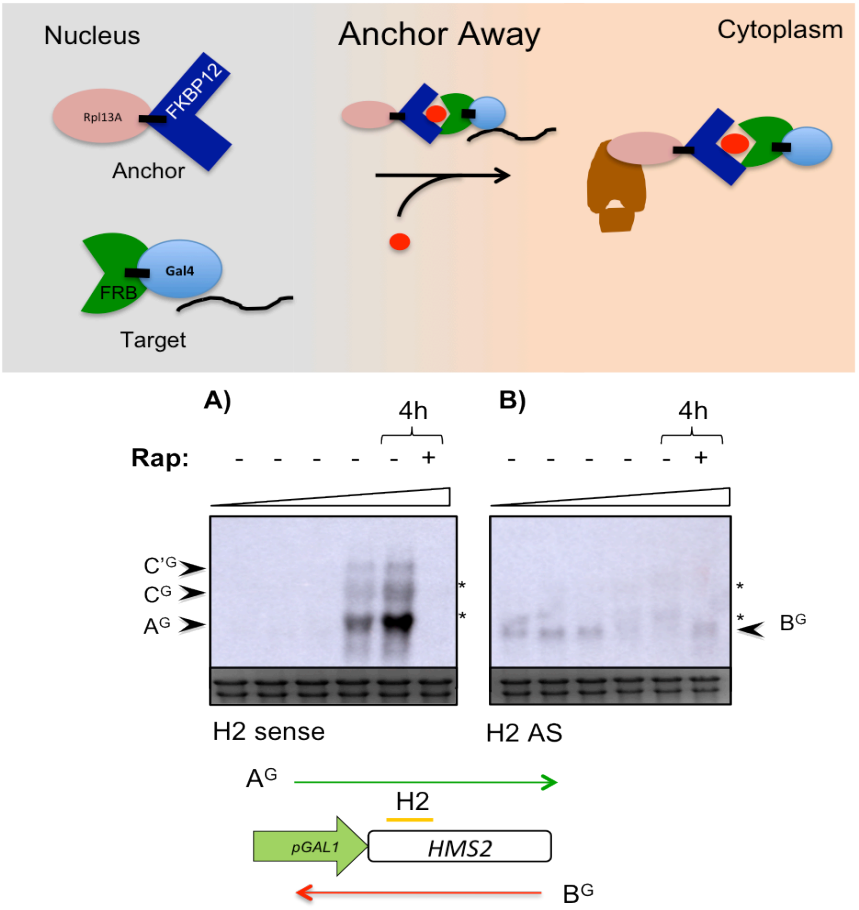


Figure 54. Induction profile of *pGAL1.HMS2* after Gal4p was depleted from the nucleus.

Northern blot analysis of the *pGAL1.HMS2* **(A)** sense and **(B)** antisense transcript profile after Gal4p was depleted from the nucleus using the Anchor Away technique (Haruki et al., 2008), represented by the schematic above. The *pGAL1.HMS2* construct was inserted into the Gal4-FRB/*RPL13A-FKBP12* background. As in the wild-type strain (BY4741), *pGAL1.HMS2* still demonstrates activation after 3 h of growth in galactose (transcript A^G). After 3 h, rapamycin was added to the growth medium in order to deplete Gal4p from the nucleus. In the control group DMSO was added. After 1 h treatment with rapamycin, level of sense transcripts A^G, C^G, and O^G were completely abolished (4 h, +Rap) in comparison to the control group (4 h, –Rap). In the antisense direction, levels of transcript B^G were recovered when Gal4p was depleted. Using formaldehyde 1% agarose gels, total RNA was separated in each lane for 2.5 h at 100 V. Probe location is shown in yellow and transcripts of interest, marked by black arrows on the blots, are shown in the schematics below; in this depiction, arrow thickness is not indicative of relative steady-state transcript levels. The asterisks represent the approximate positions of the 25S (3392 nt) and 18S (1789 nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. For full transcript maps please refer to Figure 50.

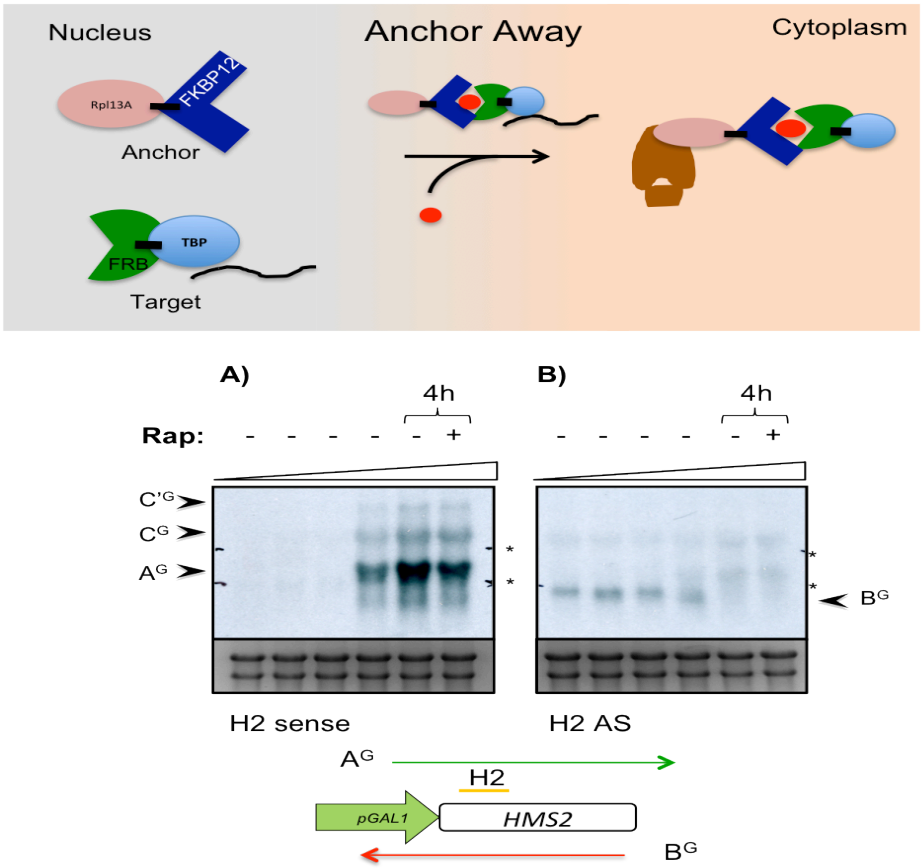


Figure 55. Induction profile of *pGAL1.HMS2* after TBP was depleted from the nucleus.

Northern blot analysis of the *pGAL1.HMS2* **(A)** sense and **(B)** antisense transcript profile after TBP was depleted from the nucleus using the Anchor Away technique (Haruki et al., 2008), represented by the above schematic. The *pGAL1.HMS2* construct was inserted into the TBP-FRB/*RPL13A-FKBP12* background. As in the wild-type strain (BY4741), *pGAL1.HMS2* still demonstrates activation after 3 h of growth in galactose (transcript A^G) and the production of longer read-through transcripts, C^G, and newly detected read-through transcript O^G. After 3 h, rapamycin was added to the growth medium in order to deplete TBP from the nucleus. In the control group DMSO was added. After 1 h treatment with rapamycin (4 h, +Rap), levels of all sense transcripts were reduced (transcript A^G, C^G, and O^G) in comparison to the control group (4 h, -Rap), while antisense transcript B^G was at undetectable levels in both groups. Using formaldehyde 1% agarose gels, total RNA was separated in each lane for 2.5 h at 100 V. Probe location is shown in yellow and transcripts of interest, marked by black arrows on the blots, are shown in the schematics below; in this depiction, arrow thickness is not indicative of relative steady-state transcript levels. The asterisks represent the approximate positions of the 25S (3392 nt) and 18S (1789 nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. For full transcript maps, please refer to Figure 50.

levels, perhaps reflecting the incapacity to form a proper PIC, which would inhibit the adoption of an antisense-dominant state (Figure 55B). Together the results from the Anchor-Away and *pGAL1.URA3* experiments demonstrate how activator binding is the principal driving force behind transcription and transcription kinetics; however, this does not rule out the potential for ancillary roles of antisense transcription in gene regulation.

5.2.3. Transcriptional state switching corresponds to changes in antisense transcript-end sites:

When the *pGAL1.HMS2* cells were moved from repressing to inducing conditions, an antisense-dominant state was maintained until after 1 h of growth in galactose-containing medium and the state switch was observed after 3 h. Further analysis of the *pGAL1.HMS2* induction profile provided evidence for the presence of two species of antisense transcript: a long species, dominant until 30 minutes after the shift to galactose-containing medium (transcript B^{G1}), and a shorter species (transcript B^{G2}), also detectable after 30 minutes of growth in galactose-containing medium (Figure 52B). This shorter species was detected at elevated levels after 1 h and constituted the dominant antisense transcript form at that time point. Initial 3' end RACE determined alternative polyadenlyation site selection as one cause for the observed size differential, with the long form terminating upstream of the Gal4p binding sites in the *GAL1* promoter and the shorter form terminating midway through the *GAL1* promoter, downstream of the Gal4p binding sites (Figure 56A). Curiously, a third species terminating midway between the upstream and downstream poly (A) sites was observed, but this form was

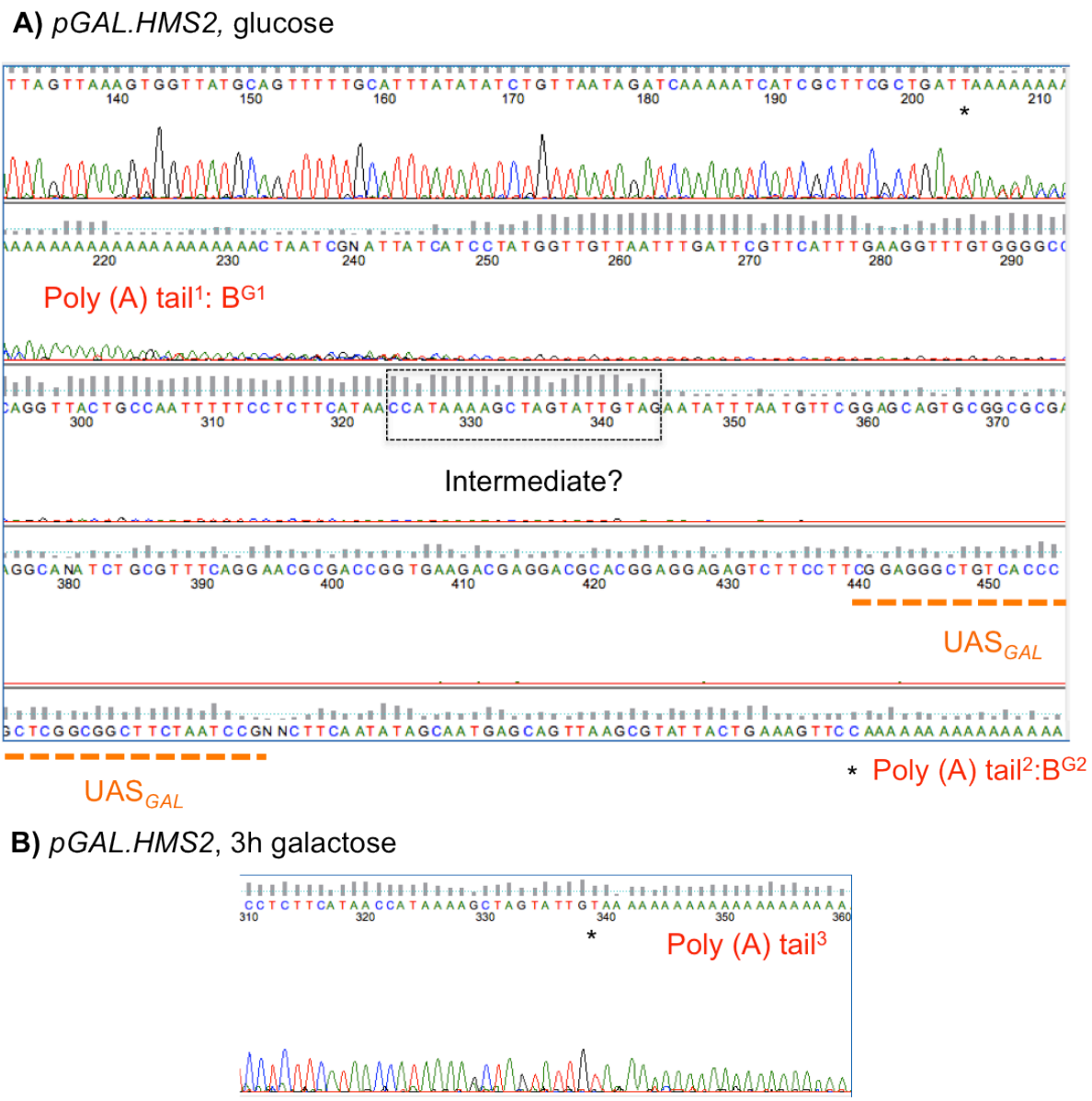


Figure 56. Mapping the 3' ends of the *pGAL1.HMS2* antisense transcript isoforms.

(A) 3' RACE performed on the galactose-inducible *HMS2* strain grown in glucose-containing medium displays evidence of two poly (A) tails, one consistent with the short, B^{G1}, AS isoform (Poly (A) tail[1]) terminating at a region downstream of the UAS_{GAL} (underlined in orange) and the long isoform, B^{G2}, terminating at a region upstream of the UAS_{GAL} (Poly (A) tail[2]). **(B)** Evidence for an intermediate species, B^{G3}, not resolvable using Northern blot analysis, was detected after 3' RACE was performed on cells grown in galactose-containing medium for 3 h. Potential end sites are marked with an asterisk.

irresolvable using Northern blot analysis (Figure 56B). Unfortunately, the 3' RACE technique proved to be intractable, potentially because of the high degree of transcript overlap in the interrogated area. Instead, short ribosomal RNA sequences were obtained in further trials (data not shown). Thus, these initial sequencing results, while noteworthy, could not be repeated due to technical failure despite multiple attempts.

However, to further gauge the contribution of these two antisense transcript bands with regards to state switching, a glucose repression experiment was conducted using the *pGAL1.HMS2* construct. When the state switch was inhibited by addition of glucose after 1 h growth in galactose-containing medium, levels of the long form of antisense transcript (transcript B^{G1}) were recovered whereas levels of the short form (transcript B^{G2}) were diminished to levels undetectable by Northern blot analysis (Figure 57). If the results from the 3' RACE are accurate, then taken together, the appearance of the short form of antisense transcript during galactose induction, and its subsequent disappearance during glucose repression, provides evidence for a relationship between early antisense transcript termination/polyadenlyation and antisense-to-sense state switching.

5.2.4. Antisense shortening is Gal4p-independent in pGAL1.HMS2 and is inherent to the HMS2 AS:

As previously discussed, the long and short antisense species observed in *pGAL1.HMS2* cells grown in galactose-containing medium terminated, notably, either upstream or downstream of the Gal4p activator binding sequences. In reference to early models of



Figure 57. Observing *pGAL1.HMS2* antisense transcript isoforms during glucose repression.

Northern blot analysis of the *pGAL1.HMS2* antisense transcript isoforms during glucose repression reveals that the long antisense form, B^{G1}, is associated with an inactive *GAL1* promoter, whereas the short form, B^{G2}, emerges during activation, but is lost when activation is prevented with glucose. Intermediate transcript isoform, B^{G3}, is not resolvable via Northern blot analysis. Using formaldehyde 1% agarose gels, total RNA was separated in each lane for 2.5 h at 100 V. Probe location is indicated by the yellow bar and transcripts of interest, marked by black arrows on the blot, and are shown in the above schematic; in this depiction, arrow thickness is not indicative of relative steady-state transcript levels. The asterisks represent the approximate positions of the 25S (3392 nt) and 18S (1789 nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below. For full transcript map, please refer to Figure 50.

transcription termination involving protein-bound "pause" elements, it was hypothesized that the observed 3' end antisense shortening was a function of Gal4p binding, which could act as a physical "roadblock" to repressive antisense transcription during galactose induction (Lang et al. 1994). To test this theory, the *pGAL1.HMS2* construct was created in a *gal4*Δ background. As expected, when Gal4p binding was perturbed, after 3 h no sense transcript was produced from the *pGAL1.HMS2* construct (Figure 58). If the previously observed antisense transcript shortening were indeed attributable to Gal4p binding to the *GAL1* promoter, then in the absence of Gal4p binding, no short transcript should be detected. Surprisingly, when *GAL4* was knocked out, the profile of the *pGAL1.HMS2* antisense transcript was identical to when Gal4p was available, indicating that antisense 3' end shortening occurs in response to environmental changes, but is independent of activator binding at the sense promoter (Figure 58, transcript B^{G1} and B^{G2}). While Gal4p binding was ruled out as the cause of alternative polyadenylation of the *HMS2* AS transcript, by effectively preventing sense activation in the *gal4*Δ strain, the consequence of overlapping sense transcription on the antisense promoter could be discerned. Strikingly, in contrast to the diminished antisense transcript levels observed when the *GAL1* promoter is properly activated (Figure 52B, transcript $B^{G1/G2}$), in the absence of the Gal4p activator, antisense levels were fully maintained after three hours of growth in galactose-containing medium (Figure 58, transcript B^{G2}). This observation establishes a direct link between sense promoter activation and antisense promoter repression. In light of evidence for transcription-mediated communication between regulatory elements in chapter 3, it is

likely that repression of the antisense promoter during sense activation is mediated through a transcription-dependent mechanism.

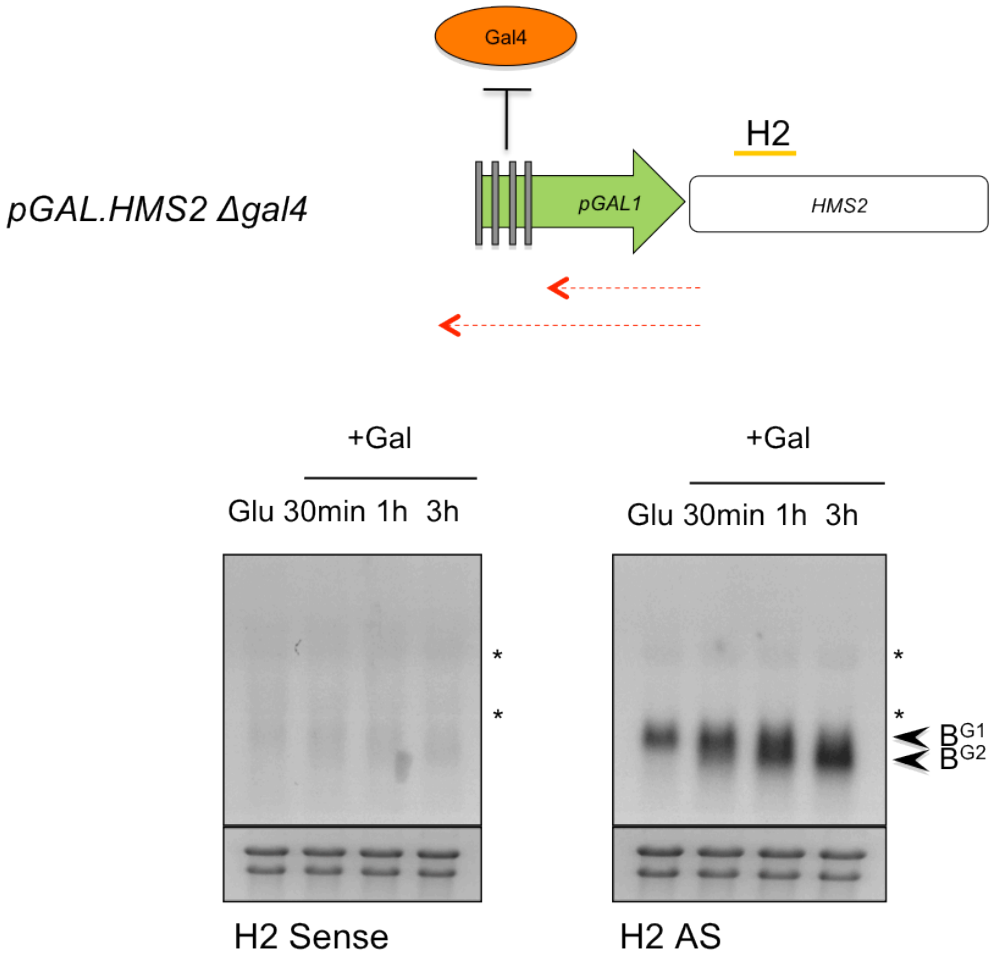


Figure 58. Observing *pGAL1.HMS2* antisense transcript isoforms in the absence of the Gal4p activator.

Northern blot analysis of the *pGAL1.HMS2* antisense transcript isoforms in the absence of Gal4p revealed that the short antisense form emerges regardless of activator binding (transcripts B^{G1} and B^{G2}). Additionally, after 3 h of growth in galactose-containing medium, antisense transcript was still detectable (transcript B^{G2}), in contrast to antisense levels in the original *pGAL1.HMS2* strain, which after 3 h are abolished. Using formaldehyde 1% agarose gels, Poly (A) + RNA was separated in each lane for 2.5 h at 100 V. Transcripts of interest are marked by black arrows. Probe location is indicated by the yellow bar in the schematic above. The asterisks represent the approximate positions of the 25S (3392 nt) and 18S (1789 nt) RNAs.

While initially confounding, the significance of transcript end heterogeneity as a general regulatory phenomenon began to emerge when data from previous chapters were re-examined. For instance, in Chapter 3 when ribosomal RNA was depleted from wild-type RNA extracts, multiple sense and antisense transcript species were resolved (Figure 16, transcripts A, A', B, B'); thus, transcript end heterogeneity is not a phenomenon unique to the *pGAL1.HMS2* construct. Similarly, depletion of rRNA from RNA extracts obtained from wild-type cells grown in glucose- and shifted to galactose-containing media, demonstrated that these isoforms were condition-specific, with the shorter form specifically manifesting in galactose-grown cells (Figure 59A, transcript B, B'). Additionally, upon closer inspection of Northern blot analyses depicting *HMS2* transcript levels in glucose versus galactose, antisense transcript shortening was apparent even in total RNA extracts (Figure 59B). In the total RNA blot represented, the juxtaposition of glucose- and galactose-grown wild-type and *URA3* cells demonstrated that this antisense shortening was not an artefact due to cross-hybridization of the antisense probe to ribosomal RNA. Thus, antisense transcript shortening is inherent to the *HMS2* antisense, and not a consequence of replacing the *HMS2* promoter with that of *GAL1*. However, with a roadblock model more or less ruled out, the source of 3' end heterogeneity and its implications, if any, on state switching remained at large. Given that antisense transcript 3' shortening occurs regardless of the nature of the 3' end sequences, an interesting possibility is that poly (A) site selection is mitigated through condition-specific factors binding to the 5' end antisense promoter; a perhaps equally attractive explanation for sense-end heterogeneity.

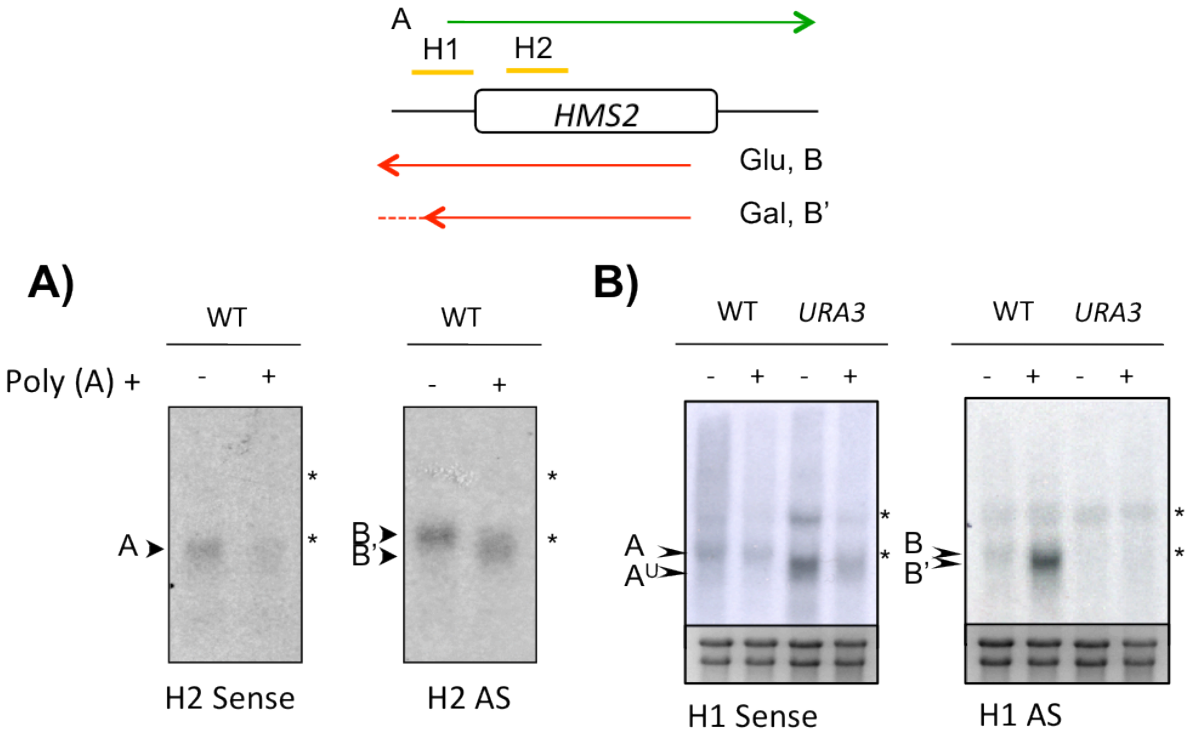


Figure 59. Antisense transcript shortening in galactose-grown cells is inherent to the *HMS2* antisense transcript.

(A) Northern blot analysis of polyadenylated *HMS2* wild-type and total *HMS2* wild-type and *URA3* RNA from cells grown in glucose- (-) and shifted to galactose-containing media (+) for 1 h demonstrated that antisense transcript shortening in galactose-grown cells is an inherent quality of the *HMS2* antisense transcript (transcripts B and B'). Lack of an antisense transcript band in the *URA3* strain demonstrates that the antisense transcript bands in wild-type cells are not the result of cross hybridization to 18S rRNA. Sense transcript shortening is not immediately apparent (transcript A). Using formaldehyde 1% agarose gels, Poly (A) + RNA or 10 µg of total RNA was separated in each lane for 2.5 h at 100 V. Transcripts of interest are marked by black arrows and depicted in the schematic. Probe locations are also indicated by the yellow bars in the schematic. The asterisks represent the approximate positions of the 25S (3392 nt) and 18S (1789 nt) RNAs. If applicable, ethidium bromide-stained ribosomal RNA loading controls are displayed below.

5.2.5. Chromatin remodelling and modification at the HMS2 locus:

Despite the ability to observe a state-switch in near real time, a mechanism for state switching remained ambiguous. However, in light of evidence from work on the *SRG1/SER3* and *IRT1/IME1* loci, the role of chromatin remodelling and histone modification in regulating the *HMS2* sense/antisense pair was investigated. By using the temperature-sensitive *spt6-1004* strain, the *HMS2* sense/antisense transcript profile was investigated at the permissive (30°C) and non-permissive (39°C) temperatures. Unlike *SER3*, which experienced derepression in the *spt6-1004* mutant, levels of the *HMS2* sense transcript remained relatively constant between growth at the permissive and non-permissive temperatures; however more read-through transcripts were detectable (Figure 60A).

The *spt6-1004* mutant had a striking effect on antisense transcript levels, which were completely abolished during growth at the non-permissive temperature (Figure 60B). Presumably this effect was due to an impaired ability to establish competent promoter architecture at the 3' end of *HMS2* as a result of Spt6p inactivation and/or increased read-through transcription repressing the promoter. So, while the *HMS2* antisense transcript does not appear to regulate the sense promoter by the same mechanism as *SRG1* regulates the *SER3* promoter, chromatin remodelling via Spt6p seems to play an important role in modulating both sense and antisense *HMS2* transcription.

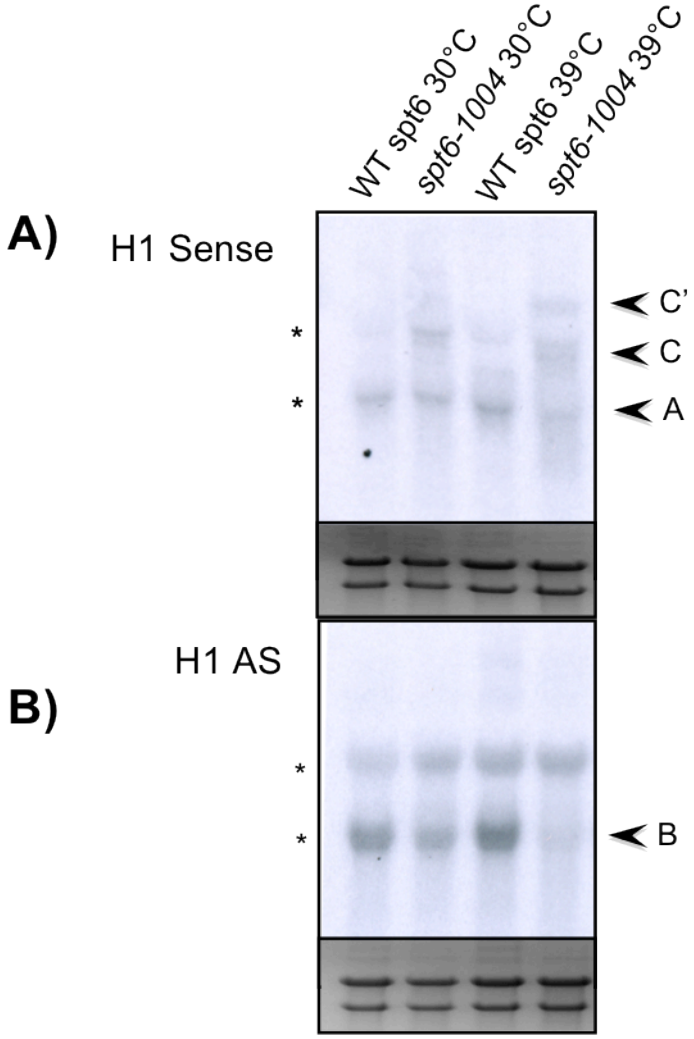


Figure 60. Inactivation of Spt6p does not result in derepression of the *HMS2* sense promoter.

Northern blot analysis depicting *HMS2* **(A)** sense (transcript A) and **(B)** antisense (transcript B) transcript profiles in wild-type and *spt6-1004* strains grown at the permissive (30°C) and non-permissive temperatures (39°C). Using formaldehyde 1% agarose gels, 10 µg of Total RNA was separated in each lane for 2.5 h at 100 V. Major transcripts are marked by black arrows. The asterisks represent the positions of the 25S (3392 nt) and 18S (1789 nt) RNAs. Ethidium bromide-stained ribosomal RNA loading controls are displayed below.

In chapter 3 it was postulated that the *HMS2* antisense transcript might define the boundaries of a transcription unit after depletion of TFIIB resulted in decreased levels of the antisense transcript, but increased detection of sense read-through transcripts (Figure 21). A similar transcript profile is evident in the *spt6-1004* mutant at the non-permissive temperature, providing further evidence that antisense transcription might help to define the margins of a transcription unit. While some of this read-through transcription could be attributed to perturbed nucleosome positioning as a result of the loss of Spt6p, thus resulting in compromised promoter and terminator chromatin architecture (which is consistent with the "complex transcription profile" hypothesis put forth in Chapter 4), increased read-through together with the loss of antisense transcription was also observed in the *URA3* strain in Chapter 4, in which the surrounding chromatin architecture and regulatory elements were undisturbed. Given Spt6's well-known functions as a histone chaperone and in modulating chromatin structure, an appealing explanation is that, rather than deleting the antisense TSS *per se*, incorporation of the *URA3* sequences created chromatin architecture at the 3' end of *HMS2* that was incapable of supporting antisense transcription.

5.2.6. Preliminary investigation of the HMS2 promoter chromatin architecture:

The H3K36me3 profile of the *HMS2* sense promoter was investigated in light of evidence that Set2-mediated H3K36 trimethylation of the *IME1* promoter, directed by *IRT1* intergenic transcription, could contribute to a repressive chromatin domain. H3K36me3 was examined by ChIP across the *HMS2* promoter in conditions of low antisense (glucose) and high antisense (galactose) transcription. In accordance with the van

Werven et al. (2012) study, *HMS2* displayed higher levels of H3K36me3 in conditions associated with higher levels of antisense transcript detection relative to conditions that yielded lower levels of antisense transcript detection (Figure 61). This trend held true for all primer sets spanning the *HMS2* promoter, however the occupancy of H3K36me3 decreased with distance upstream from the *HMS2* ATG. While more work should be done to discern whether or not increased H3K36me3 results in a repressive chromatin state at the *HMS2* promoter (presumably through Rpd3S-mediated deacetylation), thus causing impaired transcription factor binding, this preliminary analysis provides the basis for a mechanism similar to that described at the *IME1* locus. Importantly, however, earlier investigation of Set2 regulation of the *HMS2* locus revealed that there was no effect on the relative levels of *HMS2* sense and antisense transcripts when *SET2* was deleted, once again indicating that the regulatory forces functioning at the *HMS2* locus are distinct from previously published phenomena (Michael Youdell, data not shown). Thus, additional investigation into how, if at all, these chromatin states play into state-switching should be pursued further.

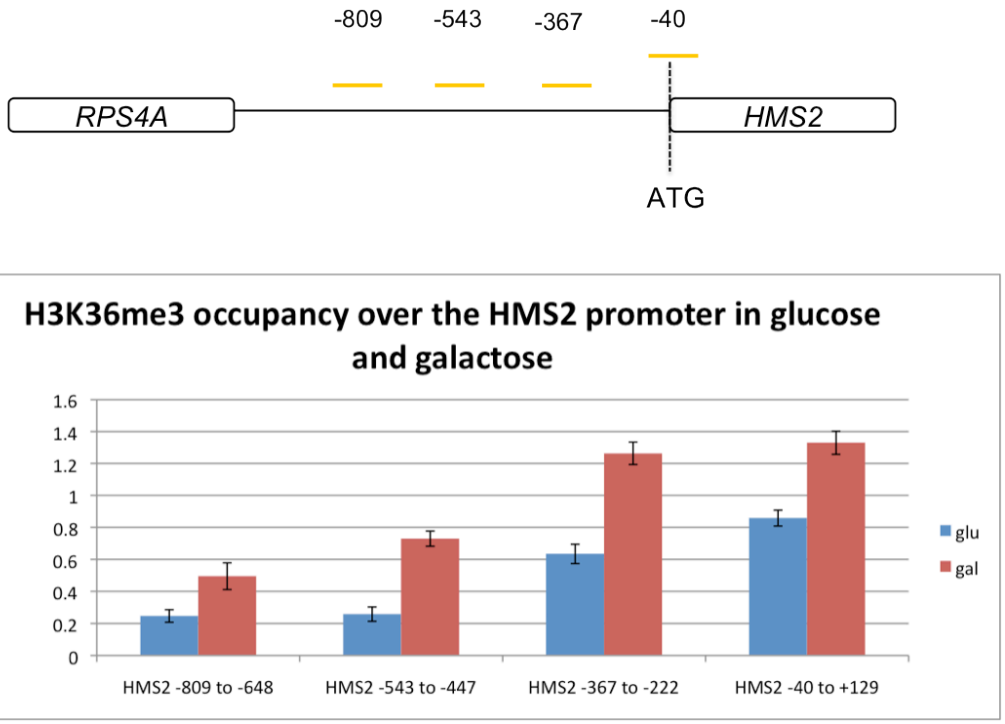


Figure 61. Distribution of H3K36 trimethylation across the *HMS2* promoter in glucose- and galactose-grown cells.

Relative occupancy of H3K36 trimethylation across the *HMS2* promoter in glucose (glu) and galactose (gal). Approximate probe locations are depicted in the above schematic of the *RPS4A-HMS2* intergenic region, with probes annotated according to the furthest distance from the *HMS2* ATG. The percent input of H3K36me3 is normalized to percent input of histone H3, which is then normalised to H3K36me3/H3 ratios in the *TUB2* coding region to control for potential genome-wide changes in H3 occupancy in the shift from glucose- to galactose-containing media. Error bars represent the compounded standard error of three real-time PCR reactions.

5.3. Discussion:

Experiments in this chapter revealed that while *trans*-acting factors are responsible for fixing a state, it is the resultant transcription that engenders the interdependency that exists between sense and antisense promoters. Furthermore, it was demonstrated that environmentally determined transcript 3' end micro-heterogeneity could be a potential means by which cells could deliberately manage the transcription interference caused by transcriptional overlap. Taken together the evidence gathered here emphasizes how a complex network of overlapping transcription is not arbitrary but can be systematically regulated at multiple stages to achieve a specific and biologically relevant transcriptional landscape.

5.3.1. Contributors to "plasticity" or high expression variability:

Identification of transcript 3' end heterogeneity in the *pGAL1.HMS2* strain signified an additional layer of complexity in an already crowded transcriptional landscape. Moreover, the apparent environmental responsiveness of the different isoforms indicated the potential for their distinct functionality. Preliminary 3' RACE studies of the antisense transcript isoforms also identified that the changes in transcript length were likely a consequence of alternative poly (A) site selection, in a manner similar to the *HMS2-BAT2* bicistronic transcript, which was shown to evade the *HMS2* poly (A) site and utilize the downstream *BAT2* poly (A) site. It was predicted that this shortening was associated with activator binding, since the alternative ends were either up- and downstream of the Gal4p activator binding sites; however, early polyadenylation and transcript shortening occurred regardless of Gal4p binding, making a roadblock model

unlikely. Even though what exactly provoked these changes in poly (A) site selection remains unclear, in any case, production of the isoforms resulted in varying degrees of overlap of the sense or antisense promoters, perhaps affording the cell the plasticity to either produce or overcome the repressive effects of overlapping transcription via highly-directed transcription interference and/or associated chromatin remodelling. Furthermore, the finding that the *HMS2* antisense transcript shortening occurs independently and prior to activation of the sense promoter, perhaps signifies that cells might anticipate state switching, and will pre-emptively manage overlapping transcription to facilitate the imminent switch.

The transcript isoforms observed in this study espouses a model in which transcript 3' end micro-heterogeneity can accommodate ultra-fine tuning in the genetic response. Thus, though there may be two states defined by sense/antisense transcription, the transcriptional aftermath of these states can differ depending on the transcript isoform produced. In accordance with this finding, previous studies have shown that antisense transcription contributes to high expression variability (Xu et al., 2011); however, exactly how overlapping transcription contributed to this plasticity remained uncertain. Here it has been shown that, in concert, overlapping transcription and its consequent transcriptional states as well as accompanying transcript 3'end heterogeneity provides cells with endless substrates for variability, or 'transcriptional potential', so that they can tailor their cellular response to their specific needs during cellular programmes or a vast array of unpredictable changes in the environment.

Chapter 6.

Discussion.

6. Discussion

Transcription is an inherently random event, contingent on the chance encounter of *cis*-acting regulatory elements and their corresponding molecular factors. So, in order to elicit a timely and appropriate genetic response in the face of ever-changing environmental and developmental requirements, cells must implement several measures of control, all feeding back onto each other in order to execute specific, highly organized gene expression profiles. The body of work presented here characterizes non-coding overlapping transcription, an ostensibly random and chaotic genomic feature, as a specialized arrangement coordinated by cells with clear biological consequences. In the last chapter, antisense contributions to the expression variability of a locus were discussed in light of new evidence that transcriptional overlap was remarkably specialized. Here, the state-switching model is reviewed but moreover, evidence from this study is interpreted holistically and the role of antisense transcription in a wider network of overlapping transcription is discussed. What emerges are two interrelated roles in regulating genomic space- and time-structures: 1) antisense transcription contributions to dissociating expression patterns of genes in close spatial proximity and 2) antisense-mediated reduction in expression noise in oscillating transcription programmes.

6.1. Reviewing the state-switching model:

A systematic dissection of the transcriptional landscape at the *HMS2* locus was performed with the aim of converging on a comprehensive model for sense/antisense relationships. The evidence strongly supported that sense/antisense expression was at odds, existing in a mutually antagonistic relationship. At a single cell level, it was found that sense/antisense overlap created an axis of "transcriptional exclusion", where convergent transcriptional events demonstrated a degree of temporal segregation; hence, the discrete sense and antisense transcription "states". From these results precipitated the state-switching model, which intended to reconcile the findings from this study with various observations related to overlapping transcription such as the architectural/organizational features of a compact, interleaved genome as well as ensuing locus interdependency. Using this model, a timeline or hierarchy of events was resolved (Figure 62). First and foremost, transcription factors such as Gal4p are united with their regulatory elements, determining a state, either sense- or antisense-dominant. The resultant transcription is intrinsically regulatory, manifesting in consequences for neighbouring genes by either interfering with a promoter, such as *HMS2* sense transcription antagonizing downstream *BAT2* expression, or alternatively, acting to abate transcription interference, such as *HMS2* antisense transcription shielding *BAT2* expression from repressive *HMS2* sense transcription. Finally, transcriptional overlap is managed within tens of base pairs in the face of changing

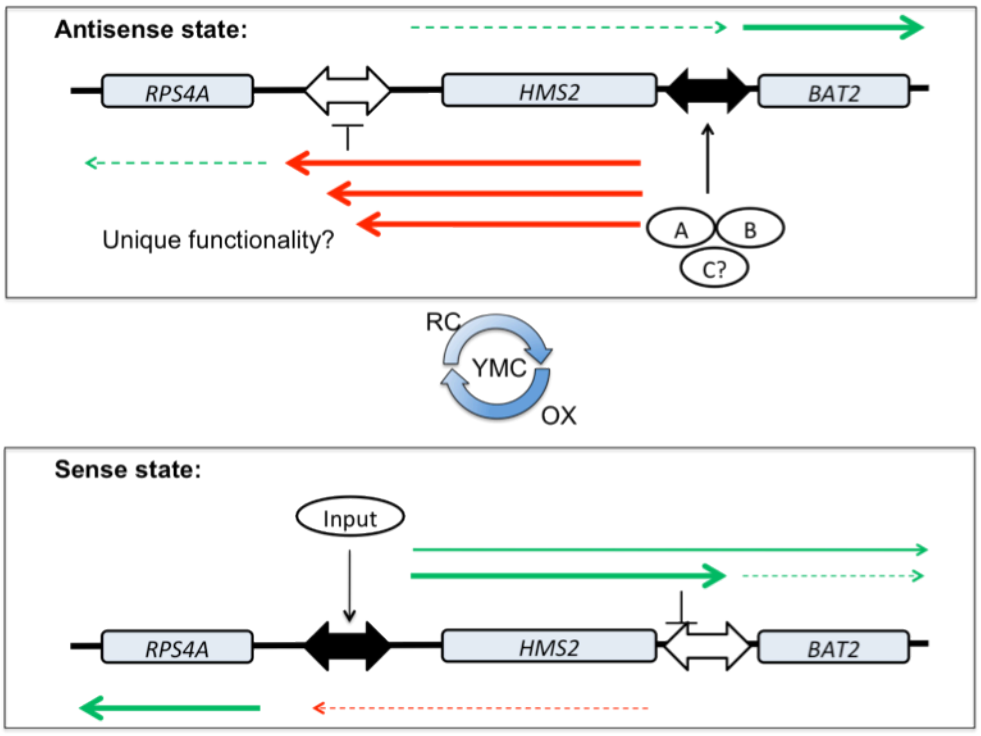


Figure 62. Revising the state-switching model.

As previously discussed, overlapping transcription results in two transcriptional states due to transcriptional exclusion. Environmental signals can be integrated at either the 5' or 3' ends of antisense-producing genes. The resultant transcription from either the sense or antisense promoter can drive the production of interfering transcripts that can act as "long range" regulators, affecting neighbouring genes. These state-deciding transcription factors may also predict poly (A) site usage, determining the degree and by extension, the consequences of transcriptional overlap.

environmental conditions, suggesting that the effects of overlap are highly directed and under cellular control. After enduring multiple tests for robustness and predictive power (Chapters 4 & 5), here the implications of this model in a broader biological context are discussed.

6.2. Sense/antisense state switching exists as a part of a larger network of interlinked overlapping transcription, making antisense-related phenotypes difficult to discern:

In terms of basic structural considerations, the model postulates that antisense-producing transcription units are necessarily flanked by two regulatory regions, or promoter-like features, capable of initiating transcription. As an isolated unit, it was thought that adopting this type of "dual-promoter" arrangement would likely confer greater flexibility to a locus in terms of responding to environmental stimuli, since regulatory signals could be integrated at either end of the gene. Thus, by virtue of transcriptional exclusion, an activator binding to the 3' end of a gene might function as a repressor of the 5' promoter and vice versa. Accordingly, in Chapter 3 it was demonstrated that at *HMS2* many of the known and projected transcription factor binding sites are found at the 3' end of the gene, suggesting that in striking contrast to traditional views on transcription regulation, a gene can be regulated via its 3' end by conventional means of *trans*-acting factors. Indeed, this conjecture is consistent with work by Murray and colleagues (2012), who demonstrated that at the 3' ends of antisense-producing units, the machinery for transcription initiation is present at levels comparable (~60%) to the 5' end.

The architecture of antisense-producing units is but one example of how conventional, linear models of genomic design principles restrict the understanding of gene regulation. For instance, in this analysis, multiple cases of multi-function promoters/terminators were documented (Chapter 4). It was also demonstrated that

transcription units are not predictably defined by two regulatory elements in tandem, since termination or poly (A) signals could be selectively circumvented to produce long read-through transcripts, such as the *HMS2-BAT2* bicistronic transcript, as well as read-through transcripts seen in the *HMS2.kan*, *HMS2.ADH*, *URA3* (Chapter 4) and *pGAL1.HMS2* strains (Chapter 5). It was put forward that these transcripts, whose generation intersected regulatory regions, could elicit inhibitory effects through transcription interference, reinforcing a second very important observation: functions of neighbouring regulatory elements are coupled through the act of transcription. For example, in Chapter 4, it was determined that by manipulating the strength of the *TEF* promoter and drastically reducing the levels of *kan*^{r} RNA, the downstream *TEF* terminator exhibited impaired capacity for bidirectional termination but potentially acquired the ability to initiate transcription. Furthermore, in Chapter 5, it was clearly demonstrated that the shut down of the *HMS2* antisense promoter in the *pGAL1.HMS2* strain was contingent on the activation of transcription from the *GAL1* promoter, as when *GAL4* was deleted to prevent *HMS2* sense induction, antisense levels were maintained in what would normally be non-permissive conditions.

Together, the evidence suggests that the interleaved architecture of the yeast genome not only lends itself to high degrees of transcriptional overlap, but more importantly, an equally high degree of transcription unit interdependency by co-opting transcription itself as a regulatory mechanism. Unfortunately, the biological implications of the functional connectedness displayed by interleaved, versatile regulatory elements would

be difficult to measure. Because a single element responding to environmental cues might not only regulate its gene, but also impact neighbouring units through long-range transcription, any calculated perturbation of the transcriptional circuitry is likely to have far-reaching indirect effects that might not manifest as clear phenotypes, but as subtle shifts in the balance of the transcriptional landscape, as seen in the 5' ^{P}*TEF* Δ strain in Chapter 4.

That is not to say that these nuances in transcription profiles would not bear any functionality. Since regulatory elements act as the interface between the environment and the ensuing genetic response, it is likely that the transcriptional overlap of these elements affects their environmental sensitivity via some physical means such as altering DNA accessibility through co-transcriptional nucleosome remodelling or histone modification, as described at the *SRG1/SER3* and *IRT1/IME1* loci, respectively. Indeed, high levels of *HMS2* antisense transcription in galactose-grown cells were correlated with increased levels of H3K36 methylation at the *HMS2* promoter; though, the downstream effects of this increased H3K36me3 occupancy have yet to be determined. What was evident from this analysis, however, was that cells carefully manage transcriptional overlap even at a small scale, implying that its effects are highly specific. This was best captured in Chapter 5, where the detection of controlled antisense transcript 3' end heterogeneity revealed a potential mechanism through which cells could micro-manage transcriptional overlap of the sense promoter in the face of environmental changes, presumably to modify the responsiveness of the overlapped

regulatory element. Given these findings, a reasonable prediction then, would be that the biological consequence of these switches in the transcriptional landscape or, more specifically, overlap would not be evident in cells growing in steady-state conditions but made apparent when cells must respond to shifts in environmental conditions.

6.3. Potential roles for overlapping transcription units in establishing relationships between genome spatial- and temporal-structures:

Overall, the results of this study have underscored the relevance of genomic organization and gene order to gene regulation, as a gene's location amidst pervasive transcription would persuade its transcriptional activity. While it has become increasingly evident that gene order in eukaryotes is non-random (Hurst et al., 2004), the basis for genomic organization remains poorly understood, and this is largely without consideration of non-coding transcription units. It has been proposed that co-functional genes might cluster together so that they may be co-expressed in "open" chromatin domains; for instance, genes in the same protein complexes (Teichmann & Veitia, 2004), or those belonging to the same functional pathway (Lee & Sonnhammer, 2003). Still, these studies have considered clusters on a large scale, and concede that functionally unrelated genes intersperse these clusters, leaving the logic behind gene order still unresolved.

On a more localized scale, the anti-correlated relationship between *HMS2* and *BAT2* expression stands in stark contrast to a large body of work that has shown that neighbouring genes, regardless of orientation and co-functionality, are more likely to be

co-expressed because of shared chromatin domains (Batada et al., 2007; Cohen et al., 2000; Ebisuya, Yamamoto, Nakajima, & Nishida, 2008; Hershberg, Yeger-Lotem, & Margalit, 2005). For instance, Cho and colleagues (2000) demonstrated that ~25% of adjacent yeast genes demonstrate co-expression in the mitotic cell cycle. Despite their adjacency, *HMS2* and *BAT2* do not display co-expression in the cell cycle (Granovskaia et al., 2010), metabolic cycle (Tu et al., 2005) or when grown in galactose-containing medium (Chapter 3). Further Gene Ontology (GO) analysis also demonstrates that *BAT2* and *HMS2* are not functionally related; however, synteny of *HMS2* and *BAT2* is maintained in three related species of budding yeast (Figure 63). While purely anecdotal, it is possible that this indicates an alternative selective advantage to maintaining adjacency between *HMS2* and *BAT2*, and perhaps their accompanying overlapping non-coding units.

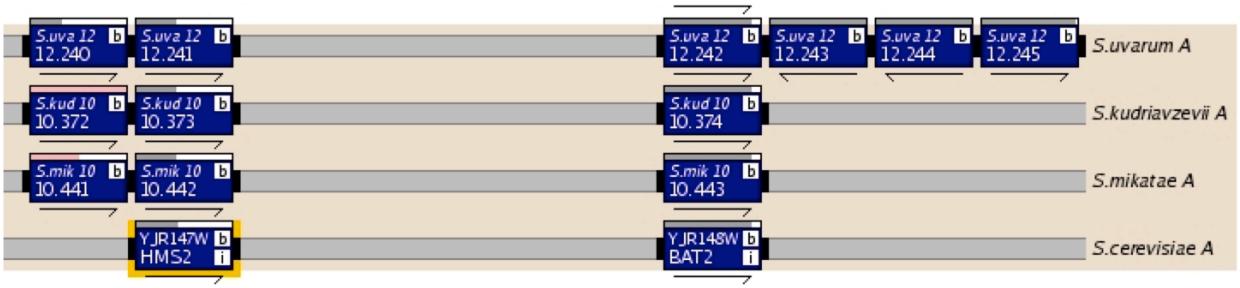


Figure 63 . Synteny of HMS2 and BAT2 in three related species of budding yeast.

Synteny of *HMS2* and *BAT2* is maintained between three related species of budding yeast: *S. uvarum A, S. kurdriavxevii A, S. mikatae A, and S. cerevisiae A.* Blue boxes represent genes, the first number from the left represents the chromosome on which the gene in located. Arrows above the gene box indicate directionality. *HMS2* in S. cerevisiae is highlighted in yellow. The image is taken from the Yeast Genome Order Browser, version August 2012 (http://ygob.ucd.ie/).

For instance, one tenable explanation for the departure from the expected relationship between *HMS2* and *BAT2* is that antisense transcription dissociates the transcription of

genes, which might normally possess a propensity for co-expression as a result of their proximity. It is also plausible that neutralizing these position effects is accomplished by creating highly localized chromatin domains, in a manner similar to the nucleosome deposition described at the SER3 locus (Hainer et al., 2011). Therefore, it is likely that the selective advantage in maintaining *HMS2* and *BAT2* adjacency lies not with the co-functionality of their protein products, but rather, the lack thereof, such that the transcription of non-coding RNAs over the gene cluster favours deliberate dissociated expression of two functionally unrelated genes. While the non-coding RNA profiles for the *HMS2* gene cluster would have to be observed across related yeast species in order to fully corroborate this theory, subjective evidence for this claim is apparent in terms of their gene biology, as *HMS2* and *BAT2* appear to have opposing functions relating to nitrogen availability, with *HMS2* likely to be optimally expressed in conditions of low nitrogen, and *BAT2* in conditions of high nitrogen, making contemporaneous expression improbable. More substantially, as previously discussed, *HMS2* and *BAT2* expression are compartmentalized in separate phases of the yeast metabolic cycle (Tu et al., 2005). So while not biologically related per se, *HMS2* and *BAT2* expression are in a sense linked through the same biological oscillator. By and large, the transcription-mediated locus interdependency described between *HMS2* and *BAT2* might be representative of a general phenomenon in which genomic space structures, including the organization of non-coding units, are intrinsically coupled to the temporal dimensions of gene regulation, such as transcription periodicity within the metabolic cycle.

6.4. Assessing the role of overlapping transcription in biological oscillatory networks:

It has been well documented that transcription takes place in the context of biological oscillators; for instance, circadian rhythms, the mitotic cell cycle, and the metabolic cycle are all interrelated physiological programmes that relate to phase-specific transcription patterns (Gillette & Sejnowski, 2005). In yeast, studies have shown that most genes display some degree of expression periodicity (~57%) within temporal windows defined by metabolic biology (Tu et al., 2005); however, in spite of its prevalence, whether non-coding transcription also demonstrates periodicity and what contributions this might have to periodic regulation of ORFs have been largely overlooked. All things considered, the *HMS2* locus provided a convenient system through which potential roles of antisense transcription in metabolic cycling could be elucidated. The *HMS2* and *BAT2* loci display clear periodicity in the metabolic cycle, with *HMS2* exhibiting expression in the HOC oxidative phase and *BAT2* in the LOC reductive charging phase. Given the co-expression displayed by the *HMS2* antisense and *BAT2* sense transcripts in this study, it was posited that they would also be co-expressed in the reductive charging phase. Assuming this logic was correct, a biological function for antisense transcription in metabolic transcription networks was pursued.

In Chapter 4, a role for antisense transcription as a chaperone of robust transcription periodicity in the metabolic cycle was deduced after elimination of antisense transcription overlap of the sense promoter resulted in apparent digression of the *HMS2*

sense transcript expression from the oxidative phase. It was further proposed that antisense transcription acted to obstruct out-of-phase sense transcription by stabilizing the "off" state of the sense promoter, potentially through transcription interference-mediated chromatin remodelling. Consistent with these conjectures, a genome-wide survey of the top 1% high amplitude, metabolically oscillating genes revealed that half of these genes possessed antisense transcripts. Additionally, the spatial and temporal relationships between *HMS2* and *BAT2* are not unique. For instance, *TKL2*, which is involved in the non-oxidative phase of the pentose-phosphate pathway (Schaff-Gernstensclager, Mannhaupt, Vetter, Zimmerman, & Feldmann, 1993) and is expressed in the reductive phase of the YMC, is found adjacent to *LYS2*, which is involved in lysine biosynthesis, and is expressed in the oxidative phase of the YMC (Figure 64A). In glucose- and galactose-grown cells, *TLK2* and *LYS2* display opposite expression patterns in a manner similar to the relationship between *HMS2* and *BAT2* (Figure 64B). Notably, *TKL2* also has an antisense pair (SUT018) that is expressed reciprocally in glucose- and galactose-grown cells. Furthermore, *TKL2* is in the top 1% of high amplitude periodic genes, so it is possible that its antisense functions not only to dissociate *TKL2* expression from *LYS2*, but also to insulate *TKL2*'s "off" state outside of the reductive charging phase. The striking similarity of *TKL2* regulation to that of *HMS2* suggests that the principles underlying the phenomena described here are perhaps not isolated to the *HMS2* locus, but representative of a more general scheme.

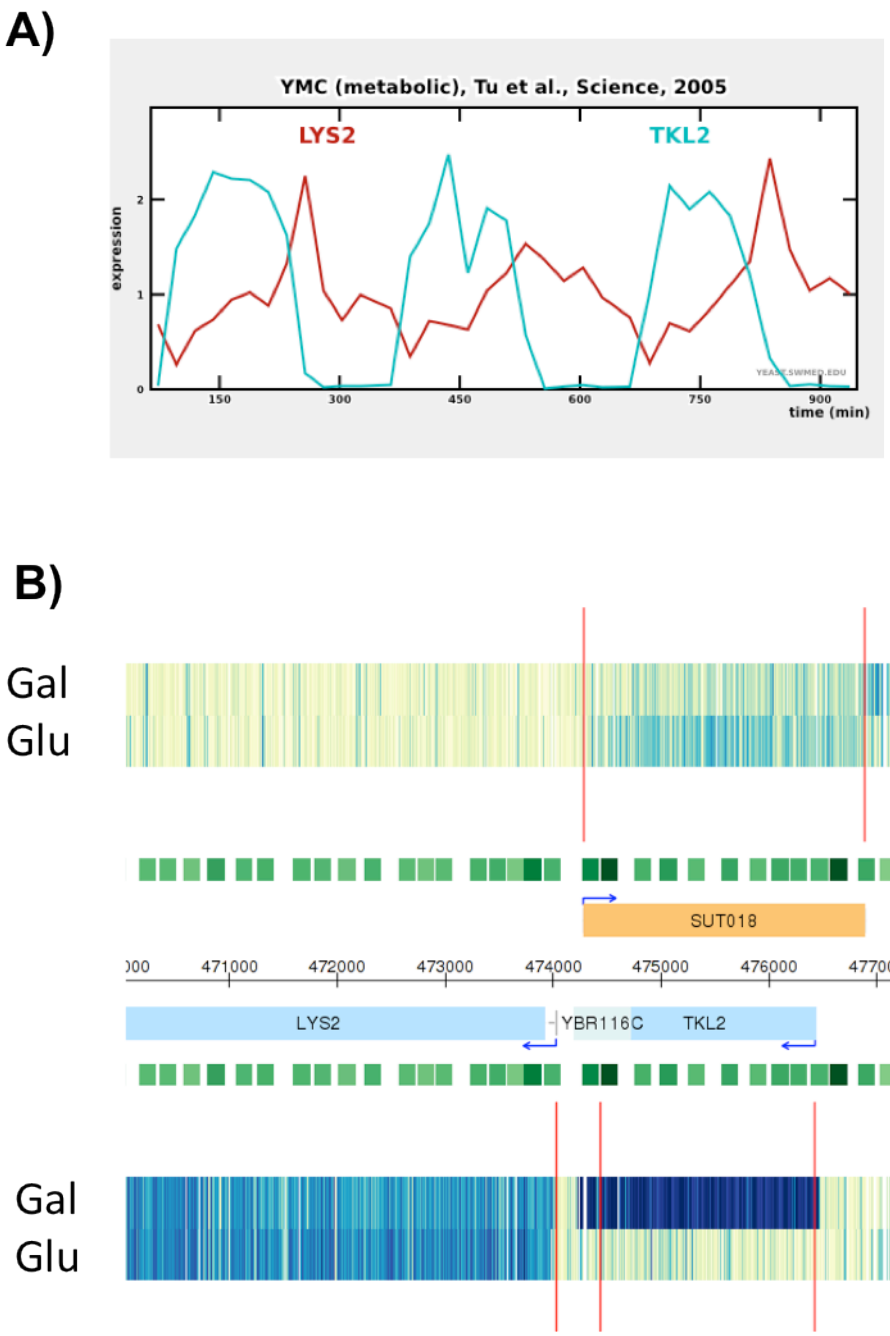


Figure 64. The relationships described at *HMS2* are not unique to the gene cluster.

The regulation of the *TKL2* locus bears striking similarity to *HMS2*. **(A)** Its expression is segregated from its downstream gene, *LYS2*, in the metabolic cycle and **(B)** it is also an antisense producing unit. Polyadenylated RNA purified by Ana Serra Barros) was hybridized to strand-specific high resolution tiling arrays in collaboration with Lars Steinmetz at the European Molecular Biology Laboratories.

The biological rationale for why cells might necessitate these additional levels of fine-tuned control is made apparent when the advantages of biological time-structures are considered. Firstly, temporal compartmentalization of phases with distinct cellular ecologies is vital to maintain cellular integrity; for instance, cell division is tightly linked with growth and oxygen consumption in the metabolic cycle, so that cells may only divide when they have achieved an appropriate level of biomass, and DNA synthesis in the RB phase is temporally sequestered from the generation of reactive oxygen species during respiration (Tu et al., 2005; Klevecz et al., 2004). Indeed, perturbation of the YMC expression circuitry has been demonstrated to yield deleterious effects on spontaneous mutation rate (Chen et al., 2007). Additionally, since transcriptional events are sensitive to changes in the stoichiometry of intracellular biomolecules, it would be advantageous to systematize transcription into temporal phases in which the expression of biologically-relevant genes coincides with optimal intracellular concentrations of their specific transcription factors, such that unproductive or potentially detrimental transcription reactions can be avoided. In view of the incentives to reduce noise within these chronological systems, it is likely that cells impose multiple degrees of control to ensure their maintenance. By reinforcing "off" switches through transcription interference, and anticipating changes through 3' end heterogeneity, it is possible that in organisms with interleaved genomes, the presence of antagonistic overlapping transcription has evolved over time to act as safeguards of these biological rhythms.

On the whole, the findings of this analysis demonstrate that cells may have adapted to the inherent incompatibility between processivity and pervasive transcription by temporally partitioning transcriptional overlap and appropriating its antagonistic effects to preserve biological clocks. Although antisense transcription as a sentinel of metabolic periodicity is an attractive suggestion that brings together the state-switching model and a general biological function, in order to fully corroborate this theory, work that maps sense transcription to specific metabolic phases, as done by Silverman and colleagues (2010), should be performed in the presence and absence of antisense overlap.

6.5. Extending the model- antisense-mediated post-transcriptional gene regulation:

Much of the work presented in this analysis focused on the consequences of transcribing antisense RNAs, when in fact, many of the regulatory effects of antisense production could be post-transcriptional. RNA species, whether coding or non-coding, undergo several post-transcriptional/co-transcriptional processing events that are tightly coupled to transcription, including the addition of the 5' 7 methylguanosine cap, the 3' poly (A) tail, and splicing of introns. Additionally, RNA is co-transcriptionally incorporated into ribonucleoprotein particles (RNP), which consist mainly of RNA export factors. These export factors demonstrate tight coupling with RNA surveillance, 3' end formation, and curiously, the initiation machinery, underscoring the feedback loops and interdependency that exist between activation, processing, and export (Vinciguerra & Stutz, 2004). One of the first studies to demonstrate these relationships without the use

of temperature-sensitive mutants, showed that perturbation of RNA processing events of the human β-globin genes, such as splicing and 3' end formation, resulted in deficient RNA export, such that mutant RNAs were retained near the site of transcription, persisting as punctate, intranuclear foci even after transcription was "shut off" (Custódio et al., 1999). This was in contrast to wild-type nuclear foci, which demonstrated rapid dissociation after transcription was inhibited (Custódio et al., 1999).

A subsequent series of studies on the *S. cerevisiae GAL10-1* cluster further elucidated the nature of these regulatory relationships and brought to light the influence of nuclear organization on different aspects of gene expression regulation. For instance, two studies by Casolari and colleagues (2004) and Cabal and colleagues (2006) demonstrated that activation of the *GAL* genes resulted in their association with the nuclear pore complex (NPC) and localization of the activated *GAL* genes to the nuclear periphery in a process called "tethering" (Cabal et al., 2006; Casolari et al., 2004). Additionally, Cabal and coworkers (2006) revealed that perinuclear confinement of the *GAL* genes during activation was dependent on components of the SAGA co-activator, Sus1p and Ada1p; the nuclear export factor Sac3p, which physically interacts with Sus1p; and the Sac3p-interacting nuclear pore complex component, Nup1p, epitomizing the interrelationships between activation, processing, and export. Furthermore, using *GAL* reporter constructs, Abruzzi and colleagues (2006) showed that sequences directing 3' end formation mediated the association of genes with the nuclear periphery during activation. Moreover, retention of RNA in nuclear foci, or "dots", was also dependent on

3’ end formation signals, with transcripts containing the *GAL1* or *TDH3* PAS manifesting as either nuclear foci, or diffuse signal, respectively (Abruzzi, Belostotsky, Chekanova, Dower, & Rosbash, 2006). The Abruzzi et al. (2006) study also revealed that RNA within these nuclear dots was largely post-transcriptional (i.e. not nascent). Like the human β-globin genes, the persistence of these dots after repression was dependent on 3’ end formation, with poly (A)- RNA dots exhibiting slower decay than the poly (A)+ dots, perhaps reflecting differences in either turnover or export to the cytoplasm (Abruzzi et al., 2006). Interestingly, both the poly (A)- and poly (A)+ dots maintained peripheral tethering post-repression, revealing that sustained transcriptional activity was not necessary for perinuclear confinement (Abruzzi et al., 2006). This result complemented earlier findings by Cabal and colleagues (2006), who demonstrated at the endogenous *GAL* locus that even when localization to the nuclear periphery was abolished with deletion of *ADA1*, intranuclear transcriptional activation could still take place unabated. Furthermore, evidence provided in a study by Green and colleagues (2012) revealed that *ADA1* mutants still exported *GAL1-GFP* tagged transcripts efficiently, indicating that export is not affected when genes are expressed from intranuclear locations (Green, Jiang, Joyner, & Weis, 2012). Evidence demonstrating that peripheral localization is seemingly independent of transcription and/or activation and unnecessary for efficient export, calls into question the popular “gene-gating” hypothesis, which asserts that active genes are found in the vicinity of NPCs in order to facilitate nuclear export (Blobel, 1985). Indeed, the Green et al. (2012) study has shown that the endogenous *GAL* genes display perinuclear localization in inducing (galactose) and non-

inducing/derepressed (raffinose/glycerol) conditions. Interestingly, ~45% of cells demonstrated peripheral localization of the *GAL* locus even in repressing conditions (glucose). Furthermore, Green and colleagues (2012) demonstrated that when interaction with the nuclear periphery was disrupted, activation was unaltered (as in Cabal et al., 2006) but kinetics of glucose repression were delayed, suggesting that perinuclear confinement acts to enhance repression of the *GAL* genes when cells are moved from inducing to repressing conditions, rather than to aid in activation. Biologically speaking, this is coherent with the premise that cells must be able to rapidly induce the *GAL* genes upon galactose exposure, but be equally poised for quick repression in the presence of glucose, the preferred carbon source (Green et al. 2012). Taken together, these studies of the *GAL* genes illustrate the role of dot RNA and perinuclear tethering in gene regulation.

In terms of a mechanism, which is at present, poorly understood, a study by Vodala and coworkers (2008) indicated that *GAL1* dot RNA was established by Rrp6p-mediated production of unadenylated transcripts accumulating near the *GAL1* gene and that this dot RNA, through interacting RNP proteins, resulted in perinuclear tethering. Deletion of *RRP6* resulted in an increase in post-transcriptional polyadenylated *GAL1* 3' ends, a decrease in dot RNA levels, and deficient perinuclear tethering (Vodala, Abruzzi, & Rosbash, 2008). A recent unrelated study by Schmid and colleagues (2012) implicated the nuclear exosome in poly (A) tail length determination, as loss of Rrp6p led to extension of poly (A) tails by the non-canonical poly (A) polymerase, Trf4p (Schmid et al.,

2012). Furthermore, the study demonstrated that poly (A) tails are differentially loaded with the poly (A) tail binding proteins (PABP) Pab1p and Nab2p, which have been shown to determine correct poly (A) tail length. The authors revealed that poly (A) tails are preferentially bound by Pab1p, since Rrp6p interaction with Nab2p displaces Nab2p from the tails, resulting in either Rrp6p-mediated decay or poly (A) tail trimming (deadenylation) of Nab2p-bound transcripts. Considering these additional processing/surveillance steps encountered by Nab2p-bound/Rrp6-trimmed transcripts, it is possible that these transcripts might display delayed export kinetics and thus nuclear retention. In terms of perinuclear tethering, RNA surveillance by the nuclear exosome has been associated with the NPC and Nab2p has been shown to target mRNPs to the nuclear pores for export (reviewed in Tutucci & Stutz, 2011). Thus, taken together, the evidence suggests that the interplay between positive Nab2 export signals counteracted by Rrp6-mediated deadenylation and/or decay might act in establishing a tether between a gene locus and the nuclear periphery.

One aspect not considered in the *GAL* localization studies was the presence of non-coding RNAs at the locus. So while sense or coding transcription might have been inactive in repressed or non-induced conditions, production of non-coding RNAs is rampant across the locus. Since non-coding RNAs are subject to many of the same forces of transcriptional regulation, co-transcriptional modification, and post-transcriptional surveillance as coding RNAs, it is likely that they are also subject to the same regulatory influences imposed by nuclear localization. In particular, given the

shared genomic space exhibited by sense/antisense pairs, 3D nuclear organization that positively affects transcription in one direction, might negatively affect transcription from the opposite strand, contributing to the switch-like relationship described in this thesis.

In light of evidence demonstrating that poly (A)+ RNA fractions appear enriched for the *HMS2* sense but not for the *HMS2* antisense transcript in comparison to total extracts (Figure 66, Appendix), it is possible that the repressive *HMS2* antisense transcripts (i.e. the long isoforms) are dependent on a Nab2p/Rrp6p-mediated 3' end processing pathway that results in the deadenylation and retention of *HMS2* antisense transcripts in perinuclear RNA dots, which act to stabilize the repressed sense state established by antisense transcription, presumably by maintaining repressive chromatin domains in conjunction with the NPC (Figure 65A). These long-form RNA dots would be expected when founding an antisense state (e.g. the reductive charging phase in the YMC). Under antisense-to-sense state-switching conditions or conditions in which a state-switch is anticipated, alternative 3' ends are chosen (i.e. transcript end micro-heterogeneity) that render the transcripts insensitive to Nab2p/Rrp6p-mediated deadenylation. Indeed, the differential deadenylation of the long and short isoforms might explain why it was difficult to detect the long transcript ends via 3' RACE, a technique dependent on the presence of an adequately long poly (A) tail. These adenylated transcripts experience faster export kinetics and are not retained in the nucleus, presumably because they are bound by Pab1p (Figure 65B). For instance, crude analysis of antisense transcripts in

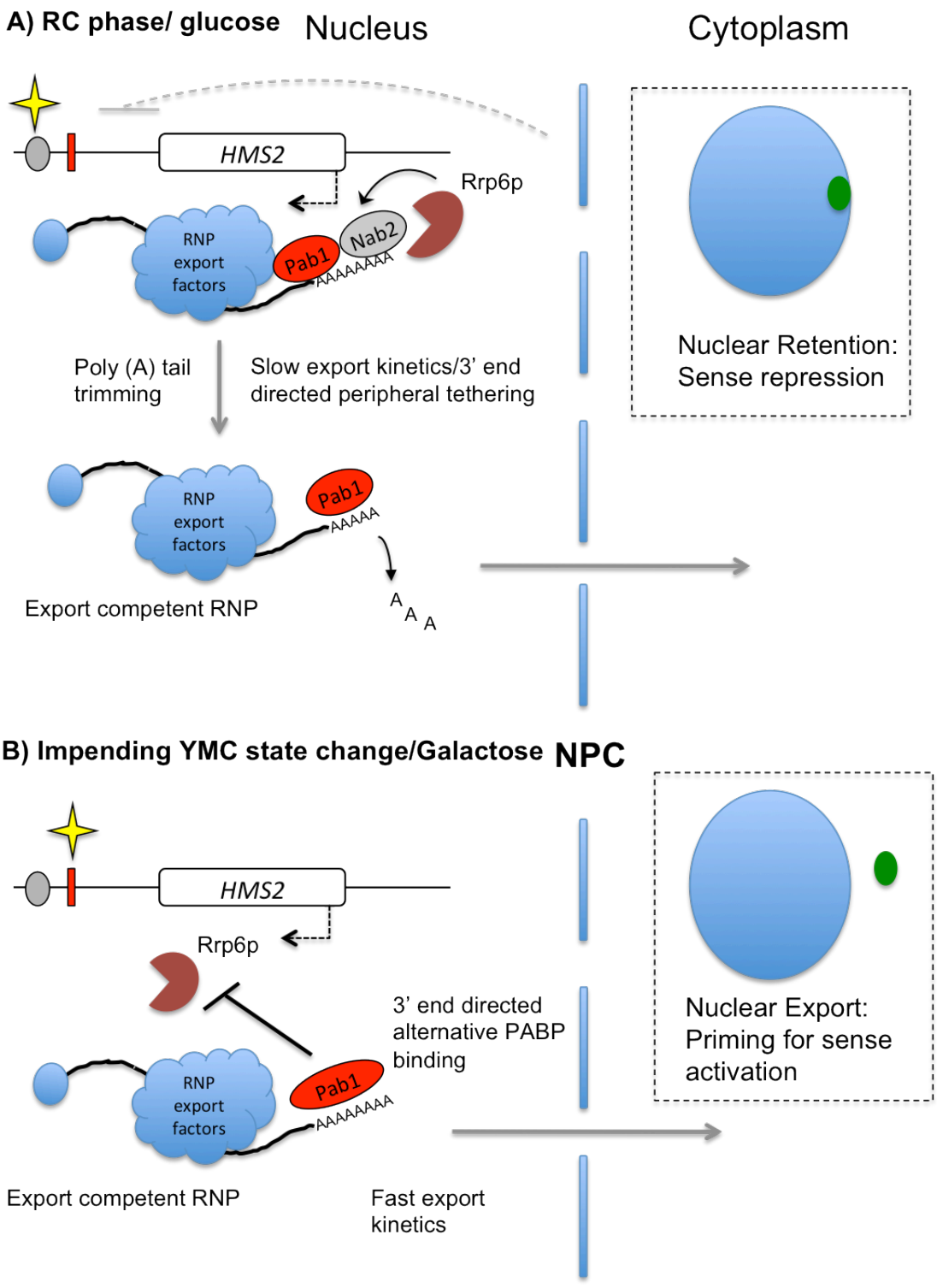


Figure 65. A potential role for perinuclear antisense RNA dots in mediating elasticity in sense/antisense state switching.

(A) In glucose, the preferred carbon source, cells in the RC phase of the YMC demonstrate formation of Rrp6p-dependent, perinuclear *HMS2* antisense RNA dots that

contribute to maintaining a repressed sense state. **(B)** When a state switch is imminent (e.g. at the end of the RC phase), or when cells are grown in unfavourable conditions, antisense RNAs are expressed, but adopt alternative 3' ends. Use of alternative ends affects downstream processing and export, abolishing the formation of antisense dots near the repressive NPC, allowing swift adoption of sense *HMS2* transcription when appropriate. This model brings together aspects of 3D nuclear organization, with the 3' end heterogeneity, alternative polyadenylation, and sense/antisense state-switching observed in this study. Yellow stars indicate the end site used. End sites are coloured (grey or red) according to the preferred bound PABP. Pab1p and Nab2p PABPs are indicated in red and grey, respectively. Rrp6p is represented in garnet.

glucose- and galactose-grown cells indicates that *HMS2* antisense transcripts are retained in the nucleus in glucose, but not after 1 h growth in galactose-containing medium. Loss of perinuclear dots may result in a 3D localization and/or chromatin context that is more amenable to a switch to sense transcription. In terms of biological practicality, by organizing sense/antisense states in domains in the nucleus, the cell can stabilize transcriptional landscapes under the appropriate conditions (e.g. sense repression in the RC phase), an idea consistent with antisense transcription acting as a chaperone of periodicity. Additionally, these relationships aid cells in eliciting the appropriate response to unfavourable conditions (e.g. galactose) but allow rapid reversion to a favourable state (e.g. glucose) as described in Green et al. (2012). Thus, while sense/antisense state-switching provides a cell with transcription variability, in conjunction, 3D organization of antisense transcripts in nuclear dots might additionally provides the cell with "elasticity" between states, such that it can respond swiftly to adopt the appropriate transcriptional landscape in highly precarious environmental conditions.

6.6. Other future work:

While this analysis has shed light on potential biological roles of antisense transcription in the metabolic cycle, evidence for this relationship was by and large anecdotal. Further experimental work must be done to assess the cyclical nature of all types of non-coding RNAs, genome-wide in order to fully understand their relevance to these biological systems. Additionally, this report focused mainly on the transcriptional effects of non-coding antisense transcription; however, given the interrelatedness between the different aspects of gene expression, including processing, nuclear localization, export, etc., it is important to broaden the scope of the analysis to consider how antisense transcription might influence each of these points of control. Most of all, the work presented here has used a multilateral approach to describing non-coding phenomena. Given the caveats each of the techniques has on their own, it is important that all future work be carried out with the same depth in order to obtain the full extent of non-coding transcription contributions to gene regulation.

6.7. Perspectives:

While this study has been predominantly conducted using a single locus approach in yeast, the implications of this work could still be far reaching. For instance, establishing wider, biologically-relevant transcriptional landscapes through sense/antisense state-switching could be pertinent to other aspects of chronobiology such as circadian rhythms and the mitotic/meiotic cell cycles, but also other developmental processes in higher eukaryotes such as genomic imprinting, where it is essential to institute and maintain transcriptional landscapes across chromosomal domains in a parent-of-origin-

dependent manner, or more broadly, in maintaining pluripotency and/or differentiation in multi-cellular organisms. Additionally, the results demonstrating the contributions of genomic spatial-structures to locus interdependency and temporal dimensions in gene expression could bear consequences on the expression and resultant phenotypes of transgenes.

6.8. Concluding remarks:

In summary, the findings of this study contribute to a more complete understanding of sense/antisense relationships, with the state-switching model bringing together multiple facets of what was initially a disparate and sometimes competing collection of observations. Furthermore, this work places sense/antisense relationships in a broader biological context and sheds light on how they might work as the nexus in more extensive networks of transcription that are organized in both spatial and temporal dimensions. How antisense relationships to 3D nuclear organization fit into establishing biologically-relevant transcriptional profiles remains a central but open question. Most of all, the multi-directional approach applied in this work comprises the most in-depth dissection of sense/antisense relationships to date, without which the breadth of understanding would have been incomplete. Given the revised knowledge of the complexity of transcriptional regulation, future work should be considerate of these intricacies and apply the same if not more comprehensive tactics.

Chapter 7.

Appendix.

7. Appendix

Strain Name	Parent Strain	Genotype	Origins
URA3.kan *[1]	*URA3*	*ura3 Δ (486-495)::kanMX6*	This Study
pGAL1.HMS2.ADH	*HMS2.ADH*	*pHMS2::HisMX6-pGAL1*	This Study
URA3.pGAL1	*URA3*	*tHMS2::pGAL-His3mx6*	This Study
pGAL1.HMS2 *[2]	W303	*pHMS2::HisMX6-pGAL1*	This Study
HHY189	W303	*MATα tor1-1 fpr1::NAT RPL13A-2×FKBP12::TRP1 SPT2-FRB::KanMX6*	Euroscarf
pGAL1.HMS2 Spt20-FRB	HHY189	*pHMS2::HisMX6-pGAL1*	This Study
Gln3-FRB	HHY168	*Gln3-FRB::KanMX6*	This Study
pGAL1.HMS2 RPL13A-2×FKBP12::TRP1 *[3]	HHY168	*pHMS2::KanMX6-pGAL1*	This Study
pGAL1.HMS2 *[4]	FY2181	*pHMS2::KanMX6-pGAL1*	This Study
pGAL1.HMS2 *[4]	FY2180	*pHMS2::KanMX6-pGAL1*	This Study

Table 7. Strains of interest not described in this thesis.

*1 The *URA3* construct was disrupted in the same manner as *HMS2* using the *kanMX6* cassette, in the absence of the *ADH1* terminator. Interestingly, in contrast to *HMS2.kan* no antisense transcript was detected initiating from the *TEF* promoter in *URA3.kan* (data not shown). Since it was estimated that the antisense start site for transcript F^K resided in the *HMS2* sequences, it is possible that the region upstream of the *kanMX6* insert in *HMS2* are capable of supporting antisense transcription from the ^{P}*TEF*, whereas *URA3* sequences upstream of the insert are not.

*2 The *pGAL1.HMS2* construct in the W303 background demonstrates the same transcription induction kinetics as the construct in the BY4741 background

*3 This construct was made to facilitate *FRB* tagging of proteins of interest in a *pGAL1.HMS2* background.

*4 These strains have not been examined by Northern analysis; however, using these strains it may be possible to look at the effects of the *spt6-1004* mutant on inducible *HMS2*.

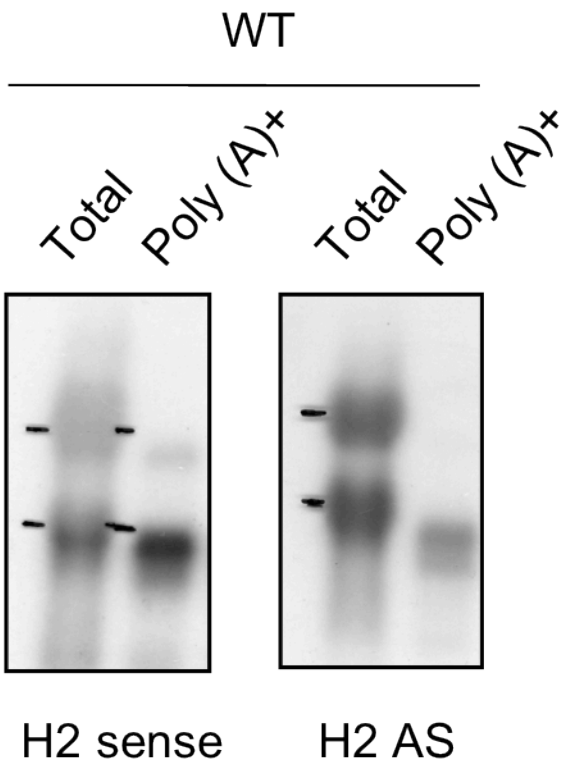


Figure 66. Evidence for enrichment of *HMS2* sense but not antisense transcripts in extracts enriched for polyadenylated RNA.

In comparison to total RNA extracts, *HMS2* sense transcripts display a slight enrichment for poly (A) + RNA, when ribosomal/poly (A)– RNA was depleted from the extracts. Polyadenylated *HMS2* antisense transcripts do not display the same enrichment relative to total RNA.

8. Bibliography

Abruzzi, K. C., Belostotsky, D. a, Chekanova, J. a, Dower, K., & Rosbash, M. (2006). 3'-end formation signals modulate the association of genes with the nuclear periphery as well as mRNP dot formation. *The EMBO journal*, *25*(18), 4253–62. doi:10.1038/sj.emboj.7601305

Adhya, S., & Gottesman, M. (1982). Promoter Occlusion : Transcription Through a Promoter May Inihibit Its Activity. *Cell*, *29*(3), 939–944.

Akhtar, M. S., Heidemann, M., Tietjen, J. R., Zhang, D. W., Chapman, R. D., Eick, D., & Ansari, A. Z. (2009). TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Molecular cell*, *34*(3), 387–93. doi:10.1016/j.molcel.2009.04.016

Aravind, L., Watanabe, H., Lipman, D. J., & Koonin, E. V. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *PNAS*, *97*(21), 11319–24. doi:10.1073/pnas.200346997

Arigo, J. T., Eyler, D. E., Carroll, K. L., & Corden, J. L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Molecular cell*, *23*(6), 841–51. doi:10.1016/j.molcel.2006.07.024

Axelrod, J. D., Reagan, M. S., & Majors, J. (1993). GAL4 disrupts a repressing nucleosome during activation of GAL1 transcription in vivo. *Genes & Development*, *7*(5), 857–869. doi:10.1101/gad.7.5.857

Bai, L., & Morozov, A. V. (2010). Gene regulation by nucleosome positioning. *Trends in genetics : TIG*, *26*(11), 476–83. doi:10.1016/j.tig.2010.08.003

Basehoar, A. D., Zanton, S. J., & Pugh, B. F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, *116*(5), 699–709. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15006352

Batada, N. N., Urrutia, A. O., & Zhang, M. Q. (2007). Chromatin remodelling is a major source of coexpression of linked genes in yeast. *Trends in genetics : TIG*, *23*(10), 480–484. doi:10.1016/j.tig.2007.08.007

Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., Rinn, J. L., et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science*, *306*(5705), 2242–6. doi:10.1126/science.1103388

Birney, E., Stamatoyannopoulos, J. a, Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799–816. doi:10.1038/nature05874

Birse, C. E. (1998). Coupling Termination of Transcription to Messenger RNA Maturation in Yeast. *Science*, *280*(5361), 298–301. doi:10.1126/science.280.5361.298

Blobel, G. (1985). Gene gating: a hypothesis. *PNAS*, *82*(24), 8527–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=390949&tool=pmcentrez&rendertype=abstract

Brantl, S. (2002). Antisense-RNA regulation and RNA interference. *Biochimica et biophysica acta*, *1575*(1-3), 15–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12020814

Brosius, J. (2005). Waste not, want not- transcript excess in multicellular eukaryotes. *Trends in genetics : TIG*, *21*(5), 287–288. doi:10.1038/nature03016

Bryant, G. O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., & Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. *PLoS biology*, *6*(12), 2928–39. doi:10.1371/journal.pbio.0060317

Buck, M. J., & Lieb, J. D. (2006). A chromatin-mediated mechanism for specification of conditional transcription factor targets. *Nature genetics*, *38*(12), 1446–51. doi:10.1038/ng1917

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Molecular cell*, *36*(4), 541–6. doi:10.1016/j.molcel.2009.10.019

Burkard, K. T. D., & Butler, J. S. (2000). A Nuclear 3′-5′ Exonuclease Involved in mRNA Degradation Interacts with Poly (A) Polymerase and the hnRNA Protein Npl3p. *Molecular and cellular biology*, *20*(2), 604–616. doi:10.1128/MCB.20.2.604-616.2000.Updated

Cabal, G. G., Genovesio, A., Rodriguez-Navarro, S., Zimmer, C., Gadal, O., Lesne, A., Buc, H., et al. (2006). SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, *441*(7094), 770–3. doi:10.1038/nature04752

Cairns, B R, Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., et al. (1996). RSC, an essential, abundant chromatin-remodeling complex. *Cell*, *87*(7), 1249–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8980231

Cairns, Bradley R. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature*, *461*(7261), 193–8. doi:10.1038/nature08450

Callen, B. P., Shearwin, K. E., & Egan, J. B. (2004). Transcriptional interference between convergent promoters caused by elongation over the promoter. *Molecular cell*, *14*(5), 647–56. doi:10.1016/j.molcel.2004.05.010

Camblong, J., Beyrouthy, N., Guffanti, E., Schlaepfer, G., Steinmetz, L. M., & Stutz, F. (2009). Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. *Genes & development*, *23*(13), 1534–45. doi:10.1101/gad.522509

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., & Stutz, F. (2007). Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. *Cell*, *131*(4), 706–17. doi:10.1016/j.cell.2007.09.014

Carey, M., Li, B., & Workman, J. L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Molecular cell*, *24*(3), 481–7. doi:10.1016/j.molcel.2006.09.012

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., et al. (2005). The transcriptional landscape of the mammalian genome. *Science*, *309*(5740), 1559–63. doi:10.1126/science.1112014

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W.-J., et al. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, *123*(4), 581–92. doi:10.1016/j.cell.2005.10.023

Casolari, J. M., Brown, C. R., Komili, S., West, J., Hieronymus, H., & Silver, P. a. (2004). Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell*, *117*(4), 427–39. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15137937

Chen, J., Zheng, X. F., Brown, E. J., & Schreiber, S. L. (1995). Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-associated protein and characterization of a critical serine residue. *PNAS*, *92*(11), 4947–51. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=41824&tool=pmcentrez&rendertype=abstract

Chen, Z., Odstrcil, E. a, Tu, B. P., & McKnight, S. L. (2007). Restriction of DNA replication to the reductive phase of the metabolic cycle protects genome integrity. *Science*, *316*(5833), 1916–9. doi:10.1126/science.1140958

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, *308*(5725), 1149–54. doi:10.1126/science.1108625

Cho, R. J., Campbell, M. J., Winzeler, E. a, Steinmetz, L., Conway, a, Wodicka, L., Wolfsberg, T. G., et al. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular cell*, *2*(1), 65–73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9702192

Churchman, L. S., & Weissman, J. S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, *469*(7330), 368–73. doi:10.1038/nature09652

Cohen, B. a, Mitra, R. D., Hughes, J. D., & Church, G. M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature genetics*, *26*(2), 183–6. doi:10.1038/79896

Collart, M. A., & Oliviero, S. (2001). Preparation of Yeast RNA. *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc. doi:10.1002/0471142727.mb1312s23

Custódio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H. S., Grosveld, F., & Antoniou, M. (1999). Inefficient processing impairs release of RNA from the site of transcription. *The EMBO journal*, *18*(10), 2855–66. doi:10.1093/emboj/18.10.2855

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., et al. (2006). A high-resolution map of transcription in the yeast genome. *PNAS*, *103*(14), 5320–5. doi:10.1073/pnas.0601091103

Davis, C. A., & Ares, M. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. *PNAS*, *103*(9), 3262–7. doi:10.1073/pnas.0507783103

Dion, M. F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., & Rando, O. J. (2007). Dynamics of replication-independent histone turnover in budding yeast. *Science*, *315*(5817), 1405–8. doi:10.1126/science.1134053

Durrin, L. K., Mann, R. K., & Grunstein, M. (1992). Nucleosome Loss Activates CUP1 and HIS3 Promoters to Fully Induced Levels in the Yeast Saccharomyces cerevisiae. *Molecular and cellular biology*, *12*(4), 1621–1629. doi:10.1128/MCB.12.4.1621.Updated

Ebisuya, M., Yamamoto, T., Nakajima, M., & Nishida, E. (2008). Ripples from neighbouring transcription. *Nature cell biology*, *10*(9), 1106–13. doi:10.1038/ncb1771

Egloff, S., O'Reilly, D., Chapman, R. D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., et al. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, *318*(5857), 1777–9. doi:10.1126/science.1145989

Fan, X., Moqtaderi, Z., Jin, Y., Zhang, Y., Liu, X. S., & Struhl, K. (2010). Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. *PNAS*, *107*(42), 17945–50. doi:10.1073/pnas.1012674107

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I. K., Sharon, E., Lubling, Y., Widom, J., et al. (2008). Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS computational biology*, *4*(11), e1000216. doi:10.1371/journal.pcbi.1000216

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, *391*(February), 806–811.

Floer, M., Wang, X., Prabhu, V., Berrozpe, G., Narayan, S., Spagna, D., Alvarez, D., et al. (2010). A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. *Cell*, *141*(3), 407–18. doi:10.1016/j.cell.2010.03.048

Geisler, S., Lojek, L., Khalil, A. M., Baker, K. E., & Coller, J. (2012). Decapping of long noncoding RNAs regulates inducible genes. *Molecular cell*, *45*(3), 279–91. doi:10.1016/j.molcel.2011.11.025

Gillette, M. U., & Sejnowski, T. J. (2005). Physiology. Biological clocks coordinately keep life on time. *Science*, *309*(5738), 1196–8. doi:10.1126/science.1111420

Graber, J. H., Cantor, C. R., Mohr, S. C., & Smith, T. F. (1999). Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic acids research*, *27*(3), 888–94. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=148262&tool=pmcentrez&rendertype=abstract

Granovskaia, M. V, Jensen, L. J., Ritchie, M. E., Toedling, J., Ning, Y., Bork, P., Huber, W., et al. (2010). High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome biology*, *11*(3), R24. doi:10.1186/gb-2010-11-3-r24

Green, E. M., Jiang, Y., Joyner, R., & Weis, K. (2012). A negative feedback loop at the nuclear periphery regulates GAL gene expression. *Molecular biology of the cell*, *23*(7), 1367–75. doi:10.1091/mbc.E11-06-0547

Greger, I. H., Aranda, A., & Proudfoot, N. (2000). Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. *PNAS*, *97*(15), 8415–20. doi:10.1073/pnas.140217697

Hahn, S., & Young, E. T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics*, *189*(3), 705–36. doi:10.1534/genetics.111.127019

Hainer, S. J., Pruneski, J. a, Mitchell, R. D., Monteverde, R. M., & Martens, J. a. (2011). Intergenic transcription causes repression by directing nucleosome assembly. *Genes & development*, *25*(1), 29–40. doi:10.1101/gad.1975011

Han, M, & Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters in vivo. *Cell*, *55*(6), 1137–45. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2849508

Han, Min, Kim, U., & Kayne, P. (1988). Depletion of the histone H4 and nucleosomes activates the PHO5 gene in Saccaromyces cerevisiae. *The EMBO journal*, *7*(7), 2221–2228.

Haruki, H., Nishikawa, J., & Laemmli, U. K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Molecular cell*, *31*(6), 925–32. doi:10.1016/j.molcel.2008.07.020

Heidmann, S., Obermaier, B., Vogel, K., & Domdey, H. (1992). Identification of pre-mRNA polyadenylation sites in Saccharomyces cerevisiae . *Molecular and cellular biology*, *12*(9), 4215–4229. doi:10.1128/MCB.12.9.4215.Updated

Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? *Genes & Development*, *7*(7b), 1291–1308. doi:10.1101/gad.7.7b.1291

Hershberg, R., Yeger-Lotem, E., & Margalit, H. (2005). Chromosomal organization is shaped by the transcription regulatory network. *Trends in genetics : TIG*, *21*(3), 138–42. doi:10.1016/j.tig.2005.01.003

Hobson, D. J., Wei, W., Steinmetz, L. M., & Svejstrup, J. Q. (2012). RNA Polymerase II Collision Interrupts Convergent Transcription. *Molecular cell*, *48*(November 9, 2012), 365–374.

Hongay, C. F., Grisafi, P. L., Galitski, T., & Fink, G. R. (2006). Antisense transcription controls cell fate in Saccharomyces cerevisiae. *Cell*, *127*(4), 735–45. doi:10.1016/j.cell.2006.09.038

Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., & Vogelauer, M. (2008). A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Molecular cell*, *32*(5), 685–95. doi:10.1016/j.molcel.2008.09.027

Huisinga, K. L., & Pugh, B. F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Molecular cell*, *13*(4), 573–85. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14992726

Hurst, L. D., Pál, C., & Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene order. *Nature reviews. Genetics*, *5*(4), 299–310. doi:10.1038/nrg1319

Ideker, T., Galitski, T., & Hood, L. (2001). A New Approach to Decoding Life: Systems Biology. *Annual Review of Genomics and Human Genetics*, *2*, 343–372.

Jansen, A., & Verstrepen, K. J. (2011). Nucleosome positioning in Saccharomyces cerevisiae. *Microbiology and molecular biology reviews : MMBR*, *75*(2), 301–20. doi:10.1128/MMBR.00046-10

Jansen, R., Greenbaum, D., & Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. *Genome research*, *12*(1), 37–46. doi:10.1101/gr.205602

Joshi, A. a, & Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Molecular cell*, *20*(6), 971–8. doi:10.1016/j.molcel.2005.11.021

Kaplan, C. D. (2013). Basic mechanisms of RNA polymerase II activity and alteration of gene expression in Saccharomyces cerevisiae ☆. *Biochimica et Biophysica Acta*, *1829*, 39–54.

Kaplan, C. D., Laprade, L., & Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, *301*(5636), 1096–9. doi:10.1126/science.1087374

Kapranov, P., Willingham, A. T., & Gingeras, T. R. (2007). Genome-wide transcription and the implications for genomic organization. *Nature reviews. Genetics*, *8*(6), 413–23. doi:10.1038/nrg2083

Kebaara, B. W., & Atkin, A. L. (2009). Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. *Nucleic acids research*, *37*(9), 2771–8. doi:10.1093/nar/gkp146

Keogh, M.-C., Kurdistani, S. K., Morris, S. a, Ahn, S. H., Podolny, V., Collins, S. R., Schuldiner, M., et al. (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell*, *123*(4), 593–605. doi:10.1016/j.cell.2005.10.025

Kim, M., Suh, H., Cho, E.-J., & Buratowski, S. (2009). Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *The Journal of biological chemistry*, *284*(39), 26421–6. doi:10.1074/jbc.M109.028993

Kitano, H. (2002). Systems biology: a brief overview. *Science*, *295*(5560), 1662–4. doi:10.1126/science.1069492

Klevecz, R. R., Bolen, J., Forrest, G., & Murray, D. B. (2004). A genomewide oscillation in transcription gates DNA replication and cell cycle. *PNAS*, *101*(5), 1200–5. doi:10.1073/pnas.0306490101

Knezetic, J. a, & Luse, D. S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell*, *45*(1), 95–104. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3955658

Koutelou, E., Hirsch, C. L., & Dent, S. Y. R. (2010). Multiple faces of the SAGA complex. *Current opinion in cell biology*, *22*(3), 374–82. doi:10.1016/j.ceb.2010.03.005

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693–705. doi:10.1016/j.cell.2007.02.005

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., & Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell*, *121*(5), 713–24. doi:10.1016/j.cell.2005.04.029

Latos, P. a, Pauler, F. M., Koerner, M. V, Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., Allhoff, W., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science*, *338*(6113), 1469–72. doi:10.1126/science.1228110

Laxman, S., Sutter, B. M., & Tu, B. P. (2010). Behavior of a metabolic cycling population at the single cell level as visualized by fluorescent gene expression reporters. *PloS one*, *5*(9), e12595. doi:10.1371/journal.pone.0012595

Lee, J. M., & Sonnhammer, E. L. L. (2003). Genomic Gene Clustering Analysis of Pathways in Eukaryotes. *Genome research*, *13*, 875–882. doi:10.1101/gr.737703.

Lee, J. T., & Bartolomei, M. S. (2013). X-Inactivation , Imprinting , and Long Noncoding RNAs in Health and Disease. *Cell*, *152*, 1308–1323.

Lemon, B., & Robert, T. (2000). Orchestrated response: a symphony of transcription factors for gene control. *Genes & Development*, *14*(20), 2551–2569. doi:10.1101/gad.831000

Liu, H., Styles, C. A., & Fink, G. R. (1996). Saccharomyces cerevisiae S288C Has a Mutation in FLO8, a Gene Required for Filamentous Growth. *Genetics society of America*, *144*(November 1996), 967–978.

Lohr, D. (1997). Nucleosome Transactions on the Promoters of the Yeast GAL and PHO Genes. *The Journal of biological chemistry*, *272*(43), 26795–26798.

Longtine, M. S., McKenzie, A., Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P., et al. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast*, *14*(10), 953–61. doi:10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U

Lorch, Y., LaPointe, J. W., & Kornberg, R. D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, *49*(2), 203–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3568125

Lorch, Y., LaPointe, J. W., & Kornberg, R. D. (1992). Initiation on chromatin templates in a yeast RNA polymerase II transcription system. *Genes & Development*, *6*(12a), 2282–2287. doi:10.1101/gad.6.12a.2282

Lorenz, M. C., & Heitman, J. (1998). Regulators of pseudohyphal differentiation in Saccharomyces cerevisiae identified through multicopy suppressor analysis in ammonium permease mutant strains. *Genetics*, *150*(4), 1443–57. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1460428&tool=pmcentrez&rendertype=abstract

Macisaac, K. D., Gordon, D. B., Nekludova, L., Odom, D. T., Schreiber, J., Gifford, D. K., Young, R. a, et al. (2006). A hypothesis-based approach for identifying the binding specificity of regulatory proteins from chromatin immunoprecipitation data. *Bioinformatics (Oxford, England)*, *22*(4), 423–9. doi:10.1093/bioinformatics/bti815

Mark Ptashne, & Alexander Gann. (1997). Transcriptional activation by recruitment, *386*(6625), 569–577.

Martens, J. a, Laprade, L., & Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature*, *429*(6991), 571–4. doi:10.1038/nature02538

Martens, J. a, Wu, P.-Y. J., & Winston, F. (2005). Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. *Genes & development*, *19*(22), 2695–704. doi:10.1101/gad.1367605

Mavrich, T. N., Ioshikhes, I. P., Venters, B. J., Jiang, C., Tomsho, L. P., Qi, J., Schuster, S. C., et al. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome research*, *18*, 1073–1083. doi:10.1101/gr.078261.108.1

Mazo, A., Hodgson, J. W., Petruk, S., Sedkov, Y., & Brock, H. W. (2007). Transcriptional interference: an unexpected layer of complexity in gene regulation. *Journal of cell science*, *120*(Pt 16), 2755–61. doi:10.1242/jcs.007633

Morillon, A., O'Sullivan, J., Azad, A., Proudfoot, N., & Mellor, J. (2003). Regulation of elongating RNA polymerase II by forkhead transcription factors in yeast. *Science*, *300*(5618), 492–5. doi:10.1126/science.1081379

Murray, S. C., Serra Barros, A., Brown, D. a, Dudek, P., Ayling, J., & Mellor, J. (2012). A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic acids research*, *40*(6), 2432–44. doi:10.1093/nar/gkr1121

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L. M., & Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, *457*(7232), 1038–42. doi:10.1038/nature07747

O'Sullivan, J. M., Tan-Wong, S. M., Morillon, A., Lee, B., Coles, J., Mellor, J., & Proudfoot, N. J. (2004). Gene loops juxtapose promoters and terminators in yeast. *Nature genetics*, *36*(9), 1014–8. doi:10.1038/ng1411

Ozdemir, A., Spicuglia, S., Lasonder, E., Vermeulen, M., Campsteijn, C., Stunnenberg, H. G., & Logie, C. (2005). Characterization of lysine 56 of histone H3 as an acetylation site in Saccharomyces cerevisiae. *The Journal of biological chemistry*, *280*(28), 25949–52. doi:10.1074/jbc.C500181200

Palmer, A. C., Egan, J. B., & Shearwin, K. E. (2011). Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. *Transcription*, *2*(1), 9–14. doi:10.4161/trns.2.1.13511

Pauler, F. M., Barlow, D. P., & Hudson, Q. J. (2012). Mechanisms of long range silencing by imprinted macro non-coding RNAs. *Current opinion in genetics & development*, *22*(3), 283–9. doi:10.1016/j.gde.2012.02.005

Pauli, A., Rinn, J. L., & Schier, A. F. (2011). Non-coding RNAs as regulators of embryogenesis. *Nature reviews. Genetics*, *12*(2), 136–49. doi:10.1038/nrg2904

Perocchi, F., Xu, Z., Clauder-Münster, S., & Steinmetz, L. M. (2007). Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic acids research*, *35*(19), e128. doi:10.1093/nar/gkm683

Prescott, E. M., & Proudfoot, N. J. (2002). Transcriptional collision between convergent genes in budding yeast. *PNAS*, *99*(13), 8796–801. doi:10.1073/pnas.132270899

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. *Current opinion in cell biology*, *16*(3), 272–8. doi:10.1016/j.ceb.2004.03.007

Proudfoot, N. J. (2011). Ending the message : poly (A) signals then and now. *Genes & development*, *25*, 1770–1782. doi:10.1101/gad.17268411.Freely

Radman-Livaja, M., & Rando, O. J. (2010). Nucleosome positioning: how is it established, and why does it matter? *Developmental biology*, *339*(2), 258–66. doi:10.1016/j.ydbio.2009.06.012

Raisner, R. M., Hartley, P. D., Meneghini, M. D., Bao, M. Z., Liu, C. L., Schreiber, S. L., Rando, O. J., et al. (2005). Histone variant H2A.Z marks the 5' ends of both active

and inactive genes in euchromatin. *Cell*, *123*(2), 233–48. doi:10.1016/j.cell.2005.10.002

Rao, B., Shibata, Y., Strahl, B. D., Lieb, D., & Lieb, J. D. (2005). Dimethylation of Histone H3 at Lysine 36 Demarcates Regulatory and Nonregulatory Chromatin Genome-Wide. *Molecular and cellular biology*, *25*(21), 9447–9459. doi:10.1128/MCB.25.21.9447

Reeves, W. M., & Hahn, S. (2005). Targets of the Gal4 Transcription Activator in Functional Transcription Complexes. *Molecular and cellular biology*, *25*(20), 9092–9102. doi:10.1128/MCB.25.20.9092

Reinberg, D., & Sims, R. J. (2006). de FACTo nucleosome dynamics. *The Journal of biological chemistry*, *281*(33), 23297–301. doi:10.1074/jbc.R600007200

Reinke, H., & Hörz, W. (2003). Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Molecular cell*, *11*(6), 1599–607. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12820972

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science*, *290*(5500), 2306–9. doi:10.1126/science.290.5500.2306

Richard, P. (2003). The rhythm of yeast. *FEMS Microbiology Reviews*, *27*(4), 547–557. doi:10.1016/S0168-6445(03)00065-2

Rufiange, A., Jacques, P.-E., Bhat, W., Robert, F., & Nourani, A. (2007). Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *Molecular cell*, *27*(3), 393–405. doi:10.1016/j.molcel.2007.07.011

Rutter, J., Reick, M., & McKnight, S. L. (2002). Metabolism and the control of circadian rhythms. *Annual review of biochemistry*, *71*, 307–31. doi:10.1146/annurev.biochem.71.090501.142857

Saha, A., Wittmeyer, J., & Cairns, B. R. (2002). Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes & development*, *16*(16), 2120–34. doi:10.1101/gad.995002

Schaff-Gernstensclager, I., Mannhaupt, G., Vetter, I., Zimmerman, F. K., & Feldmann, H. (1993). TKL2, a second transketolase gene of Saccharomyces cerevisiae. *European Journal of Biochemistry*, *217*(1), 487–492. doi:10.1111/j.1432-1033.1993.tb18268.x

Schmid, M., Poulsen, M. B., Olszewski, P., Pelechano, V., Saguez, C., Gupta, I., Steinmetz, L. M., et al. (2012). Rrp6p controls mRNA poly(A) tail length and its decoration with poly(A) binding proteins. *Molecular cell*, *47*(2), 267–80. doi:10.1016/j.molcel.2012.05.005

Shearwin, K. E., Callen, B. P., & Egan, J. B. (2005). Transcriptional interference--a crash course. *Trends in genetics : TIG*, *21*(6), 339–45. doi:10.1016/j.tig.2005.04.009

Sikorski, T. W., & Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Current opinion in cell biology*, *21*(3), 344–51. doi:10.1016/j.ceb.2009.03.006

Silverman, S. J., Petti, A. a, Slavov, N., Parsons, L., Briehof, R., Thiberge, S. Y., Zenklusen, D., et al. (2010). Metabolic cycling in single yeast cells from unsynchronized steady-

state populations limited on glucose or phosphate. *PNAS*, *107*(15), 6946–51. doi:10.1073/pnas.1002422107

Slavov, N., Airoldi, E. M., Van Oudenaarden, A., & Botstein, D. (2012). A conserved cell growth cycle can account for the environmental stress responses of divergent eukaryotes. *Molecular biology of the cell*, *23*(10), 1986–97. doi:10.1091/mbc.E11-11-0961

Slavov, N., & Botstein, D. (2011). Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast. *Molecular biology of the cell*, *22*(12), 1997–2009. doi:10.1091/mbc.E11-02-0132

Sleutels, F., Zwart, R., & Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, *415*(6873), 810–3. doi:10.1038/415810a

Song, Q., & Kumar, A. (2012). An Overview of Autophagy and Yeast Pseudohyphal Growth: Integration of Signaling Pathways during Nitrogen Stress. *Cells*, *1*(4), 263–283. doi:10.3390/cells1030263

Soranzo, N., Zampieri, M., Farina, L., & Altafini, C. (2009). mRNA stability and the unfolding of gene expression in the long-period yeast metabolic cycle. *BMC systems biology*, *3*, 18. doi:10.1186/1752-0509-3-18

Steiner, S., Philippsen, P., Gieben, U., & Str, F. (1994). Sequence and promoter analysis of the highly expressed TEF gene of the filamentous fungus Ashby gossypii. *Molecular genetics and genomics*, *242*, 263–271.

Steinmetz, E. J., Warren, C. L., Kuehner, J. N., Panbehi, B., Ansari, A. Z., & Brow, D. a. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Molecular cell*, *24*(5), 735–46. doi:10.1016/j.molcel.2006.10.023

Strahl, B. D., Grant, P. A., Briggs, S. D., Bone, J. R., Caldwell, J. A., Mollah, S., Cook, R. G., et al. (2002). Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. *Molecular and cellular biology*, *22*(5), 1298–1306. doi:10.1128/MCB.22.5.1298

Struhl, K., & Segal, E. (2013). Determinants of nucleosome positioning. *Nature structural & molecular Biology*, *20*(3), 267–273. doi:10.1038/nsmb.2506

Svaren, J., & Horz, W. (1997). Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. *Trends in biochemical science*, *22*(3), 93–97.

Tachibana, C., Yoo, J. Y., Tagne, J., Kacherovsky, N., Lee, T. I., & Young, E. T. (2005). Combined Global Localization Analysis and Transcriptome Data Identify Genes That Are Directly Coregulated by Adr1 and Cat8 Combined Global Localization Analysis and Transcriptome Data Identify Genes That Are Directly Coregulated by Adr1 and Cat8 †. *Molecular and cellular biology*, *25*(6), 2138–2146. doi:10.1128/MCB.25.6.2138

Taft, R. J., & Mattick, J. S. (2003). Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. *Genome biology*, *5*(1), PI 1–PI 22.

Teichmann, S. A., & Veitia, R. A. (2004). Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. *Genetics*, *167*(4), 2121–5. doi:10.1534/genetics.103.024505

Teixeira, M. C., Monteiro, P., Jain, P., Tenreiro, S., Fernandes, A. R., Mira, N. P., Alenquer, M., et al. (2006). The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. *Nucleic acids research*, *34*(Database issue), D446–51. doi:10.1093/nar/gkj013

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., & Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Molecular cell*, *23*(6), 853–64. doi:10.1016/j.molcel.2006.07.029

Tourrière, H., Chebli, K., & Tazi, J. (2002). mRNA degradation machines in eukaryotic cells. *Biochimie*, *84*(8), 821–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12457569

Traven, A., Jelicic, B., & Sopta, M. (2006). Yeast Gal4: a transcriptional paradigm revisited. *EMBO reports*, *7*(5), 496–9. doi:10.1038/sj.embor.7400679

Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., & Myers, R. M. (2004). An abundance of bidirectional promoters in the human genome. *Genome research*, *14*(1), 62–6. doi:10.1101/gr.1982804

Tsubota, T., Berndsen, C. E., Erkmann, J. a, Smith, C. L., Yang, L., Freitas, M. a, Denu, J. M., et al. (2007). Histone H3-K56 acetylation is catalyzed by histone chaperone-dependent complexes. *Molecular cell*, *25*(5), 703–12. doi:10.1016/j.molcel.2007.02.006

Tu, B. P., Kudlicki, A., Rowicka, M., & McKnight, S. L. (2005). Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science*, *310*(5751), 1152–8. doi:10.1126/science.1120499

Tu, B. P., & McKnight, S. L. (2006). Metabolic cycles as an underlying basis of biological oscillations. *Nature reviews. Molecular cell biology*, *7*(9), 696–701. doi:10.1038/nrm1980

Tutucci, E., & Stutz, F. (2011). Keeping mRNPs in check during assembly and nuclear export. *Nature reviews. Molecular cell biology*, *12*(6), 377–384.

Uhler, J. P., Hertel, C., & Svejstrup, J. Q. (2007). A role for noncoding transcription in activation of the yeast PHO5 gene. *PNAS*, *104*(19), 8011–6. doi:10.1073/pnas.0702431104

Van Dijk, E. L., Chen, C. L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature*, *475*(7354), 114–7. doi:10.1038/nature10118

Van Werven, F. J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., Van Oudenaarden, A., Primig, M., et al. (2012). Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell*, *150*(6), 1170–81. doi:10.1016/j.cell.2012.06.049

Vanácová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., et al. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS biology*, *3*(6), e189. doi:10.1371/journal.pbio.0030189

Vasiljeva, L., & Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Molecular cell*, *21*(2), 239–48. doi:10.1016/j.molcel.2005.11.028

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., & Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature structural & molecular biology*, *15*(8), 795–804. doi:10.1038/nsmb.1468

Venkatesh, S., Smolle, M., Li, H., Gogol, M. M., Saint, M., Kumar, S., Natarajan, K., et al. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature*, *489*(7416), 452–5. doi:10.1038/nature11326

Venters, B. J., & Pugh, B. F. (2009a). How eukaryotic genes are transcribed. *Critical reviews in biochemistry and molecular biology*, *44*(2-3), 117–41. doi:10.1080/10409230902858785

Venters, B. J., & Pugh, B. F. (2009b). A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome research*, *19*(3), 360–71. doi:10.1101/gr.084970.108

Vinciguerra, P., & Stutz, F. (2004). mRNA export: an assembly line from genes to nuclear pores. *Current opinion in cell biology*, *16*(3), 285–292. Retrieved from http://europepmc.org/abstract/MED/15145353

Vodala, S., Abruzzi, K. C., & Rosbash, M. (2008). The Nuclear Exosome and Adenylation Regulate Post-Transcriptional Tethering of Yeast GAL genes to the Nuclear Periphery. *Molecular cell*, *31*(1), 104–113. doi:10.1016/j.molcel.2008.05.015.The

Wilusz, C. J., Wormington, M., & Peltz, S. W. (2001). The cap-to-tail guide to mRNA turnover. *Nature reviews. Molecular cell biology*, *2*(4), 237–46. doi:10.1038/35067025

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*, *121*(5), 725–37. doi:10.1016/j.cell.2005.04.030

Xu, Z., Wei, W., Gagneur, J., Clauder-Münster, S., Smolik, M., Huber, W., & Steinmetz, L. M. (2011). Antisense expression increases gene expression variability and locus interdependency. *Molecular systems biology*, *7*(468), 468. doi:10.1038/msb.2011.1

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., et al. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, *457*(7232), 1033–7. doi:10.1038/nature07728

Youdell, M. L., Kizer, K. O., Kisseleva-Romanova, E., Fuchs, S. M., Duro, E., Strahl, B. D., & Mellor, J. (2008). Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. *Molecular and cellular biology*, *28*(16), 4915–26. doi:10.1128/MCB.00001-08

Young, E. T., Dombek, K. M., Tachibana, C., & Ideker, T. (2003). Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. *The Journal of biological chemistry*, *278*(28), 26146–58. doi:10.1074/jbc.M301981200

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A. P., & Cui, H. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, *451*(7175), 202–6. doi:10.1038/nature06468

Yuan, G.-C., Liu, Y.-J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J., & Rando, O. J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. *Science*, *309*(5734), 626–30. doi:10.1126/science.1112178

Zenklusen, D., Larson, D. R., & Singer, R. H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature structural & molecular biology*, *15*(12), 1263–71. doi:10.1038/nsmb.1514

Zenklusen, D., & Singer, R. H. (2010). Analyzing mRNA expression using single mRNA resolution fluorescent in situ hybridization. *Methods in enzymology*, *470*(10), 641–59. doi:10.1016/S0076-6879(10)70026-4

Zhang, H., Roberts, D. N., & Cairns, B. R. (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*, *123*(2), 219–31. doi:10.1016/j.cell.2005.08.036

Zhao, J., Herrera-diaz, J., & Gross, D. S. (2005). Domain-Wide Displacement of Histones by Activated Heat Shock Factor Occurs Independently of Swi/ Snf and Is Not Correlated with RNA Polymerase II Density. *Molecular and Cellular Biology*, *25*(20), 8985–8999. doi:10.1128/MCB.25.20.8985

Abruzzi, K. C., Belostotsky, D. a, Chekanova, J. a, Dower, K., & Rosbash, M. (2006). 3'-end formation signals modulate the association of genes with the nuclear periphery as well as mRNP dot formation. *The EMBO journal*, *25*(18), 4253–62. doi:10.1038/sj.emboj.7601305

Adhya, S., & Gottesman, M. (1982). Promoter Occlusion : Transcription Through a Promoter May Inihibit Its Activity. *Cell*, *29*(3), 939–944.

Akhtar, M. S., Heidemann, M., Tietjen, J. R., Zhang, D. W., Chapman, R. D., Eick, D., & Ansari, A. Z. (2009). TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Molecular cell*, *34*(3), 387–93. doi:10.1016/j.molcel.2009.04.016

Aravind, L., Watanabe, H., Lipman, D. J., & Koonin, E. V. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *PNAS*, *97*(21), 11319–24. doi:10.1073/pnas.200346997

Arigo, J. T., Eyler, D. E., Carroll, K. L., & Corden, J. L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Molecular cell*, *23*(6), 841–51. doi:10.1016/j.molcel.2006.07.024

Axelrod, J. D., Reagan, M. S., & Majors, J. (1993). GAL4 disrupts a repressing nucleosome during activation of GAL1 transcription in vivo. *Genes & Development*, *7*(5), 857–869. doi:10.1101/gad.7.5.857

Bai, L., & Morozov, A. V. (2010). Gene regulation by nucleosome positioning. *Trends in genetics : TIG*, *26*(11), 476–83. doi:10.1016/j.tig.2010.08.003

Basehoar, A. D., Zanton, S. J., & Pugh, B. F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, *116*(5), 699–709. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15006352

Batada, N. N., Urrutia, A. O., & Zhang, M. Q. (2007). Chromatin remodelling is a major source of coexpression of linked genes in yeast. *Trends in genetics : TIG*, *23*(10), 480–484. doi:10.1016/j.tig.2007.08.007

Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., Rinn, J. L., et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science*, *306*(5705), 2242–6. doi:10.1126/science.1103388

Birney, E., Stamatoyannopoulos, J. a, Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799–816. doi:10.1038/nature05874

Birse, C. E. (1998). Coupling Termination of Transcription to Messenger RNA Maturation in Yeast. *Science*, *280*(5361), 298–301. doi:10.1126/science.280.5361.298

Blobel, G. (1985). Gene gating: a hypothesis. *PNAS*, *82*(24), 8527–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=390949&tool=pmcentrez&rendertype=abstract

Brantl, S. (2002). Antisense-RNA regulation and RNA interference. *Biochimica et biophysica acta*, *1575*(1-3), 15–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12020814

Brosius, J. (2005). Waste not, want not- transcript excess in multicellular eukaryotes. *Trends in genetics : TIG*, *21*(5), 287–288. doi:10.1038/nature03016

Bryant, G. O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., & Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. *PLoS biology*, *6*(12), 2928–39. doi:10.1371/journal.pbio.0060317

Buck, M. J., & Lieb, J. D. (2006). A chromatin-mediated mechanism for specification of conditional transcription factor targets. *Nature genetics*, *38*(12), 1446–51. doi:10.1038/ng1917

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Molecular cell*, *36*(4), 541–6. doi:10.1016/j.molcel.2009.10.019

Burkard, K. T. D., & Butler, J. S. (2000). A Nuclear 3′-5′ Exonuclease Involved in mRNA Degradation Interacts with Poly (A) Polymerase and the hnRNA Protein Npl3p. *Molecular and cellular biology*, *20*(2), 604–616. doi:10.1128/MCB.20.2.604-616.2000.Updated

Cabal, G. G., Genovesio, A., Rodriguez-Navarro, S., Zimmer, C., Gadal, O., Lesne, A., Buc, H., et al. (2006). SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, *441*(7094), 770–3. doi:10.1038/nature04752

Cairns, B R, Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., et al. (1996). RSC, an essential, abundant chromatin-remodeling complex. *Cell*, *87*(7), 1249–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8980231

Cairns, Bradley R. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature*, *461*(7261), 193–8. doi:10.1038/nature08450

Callen, B. P., Shearwin, K. E., & Egan, J. B. (2004). Transcriptional interference between convergent promoters caused by elongation over the promoter. *Molecular cell*, *14*(5), 647–56. doi:10.1016/j.molcel.2004.05.010

Camblong, J., Beyrouthy, N., Guffanti, E., Schlaepfer, G., Steinmetz, L. M., & Stutz, F. (2009). Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. *Genes & development*, *23*(13), 1534–45. doi:10.1101/gad.522509

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., & Stutz, F. (2007). Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. *Cell*, *131*(4), 706–17. doi:10.1016/j.cell.2007.09.014

Carey, M., Li, B., & Workman, J. L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Molecular cell*, *24*(3), 481–7. doi:10.1016/j.molcel.2006.09.012

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., et al. (2005). The transcriptional landscape of the mammalian genome. *Science*, *309*(5740), 1559–63. doi:10.1126/science.1112014

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W.-J., et al. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, *123*(4), 581–92. doi:10.1016/j.cell.2005.10.023

Casolari, J. M., Brown, C. R., Komili, S., West, J., Hieronymus, H., & Silver, P. a. (2004). Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell*, *117*(4), 427–39. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15137937

Chen, J., Zheng, X. F., Brown, E. J., & Schreiber, S. L. (1995). Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-associated protein and characterization of a critical serine residue. *PNAS*, *92*(11), 4947–51. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=41824&tool=pmcentrez&rendertype=abstract

Chen, Z., Odstrcil, E. a, Tu, B. P., & McKnight, S. L. (2007). Restriction of DNA replication to the reductive phase of the metabolic cycle protects genome integrity. *Science*, *316*(5833), 1916–9. doi:10.1126/science.1140958

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, *308*(5725), 1149–54. doi:10.1126/science.1108625

Cho, R. J., Campbell, M. J., Winzeler, E. a, Steinmetz, L., Conway, a, Wodicka, L., Wolfsberg, T. G., et al. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular cell*, *2*(1), 65–73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9702192

Churchman, L. S., & Weissman, J. S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, *469*(7330), 368–73. doi:10.1038/nature09652

Cohen, B. a, Mitra, R. D., Hughes, J. D., & Church, G. M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature genetics*, *26*(2), 183–6. doi:10.1038/79896

Collart, M. A., & Oliviero, S. (2001). Preparation of Yeast RNA. *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc. doi:10.1002/0471142727.mb1312s23

Custódio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H. S., Grosveld, F., & Antoniou, M. (1999). Inefficient processing impairs release of RNA from the site of transcription. *The EMBO journal*, *18*(10), 2855–66. doi:10.1093/emboj/18.10.2855

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., et al. (2006). A high-resolution map of transcription in the yeast genome. *PNAS*, *103*(14), 5320–5. doi:10.1073/pnas.0601091103

Davis, C. A., & Ares, M. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. *PNAS*, *103*(9), 3262–7. doi:10.1073/pnas.0507783103

Dion, M. F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., & Rando, O. J. (2007). Dynamics of replication-independent histone turnover in budding yeast. *Science*, *315*(5817), 1405–8. doi:10.1126/science.1134053

Durrin, L. K., Mann, R. K., & Grunstein, M. (1992). Nucleosome Loss Activates CUP1 and HIS3 Promoters to Fully Induced Levels in the Yeast Saccharomyces cerevisiae. *Molecular and cellular biology*, *12*(4), 1621–1629. doi:10.1128/MCB.12.4.1621.Updated

Ebisuya, M., Yamamoto, T., Nakajima, M., & Nishida, E. (2008). Ripples from neighbouring transcription. *Nature cell biology*, *10*(9), 1106–13. doi:10.1038/ncb1771

Egloff, S., O'Reilly, D., Chapman, R. D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., et al. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, *318*(5857), 1777–9. doi:10.1126/science.1145989

Fan, X., Moqtaderi, Z., Jin, Y., Zhang, Y., Liu, X. S., & Struhl, K. (2010). Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. *PNAS*, *107*(42), 17945–50. doi:10.1073/pnas.1012674107

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I. K., Sharon, E., Lubling, Y., Widom, J., et al. (2008). Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS computational biology*, *4*(11), e1000216. doi:10.1371/journal.pcbi.1000216

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, *391*(February), 806–811.

Floer, M., Wang, X., Prabhu, V., Berrozpe, G., Narayan, S., Spagna, D., Alvarez, D., et al. (2010). A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. *Cell*, *141*(3), 407–18. doi:10.1016/j.cell.2010.03.048

Geisler, S., Lojek, L., Khalil, A. M., Baker, K. E., & Coller, J. (2012). Decapping of long noncoding RNAs regulates inducible genes. *Molecular cell*, *45*(3), 279–91. doi:10.1016/j.molcel.2011.11.025

Gillette, M. U., & Sejnowski, T. J. (2005). Physiology. Biological clocks coordinately keep life on time. *Science*, *309*(5738), 1196–8. doi:10.1126/science.1111420

Graber, J. H., Cantor, C. R., Mohr, S. C., & Smith, T. F. (1999). Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic acids research*, *27*(3), 888–94. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=148262&tool=pmcentrez&rendertype=abstract

Granovskaia, M. V, Jensen, L. J., Ritchie, M. E., Toedling, J., Ning, Y., Bork, P., Huber, W., et al. (2010). High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome biology*, *11*(3), R24. doi:10.1186/gb-2010-11-3-r24

Green, E. M., Jiang, Y., Joyner, R., & Weis, K. (2012). A negative feedback loop at the nuclear periphery regulates GAL gene expression. *Molecular biology of the cell*, *23*(7), 1367–75. doi:10.1091/mbc.E11-06-0547

Greger, I. H., Aranda, A., & Proudfoot, N. (2000). Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. *PNAS*, *97*(15), 8415–20. doi:10.1073/pnas.140217697

Hahn, S., & Young, E. T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics*, *189*(3), 705–36. doi:10.1534/genetics.111.127019

Hainer, S. J., Pruneski, J. a, Mitchell, R. D., Monteverde, R. M., & Martens, J. a. (2011). Intergenic transcription causes repression by directing nucleosome assembly. *Genes & development*, *25*(1), 29–40. doi:10.1101/gad.1975011

Han, M, & Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters in vivo. *Cell*, *55*(6), 1137–45. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2849508

Han, Min, Kim, U., & Kayne, P. (1988). Depletion of the histone H4 and nucleosomes activates the PHO5 gene in Saccaromyces cerevisiae. *The EMBO journal*, *7*(7), 2221–2228.

Haruki, H., Nishikawa, J., & Laemmli, U. K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Molecular cell*, *31*(6), 925–32. doi:10.1016/j.molcel.2008.07.020

Heidmann, S., Obermaier, B., Vogel, K., & Domdey, H. (1992). Identification of pre-mRNA polyadenylation sites in Saccharomyces cerevisiae . *Molecular and cellular biology*, *12*(9), 4215–4229. doi:10.1128/MCB.12.9.4215.Updated

Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? *Genes & Development*, *7*(7b), 1291–1308. doi:10.1101/gad.7.7b.1291

Hershberg, R., Yeger-Lotem, E., & Margalit, H. (2005). Chromosomal organization is shaped by the transcription regulatory network. *Trends in genetics : TIG*, *21*(3), 138–42. doi:10.1016/j.tig.2005.01.003

Hobson, D. J., Wei, W., Steinmetz, L. M., & Svejstrup, J. Q. (2012). RNA Polymerase II Collision Interrupts Convergent Transcription. *Molecular cell*, *48*(November 9, 2012), 365–374.

Hongay, C. F., Grisafi, P. L., Galitski, T., & Fink, G. R. (2006). Antisense transcription controls cell fate in Saccharomyces cerevisiae. *Cell*, *127*(4), 735–45. doi:10.1016/j.cell.2006.09.038

Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., & Vogelauer, M. (2008). A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Molecular cell*, *32*(5), 685–95. doi:10.1016/j.molcel.2008.09.027

Huisinga, K. L., & Pugh, B. F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae.

Molecular cell, *13*(4), 573–85. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14992726

Hurst, L. D., Pál, C., & Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene order. *Nature reviews. Genetics*, *5*(4), 299–310. doi:10.1038/nrg1319

Ideker, T., Galitski, T., & Hood, L. (2001). A New Approach to Decoding Life: Systems Biology. *Annual Review of Genomics and Human Genetics*, *2*, 343–372.

Jansen, A., & Verstrepen, K. J. (2011). Nucleosome positioning in Saccharomyces cerevisiae. *Microbiology and molecular biology reviews : MMBR*, *75*(2), 301–20. doi:10.1128/MMBR.00046-10

Jansen, R., Greenbaum, D., & Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. *Genome research*, *12*(1), 37–46. doi:10.1101/gr.205602

Joshi, A. a, & Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Molecular cell*, *20*(6), 971–8. doi:10.1016/j.molcel.2005.11.021

Kaplan, C. D. (2013). Basic mechanisms of RNA polymerase II activity and alteration of gene expression in Saccharomyces cerevisiae ☆. *Biochimica et Biophysica Acta*, *1829*, 39–54.

Kaplan, C. D., Laprade, L., & Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, *301*(5636), 1096–9. doi:10.1126/science.1087374

Kapranov, P., Willingham, A. T., & Gingeras, T. R. (2007). Genome-wide transcription and the implications for genomic organization. *Nature reviews. Genetics*, *8*(6), 413–23. doi:10.1038/nrg2083

Kebaara, B. W., & Atkin, A. L. (2009). Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. *Nucleic acids research*, *37*(9), 2771–8. doi:10.1093/nar/gkp146

Keogh, M.-C., Kurdistani, S. K., Morris, S. a, Ahn, S. H., Podolny, V., Collins, S. R., Schuldiner, M., et al. (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell*, *123*(4), 593–605. doi:10.1016/j.cell.2005.10.025

Kim, M., Suh, H., Cho, E.-J., & Buratowski, S. (2009). Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *The Journal of biological chemistry*, *284*(39), 26421–6. doi:10.1074/jbc.M109.028993

Kitano, H. (2002). Systems biology: a brief overview. *Science*, *295*(5560), 1662–4. doi:10.1126/science.1069492

Klevecz, R. R., Bolen, J., Forrest, G., & Murray, D. B. (2004). A genomewide oscillation in transcription gates DNA replication and cell cycle. *PNAS*, *101*(5), 1200–5. doi:10.1073/pnas.0306490101

Knezetic, J. a, & Luse, D. S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell*, *45*(1), 95–104. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3955658

Koutelou, E., Hirsch, C. L., & Dent, S. Y. R. (2010). Multiple faces of the SAGA complex. *Current opinion in cell biology*, *22*(3), 374–82. doi:10.1016/j.ceb.2010.03.005

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693–705. doi:10.1016/j.cell.2007.02.005

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., & Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell*, *121*(5), 713–24. doi:10.1016/j.cell.2005.04.029

Latos, P. a, Pauler, F. M., Koerner, M. V, Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., Allhoff, W., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science*, *338*(6113), 1469–72. doi:10.1126/science.1228110

Laxman, S., Sutter, B. M., & Tu, B. P. (2010). Behavior of a metabolic cycling population at the single cell level as visualized by fluorescent gene expression reporters. *PloS one*, *5*(9), e12595. doi:10.1371/journal.pone.0012595

Lee, J. M., & Sonnhammer, E. L. L. (2003). Genomic Gene Clustering Analysis of Pathways in Eukaryotes. *Genome research*, *13*, 875–882. doi:10.1101/gr.737703.

Lee, J. T., & Bartolomei, M. S. (2013). X-Inactivation , Imprinting , and Long Noncoding RNAs in Health and Disease. *Cell*, *152*, 1308–1323.

Lemon, B., & Robert, T. (2000). Orchestrated response: a symphony of transcription factors for gene control. *Genes & Development*, *14*(20), 2551–2569. doi:10.1101/gad.831000

Liu, H., Styles, C. A., & Fink, G. R. (1996). Saccharomyces cerevisiae S288C Has a Mutation in FLO8, a Gene Required for Filamentous Growth. *Genetics society of America*, *144*(November 1996), 967–978.

Lohr, D. (1997). Nucleosome Transactions on the Promoters of the Yeast GAL and PHO Genes. *The Journal of biological chemistry*, *272*(43), 26795–26798.

Longtine, M. S., McKenzie, A., Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P., et al. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast*, *14*(10), 953–61. doi:10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U

Lorch, Y., LaPointe, J. W., & Kornberg, R. D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, *49*(2), 203–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3568125

Lorch, Y., LaPointe, J. W., & Kornberg, R. D. (1992). Initiation on chromatin templates in a yeast RNA polymerase II transcription system. *Genes & Development*, *6*(12a), 2282–2287. doi:10.1101/gad.6.12a.2282

Lorenz, M. C., & Heitman, J. (1998). Regulators of pseudohyphal differentiation in Saccharomyces cerevisiae identified through multicopy suppressor analysis in ammonium permease mutant strains. *Genetics*, *150*(4), 1443–57. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1460428&tool=pmcentrez&rendertype=abstract

Macisaac, K. D., Gordon, D. B., Nekludova, L., Odom, D. T., Schreiber, J., Gifford, D. K., Young, R. a, et al. (2006). A hypothesis-based approach for identifying the binding specificity of regulatory proteins from chromatin immunoprecipitation data. *Bioinformatics (Oxford, England)*, *22*(4), 423–9. doi:10.1093/bioinformatics/bti815

Mark Ptashne, & Alexander Gann. (1997). Transcriptional activation by recruitment, *386*(6625), 569–577.

Martens, J. a, Laprade, L., & Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature*, *429*(6991), 571–4. doi:10.1038/nature02538

Martens, J. a, Wu, P.-Y. J., & Winston, F. (2005). Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. *Genes & development*, *19*(22), 2695–704. doi:10.1101/gad.1367605

Mavrich, T. N., Ioshikhes, I. P., Venters, B. J., Jiang, C., Tomsho, L. P., Qi, J., Schuster, S. C., et al. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome research*, *18*, 1073–1083. doi:10.1101/gr.078261.108.1

Mazo, A., Hodgson, J. W., Petruk, S., Sedkov, Y., & Brock, H. W. (2007). Transcriptional interference: an unexpected layer of complexity in gene regulation. *Journal of cell science*, *120*(Pt 16), 2755–61. doi:10.1242/jcs.007633

Morillon, A., O'Sullivan, J., Azad, A., Proudfoot, N., & Mellor, J. (2003). Regulation of elongating RNA polymerase II by forkhead transcription factors in yeast. *Science*, *300*(5618), 492–5. doi:10.1126/science.1081379

Murray, S. C., Serra Barros, A., Brown, D. a, Dudek, P., Ayling, J., & Mellor, J. (2012). A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic acids research*, *40*(6), 2432–44. doi:10.1093/nar/gkr1121

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L. M., & Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, *457*(7232), 1038–42. doi:10.1038/nature07747

O'Sullivan, J. M., Tan-Wong, S. M., Morillon, A., Lee, B., Coles, J., Mellor, J., & Proudfoot, N. J. (2004). Gene loops juxtapose promoters and terminators in yeast. *Nature genetics*, *36*(9), 1014–8. doi:10.1038/ng1411

Ozdemir, A., Spicuglia, S., Lasonder, E., Vermeulen, M., Campsteijn, C., Stunnenberg, H. G., & Logie, C. (2005). Characterization of lysine 56 of histone H3 as an acetylation site in Saccharomyces cerevisiae. *The Journal of biological chemistry*, *280*(28), 25949–52. doi:10.1074/jbc.C500181200

Palmer, A. C., Egan, J. B., & Shearwin, K. E. (2011). Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. *Transcription*, *2*(1), 9–14. doi:10.4161/trns.2.1.13511

Pauler, F. M., Barlow, D. P., & Hudson, Q. J. (2012). Mechanisms of long range silencing by imprinted macro non-coding RNAs. *Current opinion in genetics & development*, *22*(3), 283–9. doi:10.1016/j.gde.2012.02.005

Pauli, A., Rinn, J. L., & Schier, A. F. (2011). Non-coding RNAs as regulators of embryogenesis. *Nature reviews. Genetics*, *12*(2), 136–49. doi:10.1038/nrg2904

Perocchi, F., Xu, Z., Clauder-Münster, S., & Steinmetz, L. M. (2007). Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic acids research*, *35*(19), e128. doi:10.1093/nar/gkm683

Prescott, E. M., & Proudfoot, N. J. (2002). Transcriptional collision between convergent genes in budding yeast. *PNAS*, *99*(13), 8796–801. doi:10.1073/pnas.132270899

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. *Current opinion in cell biology*, *16*(3), 272–8. doi:10.1016/j.ceb.2004.03.007

Proudfoot, N. J. (2011). Ending the message : poly (A) signals then and now. *Genes & development*, *25*, 1770–1782. doi:10.1101/gad.17268411.Freely

Radman-Livaja, M., & Rando, O. J. (2010). Nucleosome positioning: how is it established, and why does it matter? *Developmental biology*, *339*(2), 258–66. doi:10.1016/j.ydbio.2009.06.012

Raisner, R. M., Hartley, P. D., Meneghini, M. D., Bao, M. Z., Liu, C. L., Schreiber, S. L., Rando, O. J., et al. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell*, *123*(2), 233–48. doi:10.1016/j.cell.2005.10.002

Rao, B., Shibata, Y., Strahl, B. D., Lieb, D., & Lieb, J. D. (2005). Dimethylation of Histone H3 at Lysine 36 Demarcates Regulatory and Nonregulatory Chromatin Genome-Wide. *Molecular and cellular biology*, *25*(21), 9447–9459. doi:10.1128/MCB.25.21.9447

Reeves, W. M., & Hahn, S. (2005). Targets of the Gal4 Transcription Activator in Functional Transcription Complexes. *Molecular and cellular biology*, *25*(20), 9092–9102. doi:10.1128/MCB.25.20.9092

Reinberg, D., & Sims, R. J. (2006). de FACTo nucleosome dynamics. *The Journal of biological chemistry*, *281*(33), 23297–301. doi:10.1074/jbc.R600007200

Reinke, H., & Hörz, W. (2003). Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Molecular cell*, *11*(6), 1599–607. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12820972

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science*, *290*(5500), 2306–9. doi:10.1126/science.290.5500.2306

Richard, P. (2003). The rhythm of yeast. *FEMS Microbiology Reviews*, *27*(4), 547–557. doi:10.1016/S0168-6445(03)00065-2

Rufiange, A., Jacques, P.-E., Bhat, W., Robert, F., & Nourani, A. (2007). Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *Molecular cell*, *27*(3), 393–405. doi:10.1016/j.molcel.2007.07.011

Rutter, J., Reick, M., & McKnight, S. L. (2002). Metabolism and the control of circadian rhythms. *Annual review of biochemistry*, *71*, 307–31. doi:10.1146/annurev.biochem.71.090501.142857

Saha, A., Wittmeyer, J., & Cairns, B. R. (2002). Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes & development*, *16*(16), 2120–34. doi:10.1101/gad.995002

Schaff-Gernstensclager, I., Mannhaupt, G., Vetter, I., Zimmerman, F. K., & Feldmann, H. (1993). TKL2, a second transketolase gene of Saccharomyces cerevisiae. *European Journal of Biochemistry*, *217*(1), 487–492. doi:10.1111/j.1432-1033.1993.tb18268.x

Schmid, M., Poulsen, M. B., Olszewski, P., Pelechano, V., Saguez, C., Gupta, I., Steinmetz, L. M., et al. (2012). Rrp6p controls mRNA poly(A) tail length and its decoration with poly(A) binding proteins. *Molecular cell*, *47*(2), 267–80. doi:10.1016/j.molcel.2012.05.005

Shearwin, K. E., Callen, B. P., & Egan, J. B. (2005). Transcriptional interference--a crash course. *Trends in genetics : TIG*, *21*(6), 339–45. doi:10.1016/j.tig.2005.04.009

Sikorski, T. W., & Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Current opinion in cell biology*, *21*(3), 344–51. doi:10.1016/j.ceb.2009.03.006

Silverman, S. J., Petti, A. a, Slavov, N., Parsons, L., Briehof, R., Thiberge, S. Y., Zenklusen, D., et al. (2010). Metabolic cycling in single yeast cells from unsynchronized steady-state populations limited on glucose or phosphate. *PNAS*, *107*(15), 6946–51. doi:10.1073/pnas.1002422107

Slavov, N., Airoldi, E. M., Van Oudenaarden, A., & Botstein, D. (2012). A conserved cell growth cycle can account for the environmental stress responses of divergent eukaryotes. *Molecular biology of the cell*, *23*(10), 1986–97. doi:10.1091/mbc.E11-11-0961

Slavov, N., & Botstein, D. (2011). Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast. *Molecular biology of the cell*, *22*(12), 1997–2009. doi:10.1091/mbc.E11-02-0132

Sleutels, F., Zwart, R., & Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, *415*(6873), 810–3. doi:10.1038/415810a

Song, Q., & Kumar, A. (2012). An Overview of Autophagy and Yeast Pseudohyphal Growth: Integration of Signaling Pathways during Nitrogen Stress. *Cells*, *1*(4), 263–283. doi:10.3390/cells1030263

Soranzo, N., Zampieri, M., Farina, L., & Altafini, C. (2009). mRNA stability and the unfolding of gene expression in the long-period yeast metabolic cycle. *BMC systems biology*, *3*, 18. doi:10.1186/1752-0509-3-18

Steiner, S., Philippsen, P., Gieben, U., & Str, F. (1994). Sequence and promoter analysis of the highly expressed TEF gene of the filamentous fungus Ashby gossypii. *Molecular genetics and genomics*, *242*, 263–271.

Steinmetz, E. J., Warren, C. L., Kuehner, J. N., Panbehi, B., Ansari, A. Z., & Brow, D. a. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Molecular cell*, *24*(5), 735–46. doi:10.1016/j.molcel.2006.10.023

Strahl, B. D., Grant, P. A., Briggs, S. D., Bone, J. R., Caldwell, J. A., Mollah, S., Cook, R. G., et al. (2002). Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. *Molecular and cellular biology*, *22*(5), 1298–1306. doi:10.1128/MCB.22.5.1298

Struhl, K., & Segal, E. (2013). Determinants of nucleosome positioning. *Nature structural & molecular Biology*, *20*(3), 267–273. doi:10.1038/nsmb.2506

Svaren, J., & Horz, W. (1997). Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. *Trends in biochemical science*, *22*(3), 93–97.

Tachibana, C., Yoo, J. Y., Tagne, J., Kacherovsky, N., Lee, T. I., & Young, E. T. (2005). Combined Global Localization Analysis and Transcriptome Data Identify Genes That Are Directly Coregulated by Adr1 and Cat8 Combined Global Localization Analysis and Transcriptome Data Identify Genes That Are Directly Coregulated by Adr1 and Cat8 †. *Molecular and cellular biology*, *25*(6), 2138–2146. doi:10.1128/MCB.25.6.2138

Taft, R. J., & Mattick, J. S. (2003). Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. *Genome biology*, *5*(1), PI 1–PI 22.

Teichmann, S. A., & Veitia, R. A. (2004). Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. *Genetics*, *167*(4), 2121–5. doi:10.1534/genetics.103.024505

Teixeira, M. C., Monteiro, P., Jain, P., Tenreiro, S., Fernandes, A. R., Mira, N. P., Alenquer, M., et al. (2006). The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. *Nucleic acids research*, *34*(Database issue), D446–51. doi:10.1093/nar/gkj013

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., & Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Molecular cell*, *23*(6), 853–64. doi:10.1016/j.molcel.2006.07.029

Tourrière, H., Chebli, K., & Tazi, J. (2002). mRNA degradation machines in eukaryotic cells. *Biochimie*, *84*(8), 821–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12457569

Traven, A., Jelicic, B., & Sopta, M. (2006). Yeast Gal4: a transcriptional paradigm revisited. *EMBO reports*, *7*(5), 496–9. doi:10.1038/sj.embor.7400679

Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., & Myers, R. M. (2004). An abundance of bidirectional promoters in the human genome. *Genome research*, *14*(1), 62–6. doi:10.1101/gr.1982804

Tsubota, T., Berndsen, C. E., Erkmann, J. a, Smith, C. L., Yang, L., Freitas, M. a, Denu, J. M., et al. (2007). Histone H3-K56 acetylation is catalyzed by histone chaperone-dependent complexes. *Molecular cell*, *25*(5), 703–12. doi:10.1016/j.molcel.2007.02.006

Tu, B. P., Kudlicki, A., Rowicka, M., & McKnight, S. L. (2005). Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science*, *310*(5751), 1152–8. doi:10.1126/science.1120499

Tu, B. P., & McKnight, S. L. (2006). Metabolic cycles as an underlying basis of biological oscillations. *Nature reviews. Molecular cell biology*, *7*(9), 696–701. doi:10.1038/nrm1980

Tutucci, E., & Stutz, F. (2011). Keeping mRNPs in check during assembly and nuclear export. *Nature reviews. Molecular cell biology*, *12*(6), 377–384.

Uhler, J. P., Hertel, C., & Svejstrup, J. Q. (2007). A role for noncoding transcription in activation of the yeast PHO5 gene. *PNAS*, *104*(19), 8011–6. doi:10.1073/pnas.0702431104

Van Dijk, E. L., Chen, C. L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature*, *475*(7354), 114–7. doi:10.1038/nature10118

Van Werven, F. J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., Van Oudenaarden, A., Primig, M., et al. (2012). Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell*, *150*(6), 1170–81. doi:10.1016/j.cell.2012.06.049

Vanácová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., et al. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS biology*, *3*(6), e189. doi:10.1371/journal.pbio.0030189

Vasiljeva, L., & Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Molecular cell*, *21*(2), 239–48. doi:10.1016/j.molcel.2005.11.028

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., & Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature structural & molecular biology*, *15*(8), 795–804. doi:10.1038/nsmb.1468

Venkatesh, S., Smolle, M., Li, H., Gogol, M. M., Saint, M., Kumar, S., Natarajan, K., et al. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature*, *489*(7416), 452–5. doi:10.1038/nature11326

Venters, B. J., & Pugh, B. F. (2009a). How eukaryotic genes are transcribed. *Critical reviews in biochemistry and molecular biology*, *44*(2-3), 117–41. doi:10.1080/10409230902858785

Venters, B. J., & Pugh, B. F. (2009b). A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome research*, *19*(3), 360–71. doi:10.1101/gr.084970.108

Vinciguerra, P., & Stutz, F. (2004). mRNA export: an assembly line from genes to nuclear pores. *Current opinion in cell biology*, *16*(3), 285–292. Retrieved from http://europepmc.org/abstract/MED/15145353

Vodala, S., Abruzzi, K. C., & Rosbash, M. (2008). The Nuclear Exosome and Adenylation Regulate Post-Transcriptional Tethering of Yeast GAL genes to the Nuclear Periphery. *Molecular cell*, *31*(1), 104–113. doi:10.1016/j.molcel.2008.05.015.The

Wilusz, C. J., Wormington, M., & Peltz, S. W. (2001). The cap-to-tail guide to mRNA turnover. *Nature reviews. Molecular cell biology*, *2*(4), 237–46. doi:10.1038/35067025

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*, *121*(5), 725–37. doi:10.1016/j.cell.2005.04.030

Xu, Z., Wei, W., Gagneur, J., Clauder-Münster, S., Smolik, M., Huber, W., & Steinmetz, L. M. (2011). Antisense expression increases gene expression variability and locus interdependency. *Molecular systems biology*, *7*(468), 468. doi:10.1038/msb.2011.1

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., et al. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, *457*(7232), 1033–7. doi:10.1038/nature07728

Youdell, M. L., Kizer, K. O., Kisseleva-Romanova, E., Fuchs, S. M., Duro, E., Strahl, B. D., & Mellor, J. (2008). Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. *Molecular and cellular biology*, *28*(16), 4915–26. doi:10.1128/MCB.00001-08

Young, E. T., Dombek, K. M., Tachibana, C., & Ideker, T. (2003). Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. *The Journal of biological chemistry*, *278*(28), 26146–58. doi:10.1074/jbc.M301981200

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A. P., & Cui, H. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, *451*(7175), 202–6. doi:10.1038/nature06468

Yuan, G.-C., Liu, Y.-J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J., & Rando, O. J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. *Science*, *309*(5734), 626–30. doi:10.1126/science.1112178

Zenklusen, D., Larson, D. R., & Singer, R. H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature structural & molecular biology*, *15*(12), 1263–71. doi:10.1038/nsmb.1514

Zenklusen, D., & Singer, R. H. (2010). Analyzing mRNA expression using single mRNA resolution fluorescent in situ hybridization. *Methods in enzymology*, *470*(10), 641–59. doi:10.1016/S0076-6879(10)70026-4

Zhang, H., Roberts, D. N., & Cairns, B. R. (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*, *123*(2), 219–31. doi:10.1016/j.cell.2005.08.036

Zhao, J., Herrera-diaz, J., & Gross, D. S. (2005). Domain-Wide Displacement of Histones by Activated Heat Shock Factor Occurs Independently of Swi/ Snf and Is Not Correlated with RNA Polymerase II Density. *Molecular and Cellular Biology*, *25*(20), 8985–8999. doi:10.1128/MCB.25.20.8985