ANTISENSE TRANSCRIPTION REGULATES GENE EXPRESSION IN

SACCHAROMYCES CEREVISIAE

By

BRIAN NATHAN GELFAND

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Program in Microbiology and Molecular Genetics

written under the direction of Professor Andrew K. Vershon

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2011

ABSTRACT OF THE DISSERTATION

Antisense Transcription Regulates Gene Expression in *Saccharomyces cerevisiae*

By Brian Gelfand

Dissertation Director:
Dr. Andrew K. Vershon

The yeast, *Saccharomyces cerevisiae,* has previously been shown to have a transcriptome comprising over 6,000 protein coding genes, as well as over 900 non-coding RNAs (ncRNAs). This work focuses on the regulation of the gene *IME4*, which is normally expressed only in **a**/α diploid cells. In haploid cells, an ncRNA, which I have termed Regulator of Meiosis 2, (*RME2*) is expressed from the antisense strand of *IME4*. *RME2* has a direct role in the haploid-specific repression of *IME4*. I have shown that *RME2* represents a novel class of *cis*-acting non-coding RNA regulators, as it does not regulate in a *trans*-acting mechanism like microRNA. Furthermore, *RME2* represses *IME4* transcription in a promoter-independent mechanism, as transcription factor binding at the *IME4* promoter is not perturbed. Regulation by *IME4* does appear to require transcription across specific sequences within the *IME4* ORF itself; in absence of the required sequences, *IME4* and *RME2* are co-expressed. In addition to *IME4*, this work details another meiotic gene, *ZIP2*, which was found to have a similar regulatory antisense transcript, *RME3*. Like *RME2, RME3* represses in a *cis*- acting mechanism.

To determine if other genes are regulated in a manner similar to *IME4* and *ZIP2*, strand-specific RNA-sequence analysis was used to compare the antisense transcriptomes of *MAT***a** and **a**/α cells grown vegetatively and in early meiotic conditions. This analysis identified over 1400 antisense ncRNAs, including 147 cases where antisense ncRNAs are differentially expressed in relation to the sense transcripts in different cell-types or

growth conditions.   Another subset of 65 genes express antisense ncRNA but not the sense transcript in the four conditions assayed here.   These genes may be regulated by antisense transcription, and derepressed under other environmental conditions. There are also over 300 examples of genes that express both sense and antisense transcripts at similar levels. In these cases, the antisense transcript may not have a role in regulating coding expression. This work shows that the non-coding transcriptome has an important role in differential cellular responses, and suggests *cis*-acting antisense transcription may be a widespread mechanism of regulation.

ACKNOWLEDGEMENTS

DEDICATION

This thesis is dedicated to my parents, Arthur and Barbara Gelfand, who have supported

me in my work, and pushed me to succeed and believed in me, even in times when I did

not think I could.

LIST OF ABBREVIATIONS

bp	Base Pair
cDNA	complementary DNA
ChIP	Chromatin Immunopreciptiation
DEPC	Diethyl Pyrocarbonate
EDTA	Ethylenediaminetetraacetic acid
IP	Immunoprecipitate
Kb	Kilobase
miRNA	microRNA
MOPS	3-(N-morpholino)propanesulfonic acid
ncRNA	non-coding RNA
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
RISC	RNA-induced silencing complex
RT-PCR	Reverse Transcriptase PCR
SDS	Sodium Dodecyl Sulfate
siRNA	Small-interfering RNA
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
UV	ultraviolet
TAE	Tris Acetate EDTA
TBP	TATA Binding Protein
TC	Total Chromatin
TES	Tris -EDTA -Sodium Dodecyl Sulfate

YEPD   Yeast extract peptone dextrose

YEPA   Yeast extract peptone acetate

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

The yeast *S. cerevisiae* is useful as a model eukaryotic system for its simplicity of growth and ease of genetic manipulation. *S. cerevisiae* possesses many of the same basic cellular processes found in higher eukaryotes, such as transcriptional regulation, RNA processing, gene splicing, and a meiotic phase of their life cycle. Many of the genes and proteins involved in these pathways are highly conserved in structure and function from *S. cerevisiae* into higher eukaryotes. Therefore, the molecular mechanisms of gene regulation in yeast may be extrapolated to function in other eukaryotes. This allows us to use yeast as a molecular model for human diseases caused by genetic dysfunction.

1. Yeast life cycle.

S. cerevisiae lives as three cell types; haploid (1n) **a** and α, and diploid (2n) **a**/α. The **a** and α haploid cells express different cell-type specific genes involved in mating. When two haploid cells of opposite mating type come into contact, they respond to the opposite-type mating pheromone, and undergo conjugation, forming the **a**/α diploid cell (Fig. 1). All three cell-types grow and divide mitotically under conditions where carbon and nitrogen are plentiful. When cells are starved, haploid cells will enter cell-cycle arrest and eventually die. In contrast, **a**/α diploid cells have a different response to the insufficient carbon and nitrogen. Cell-type specific gene expression changes unique to **a**/α diploid cells permit them to enter the meiosis and sporulation pathway. This leads to the production of four haploid spores, two *MATa* and two *MATα,* that are resistant to environmental stress. The spores will germinate when conditions improve, meaning sufficient water, carbon, and nitrogen, allowing them to

Figure 1: Yeast life cycle. Haploid *MAT*a**, *MAT*α*, and **a**/α diploid cells grow and divide mitotically in rich media. Mitotic division continues until starvation conditions (low carbon/low nitrogen). **a**/α diploid cells can undergo meiosis and sporulation, and their haploid spores start the cycle again when carbon, nitrogen, and water are detected.

α
(1N)

a
(1N)

Mating
+Conjugation

a/α

a/α
(2N)

Rich Media

Poor Media

Meiosis

Sporulation

Germination

Spores

Ascus
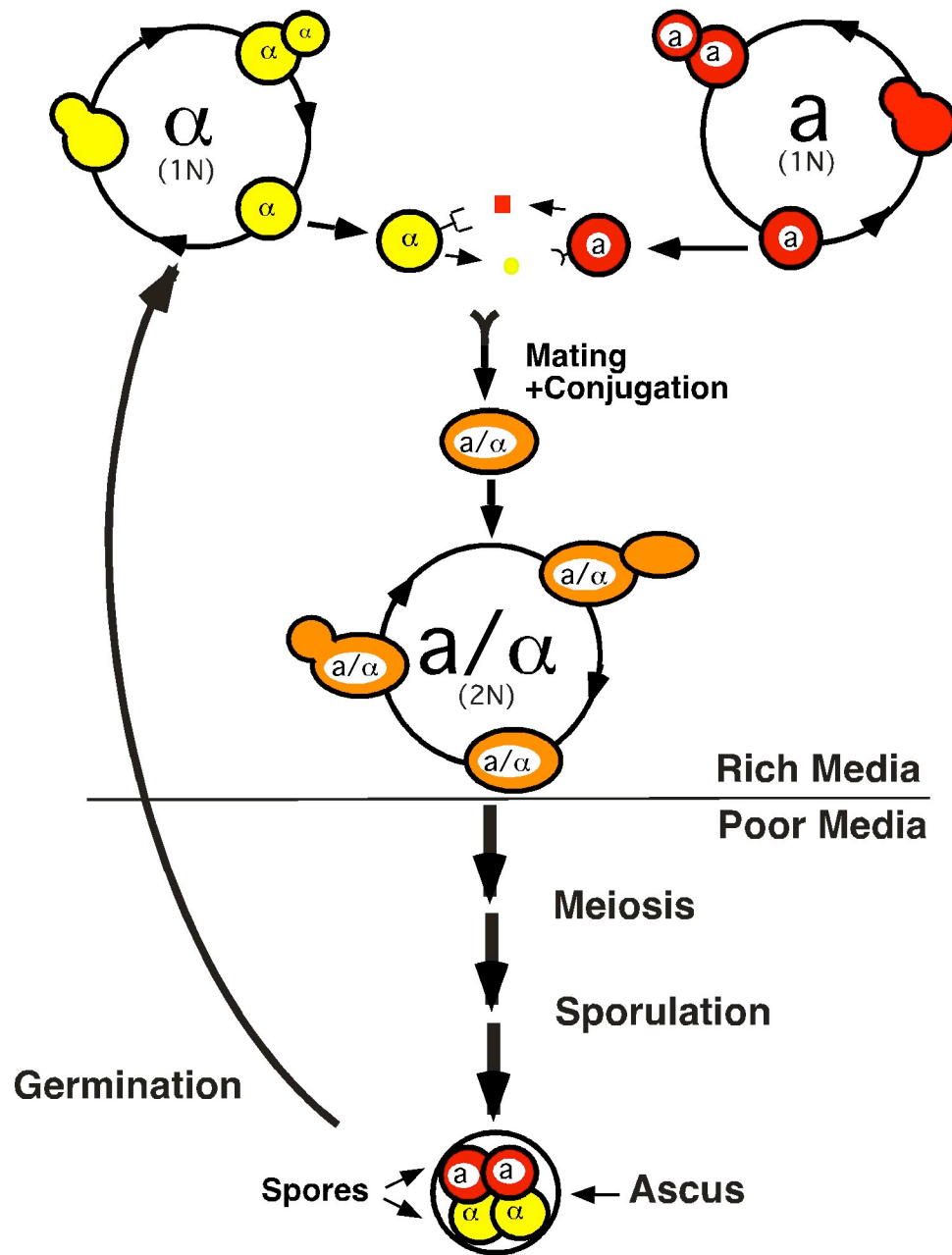
resume mitotic division, and starting the life cycle again. Many of the proteins and regulatory mechanisms involved in the different stages of cell cycle, cell-type, differentiation, and meiosis are similar to those in found higher eukaryotes. Therefore, studies of *S. cerevisiae* can provide information on the conserved systems found in higher eukaryotes, with the benefit of simpler genetic manipulation and faster growth.

While the three cell types in yeast express a large number of genes in common, each one also has a set of genes unique to a particular cell type, which are regulated by cell-type specific transcription factors. Mcm1, as a homodimer, activates **a**-specific genes in **a** cells; these same genes are repressed in the other two cell types by the Mcm1-α2 heterotetramer complex (Fig. 2) The α–specific genes are activated by the Mcm1-α1 complex, and are not expressed in the other cell types (Galgoczy, et al., 2004). Diploid **a**/α cells express both the **a**1 and α2 proteins to form the diploid-specific heterodimer repressor complex, **a**1-α2. This complex binds to a specific sequence, which is found in the promoters of many haploid-specific genes (Fig. 2) (Goutte and Johnson, 1988; Jin, et al., 1995; Nagaraj, et al., 2004). The **a**1-α2 complex recruits other proteins (including Tup1, Ssn6, and Sin4) that function in concert in **a**/α diploid cells to repress this of group haploid-specific genes, which includes *HO* (mating-type switching); *STE4, STE5, STE18* (mating); and *GPA1* (pheromone response). This prevents diploid cells from engaging in processes (i.e. mating) which would be biologically disadvantageous to the cell (Goutte and Johnson, 1988; Oshima and Campisi, 1991; Strathern, et al., 1981).

Figure 2: Transcriptional regulatory protein complexes that regulate cell-type specific genes in yeast. DNA-binding activator complexes are indicated by an arrow in the direction of transcription. Repression is indicated by a -| in the direction of transcription. Mcm1 activates **a**-specific genes in *MAT***a** cells, and α2-Mcm1 represses **a**-specific genes in *MAT*α and **a**/α cells. The α1-Mcm1 complex activates α–specific genes in *MAT*α cells. Haploid-specific genes are repressed by the **a**1-α2 complex in **a**/α cells.
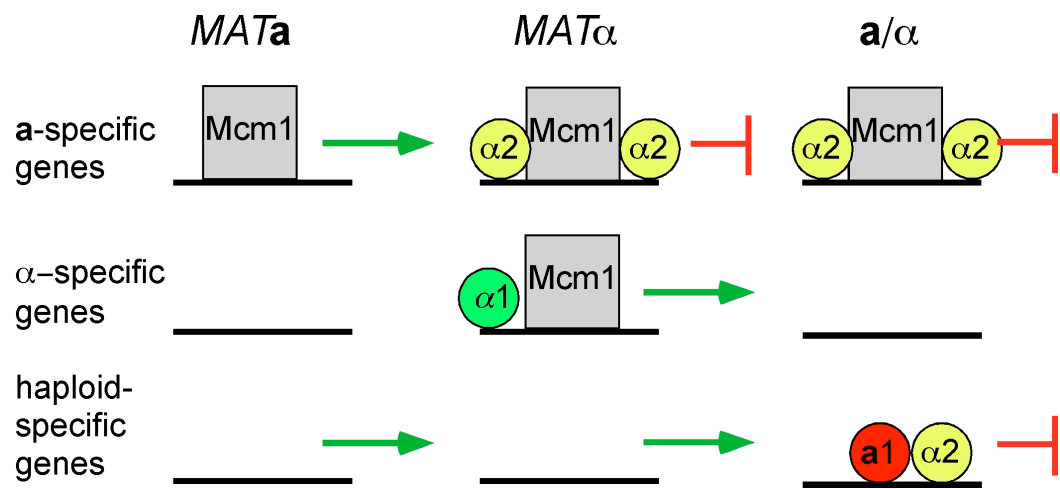
2. Meiosis and Sporulation

The **a**1-α2 complex is also involved in the regulation of meiosis-specific genes in yeast. The process of meiosis and sporulation in yeast involves a cascade of expression changes for over 500 genes (Chu, et al., 1998). It is also one of the most highly regulated processes in the cell. Sporulation must be prevented in haploid cells because they will not undergo proper chromosomal segregation for meiosis, which will kill the daughter cells. Entry into sporulation in diploid cells must be prevented when nutrient conditions are good, because growth and mitotic division are favorable.

In diploid cells, entry into the sporulation pathway is controlled by expression of diploid-specific kinase, *IME1*, which is the master initiator of meiosis (Covitz, et al., 1991; Kassir, et al., 1988; Mitchell, et al., 1990). Two classes of inputs regulate *IME1*: starvation signals, which are activated by the nutritional status of the cell, and cell type-specific responses (Fig. 3). The **a**1-α2 complex regulates cell type-specific control of *IME1* through two different pathways. One pathway is regulated via the haploid-specific gene *RME1* (Covitz, et al., 1991; Mitchell and Herskowitz, 1986). In haploid cells, the Rme1 repressor is expressed and binds to the promoter of *IME1,* acting as a transcriptional repressor (Mitchell, et al., 1990). This prevents haploid cells from entering into the meiotic pathway. In diploid cells, **a**1-α2 complex binds to the promoter of *RME1,* repressing its transcription, thereby relieving repression of *IME1* (Covitz, et al., 1991).

Figure 3. Nutritional and cell-type specific inputs regulate meiotic entry. Genes

expressed in haploid cells are indicated in red. Diploid (**a**/α) –specific genes are indicated in

orange. Activation is indicated by an arrow, and repression of the target gene is indicated

with a -|. *IME1*, the master regulator of meiosis is regulated by two classes of input:

nutritional and cell-type specific. In *MAT***a** and *MAT*α cells, *RME1* binds to the *IME1*

promoter, acting as a transcriptional repressor. In **a**/α cells, *RME1* is repressed by **a**1-α2,

allowing *IME1* activation. *IME1* activation also requires the **a**1-α2-regulated gene *IME4*. It

was speculated that a haploid-specific gene "X", repressed by **a**1-α2, represses *IME4* (Shah

and Clancy, 1992).

Sporulation and meiotic genes
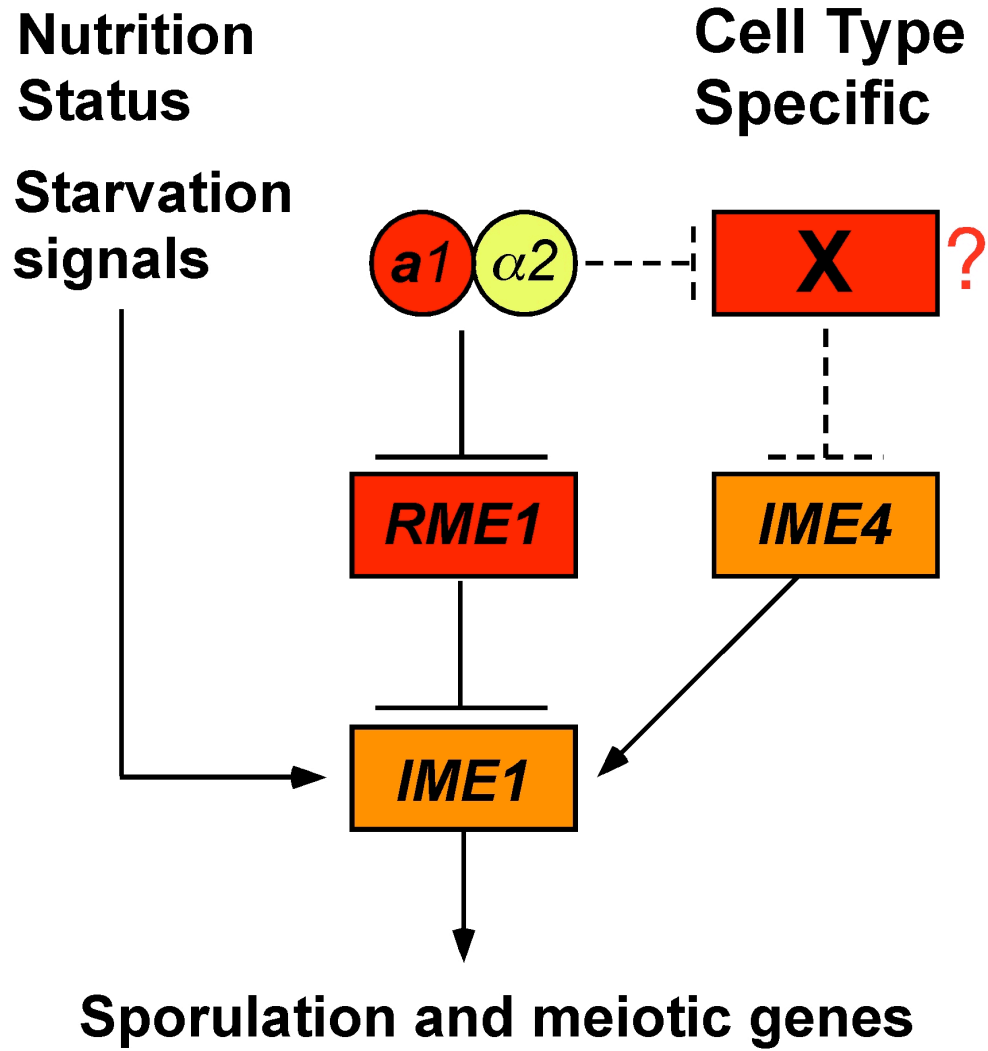
There is a second cell type-specific input required to initiate sporulation. The *IME4* gene

was originally identified as a mutation called *spo8*, which was defective in the diploid cell's

ability to sporulate (Esposito and Esposito, 1974; Esposito and Esposito, 1975). It was later

identified as *IME4* (**I**nitiator of **Me**iosis), an activator of meiosis in diploid cells (Shah and

Clancy, 1992). The *IME4* gene is only expressed at high levels in **a**/α diploid cells, and not

in cells which lack the **a**1-α2 complex (haploid, **a**/**a** diploids, or α/α diploids). Therefore, it

appeared as though the activation of *IME4* required the **a**1-α2 complex. However, the

biological function of the **a**1-α2 complex had been previously established as being a

transcriptional repressor, of haploid-specific genes (Goutte and Johnson, 1988). It was

assumed that a second intermediary haploid-specific repressor, was acting in between *IME4*

and **a**1-α2 in the pathway to repress *IME4* in haploid cells. This repressor, indicated by the

"**X**" in Figure 3, in turn would be directly repressed by the **a**1-α2 complex, similar to *RME1*,

in diploid cells under conditions of sporulation. However, work to identify a protein cofactor

did not identify what this mystery factor "**X**" was.   The question remained for how the **a**1-

α2 complex acts as both an activator and a repressor of different genes.

3. Identification of **a**1-α2 sites in the yeast genome.

Identification of **a**1-α2 sites in the yeast genome provided further information on

the regulation of *IME4* (Nagaraj, et al., 2004). To identify the target sites, an algorithm

that combined **a**1-α2 binding site preference data, with microarray data showing genes

strongly expressed in haploid cells, but not in diploid cells (Nagaraj, et al., 2004). This

screen correctly identified many of the known targets of **a**1-α2 repression, such as *HO,*

STE5, and *FUS3*. The screen also identified a number of novel target genes, including

RDH54 and *MET31,* which are expressed at comparatively lower levels in diploid cells, due to the control of **a**1-α2 binding in their promoter regions.

Additionally, the screen identified an **a**1-α2 binding site upstream of the *COX13* gene. *COX13* encodes a cytochrome C oxidase that did not appear to be differentially expressed in haploid and diploid cell types (Galitski, et al., 1999; Yin, et al., 2003). An Electrophoretic Mobility Shift Assay (EMSA) illustrated that the **a**1-α2 complex is able to bind to this DNA sequence *in vitro,* as evidenced by the shifted bands in lanes 8-11 (Fig.4A). The amount of shifted DNA appears to be similar to the amount seen for the known strong binding site *HO(6)* (Fig. 4A, lanes 4-7) (Mathias, et al., 2004). To test whether the **a**1-α2 complex binds to this site *in vivo*, a Chromatin Immunoprecipitation (ChIP) assay using antibody to α2 was performed (Fig 4B). A band corresponding to the *COX13* promoter site was amplified from ChIP material from the **a**/α strain, but not the Δ/α strain, as seen in the *COX13* test bands from the IP lanes of Fig 4B. Because the Δ/α IP did not amplify the *COX13* site, it indicates that α2 is only bound to the site in complex with **a**1, and not in complex with Mcm1. Finally, to test if this site was functional as a repressor site, it was cloned into a heterologous promoter that was driving expression of *lacZ*. A β-galactosidase assay showed that this site functions as a very strong repressor site in the context of a heterologous reporter, with 143x fold repression of the downstream gene, compared to a *lacZ* construct without an **a**1-α2 binding site in the promoter (Fig. 4C). Taken together, this data showed that the *COX13* site is a functional **a**1-α2 repressor site. When analyzed in the context of the adjacent loci, we

Figure 4. The a1-α2 complex binds in the *COX13/IME4 i*ntergenic region. (A)

EMSA of DNA containing the **a**1-α2 binding site found upstream of *HO(6)* (lanes 4-7) or

COX13 (lanes 8-11). The shifts for unbound radiolabeled DNA and bound fraction are

indicated. The upper band is a degradation product. Lane 1 is a no-probe control, lane 2

is **a**1 alone, and lane 3 is α2 alone . (B) ChIP assay of the *COX13* site. TC is total

Chromatin, IP is DNA immunoprecipitated with anti-α2 antibody. *STE6* is a known

target of α2-Mcm1 and *YDL223C* is a negative control for -α2 binding. (C) β-

galactosidase assay of a heterologous reporter containing the *HO(6)* or *COX13* **a**1-α2

site, transformed in an **a**/α cell, as compared to the vector without a site.

A

HO(6) COX13

a1 α2

a1α2

a1 α2

*

*

1 2 3 4 5 6 7 8 9 10 11

B

a/α Δ/α

TC IP IP TC

COX13

STE6

YDL223C

C

lacZ

Fold repress. | *HO(6)* 43x | *COX13* 143x |
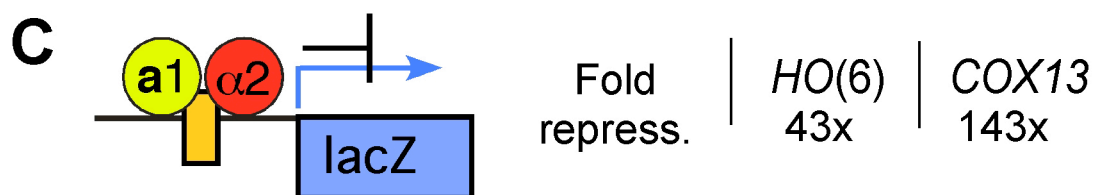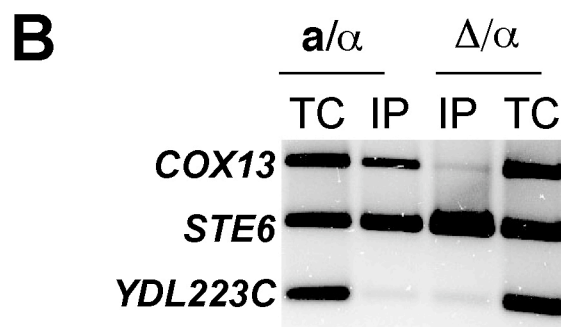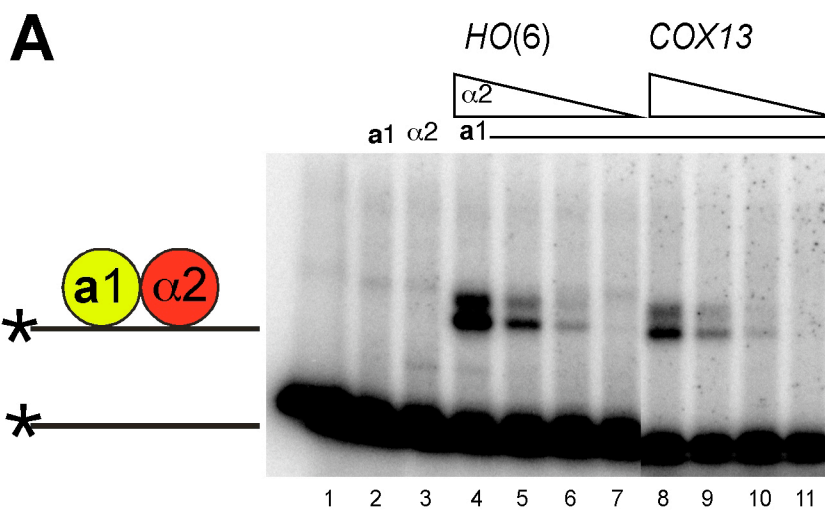
found the site listed as upstream of *COX13* is actually closer to the 3' end of the *IME4* open reading frame, than to the 5' end of *COX13* (125 vs. 576 base pairs) (Fig. 5). This data was the first suggestion of a direct relationship between *IME4* and **a**1-α2.

The possibility that the downstream **a**1-α2 site somehow acts as a direct activator of *IME4* by looping or binding to the promoter was tested. It was found that there was no discernable interaction between **a**1-α2 and the promoter DNA of *IME4 (*Nagaraj, unpublished). In reexamining the previous work on *IME4,* it was found that strand-specific Northern Blotting detected the presence of a second transcript, expressed from the antisense strand of *IME4* (Shah and Clancy, 1992). Analysis of the antisense strand of *IME4* using the ORF finder (**http://www.ncbi.nlm.nih.gov/projects/gorf/orfig.cgi**) did not show any significant ORFs in this direction, so it is unlikely that this transcript codes for a protein, which could have been a potential candidate for the repressor "X". We hypothesize that expression of this antisense transcript itself, which we refer to as *RME2* (Repressor of Meiosis 2), is the inhibitor of *IME4* expression in haploid cells, and not a regulatory protein. The position of the **a**1-α2 binding site in the *IME4* intergenic region suggests that it may be in the *RME2* promoter region; therefore, **a**1-α2 would repress *RME2*. If this is true, when *RME2* is expressed it may repress *IME4*, which in turn prevents induction of *IME1*. In diploid cells, the **a**1-α2 complex binds at the *RME2* promoter, repressing *RME2*, and allowing *IME4* expression. Chapter 3 describes the analysis of this ncRNA and its role in regulating *IME4*.

Figure 5: Cartoon illustrating the *IME4* and *COX13* loci, and proximity to the a1-α2 binding site. The *IME4* and *COX13* ORFs are indicated in yellow. RNA transcripts are indicated with green arrows. The **a**1-α2 site is 125 bp from the end of *IME4,* and 576 bp from the end of *COX13*.

IME4 mRNA

COX13 mRNA

IME4

COX13

a1-α2 binding site
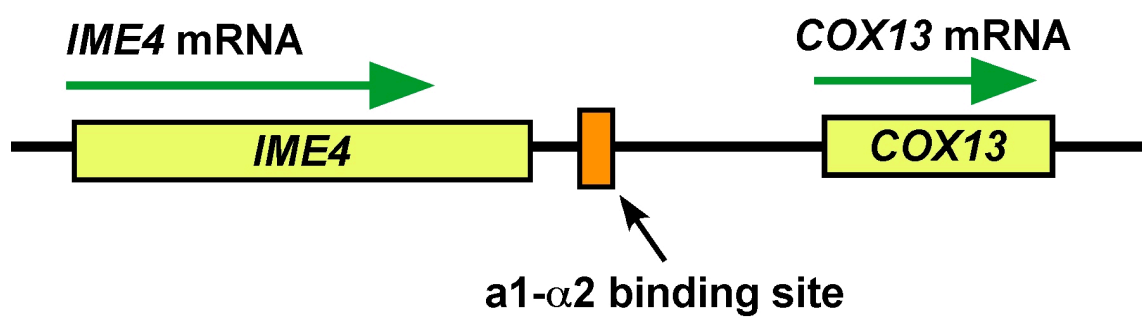
4. RNA mediated regulation

One of the canonical mechanisms of gene regulation in eukaryotes has been is the RNA interference (RNAi) model, or quelling, as it is known in filamentous fungi (He and Hannon, 2004; Pickford, et al., 2002). In these systems, 22-25 bp antisense RNA fragments known as microRNAs (miRNAs) anneal to coding RNA, leading to targeted degradation (Fig. 6A) (Cogoni and Macino, 1997; He and Hannon, 2004; Mallory, et al., 2004). This process also requires the presence of three proteins. Drosha and Dicer cleave the initial pre-miRNA transcript to a mature 22-25 bp miRNA (He and Hannon, 2004). RISC is a complex that cleaves the miRNA-mRNA duplex. RISC also activates gene silencing when bound by the miRNA-mRNA duplex (Couzin, 2002; He and Hannon, 2004). Interestingly, while Dicer, Drosha, and RISC homologues have been identified in many eukaryotes, including the yeast *Schizosaccharomyces pombe*, they have not been identified in *S. cerevisiae* (Drinnenberg, et al., 2009). If *trans*-acting antisense-mediated regulation occurs in *S. cerevisiae,* it could utilize proteins with similar biological functions to RNAi machinery (such as RNAses) in a non-canonical mechanism. It is also possible that dsRNA-mediated gene repression works through an entirely different *trans*-acting mechanism which is not similar to RNAi.

In addition to the miRNA and siRNA pathways, there are *trans*- acting ncRNAs which regulate gene expression in other ways. One example is *kcnq1ot1,* an antisense RNA which has a role in the silencing of *Kcnq1* from the paternal locus in mice and humans (Fig. 6B.) (Kanduri, et al., 2006; Rump, et al., 2005; Thakur, et al., 2004). Only the

Figure 6. Antisense mediated regulation (A) The *LET7* microRNA, a 22 bp antisense RNA, binds in *–trans* to the *LIN-41* mRNA, leading to targeted degradation and silencing. (B) In mouse maternal cells, *KCNQ1* is expressed (Top). In paternal cells, an intronic antisense ncRNA, *Kcnq1ot1* is expressed instead. Transcription of *Kcnq1ot1* silences *Kcnq1* and surrounding genes due to methylation changes.

A

lin-41-mRNA

lin-41

LET7 pre-miRNA

LET7

lin41-mRNA

hairpin fold

Drosha

Dicer

LET7 miRNA

dsRNA

target for:
degradation
silencing

B

Maternal Chromosome

KCNQ1 mRNA

KCNQ1

Paternal Chromosome

KCNQ1

KCNQ1ot1 nc-RNA
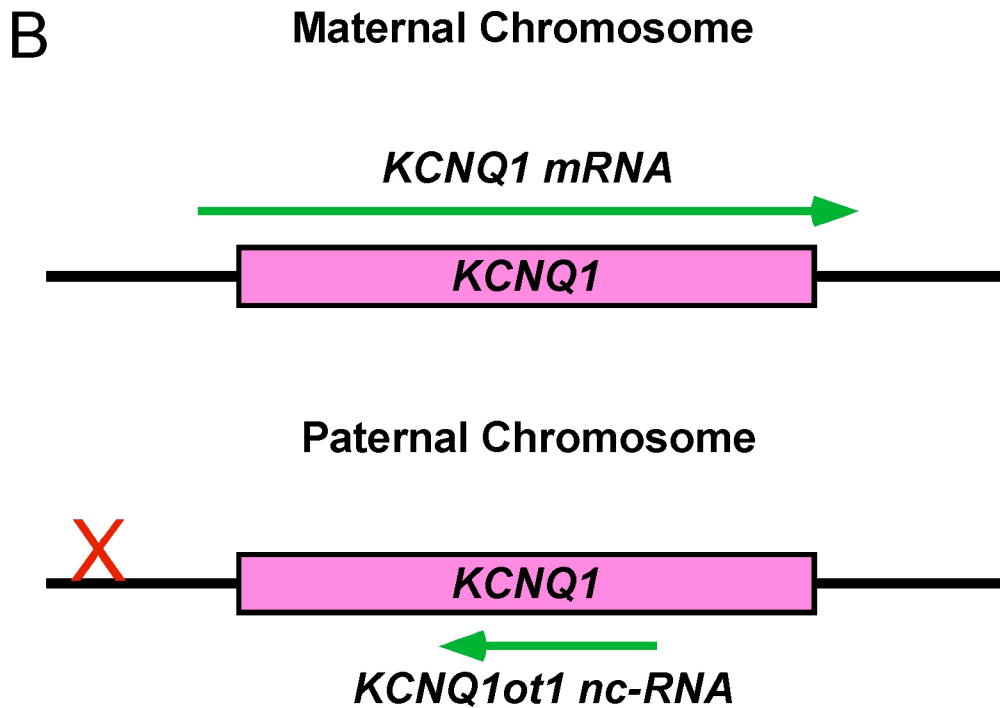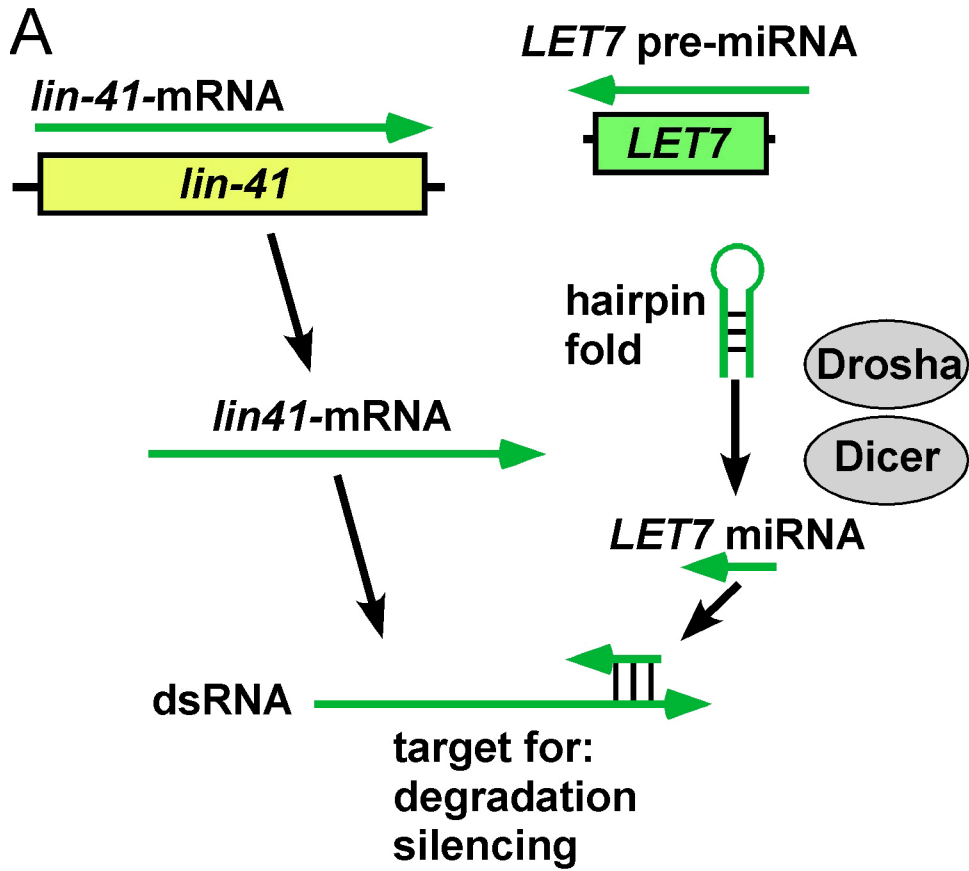
maternal allele expresses *Kcnq1;* failure to silence the paternal copy is associated with Beckwith-Wiedemann Syndrome and colonic cancers (Rump, et al., 2005).

Curiously, *S. cerevisiae* lacks the endogenous ability to perform RNAi regulation utilizing short ncRNA, such as miRNAs or siRNAs (Drinnenberg, et al., 2009). However, there are examples known in which antisense transcripts affect coding gene expression in yeast. RNA fragments antisense to the cell-cycle gene *YBR1012* were found to reduce sense RNA levels (Nasr, et al., 1994). This *trans-* acting antisense RNA led to inhibited cell cycle progress and a slow growth phenotype (Nasr, et al., 1995; Nasr, et al., 1994). A similar artificial antisense methodology was employed to repress the *CAR1* gene. Expression of a *CAR1* antisense transcript in *trans* reduced *CAR1* expression nearly six-fold (Park, et al., 2001). However, attempts to use a similar method to repress the gene *MIG1* were not successful (Olsson, et al., 1997). This indicates that there is more to this regulatory mechanism than simply the presence of an antisense RNA.

More recent studies have found endogenous examples of *trans-* acting RNA in *S. cerevisiae*. A transcript antisense to the *TY1-* locus is expressed from distant loci in *S. cereivisiae* (Berretta, et al., 2008). This ncRNA binds to the *TY1* retrotransposon, inducing changes in chromatin structures that repress transcription and impair transposition mobility. The *PHO84* locus also expresses an antisense transcript in wild-type cells that reduce the amount of *PHO84* mRNA, after long-term storage at low temperatures (Camblong, et al., 2009). Interestingly, high copy, *trans*-expressed *PHO84-antisense* can also repress *PHO84* in the same manner (Camblong, et al., 2009). Regulation in *trans-* therefore may be dependent on specific RNA sequences that form RNA-RNA or RNA-DNA hybrids. There may also be as-yet unknown cofactors required to degrade dsRNA molecules.

It is also possible that ncRNAs work in *cis-* to regulate coding gene expression. For example, transcription of *RME2* may alter the surrounding chromatin structure of the promoter in a manner that prevents sense transcription. Alternately, the antisense RNA itself may act as a physical block to extending sense transcription. This form of regulation has already been proposed for the mouse gene *Kcnq1ot1* (Fig. 6B) (Kanduri, et al., 2006). Transcription of the antisense transcript through the coding region of *Kcnq1* alters the chromatin structure, rendering sense transcription inactive (Kanduri, 2006).

When yeast cells are starved for zinc, the Zap1 transcription factor is induced, and it binds to the promoters of Zinc-response genes to activate transcription (Zhao, et al., 1998). Zap1 was shown to bind upstream of *ADH1*, but *ADH1* transcription is reduced during Zinc depletion (Fig. 7A)(Bird, et al., 2006), It was shown that Zap1 induced expression of the *ZRR1* ncRNA which acts in *cis-* to block access to the *ADH1* promoter (Bird, et al., 2006). The *ADH3* gene was shown to be similarly regulated by the ncRNA *ZRR3*.

A c*is-* acting mechanism of RNA-mediated regulation also controls expression of the *SER3* gene (Fig. 7B). When serine is limited, the *SER3* gene is expressed through the action of an activator that binds to the *SER3* promoter. When serine is plentiful, an upstream, non-coding transcript, *SRG1*, is activated by Swi/Snf bound upstream of *SER3* (Martens, et al., 2004). Transcription of *SRG1* across the *SER3* promoter was found to block binding of TBP, which in turn prevents *SER3* expression (Fig. 7B) (Martens, et al., 2004) .

Figure 7. *ADH1, ADH3,* **and** *SER3* **are regulated in** *-cis* **by ncRNAs.** A) *ADH1* is

expressed in vegetative media. When Zinc is depleted, Zap1 binds to a site upstream of

the promoter, activating *ZRR1*. Transcription of *ZRR1* blocks access to the coding

promoter. *ADH3* is regulated by *ZRR3* in the same manner. B) The *SER3* gene is

expressed when low levels of serine are present in the media. When serine levels are

high, Swi/Snf binds upstream of *SER3*, and activates transcription of *SRG1*. Expression

of *SRG1* blocks access to the *SER3* promoter.

A

Vegetative:

Zinc-depletion:

B

Serine poor media:

Serine rich media:
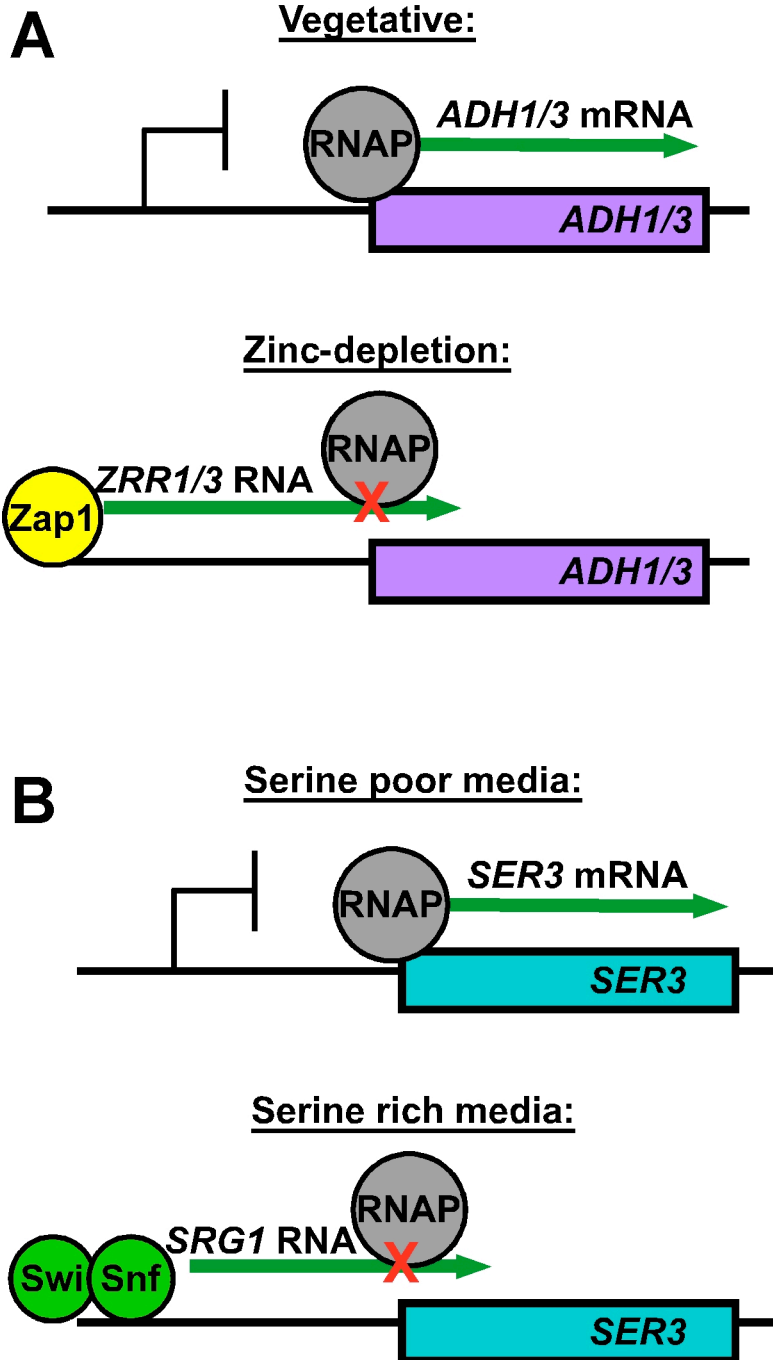
Interestingly, the mechanism of repression of *SER3* was discovered because prior research had shown that the Swi/Snf activator complex was required to repress *SER3* (Martens, et al., 2004). This activator complex binds at the *SRG1* promoter, increasing *SRG1* transcription under conditions of plentiful serine; in turn, *SRG1* represses *SER3* (Martens, et al., 2004; Martens, et al., 2005). This mechanism could be considered analogous to the model for *IME4,* where the **a**1-α2 complex appears to act as an activator for *IME4*. In haploid cells, *RME2* transcription prevents *IME4* activators from binding. Downstream binding of the **a**1-α2 complex in diploid cells represses *RME2,* allowing *IME4* activation.

5. Transcriptome Studies

Detailed RNA profiling using tiled microarrays has shown that there are over 700 examples of overlapping antisense transcripts. It highly likely that the expression of some of these transcripts affects expression of the other strand (David, et al., 2006; Steigele and Nieselt, 2005).   The more recent studies of the yeast transcriptome estimate that there are many different kinds of non-coding genes present, including antisense RNAs, intergenic transcripts, and unstable non-coding transcripts (Gagneur, et al., 2009; Xu, et al., 2009).

Use of existing microarray datasets was able to identify the **a**1-α2 complex binding downstream of *IME4,* which regulates the *RME2* transcript that represses it in haploid cells; this work is detailed in Chapter III.   By revisiting the same **a**1-α2 ChIP dataset and focusing on sites near the 3' ends of genes, instead of 5' ends, and comparing it to the tiled array transcriptome produced by the Huber and Steinmetz laboratories, I

identified the ncRNA *RME3,* which regulates *ZIP2* in a cell-type specific manner, similar

to *IME4.* These experiments are detailed in Chapter IV. Finally, in Chapter 5, I detail a

stranded RNA-seq analysis that compares the transcriptomes of *MAT***a** and **a**/α cell-types

under vegetative and meiosis inducing conditions, to identify an extensive array of cell-

type and condition-specific examples of ncRNA-mediated regulation.

II. MATERIALS AND METHODS

A. Materials

Plasmid construction. A list of the plasmids used in this study is shown in Table 1.

Plasmid pBG1 contains a 600 bp PCR generated fragment of bp +1400 to +2000 bp from

IME4 genomic DNA cloned into the TOPO TA vector (Invitrogen). Site-directed

mutagenesis was used to change four base pairs in the **a**1-α2 binding site (WT:

GTGTATTTTTTACATCA; mu: GTcgATTTTTTTACggCA) to produce plasmid

pBG7. Plasmid pBG113 contains a 2.9-kb PCR generated fragment of bp -450 to +400

bp flanking *IME4* genomic DNA cloned into the TOPO TA vector (Invitrogen). This

plasmid was digested with XbaI and HindIII to clone into the same sites in pRS415 to

generate pBG112, and it was digested with XhoI and HindIII to clone into the same sites

in pRS405 to generate pBG129 (Sikorski and Hieter, 1989).

The *HOP1-urs1* mutant (pBG157) was generated by gap-repair of BglII digested

pBG112 using a 216 bp PCR fragment spanning -216 to -1 of *HOP1* amplified from

pAV124. To generate pBG167, mutagenic PCR was used to change four base pairs in

the **a**1-α2 binding site (WT: TGTATTTTTTACATC; mu: TcgATTTTTTTACggC)

The 225-675 Flip plasmid, pBG166, was generated by gap repair of NdeI digested

pBG112 using a PCR fragment of bp 225-675 of IME4 with 50 bp flanking overhangs of

the opposite end (Muhlrad, et al., 1992). The *IME4* deletion analysis mutants, *rme2-s1*,

and *rme2-s2* were generated by site-directed mutagenesis of pBG129.

Plasmid pJM532 contains a 3.1-kb PCR generated fragment of bp -450 to +400

bp flanking *ZIP2* genomic DNA cloned into the TOPO TA vector (Invitrogen). This

plasmid was digested with SpeI and ApaI to clone *ZIP2* into the same sites in pRS405, to

generate pJM533. Site directed mutagenesis was used to change four base pairs in the

Table 1: Plasmids used in this study.

pAV113	*MAT*α cassette in 2μ *URA3* vector
pAV124	*HOP1-LacZ* in which UAS$_H$ is mutated with a XhoI site
pUG6	*LoxP-KanMX-LoxP* cassette
pSH47	Cre recombinase, 2μ, *URA3*
pJM530	pBG112 with *ime4::URA3*
pJM533	*ZIP2* with -400/+400 bp from *ZIP2* cloned into XhoI/HindIII sites in RS405
pJM534	pJM533 with **a**1-α2 site mutant with 4 bp changes
pBG1	PCR-TOPO 2.1 with bp +1600 to +2200 of *IME4*
pBG7	pBG1 with **a**1-α2 site mutant with 4 bp changes
pBG110	*IME4* with -400/+400 bp from pBG113 cloned into XbaI/HindIII sites in RS415
pBG111	*IME4* with -400/+400 bp from pBG114 cloned into XbaI/HindIII sites in RS415
pBG112	isogenic to pBG111
pBG113	2.9 kb *IME4* fragment with -400/+400 bp flanking in PCR TOPO 2.1 , T7 direction
pBG114	2.9 kb *IME4* fragment with -400/+400 bp flanking in PCR TOPO 2.1 , Sp6 direction
pBG129	XbaI/HindIII fragment from pBG112 cloned into RS405
pBG132	pBG129 with *rme2-s1* mutation
pBG133	pBG129 with *ime4Δ800-1700* mutation
pBG134	pBG129 with *ime4Δ1-900* mutation
pBG135	pBG129 with *ime4Δ1-451* mutation
pBG136	pBG129 with *ime4Δ451-900* mutation
pBG151	pBG129 with *ime4Δ1-224* mutation
pBG152	pBG129 with *ime4Δ676-900* mutation
pBG157	pBG112 with *HOP1-urs1* promoter from pAV124
pBG165	pBG112 with MluI site introduced at +148 bp
pBG166	pBG112 with *ime4::225-675 Flipped*
pBG167	pBG157 with **a**1-α2 site mutant with 4 bp changes
pBG173	pBG129 with *rme2-s2* mutation
pBG200	*HIM1* -387/+336 in pGEM Easy-T
pBG201	isogenic to pBG200
pBG202	*HIM1* -387/+336 in SacI/SacII sites in RS405
pBG203	*HIM1* -387/+336 in SacI/SacII sites in RS303
pBG204	*him1Δpr* in pBG202
pBG205	*rhi1Δpr* in pBG203

downstream **a**1-α2 binding site (as indicated above, for the *IME4* site) to produce

pJM535.

Plasmid pBG200 and pBG201 contain a PCR cassette of +387 to -336 of *HIM1*,

amplified by PCR from genomic DNA of JRY118, cloned into pGEM-Easy T

(Promega). Plasmid pBG201 was digested with SacI and SacII to liberate the insert with

sticky ends, to clone in the SacI/SacII sites of RS405 and RS303 to create PBG202 and

pBG203, respectively. Site directed mutagenesis was used to delete 250 bp upstream of

HIM1 in pBG202 to create pBG204, and 250 bp downstream of *HIM1* in pBG203 to

create pBG205; 250 bp downstream of *HIM1* was deleted in pBG203 to generate

pBG205.

Strain construction. Strains YBG111 and YBG112 were constructed by transforming

W303 derivative strains LNY315 and LNY316 with a PCR fragment amplified from

pFA6a-KanMX6 with the *KanMX* cassette and 50 bp flanking homology, to delete the

IME4 ORF and -450 to +400 bp flanking ORF (Wach, et al., 1994). These strains were

mated to produce the diploid strain *ime4Δ/ime4Δ* strain YBG115. Plasmids were

transformed via the lithium acetate method. *IME4* deletion and polyA terminator

mutants were integrated at the *LEU2* locus by digesting the listed plasmids with XcmI

and transforming the linearized DNA into yBG111. Transformants were selected on

media lacking Leucine.

Strains JMY076 and JMY077 were constructed by using a *Candida albicans*

URA3 cassette from pGEM-*CaURA3* (A gift from Chuck Martin) to delete the **a**1-α2

binding site. These strains were transformed with a PCR product from pBG7 in the

presence of 5-FOA to recombine the mutant binding site at the wild-type locus, to

generate strains JMY081 and JMY082.   JMY081 and JMY082 were mated to generate JMY084, a homozygous diploid with the **a**1-α2 mutation downstream of *IME4*.

The native *ZIP2* gene was deleted by transformation with a KanMX PCR fragment amplified from the Yeast Deletion Strain collection (Research Genetics) in strains LNY392 and LNY433 to generate JMY104 and JMY105. JMY104 and JMY105 were transformed with XcmI-linearized pJM533 to generate JMY108 and JMY109, which were mated to produce JMY110. JMY104 and JMY105 were also transformed with XcmI-linearized pJM535 to generate JMY111 and JMY112, which were mated to produce JMY113.   All genomic integrations were confirmed by PCR.

Strains YBG200 and YBG201 were generated by first transforming JRY116 with plasmid pAV113 (MATα, 2μ, *URA3*) and JRY102 with pRS423 (2μ , *HIS3*). Transformants were cross-streaked on –His, –Ura media to select for mating cells. Mated single colonies were grown overnight in YEPD to facilitate plasmid loss, and then struck onto 5-fluoroorotic acid (5-FOA) to select for loss of pAV113, and on –HIS media to confirm loss of pRS423. Mating tests against **a**- and α- testers were done to confirm *MAT***a** mating type.

The *MATa* and *MATα him1Δ* strains were constructed by transforming LNY433 and LNY392 with a *him1::KanMX* cassette from the Yeast Deletion Library to knockout the native *HIM1* gene, generating strains YBG204 and YBG205. YBG204 was transformed with XcmI digested pBG202 to generate YBG206, with XcmI digested pBG204 to generate YBG208, and with NdeI digested pBG205 to generate YBG210. YBG208 was transformed with NdeI digested pBG205 to generate YBG212. YBG205 was transformed with XcmI digested pBG202 to generate YBG207, with XcmI digested

pBG204 to generate YBG209, and with NdeI digested pBG205 to generate YBG211. YBG208 was transformed with NdeI digested YBG205 to generate pBG213.

Strain YBG225 was created by transforming LNY433 with a PCR generated *LoxP-KANMX-LoxP* cassette from pUG6 which was targeted to delete the 250 bp immediately downstream of *LAP4,* selecting on YEPD+G415 (Guldener, et al., 1996). KanMX integration was confirmed by PCR. Strain YBG227 was then transformed with pSH47 and selected on -Ura media, then grown for 14 hours in Synthetic Galactose –Ura media to induce the Cre-recombinase on the plasmid. Cells were streaked for single colonies, and replica patched on YEPD+G415 to select for loss of the *LoxP-KANMX-LoxP* cassette.

Strain YBG226 was created by transforming LNY433 with a PCR generated *LoxP-KANMX-LoxP* cassette from pUG6 which was targeted to delete the 250 bp immediately downstream of *HPF1,* selecting on YEPD+G415 (Guldener, et al., 1996). KanMX integration was confirmed by PCR. Strain YBG228 was then transformed with pSH47 and selected on -Ura media, then grown for 14 hours in Synthetic Galactose –Ura media to induce the Cre-recombinase on the plasmid. Cells were streaked for single colonies, and replica patched on YEPD+G415 to select for loss of the *LoxP-KANMX-LoxP* cassette.

TABLE 2. Yeast Strains

Strain	Genotype	Source
LNY315	*MAT**a** ade2-1 trp1-1 his3-11,15 can1-100 ura3-1 leu2-3,112*	L. Neigeborn
LNY316	*MATα ade2-1 trp1-1 his3-11,15 can1-100 ura3-1 leu2-3,112*	L. Neigeborn
LNY392	*MATα ade2-1 TRP1 his3-11,15 can1-100 ura3-1 leu2-3,112*	L. Neigeborn
LNY433	*MAT**a** ADE2 trp1-1 his3-11,15 can1-100 ura3-1 leu2-3,112*	L. Neigeborn
LNY435	*MAT**a**/MATα ade2-1/ade2-1 trp1-1/trp1-1 his3-11,15/ his3-11,15 can1-100/ can1-100 ura3-1/ ura3-1 leu2-3,112/ leu2-3,112*	L. Neigeborn
FY2180	*MAT**a** leu2Δ1 his4-912s lys2-128σ FLAG-spt6-1004*	F. Winston
FY346	*MAT**a** spt16-197 ura3-52 leu2Δ1 lys2-128σ*	F. Winston
DN1359	*MAT**a** leu2Δ1 ura3-52 his3Δ200 trp1Δ63 ade2-101 lys2-801*	D. Norris
JRY103	*MAT**a**/MATα trp1-1/TRP his3-11,15/ his3-11,15 can1-100/ can1-100 ura3-1/ ura3-1 leu2-3,112/ leu2-3,112 ash1Δ::LEU2 / ash1Δ::LEU2*	J. Matthias
JRY116	*mat**a**Δ::TRP1 ADE2 his3 leu2 trp1 ura3 ash1Δ::LEU2*	J. Matthias
JRY118	*MATΔ/MATα trp1-1/TRP his3-11,15/ his3-11,15 can1-100/ can1-100 ura3-1/ ura3-1 leu2-3,112/ leu2-3,112 ash1Δ::LEU2 / ash1Δ::LEU2*	J. Matthias
YBG111	*MAT**a** ade2-1 trp1-1 his3-11,15 can1-100 ura3-1 leu2-3,112 ime4Δ::kanMX4*	this study
YBG112	*MATα ade2-1 trp1-1 his3-11,15 can1-100 ura3-1 leu2-3,112 ime4Δ::kanMX4*	this study
YBG115	*MAT**a**/MATα ade2-1/ ade2-1 trp1-1/ trp1-1 his3-11,15/ his3-11,15 can1-100/ can1-100 ura3-1/ ura3-1 leu2-3,112/ leu2-3,112 ime4Δ::kanMX4/ ime4Δ::kanMX4*	this study
YBG144	*MAT**a**/MATα ade2-1ADE trp1-1/TRP his3-11,15/ his3-11,15 can1-100/ can1-100 ura3-1/ ura3-1 leu2-3,112/ leu2-3,112*	this study
JMY076	LNY392 with *IME4—**a**1-α2::CaURA3*	this study
JMY077	LNY433 with *IME4—**a**1-α2::CaURA3*	this study
JMY081	LNY392 with *IME4—**a**1-α2 mut*	this study
JMY082	LNY433 with *IME4—**a**1-α2 mut*	this study
JMY084	YBG144 with *IME4—**a**1-α2 mut/ IME4—**a**1-α2 mut*	this study
JMY104	LNY392 with *zip2:: kanMX4*	this study
JMY105	LNY433 with *zip2:: kanMX4*	this study
JMY108	LNY392 with *zip2:: kanMX4+ ZIP2::LEU2*	this study
JMY109	LNY433 with *zip2:: kanMX4 + ZIP2::LEU2*	this study
JMY110	YBG144 with *zip2:: kanMX4/ zip2:: kanMX4 leu2::ZIP2 / leu2::ZIP2*	this study
JMY111	LNY392 with *zip2:: kanMX4 leu2::ZIP2-**a**1-α2 mut*	this study
JMY112	LNY433 with *zip2:: kanMX4 leu2::ZIP2-**a**1-α2 mut*	this study
JMY113	YBG144 with *zip2:: kanMX4 leu2::ZIP2-**a**1-α2 mut*	this study
YBG145	YBG111 with *leu2::ime4Δ3'*	this study
YBG147	YBG111 with *leu2::IME4*	this study
YBG149	YBG115 with *leu2::IME4 / leu2::IME4*	this study

YBG150	LNY315 with *leu2::ime4Δ3'*	this study
YBG158	YBG144 with *zip2:: kanMX4/ zip2:: kanMX4*	
	*leu2::ZIP2-**a**1-α2 mut / leu2::ZIP2*	this study
YBG159	YBG111 with *leu2::rme2-s1*	this study
YBG160	YBG111 with *leu2::ime4Δ1-900*	this study
YBG161	YBG111 with *leu2::ime4Δ1-450*	this study
YBG162	YBG111 with *leu2::ime4Δ451-900*	this study
YBG176	YBG111 with *ime4::Hgh*	this study
YBG183	YBG111 with *leu2::ime4Δ1-224*	this study
YBG184	YBG111 with *leu2::ime4Δ676-900*	this study
YBG200	*MAT**a**/MATΔ::TRP1 trp1-1/TRP his3-11,15/ his3-11,15*	this study
	can1-100/ can1-100 ura3-1/ ura3-1 leu2-3,112/ leu2-3,112	
	ash1Δ::LEU2 / ash1Δ::LEU2	
YBG202	YBG111 with *leu2::rme2-s2*	this study
YBG204	LNY392 with *him1Δ::kanMX4*	this study
YBG205	LNY433 with *him1Δ::kanMX4*	this study
YBG206	YBG204 with *HIM1::LEU2*	this study
YBG207	YBG205 with *HIM1::LEU2*	this study
YBG208	YBG204 with *him1Δpr::LEU2*	this study
YBG209	YBG204 with r*hi1Δpr::HIS3*	this study
YBG210	YBG204 with r*hi1Δpr::HIS3, him1Δpr::LEU2*	this study
YBG211	YBG205 with *him1Δpr::LEU2*	this study
YBG212	YBG205 with r*hi1Δpr::HIS3*	this study
YBG213	YBG205 with r*hi1Δpr::HIS3, him1Δpr::LEU2*	this study
YBG225	LNY433 with *alf4Δpr::LoxP-KanMX-LoxP*	this study
YBG226	LNY433 with *hzr1Δpr::LoxP-KanMX-LoxP*	this study
YBG227	LNY433 with *alf42prdown::LoxP*	this study
YBG228	LNY433 with *hzr1Δpr::LoxP*	this study

B. Methods

RNA extraction. RNA was extracted from cells using the hot acid phenol method

previously described (Abraham and Vershon, 2005; Ausubel, 1987). Cell pellets of

between 50-150 μl in volume (described below) were frozen overnight at -80°C. Pellets

were resuspended in 400 μl TES+DEPC and 400 μl of acid-phenol chloroform (Sigma),

and incubated for 45' at 65°C, vortexing several times during heating. Mixtures are

centrifuged 5' and the aqueous phase was transferred and re-treated with acid phenol-

chloroform twice. The final aqueous phase was recovered with 1 ml Ethanol (100% non-

denatured) and 40 μl of 3M NaOAc + DEPC, pH 5.2. RNA samples were resuspended in

75-150 μl of DEPC-treated water, and quantitated by UV.

Northern Blot. Yeast strains LNY315 (*MAT*a) and LNY435 (a/α) were grown in

volumes of 50 ml at 30° C in YEPD media, to an OD_{600} range of .400 - .800. At this

point, 49 ml of the culture was spun down at 1000 rpm, washed with DEPC water, and

the cell pellet frozen at -80°C. A 1 ml aliquot of the YEPD culture was used to inoculate

a total of 100 ml of YEPA media and grown for 12-16 hours. At the end of YEPA

growth, 50 ml were spun down, and washed with DEPC water and frozen as above. The

remaining 50 ml YEPA culture was spun down, washed with water, and resuspended in

50 ml SPM media. The SPM cells were incubated for 3 hours at 30°C, and harvested as

above. RNA was extracted from the cell pellets as described above (Abraham and

Vershon, 2005; Ausubel, 1987). RNA was separated by molecular weight on a 1.5%

MOPS-agarose gel containing formaldehyde and transferred to a Nytran Membrane

(Scleicher and Schuell) by capillary transfer. The membrane was pre-hybridized in 10 ml

hybridization solution for 3 hours (50% Formamide, 5x SSC, 50 mM $NaPO_4$ pH 7.0, 1x

Denhardt's reagent, 0.1% SDS, 100 μg/ml sheared single stranded salmon sperm DNA).

Double-stranded DNA probes were generated by PCR and labeled with $[\alpha\text{-}^{32}P]$ dCTP by

random priming (MegaPrime DNA Labeling Kit, Amersham/GE). After overnight

hybridization to radiolabeled DNA probes and washing with 2X SSC and 0.1% SDS at

42°, RNA expression was detected by exposure to a phosphorscreen and imaging on a

Molecular Dynamics phosphoimager.

RT-PCR assays. Expression of the RNA transcripts was assayed by Reverse-

Transcriptase PCR (RT-PCR). Yeast strains were grown under rich nutrient (YEPD) or

sporulation-inducing conditions (SPM; 3 hours for *IME4* assays, and 5 hours for *ZIP2*

assays) and total RNA was extracted by hot acid-phenol extraction, as previously

described (Abraham and Vershon, 2005). Normalized RNA samples were treated with

Turbo DNA Free DNase (Ambion) and PCR amplification of the DNase treated RNA

with the *IME4*/*RME2* or *ZIP2*/*RME3* primer set was performed to verify the absence of

contaminating DNA. Sense and antisense cDNA of the *IME4, URA3, ZIP2, YFL012W,*

HIM1, HPF1, SUL1, LAP4 and *ACT1* genes was synthesized for each RNA sample with

sense and antisense specific primers, using Omniscript RT (Qiagen). Different cDNA

sample concentrations (1-4 μl) were assayed to verify that the reaction was in the linear

range. Nested PCR reactions using different concentrations of cDNA as a template were

amplified in 50 μl reactions containing 10 pmols of each primer, 1X Amplitaq Taq Buffer

II, 2.5 U of Amplitaq Taq polymerase (Applied Biosystems). The amplifications were

carried out at 94°C for 3 min, followed by 30 cycles of 94°C for 30 secs, 52°C for 1 min,

and 72°C for 30 secs and a final extension step of 7 min at 72°C.   Samples were run on

1.4% Agarose TAE gels, and images were photographed with a Fluorochem 8800

camera.

Chromatin Immunoprecipitation Assay (ChIP).   ChIP assays were performed using a

modified version of protocols described elsewhere (Mathias, et al., 2004; Nagaraj, et al.,

2004; Rusche and Rine, 2001; Shang, et al., 2000). Cultures of LNY392 and YBG144, or

YBG111 and YBG115 strains carrying the indicated *HOP1pr::IME4* constructs were

grown to midlog (O.D. $_{600}$ = .500) in the appropriate media (YEPD, SD-Leu, or SPM); 50

ml of the cultures was fixed with a final concentration of 1% formaldehyde for 15 min at

22°C, washed with 1 ml TBS, and frozen at -80°C for a minimum of 12 h. Cell pellets

were resuspended with 400 µl FA lysis buffer (50 mM HEPES, pH 7.4, 150 mM NaCl, 1

mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate) plus 1mM PMSF, 1X Protease

Inhibitor Cocktail from Roche (1873580), and 50 µl Sigma Protease Inhibitor Cockail

(P8215). To this, 200 µl of glass beads was added and cells were lysed by vortexing at

full speed for 40 min at 4°C. The lysate was then centrifuged for 5 min at 12,000 x *g* at

4°C. The supernatant was transferred to a new tube, and the beads were washed with 500

µl of FA lysis buffer; the supernatant from this wash was added to the original

supernatant fraction. DNA was sonicated at 30% output for 6 x 5 sec cycles to give an

average chromatin fragment size of 500 bp. Sonicated lysates were centrifuged at 12,000

x *g*, for 5 min and 50 µl of the sonicated DNA was reserved as a Total Chromatin (TC)

sample.   The remaining DNA was pre-cleared by the addition of 25 µl of protein G

agarose beads, nutated for 1 hour at 4°C; and the lysates were cleared by centrifugation at

12,000 x *g* for 5 min at 4°C. To immunoprecipitate TBP- or Abf1-bound DNA, 10 μl of

TBP yN-20 polyclonal antibody (Santa Cruz, sc-26141) or Abf-1 yC-20 polyclonal

antibody (Santa Cruz, sc-6679) was added and nutated for 16 h at 4°C. The protein G

beads centrifugation, washes, DNA elution, crosslinking reversal, and proteinase

treatment, were all performed as previously described (Nagaraj, et al., 2004; Rusche and

Rine, 2001). Frozen TC samples were brought up to 500 μl in volume with elution

buffer.   To purify DNA, a Qiagen PCR Purification kit was used, adding 50-80 μl of 3M

Sodium Acetate, pH 5.2 per sample, and following the manufacturer's protocol, eluting in

50 μl 10 mM Tries, pH 8.0. For PCR, TC samples were diluted 50-fold, (IP samples

were undiluted) and 2 μl of DNA was used in each reaction.   PCR was performed in 50

μl reactions containing 10 pmols of each primer, 1X Amplitaq Taq Buffer II, 2.5 U of

Amplitaq Taq polymerase (Applied Biosystems). The amplifications were carried out at

94°C for 3 min, followed by 30 cycles of 94°C for 30 sec, 52°C for 1 min, and 72°C for

30 secs and a final extension step of 7 min at 72°C. The PCR products were run on 1.5%

agarose gels.

 3' **RACE (Rapid Amplification of cDNA Ends).** The 3' RACE assay was performed as

described in Molecular Cloning (Sambrook and Russell), using an oligo(dT) primer with

a 15 bp tag on the 5' end to generate poly(A) cDNA. This cDNA was amplified using a

primer downstream from the *IME4* start codon that would anneal specifically to *RME2*,

and a primer that hybridizes to the tag sequence on the oligo(dT) primer used for cDNA

synthesis. The PCR products of this reaction were gel purified (Qiagen), and the PCR

phase was repeated with a second *RME2*-specific primer and the tagged primer to

eliminate false positives. These products were ligated into TOPO TA (Invitrogen) and

sequenced to determine the 3' end of the *RME2* transcript.

Biological assays. To test *URA* expression from the *IME4-RME2* promoters strains

LNY392 (1n) and yBG144 (2n) were transformed with pJM530, pRS415, or pRS415 and

pRS426, and grown in –Leu or –Leu –Ura media to saturation. Concentrations were

normalized based on O.D. $_{600}$ values. Five-fold serial dilutions were spotted onto –Leu, –

Leu –Ura, or –Leu 5-FOA media, and grown at 30 degree for the indicated amounts of

time; photographs were taken with a Fluorochem 8800 camera.

Preparing and Purifying RNA for SOLiD Library. Strains YBG200, and JRY103

were streaked for single colonies on YEPD plates. For glucose samples, three different

single colonies were inoculated and grown in separate 50 ml cultures of YEPD at 30° to

OD_{600} = .600. Sporulation samples were made by first growing in YEPD as above,

followed by centrifugation, washing, and transferring the cell pellets into 50 ml of YEPA

for 16 hours growth at 30°. The YEPA sample was spun down, washed with water, 40

OD units of cells were transferred to SPM (3% Potassium Acetate, .2% Raffinose) media

for 3 hours. Upon completion of the growth phase, cells were pelleted by centrifugation,

washed with DEPC water, and stored at -80C. RNA was extracted as described above by

the hot-acid phenol method, and quantitated by UV; quality and relative quantity of

samples was confirmed by running 10 mg on a 1X MOPS-Formaldehyde-DEPC-2%

Agarose gel (Abraham and Vershon, 2005; Ausubel, 1987). For each biological replicate

library, 30 mg of total RNA was treated with RiboMinus (Invitrogen), and recovered

using ethanol precipitation in the presence of glycogen and 3M sodium acetate, as

described by the manufacturer. The RiboMinus RNA sample was resuspended in 60 μl of nuclease-free, DEPC treated water, and ribosomal RNA removal was verified by running 2 μl (estimated .01 μg) on a LONZA 1.2% RNA gel, versus an equivalent amount of total yeast RNA.

RNA was quantitated on a NanoDrop ND-1000 spectrophotometer, and subsequent quality analysis was done via Bioanalyzer E2100 Expert Pico, to verify quantitation and reduction of Ribosomal RNA to under 2.0% of total sample. Samples which failed the initial analysis (rRNA > 2.0% of total sample) were re-treated with RiboMinus to reduce contaminant rRNA to below this level.

SOLiD Library Assembly. After RiboMinus treatment and purification, roughly 300 ng of RNA was submitted to the Waksman Genomics Facility for library assembly. The standard Fragment Library RNA-seq protocol was used, as described by the manufacturer (Applied Biosystems). All samples were quality-controlled and quantified by Q-bit analysis and real-time PCR prior to the SOLiD Emulsion PCR protocol.

SOLiD Emulsion PCR and Sequencing-by-Ligation. SOLiD emulsion PCR was performed by the manufacturer's protocol for 50 bp fragment library reads using SOLiD version 4.0 (Applied Biosystems) at the Waksman Genomics Facility. WPA analysis was performed prior to sequencing by ligation to verify samples were not overamplified during Emulsion PCR.

Mapping of SOLiD reads and comparative alignment. All libraries were mapped with

Bioscope 1.2 at the Cancer Institute of New Jersey, allowing for up to two mismatches

per 50 bp read, against the sacCer2 yeast reference genome. Bioscope was chosen over

other reads mappers due to its native support for color-space and handling of strand-

specific reads. Read-Per Kilobase per Million Reads Sequenced (RPKM) values for each

gene from each experiment were calculated using custom Perl scripts based on the CDS

annotation from sacCer2. Reads that mapped outside of CDS regions or mapped to

opposite strands of CDS regions were classified as unannotated, mapped reads. Reads

antisense to coding genes were measured by RPKM values relative to the overlapped

coding gene. Unannotated mapped reads that mapped within ±100bp of a CDS feature

on the same strand as the CDS feature were discarded as they are indicative of

unannotated UTRs. The remaining unannotated mapped reads were split into non-

overlapping 1000bp bins along the genome. These bins were used to identify genomic

regions showing unannotated expression.

III. CHARACTERIZATION OF *RME2* AS A CELL-TYPE SPECIFIC

REGULATOR OF *IME4*

A. Results
1. *IME4* is regulated by *RME2*

i. a1-α2 binding in the *IME4-COX13* intergenic region affects *IME4* expression, and not *COX13*

IME4, an N6-methyladenosine transferase required for sporulation, was shown to be expressed only in **a**/α diploid cells, which express the **a**1-α2 complex (Clancy, et al., 2002; Shah and Clancy, 1992). This suggested that the **a**1-α2 repressor complex could also act as an activator.   Previous work had identified that **a**1-α2 was bound in-*vivo* to a functional site in the *IME4-COX13* intergenic region (Galgoczy, et al., 2004; Nagaraj, et al., 2004).   While *IME4* expression was known to be cell-type regulated, *COX13*, a component of the cytochrome C oxidase complex, was not regulated in a cell-type specific manner (Galitski, et al., 1999; Shah and Clancy, 1992; Yin, et al., 2003). Because *IME4* is upregulated in response to starvation, we examined expression in rich (YEP+2% Glucose), glucose-limited (YEP+2% Potassium Acetate) and Sporulation-inducing (SPM) media. To further examine transcription of the *IME4* and *COX13* loci, I performed Northern Blot Analysis using double-stranded, random primed probes in both *MAT***a** and **a**/α cells.   The expression pattern of *COX13* appears to be the same for both 1n and 2n cells, indicating that *COX13* expression is not affected by the binding of **a**1-α2 (Fig. 8, lanes 1-3 vs. 4-6) at this site.   In contrast, *IME4* (indicated by the + for both probes) is not expressed in 1n cells, and the highest expression is seen in 2n cells under sporulation, consistent with previous findings (Shah and Clancy, 1992). Interestingly, both of the probes used for *IME4,* as well as a probe to dubious ORF *YGL193c* were found to anneal to a haploid-specific

Figure 8. Northern blots of *IME4* and *COX13* in 1n (*MAT*a) and 2n (a/α) cells.

Northern blots of 1n (LNY315, lanes 1-3) and 2n (LNY435, lanes 4-6) cells grown to log

phase in YEP + 2% Glucose (Glu, lanes 1,4), YEP+2% Potassium Acetate (Ace, lanes

2,5) and Sporulation Inducing (SPM, lanes 3,6) media. Blots were probed for 24-48

hours before exposure on a phosphor screen, and then stripped before re-probing.   All

probes, except pC4, which is random primed from the pC4 plasmid with ^{32}P-dCTP, are

double stranded PCR products labeled by random priming with ^{32}P-dCTP.
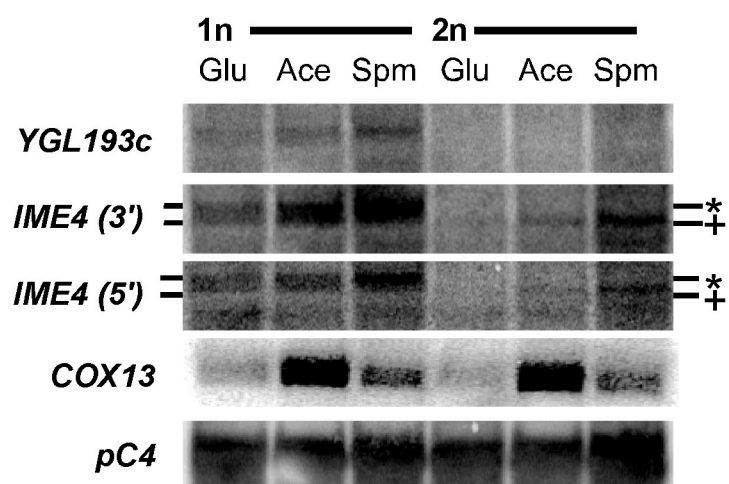
RNA larger than the 1800 bp – *IME4* transcript (Fig. 8, *IME4 3', IME4 5', and YGL193c).*

ii. The *IME4* locus expresses an antisense transcript.

Previous investigations of *IME4* indicated the presence of a second transcript, named *RME2*, for <u>R</u>epressor of <u>Me</u>iosis <u>2</u>, which is expressed from the antisense strand of *IME4* (Shah and Clancy, 1992). I hypothesized that it is possible that the **a**1-α2 site downstream of *IME4* was regulating expression of the *RME2* transcript, and that *RME2* expression may be cell-type specific. To assay for *RME2* in a strand-specific manner, RT-PCR (Reverse Transcriptase PCR) was performed using primers that hybridize to the *IME4* sense and antisense strands to synthesize the cDNA of each transcript (Fig. 9A). The cDNAs were then used as templates in PCR, with a second primer set which was nested relative to the cDNA primers, to independently detect expression of *IME4* and *RME2*. The expression of *IME4* and *RME2* was assayed in haploid and diploid cells grown in rich or sporulation-inducing media (Fig. 9B). Consistent with previous observations, the *IME4* transcript is only expressed at high levels in diploid cells under conditions inducing sporulation (Fig. 9B, lane 4) (Chu, et al., 1998; Shah and Clancy, 1992). In contrast, the *RME2* transcript is expressed in both rich and sporulation media in haploid, but not **a**/α diploid cells (Fig. 9B, lanes 1 and 3). This data is consistent with the hypothesis that the **a**1-α2 binding site downstream of *IME4* acts as a repressor of *RME2* expression. Furthermore, the comparative lack of *RME2* expression in diploid cells is consistent with the hypothesis that *RME2* must be repressed to permit *IME4* expression.

Figure 9. *IME4* and *RME2* expression in wild-type cells assayed by RT-PCR.   A)

Schematic of the nested RT-PCR reaction used to assay for *IME4* and *RME2* transcripts.

A strand-specific primer is used to generate cDNA, and the internal primer set is used in

PCR reactions to amplify.   B) An RT-PCR assay of *IME4* and *RME2* in both wild-type

MATa (LNY392, lanes 1,3) and **a**/α (YBG144, lanes 2,4) cells under glucose (lanes 1,2)

and sporulation-inducing (lanes 3,4) conditions.   –RT is a PCR-amplification of the

DNAse-treated RNA with the nested primer set. Genomic DNA is amplified as a

positive control for PCR amplification. *ACT1* is a positive loading control.

A

B
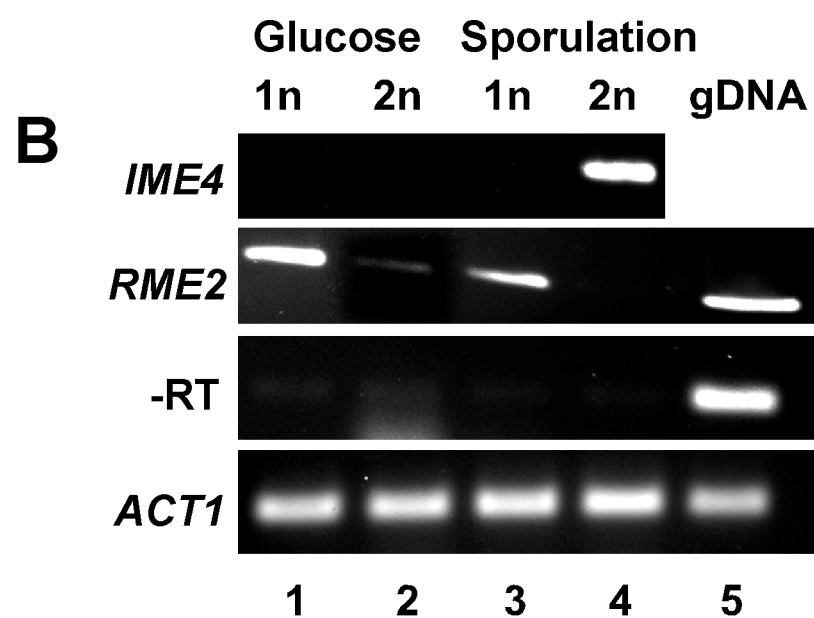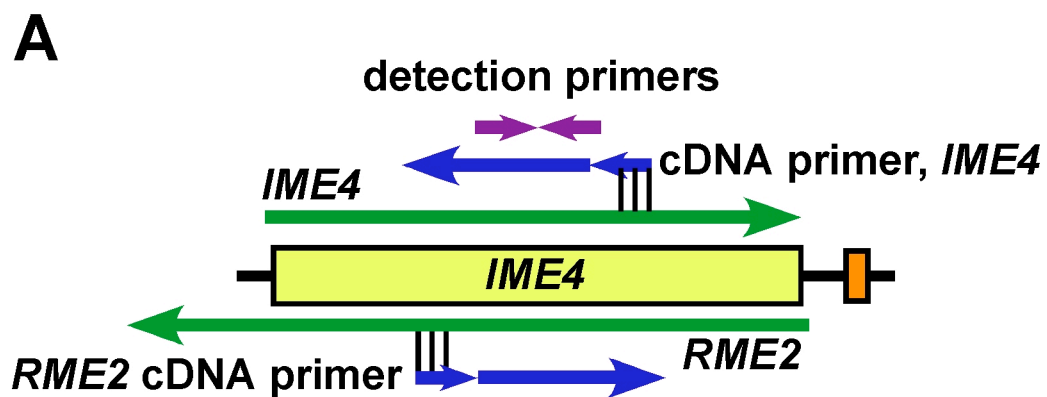
iii. Loss of Downstream a1-α2 binding prevents *IME4* expression

To determine if **a**1-α2 binding site downstream of *IME4* represses *RME2* in a diploid-specific manner, I introduced base pair substitutions at four positions in the site that are contacted by the **a**1 and α2 proteins in the site (Fig. 10A). These bases have previously been shown to individually reduce **a**1-α2-mediated repression by greater than 20-fold (Jin, et al., 1995; Vershon, et al., 1995). As predicted, this mutation did not appear to affect *IME4* or *RME2* expression in haploid cells (Fig. 10B, lanes 1 vs. 3; and 5 vs. 7). In comparison to the wild-type diploid, where *RME2* is repressed, *RME2* was expressed in **a**1-α2 mutant diploid cells (Fig. 10B, *RME2*- lane 6 vs. lane 8). These results indicated that the **a**1-α2 complex directly repressed *RME2* expression in diploid cells. More interestingly, there was a complete loss of *IME4* expression in the mutant diploid cells under sporulation conditions (Fig. 10B, *IME4*- lane 8). These results, taken together show that *RME2* expression in diploids prevents *IME4* expression.

These results suggested that the *RME2* transcript, rather than a protein cofactor, may be the haploid-specific repressor of *IME4* in haploid cells ("X" in Fig. 3). To verify this, I constructed a mutant version of *IME4* that has a 400 bp deletion (+1809 to +2209 relative to the *IME4* ATG) that removes the presumptive *RME2* promoter region. If *RME2* expression regulates *IME4,* loss of *RME2* should permit *IME4* expression. Haploid cells with this mutation failed to express *RME2* (Fig. 10C lane *RME2*). More importantly, this mutation caused *IME4* to be derepressed (Fig 10C, lane 2, *IME4*). This verified that haploid-specific expression of *RME2* is required to repress *IME4*.

Figure 10. *IME4* **is regulated by** *RME2,* **and** *RME2* **by a1- α2.** (A) Mutation of the

a1-α2 site downstream of *IME4*. The contacts with residues of **a**1 (underlined) and α2

are shown. Arrows are hydrogen bonds; diamonds denote water-mediated contacts; lines

ending in circles are van der Waals interactions. The bottom shows the mutated binding

site, with altered DNA bases in red. The retained contacts in the mutant are shown. (B)

RT-PCR assays for the presence of the *IME4* (sense) and *RME2* (antisense) transcripts

from strains *MATa* haploid (LNY392, lanes 1, 5), **a**1-α2 mutant haploid (JMY082, lanes

3, 7) WT **a**/α diploid (YBG144, lanes 2, 6), and **a**1-α2 mutant diploid (JMY084 lanes 4,

8) grown in vegetative (lanes 1-4) or sporulation-inducing conditions (lanes 5-8). –RT

row is a PCR amplification of DNase treated RNA with the *IME4* primer set to control

for contamination of genomic DNA. RT-PCR of *ACT1* expression is used as a loading

control. Genomic DNA (gDNA, lane 9) was included as a control for PCR amplification.

C) RT-PCR of *IME4* and *RME2* expressed in haploid cells with either wild-type *IME4*

(YBG147; lane 1) or *ime4Δ3'(*YBG150; lane 2) integrated at *LEU2*.

A

Wild Type

S50 R54 N51   I50   M54 N51 R55

C C G T G T A T T T T T T T A C A T C A

G G C A C A T A A A A A A T G T A G T

R4   G5   R7   V47

Mutant

S50   I50   M54

C C G T *C G* A T T T T T T T A C *G G* C A

G G C A C *A* T A A A A A A T G C *C G* T

R4 G5 R7

B

Glucose ——— Sporulation ———

| 1n | 2n | 1n | 2n | 1n | 2n | 1n | 2n | |
| WT | WT | Mut | Mut | WT | WT | Mut | Mut | gDNA |

IME4

RME2

-RT

ACT1

1   2   3   4   5   6   7   8   9

C

1n   1n
WT   *imeΔ3'*

1   2
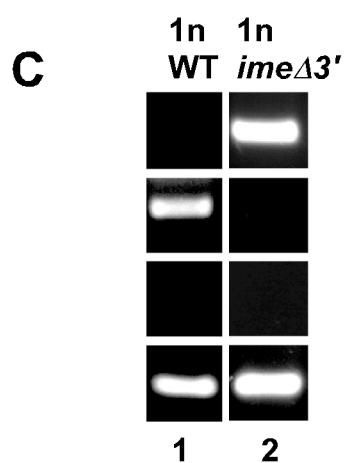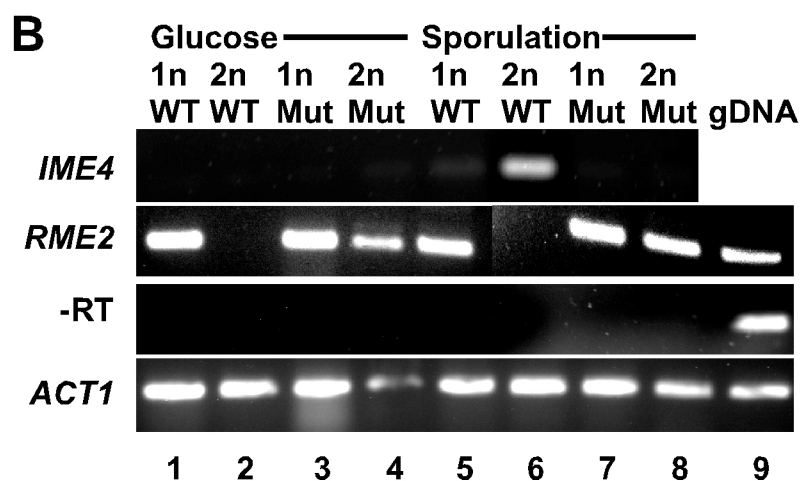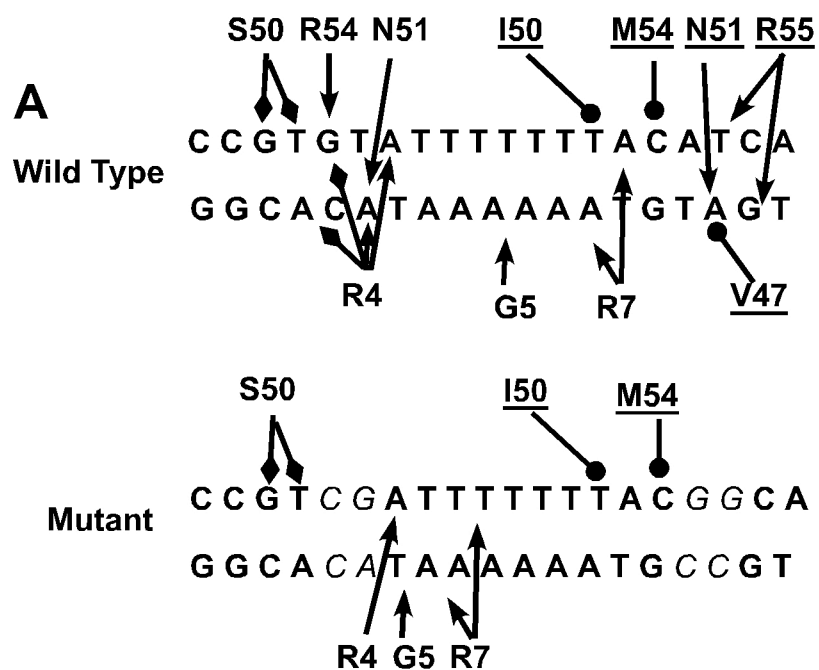
b. *RME2* represses *IME4* in *cis-*

There are several possible models for how *RME2* may function to repress *IME4*. It is possible that expression of the complementary *IME4* and *RME2* transcripts allows the formation of double stranded RNA (dsRNA). The formation of dsRNA by the transcripts could serve as a target for degradation by nucleases, such as Rnt1, or trigger silencing of *IME4* expression through a mechanism similar to RNAi-mediated regulation found in *C. elegans*, *Arabidopsis*, and *S. pombe* (Ge, et al., 2005; He and Hannon, 2004; Matzke and Birchler, 2005). Alternatively, it is possible that the process of transcription of *RME2* interferes with expression of *IME4,* similar to how transcription of *SRG1* interferes with expression of *SER3* in yeast (Martens, et al., 2004). To test between these alternatives, we examined if expression of *RME2* is able to repress *IME4* at another location in the chromosome.

As shown in above, the *ime4Δ3'* mutant is unable to express the *RME2* transcript, and instead, expresses *IME4* constitutively (Fig. 10C, lane 2). The *ime4Δ3'* construct was integrated at *LEU2* into a haploid strain with a wild-type *IME4* gene. This strain was assayed for the ability of the endogenous *RME2* to repress *IME4* expression from the *ime4Δ3'* locus. If *RME2* can repress *IME4* in *trans-*, this strain would have an expression pattern like a wild-type haploid. Expression of miRNA, in *trans-*, has been shown to reduce coding gene expression to undetectable levels (Couzin, 2002). However, if *RME2* can only repress *IME4* in *cis*, then the *IME4* transcript expressed from *ime4Δ3'* would still be present. Both *RME2* and *IME4* were expressed in this strain, and there was no appreciable loss of *IME4* expression when compared to the *ime4Δ3'* strain. (Fig. 11, compare lane 3 vs. 4). These results indicate that *RME2* is unable to repress expression

Figure 11. Wild-type *RME2* only represses *IME4* in *–cis*. Deletion of the *RME2*

promoter region allows expression of *IME4* in haploid cells, and wild-type *RME2*

expression in *trans* fails to repress the *ime4Δ3'* mutation. RT-PCR assays of the *IME4*

and *RME2* transcripts were performed on strains *ime4Δ* (YBG111, lane 1), WT

(YBG147, lane 2), *ime4Δ3'* (YBG145, lane 3) and WT + *ime4Δ3'* (YBG150, lane 4) in

haploid strains, under sporulation-inducing conditions. The –RT row are PCR

amplifications of DNase treated RNA with the *IME4* primer set to control for

contamination of genomic DNA. RT-PCR of *ACT1* expression is used as a loading

control. Genomic DNA (gDNA, lane 5) was included as a control for PCR amplification.
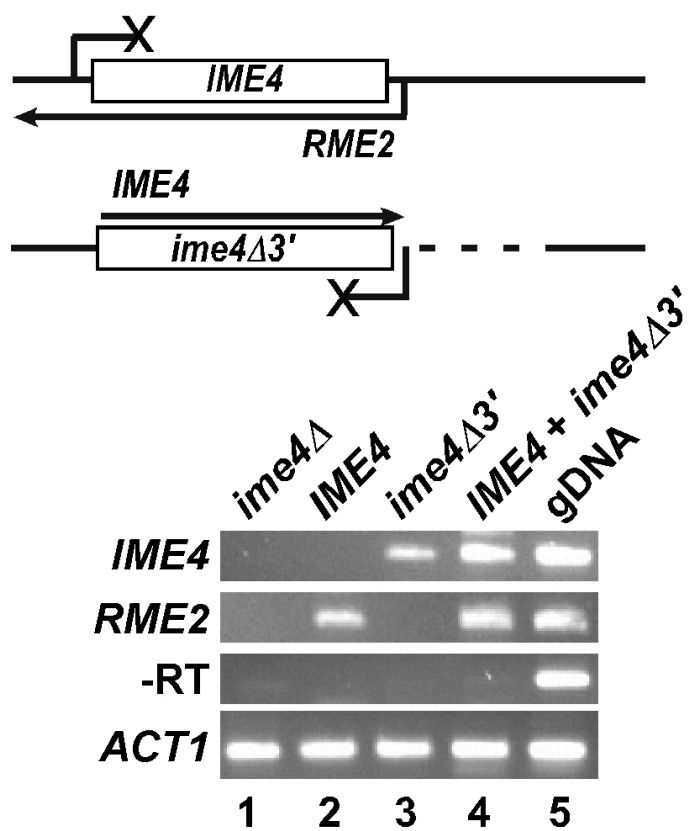
of *IME4* if supplied in *trans-,* ruling out mechanisms that involve post-transcriptional

formation of dsRNA.   Therefore, it appears that *S. cerevisiae* uses a novel *cis-*

mechanism for antisense-mediated repression.

3. ROLE OF THE SENSE PROMOTER REGION IN REGULATION

i. *RME2* is polyadenylated and extends into the *IME4* sense promoter

It has been shown that regulated transcription of the *SRG1* gene, which overlaps

the *SER3* promoter, interferes with transcription factors binding to the coding promoter

(Martens, et al., 2004; Martens, et al., 2005). If regulation of *IME4* is working through a

similar mechanism then we would expect that the *RME2* transcript to extend through the

IME4 promoter. Additionally, knowing what type of transcript *RME2* is (PolI, PolII,

PolII) would give a better idea how it might be regulated, and how it might be possible to

create different mutations of *RME2.* To determine if *RME2* is a PolII transcript, I

performed parallel cDNA syntheses of *IME4* and *RME2* using both the gene-specific

internal primer set described above, and with oligo-dT primer. Because *RME2* is only

expressed in haploid cells, PCR-detection of cDNA synthesized with the oligo-dT primer

could only be *RME2*, and not *IME4,* which is polyadenylated.   *RME2* cDNA made using

oligo-dT primers was detected in the PCR phase, indicating that the *RME2* ncRNA is a

RNA Pol II transcript (Fig. 12A, lane 2).

Because *RME2* is polyadenylated, the 3' end was mapped using Rapid

Amplification of cDNA End (RACE) experiments with an oligo-dT primer. Sequence

analysis of several independent clones showed the wild-type *RME2* transcript ends at bp

+384 upstream of the start of the *IME4* ORF (Fig. 12B). The *RME2* transcript therefore

likely overlaps most of the *IME4* promoter region, and could interfere with initiation of transcription or the binding of activators and general transcriptional machinery. If this is the case, then alteration of the 3' end of the *RME2* transcript could inhibit its ability to silence the *IME4* transcript in haploid cells.

ii. Termination site of *RME2* determines ability to repress *IME4*

The observation that *IME4* is repressed in haploid cells through antisense transcription raises the question of how does expression of an antisense transcript in *cis* prevents expression of the sense transcript? One of the most economical models is that antisense transcription through the sense promoter prevents transcription factor binding and activation of the sense promoter. This model is similar to the mechanisms proposed for the *SER3* and *ADH1* genes, which are regulated by the upstream ncRNAs *SRG1* and *ZRR1*, respectively (Bird, et al., 2006; Martens, et al., 2004; Martens, et al., 2005). Transcription of the upstream ncRNAs through the coding gene promoter prevents binding of the transcriptional activators required for expression. If *IME4* is regulated by a similar mechanism, then I predicted that premature termination of *RME2* transcript may allow expression of the *IME4* transcript in haploid cells.

To test this model, two mutations were constructed which are silent with respect to *IME4*, but introduce signal sequences for premature polyadenylation at 800 or 400 bp upstream of the normal 3-prime end of the *RME2* transcript (mutant *rme2-s1* and *rme2-s2* respectively, Fig. 13A). Premature polyadenylation was similarly used to shorten the *Kcnq1ot1* transcript (Kanduri, et al., 2006). The placement of the *rme2-s1* signal, at approximately 440 bp downstream from the *IME4* ATG, was predicted to truncate the

Figure 12. ***RME2* is a polyadenylated transcript that extends into the *IME4***

promoter. A) RT-PCR assays of *MATa* haploid (LNY433) and **a**/α diploid (YBG144)

cells grown in sporulation-inducing media. GS is cDNA made with a stranded, gene-

specific primer, and dT is cDNA made with an oligo-DT primer; both are amplified with

the same detection primer set. *ACT1* is a positive control for oligo-dT priming and

loading, the U1 SNRP is a negative control for –dT priming. B) 3' RACE assay of

RME2 indicates that the transcript extends to 384 bp upstream of the *IME4* ATG. cDNA

was synthesized using adaptor-oligo-dT and amplified with a primer antisense to *RME2*

and the adaptor sequence (top, PCR). Bands from the first PCR were extracted and

reamplified in the Nested PCR. The indicated band from this reaction was cloned into a

TOPO vector (Invitrogen) and sequenced to determine the 3' end of *RME2*.

A

B
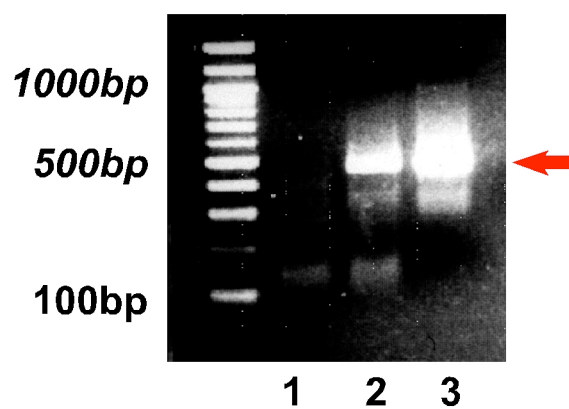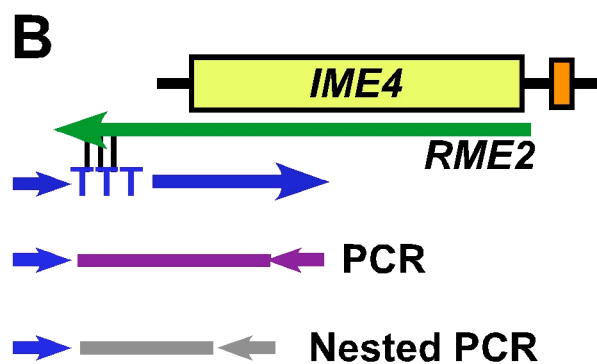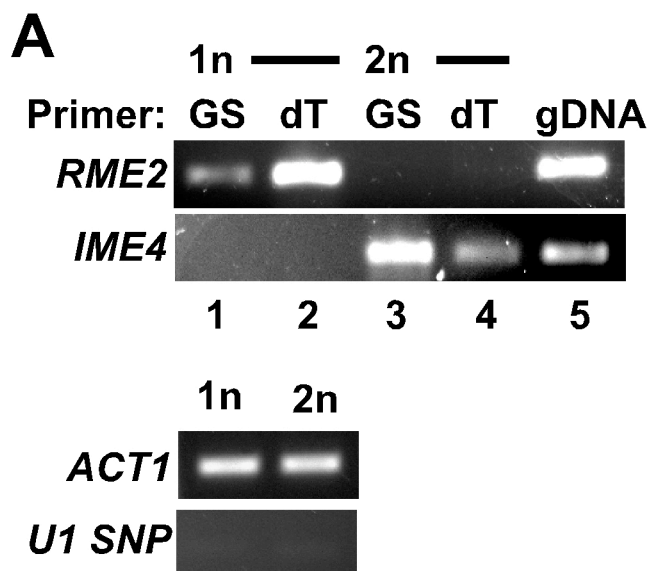
RME2 transcript before it extends through the 5' end of the *IME4* ORF. The *rme2-s2*

mutation, at 23 bp upstream of the *IME4* ATG, terminates transcription of *RME2* through

the whole *IME4* sense promoter, but not through the 5' end of the ORF. To assay for

premature termination of *RME2,* primer sets within the *IME4* promoter (Pr), *IME4* 5-

prime end (5'), and the middle of the ORF (Mid) were used (Fig. 13A).   These alternate

primers sets should differentially detect the wild-type, *rme2-s1* and *rme2-s2* mutants.

In wild-type cells, all three primer sets detected the *RME2* transcript (Fig. 13B

lane 1). In contrast, the *rme2-s1* transcript was only detected by the Mid- primer set,

indicating that the transcript was prematurely terminated; similarly, *rme2-s2* is only

detected by the Mid - and 5' primer sets. (Fig. 13B, *RME2* lane 3).   The *rme2-s2*

mutation does not appear to effect the repression of *IME4*, which is undetectable in this

strain, as in wild-type. In contrast, expression of the *rme2-s1* transcript, which does not

extend through the *IME4* 5' end, does result in loss of *IME4* repression (Fig. 13B, *IME4*

lane 2). These results suggest that a minimum length of antisense transcript is required to

repress the *IME4* gene.   This may be somewhat similar to the mechanism of repression

of *SER3* by *SRG1*. A terminator was used to prevent the *SRG1* ncRNA from extending

into the promoter caused repression to be lost (Martens, et al., 2004). In the case of

IME4, it appears that extension of the antisense transcript through the 5' end of the gene

is required for repression, but fully crossing the promoter region is not.

iii. Substitution of *HOP1-urs1* promoter

Previous research on the regulation of *IME4* showed that expression of *IME4*

transcripts from the *GAL1* promoter under conditions of galactose induction could

Figure 13. Altering the polyadenylation and termination of the *RME2* transcript alters its ability to repress *IME4* in haploid cells. (A) Illustration of the three PCR amplicons *(IME4* Promoter, Pr; 5' end of the sense ORF, 5'; and middle of the ORF, Mid) used to detect both *IME4* and *RME2* transcripts, up- and downstream of the introduced termination sites in *rme2-s1* and *rme2-s2*. Primers downstream (relative to *RME2*) of each amplicon were used to generate strand-specific cDNA. (B) RT-PCR assays of wild type (YBG147, lane 1), *rme2-s1* (YBG159, lane 2) and *rme2-s2* haploid strains (YBG202, lane 3). Assays and controls were performed as described in Fig. 2. The -RT control was performed for all assay primer pairs.
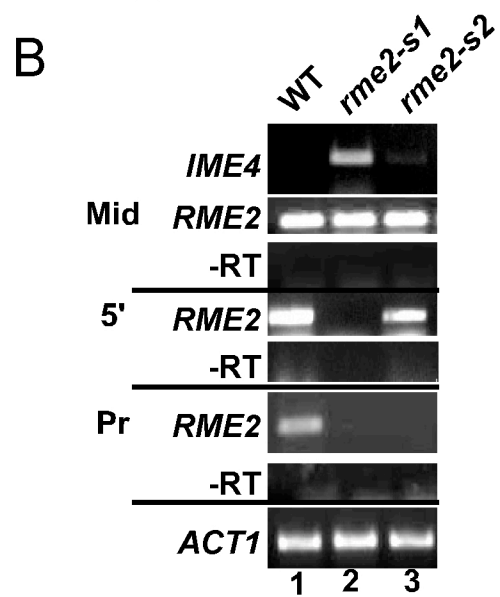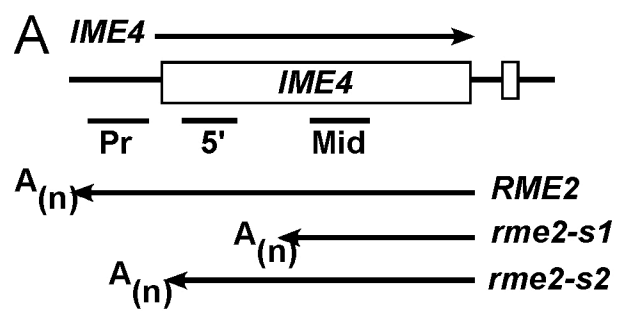
override repression by *RME2* in haploid cells (Hongay, et al., 2006). It is possible that the *GAL1* promoter was not repressed by *RME2* transcription due to differences in promoter specificity. For example, it is possible that the binding of specific transcription factors to the *IME4* promoter is sensitive to disruption by antisense transcription. In contrast, the transcriptional activator of *GAL1,* the Gal4 protein and its cofactors may be insensitive to this form of regulation. To further test this model we assayed for the ability of *RME2* to repress a heterologous promoter with similar activity to the native *IME4* promoter. We constructed a strain in which the *IME4* promoter was substituted with a mutant version of the *HOP1* promoter. The *HOP1* promoter is normally repressed in haploid and diploid cells in vegetative media by the Ume6 repressor protein binding to the URS1 site (Strich, et al., 1994). A mutation in the URS1 site allows constitutive expression from this promoter in both haploid and diploid cells (Vershon, et al., 1992). Diploid cells in sporulation-inducing media expressed similar amounts of *IME4* under the control of either the wild-type *IME4* or *HOP1-urs1* promoter (Fig. 15A, lane 6 vs. lane 8). However, in haploid cells, the *HOP1-urs1* promoter was repressed by *RME2* transcription (Fig. 14A, lanes 5, 7).

To further test this result, we mutated the **a**1-α2 site downstream of *IME4* in the context of the *HOP1-urs1* promoter. If *RME2* transcription is able to repress transcription from the *HOP1-urs1* promoter, then both haploid and diploid cells containing the mutant **a**1-α2 site should be unable to express *IME4.* The *HOP1-urs1* promoter was repressed in diploid cells when the **a**1-α2 site is mutated (Fig. 14B, lane 6 vs. lane 8). These results indicate that *RME2* transcription is able to repress a

heterologous promoter and that there are unlikely to be specific elements or factors bound

at the *IME4* promoter that make it sensitive to antisense transcription.

iv. ChIP for Abf1 protein

The specific transcriptional activators of *IME4* are unknown. However, the

HOP1-urs1 promoter, which is repressed by *RME2* in a manner similar to the *IME4*

promoter, is activated by the transcription factor Abf1 (Gailus-Durner, et al., 1996). We

used this promoter fusion to determine if *RME2* transcription prevents Abf1 from binding

to the DNA. If the promoter interference model is correct, we would expect to see Abf1

bound to the *HOP1-urs1* promoter only in diploid cells, where *RME2* is repressed.

Surprisingly, we found that Abf1 remains bound to the *HOP1-IME4* promoter in cell

types where *IME4* transcription is repressed by *RME2* (Fig. 15). This result suggests that

RME2 does not repress *IME4* through a mechanism of interference with any general

transcription factors binding to the promoter.

v. ChIP for TBP

The inability of the shortened *rme2-s1* mutant to repress *IME4* suggested that it was

possible that antisense expression represses transcription by interfering with factors

binding to the *IME4* promoter. However, the above experiment suggested that this was

not the case in for *HOP1-urs1-IME4,* as Abf1 was still bound, even when *RME2* was

being expressed, and *IME4* repressed. This could be due to constitutive binding of Abf1,

which may render it insensitive to disruption by antisense transcription (Buchman, et al.,

1988). To account for this, I also looked at general transcription factor binding at both

Figure 14. Expression of *IME4* from the *HOP1-urs1* promoter is regulated by antisense transcription in a manner similar to the native *IME4* promoter.   (A) RT-PCR assays of *IME4* and *RME2* from haploid (1n) and diploid (2n) cells, grown in YEPD (Veg) or Sporulation inducing (Spo) media.   *IME4* (sense) transcription is driven by either the native *IME4* promoter (WT, lanes 1, 2, 5, and 6), or *HOP1* promoter containing a mutation in the URS1 site (lanes 3, 4, 7, and 8). (B) RT-PCR assays of *IME4* and *RME2* from 1n and 2n cells with *IME4* (sense) transcription is driven by either the native *IME4* promoter (lanes 1, 2, 5, and 6) or the constitutive *HOP1-urs1* mutant promoter (lanes 3, 4, 7, and 8) with a mutant (Mu) **a**1-α2 binding site downstream of the ORF. Assays and controls were performed as descried in Fig. 9.

A

Veg. ————————— Spo. —————————

Promoter: *IME4*   *HOP1-urs1*   *IME4*   *HOP1-urs1*

1n 2n 1n 2n 1n 2n 1n 2n gDNA

IME4

RME2

-RT

ACT1

1   2   3   4   5   6   7   8   9

B

Veg. ————————— Spo. —————————

Promoter: *IME4*   *HOP1-urs1*   *IME4*   *HOP1-urs1*

1n 2n 1n 2n 1n 2n 1n 2n gDNA

IME4

RME2

-RT

ACT1

1   2   3   4   5   6   7   8   9
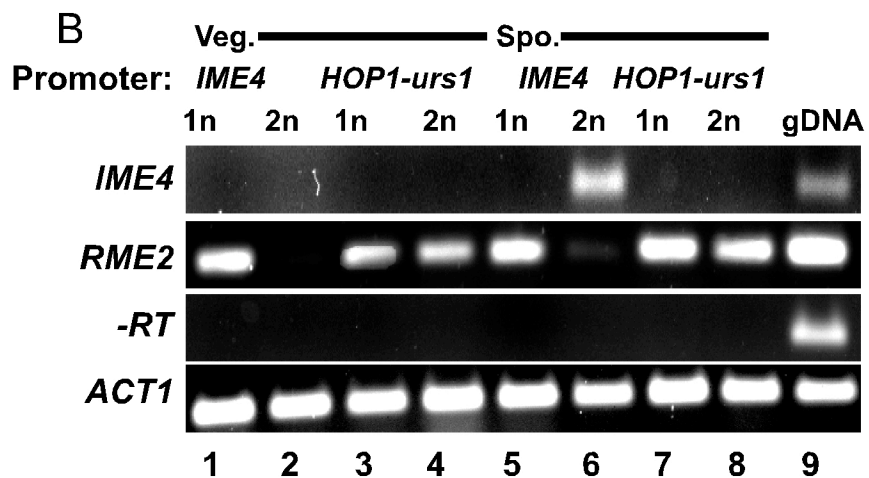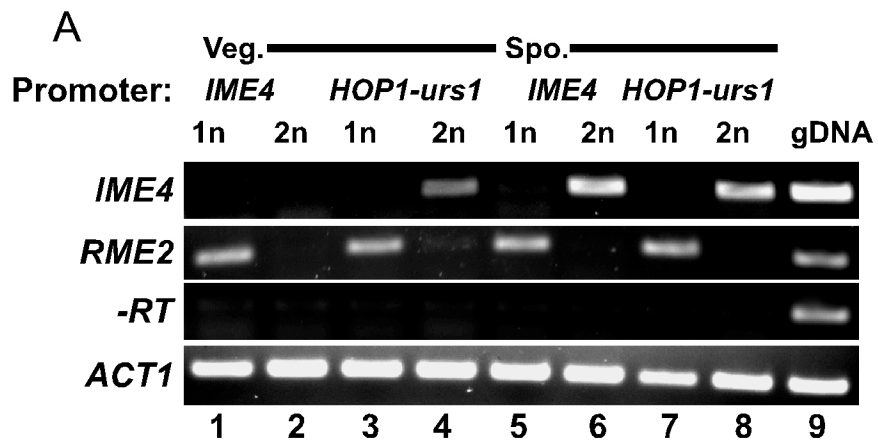
Figure 15. Transcription of *RME2* does not prevent the Abf1 transcription factor from binding at the *HOP1-urs1-IME4* promoter. Relative position of primer set used in a ChIP assays for the presence of Abf1 transcription factor bound at the promoter of *HOP1-urs1::IME4* is indicated at top. Binding was compared to the occupancy of the native *HOP1* gene as a positive control, and to the *RPL19B* promoter, which is not regulated by Abf1, as negative control. ChIP assays to detect the binding of Abf1 at the *HOP1-urs1::IME4* promoter were performed on both haploid (lanes 1, 2, 5, and 6) and diploid (lanes 3, 4, 7, and 8) strains, and in the context of WT (lanes 1-4) and mutant **a**1-α2 binding sites (lanes 5-8) downstream of the *IME4* gene. TC is the Total Chromatin sample (lanes 1, 3, 5, and 7), and IP is DNA Immunoprecipitated with Abf1 yC-20 antibody (lanes 2, 4, 6, and 8).
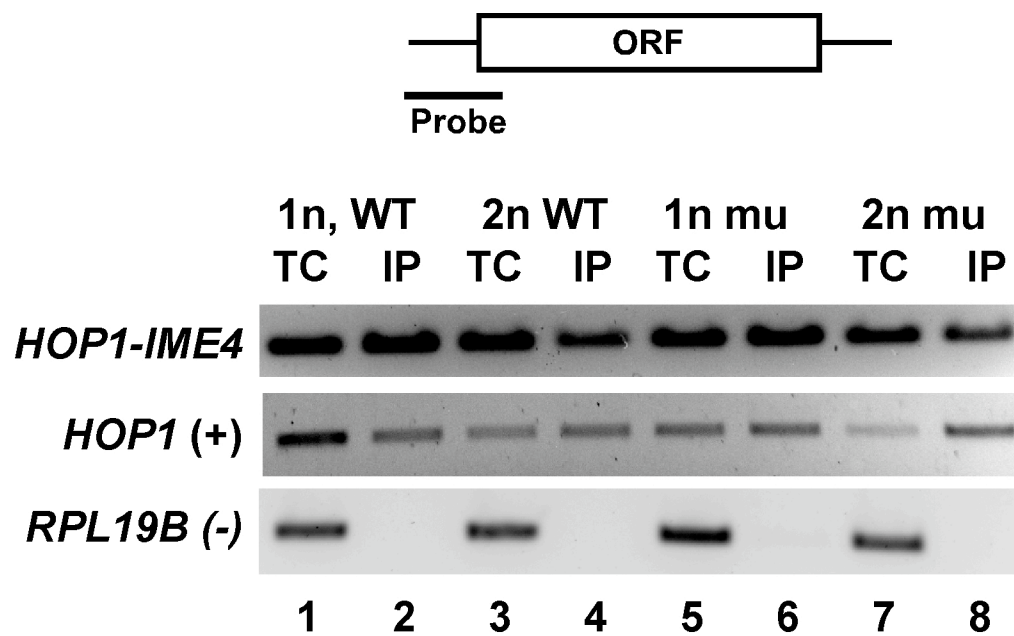
sense and antisense promoter regions. TATA-Binding Protein (TBP) recruitment has previously been used to determine if a non-coding RNA acts to disrupt transcription factor binding at another undetermined promoter (Martens, et al., 2004). The wild-type *IME4* and hybrid *HOP1-urs1-IME4* promoters, were both analyzed for TBP binding, under sporulation-inducing and vegetative growth conditions respectively. TBP binding to the *IME4* promoter (5'), middle of the ORF (Mid) and to the *RME2* promoter (3') was assayed in haploid and diploid cells. The *RME2* promoter was bound by TBP in haploid cells with 16.9-fold higher affinity than in diploid cells (Fig. 16A, 3' *IME4,* lanes 3 vs. 4). As expected, TBP was bound to the *IME4* promoter in diploid cells (Fig. 16A, 5' *IME4,* lane 4). Surprisingly, despite repression of *IME4* by *RME2*, TBP was also bound to the *IME4* promoter with almost equal affinity (0.90-fold difference) in haploid and diploid cells (Fig. 16A, 5' *IME4,* lane 3). Similar patterns of TBP binding were observed for *HOP1-urs1-IME4* (Fig. 16B). These observations show tha*t RME2* transcription is not blocking binding of transcription factors or actively knocking them off the DNA. This shows that *RME2* represses *IME4* through a completely different mechanism than the regulation of *SER3* and *ADH1*, by *SRG1* and *ZRR1*, respectively.

4. ROLE OF THE *IME4* ORF IN REGULATION

i. Substitution of the *URA3* ORF

Regulation of the *IME4* gene by expression of the antisense transcript may be a function of the relative strengths of the promoters (Hongay, et al., 2006). In haploid cells the stronger expression of the *RME2* promoter overcomes the weaker *IME4* promoter. In diploid cells, where **a**1-α2 is bound downstream of the gene, *RME2* promoter is

Figure 16. TBP is not blocked from the *IME4* promoter by expression of *RME2*.

Antisense transcription does not disrupt TBP binding at the *IME4* promoter. (A) ChIP

assays for TBP bound at both the sense (5') and antisense (3') promoters in haploid wild-

type (LNY392 lanes 1,3); diploid wild-type(YBG144 lanes 2,4) cells.   Assays were

performed after 3 hours in SPM media. The middle of the *IME4* and *ACT1* ORFs, which

are not precipitated with TBP-antibody were included as negative controls, and the

constitutive *ACT1* promoter is included as a positive control for antibody binding. TC is

the Total Chromatin sample (lanes 1-2), and IP is DNA Immunoprecipitated with TBP

yN-20 antibody (lanes 3,4). Numbers to the right represent the fold-change in TBP

binding in haploid over diploid cells. B) ChIP assays for TBP bound at both the sense

(5') and antisense (3') promoters in the context of haploid *HOP1-urs1-IME4* (YBG111+

pBG157) or diploid *HOP1-urs1-IME4* (YBG115+ pBG157). The assays are performed

as above.
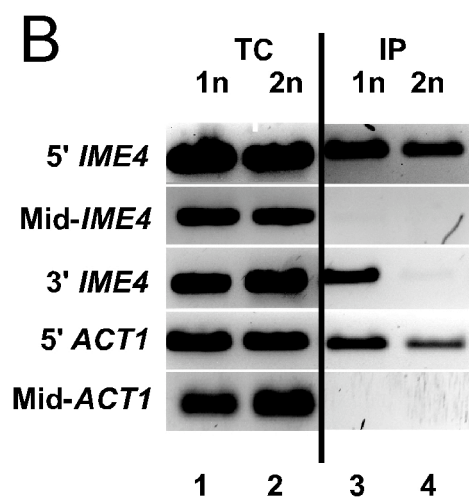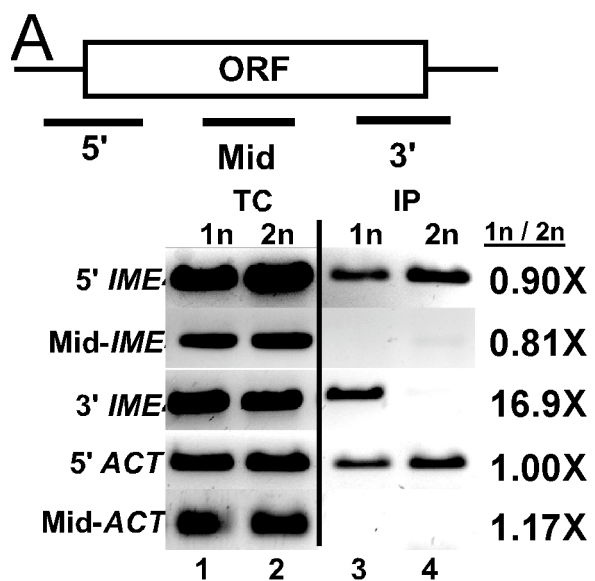
repressed, allowing expression of the weaker *IME4* promoter.   If the mechanism of

antisense-mediated repression is strictly through the relative strengths of the promoters,

then one would predict that if another gene were substituted for the *IME4* ORF it would

be expressed in a similar cell-type dependant manner that is similar to *IME4*.

    To test this model, we substituted the *URA3* ORF for the *IME4* ORF, and assayed

for the ability of haploid and diploid cells harboring this construct regulate *URA3*

expression by examining growth on media lacking uracil (Fig. 17A).   While diploid cells

express the *ime4::URA3* gene and grow as predicted, haploid cells containing the

ime4::URA3 construct are also able grow on media lacking uracil.   This suggests that the

ime4::URA3 construct is not repressed (Fig. 17A). RT-PCR assays for the expression of

the sense and antisense *ime4::URA3* transcripts confirmed that haploid

cells express both the antisense-*ime4::URA3* and *ime4::URA3* sense transcripts (Fig. 17B

lane 1).   In contrast, the native *IME4* gene is properly repressed by the presence of

RME2 in this same strain. These results suggest the possibility that there are sequences

within the *IME4* ORF are required for antisense-mediated repression.

Deletion Analysis of *IME4*

    Because the *HOP1* promoter is correctly repressed by transcription of *RME2,* it

suggested the possibility that the *IME4* ORF sequence has a role in regulation by *RME2*.

This could include sequence that is more sensitive to permitting transcription in the sense

direction to be interrupted by transcription in the antisense direction. To identify possible

regions within the gene that are required for antisense mediated repression, I performed a

series of deletion analyses of the *IME4* ORF.   The different deletions are summarized in

Figure 17. The *URA3* ORF in the context of *IME4* is not regulated by antisense transcription. A) Growth assays of haploid (LNY392) or diploid (YBG144) cells transformed with *ime4::URA3* (pJM530), ura3- (PRS415) or URA3+ (PRS416) plasmids. Five-fold serial dilutions were grown for 3 days, on media with and without Uracil. B) RT-PCR assays of i*me4::URA3* in 1n (LNY392) or 2n (YBG144) cells carrying pJM530 (+) or RS415 vector (-) . The *ime4::URA3* cDNA was primed in the flanking (*IME4*-specific) region, and amplified with *URA3*-specific primers.
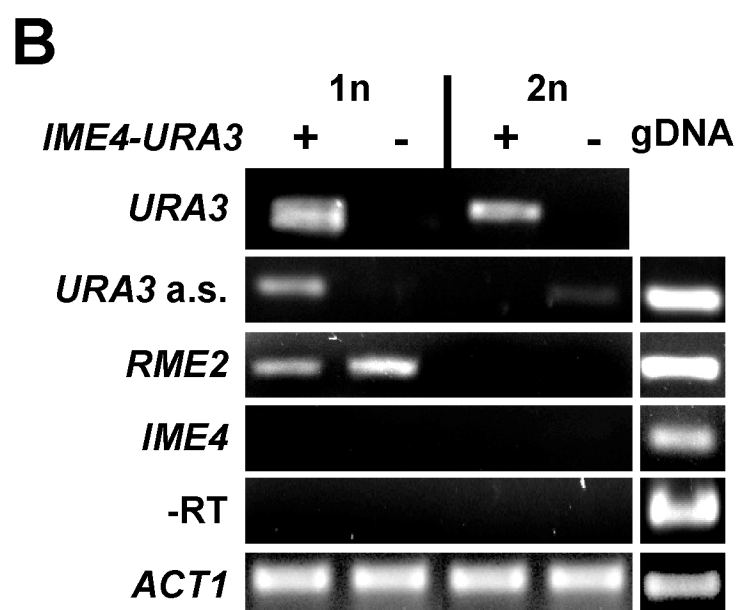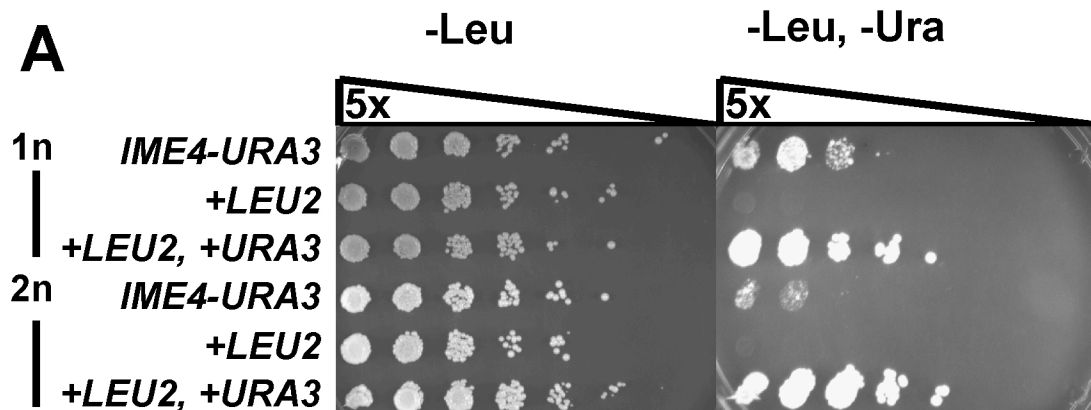
Figure 19A. These deletions were integrated into haploid cells with the native *IME4*

locus deleted, and expression of the *IME4* and *RME2* transcripts was assayed in

comparison to both wild-type haploid and diploid cells. Deletion of base pairs 1-900

derepressed the *IME4* sense transcript in haploid cells, compared to a WT haploid (Fig.

18B, lane 1 vs. lane 3). In contrast, a deletion of bp 901-1800 had no effect on *IME4*

repression in haploid cells *(data not shown)*. It was possible that the loss of *IME4*

transcriptional repression in the *ime4Δ1-900* mutant was due to premature termination of

the *RME2* transcript. However, RT-PCR assays confirmed that *RME2* was expressed

and that the transcript extended across the *IME4* promoter region (Fig. 18B, lane 3).

Taken together, these results suggest that the first 900 bp of the *IME4* ORF is required for

antisense-mediated repression.

To further define what region is required for repression of *IME4* by *RME2*, a

series of smaller deletions were made within the first 900 bp of the *IME4* ORF. Deletion

of bp 1-224 or bp 676-900 had no effect on the repression of *IME4* in haploid cells (Fig.

18B, lanes 6 and 7). In contrast, deletions of bp 1-450 or 451-900 caused derepression of

the *IME4* sense transcript in haploid cells (Fig. 18B, lanes 4 and 5). This indicates that a

DNA element within bp 225-675 is required for proper antisense-mediated repression.

Orientation switch of the required region

Deletion analysis of the *IME4* ORF showed that a region from bp 225-675 is

required for *RME2*-mediated repression of *IME4* in haploid cells. It is possible that

transcription of this region in the antisense direction creates a chromatin structure that

Figure 18. Deletion analysis of *IME4* reveals a region within the ORF is required for antisense-mediated regulation. The cartoon on top shows the location of the deletions and how deletion of the hatched region leads to loss of *IME4* repression. RT-PCR assays of *IME4* and *RME2* expression are shown from haploid (lane 1), diploid (lane 2), and haploid cells with the indicted bp of the *IME4* ORF deleted (lanes 3-7), grown in Spo media. *IME4* expression was monitored using an amplicon within the ORF and *RME2* expression was monitored using an amplicon in the *IME4* promoter region. Deletion strains are listed in Table 2.
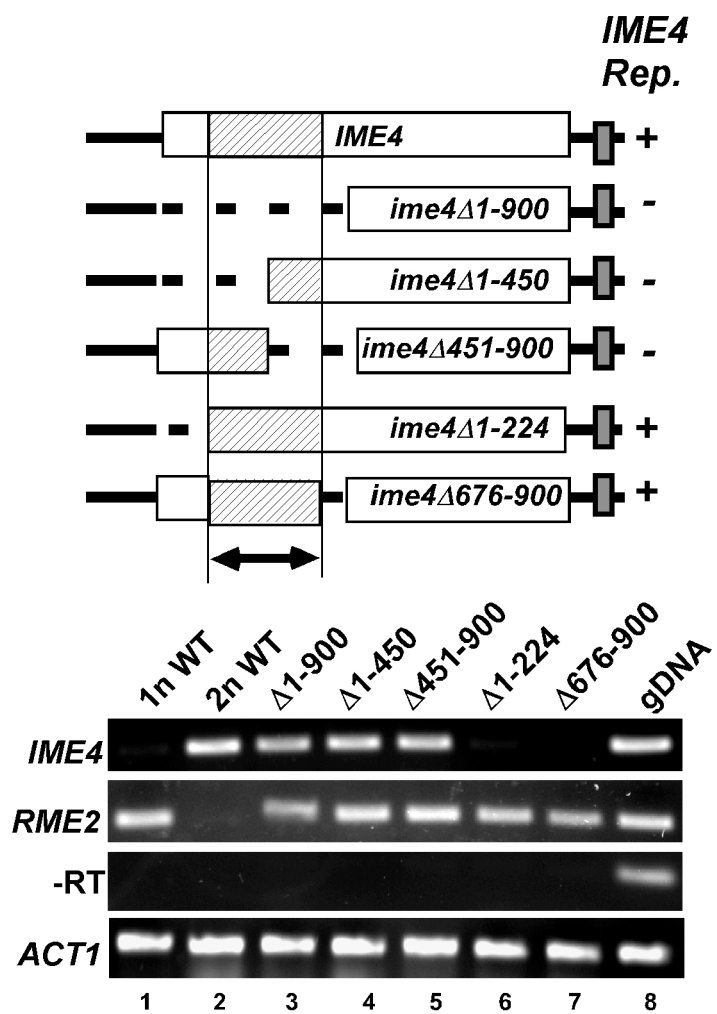
prevents effective extension of the full-length sense transcript. If this occurs, then there may be an orientation-specific requirement of this DNA element.

To test for this requirement, the segment of DNA from bp 225-675 was flipped in the opposite orientation within the context of the *IME4* ORF. While this mutation had no effect on the expression of *IME4* in diploid cells, it caused derepression of the *IME4* transcript in haploid cells (Fig. 19 lane 1 vs. lane 3). Interestingly, assays for the presence of the *RME2* transcript up- and downstream of this region, showed that there was premature termination of the *RME2* transcript in the flipped orientation in contrast to the wild-type orientation (Fig. 19, *RME2- end,* compare lane 1 vs. lane 3). This result suggests that transcription of *RME2* across this element may set up orientation-specific termination of the sense transcript.

Protein cofactors tested for role in regulation

In addition to the DNA element within bp225-675, it is possible there are additional components required for the antisense mediated repression of *IME4*. For example, protein cofactors may bind to this element during the transcription of *RME2* to block transcription of full-length *IME4*. Alternately, there might be chromatin remodeling enzymes that bind at this element and spread, permitting transcription only in the direction of *RME2*. In addition, I had wanted to identify what the specific activators of *IME4* are to determine if these have any role in antisense-mediated regulation.

To identify potential proteins involved in the activation of *IME4*, a search using the TRANSFAC database was performed of *IME4* and its upstream sequence. The

Figure 19. Reversing the orientation of the bp 225-675 region of *IME4* prevents proper regulation by *RME2*. RT-PCR assays of *IME4* and *RME2* from haploid (YBG111, lanes 1,2) or diploid (YBG115, lanes 3,4) cells, carrying *IME4* on a plasmid with bp 225-675 in either the wild-type (pBG112; lanes 1,3) or "flipped" (antiparallel, relative to WT) orientation (pBG166; lanes 2,4).
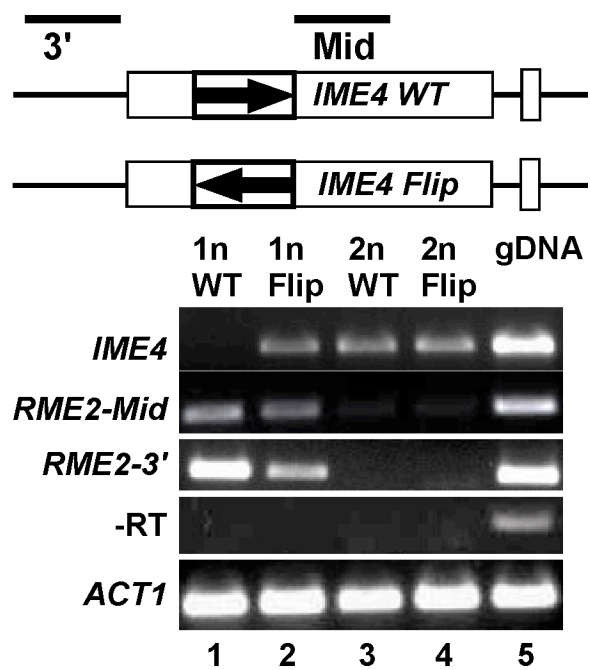
search suggested that there were binding motifs for *ADR1* and *GCN4* within the *IME4*

promoter region. To test if these proteins had a role in either *IME4* activation or *RME2-*

mediated repression, I made both haploid and diploid deletions of these genes, and

assayed for *IME4* and *RME2* expression. There was no significant effect on transcription

of either *IME4* or *RME2* observed for either gene deletion (Fig. 20A and 20B).

Therefore, it is unlikely that either of these transcription factors individually plays a role

in the activation of *IME4*.

Additionally, this study also tried to identify cofactors that may be bound to the

DNA during antisense transcription, and have a role in preventing simultaneous sense

transcription. The elongation factors Spt6 and Spt16 have been previously described as

having roles in the establishment of the boundaries of coding genes, preventing the

expression of ncRNAs from cryptic sites (Duina, et al., 2007). Furthermore, Spt6 and

Spt16 –*ts* mutants have been shown to be defective in repression of *SER3* by *SRG1*

(Hainer, et al., 2011; Thebault, et al., 2011).   If Spt6 and/or Spt16 do have roles in

repression of *IME4*, it would be expected that induction of the –*ts* mutation would result

in expression of both transcripts   Haploid strains carrying the *spt16-197* and *spt6-1004 –*

ts mutants were assayed under both uninduced conditions (Fig. 20C, lanes 1-3) and

induced by a 1-hour shift to 37' C (Fig. 20C, lanes 4-6) for both *IME4* and *RME2*

transcription. Under both conditions, there was no discernable effect on repression of

IME4 by *RME2* (Fig. 20C, lane 4 vs. 5,6). Therefore, it can be concluded that regulation

of *IME4* does not involve these cofactors, unlike *SER3*.

Figure 20.   Testing of Cofactors that may have a role in the regulation of *IME4* and

***RME2*.** A) RT-PCR assays of wild-type haploid (LNY392) and diploid (YBG144) cells

compared to *adr1Δ* haploids (YBG151) and diploids (YBG152). B) RT-PCR assays of

wild-type haploid (LNY392) and diploid (YBG144) cells compared to *gcn4Δ* haploids

(YBG153) and diploids (YBG155).   C) RT-PCR assays of wild-type haploid (DN1359,

lane 1, 4) cells compared *spt6-197* haploids (FY346, lane 2,) and *spt16-1004* haploids

(FY2180 lane 3, 6).   Cells were grown in sporulation media for 3 hours as described

above (SPM), or grown in sporulation media for 2 hours, then grown 1 hour in the same

media at 37 degrees to induce the TS- mutation. All assays and controls are performed as

described in Figure 9.
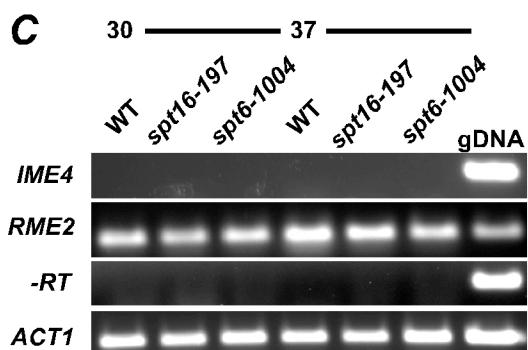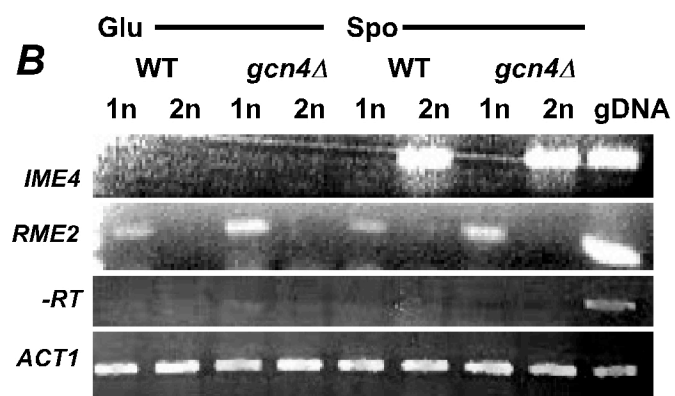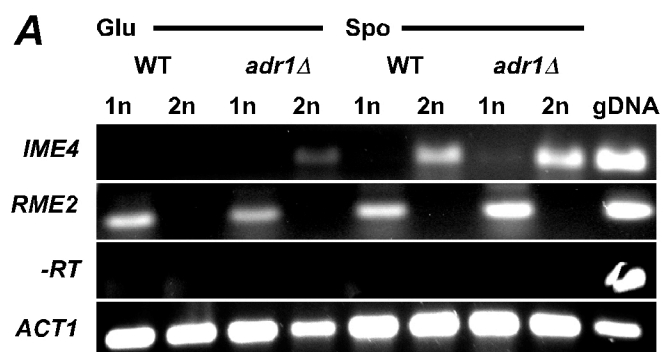
C. Discussion

In this chapter, the cell-type specific regulation of *IME4* by the ncRNA expressed from its antisense strand, *RME2,* was explored as a novel form of gene regulation. It had previously been shown that *IME4* expression required the presence of **a**1-α2, which is known as a transcriptional repressor (Shah and Clancy, 1992; Strathern, et al., 1981). The research presented here shows that the haploid-specific repressor of *IME4*, indicated by the "X" in Figure 3, is in fact the *RME2* ncRNA, and not a protein cofactor. Diploid cells express *IME4* when starved; haploid cells constitutively express *RME2* (Fig. 21A). The connection of **a**1-α2 to *IME4* is through the binding site 125 bp downstream of the ORF (Nagaraj, et al., 2004). Rather than activating *IME4* expression, *RME2* is repressed in diploid cells by **a**1-α2 binding downstream of the *IME4* ORF (Fig. 21B). Loss of a functional site results in diploid cells which constitutively express *RME2*. Expression of *RME2* in this mutant repressed *IME4,* which strongly suggested that *RME2* is the haploid specific repressor of *IME4* (Fig. 21D). The role of *RME2* as a direct repressor of *IME4* was verified by the *ime4Δ3'* mutation, which abolished *RME2* expression in haploid cells and allowed *IME4* expression (Fig. 21C). This circuit is very similar to the regulation of *SER3* by *SRG1*. Repression of *SER3* depends on the Swi-Snf activator complex, which activates the intermediary ncRNA *SRG1*. The fact that **a**1-α2 repressor complex can function as an activator through the intermediary *RME2* suggests the possibility that other **a**1-α2 sites may have similar functionality. The activation of diploid-specific genes may require **a**1-α2 repression of haploid-specific ncRNAs.

The constitutive expression of *IME4* from *ime4Δ3'* allowed me to determine if the interaction between *IME4* and *RME2* occurs in *cis-* or *trans-*. *Trans-* interactions

Figure 21 Repression of *IME4* by *RME2* is cell-type specific. A) In haploid cells, **a**1-α2 is not expressed. The *RME2* transcript is expressed, which prevents expression. In diploid cells (**a**/α), **a**1-α2 is bound downstream of *IME4*. This represses *RME2*, and in turn, permits *IME4* expression. C) Haploid cells which have the *RME2* promoter region deleted do not express *RME2*, and are able to express *IME4*. D) Diploid cells with a mutant **a**1-α2 binding site downstream of *IME4* fail to repress *RME2* in a cell-type specific manner. The expression of *RME2* in diploid cells prevents expression of *IME4* from that locus.

A haploid

B a/α diploid

C promoter mutant haploid

D a1-α2 site mutant diploid
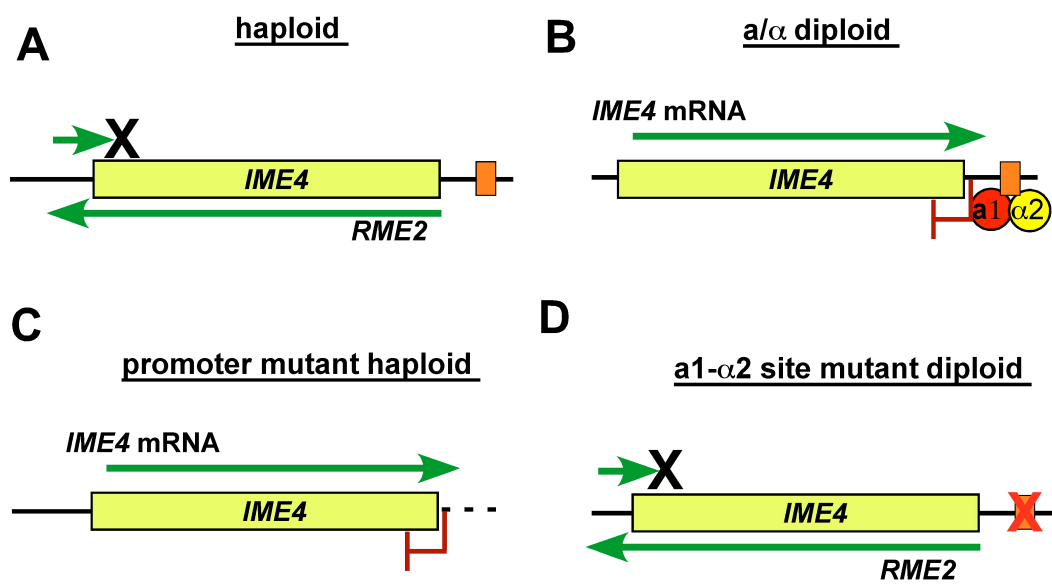
between sense and antisense RNAs have been identified in numerous eukaryotes. This includes the miRNAs and siRNAs (Couzin, 2002; He and Hannon, 2004). However, *S. cerevisiae*, unlike other eukaryotes, lacks the endogenous Dicer, Drosha, and Argonaute homologs needed to carry out siRNA-mediated repression (Drinnenberg, et al., 2009). Prior to this work, there was only one case of a trans- acting RNA known in yeast; all other endogenous regulatory ncRNAs function in *cis*- (Berretta, et al., 2008; Bird, et al., 2006; Martens, et al., 2005). The fact that the constitutive expression of *IME4* from *ime4Δ3'* was not repressed by *RME2* expressed from the *IME4* locus or a heterologous construct even in high copy, is strong evidence that this form of regulation is *cis*- acting; this has been shown by other sources (Hongay, et al., 2006).

The mechanism of how a *cis*- acting antisense transcript functions to repress transcription was also examined. *SRG1* and *ZRR1,* the upstream regulatory ncRNAs , repress by blocking access of transcription factors to the coding gene's promoter. *RME2* also extends across the promoter region of *IME4*, in the antisense direction, so it is formally possible that it regulates in a similar manner. However, analysis of the *rme2-s2* mutant, which does not cross the promoter of *IME4*, showed that transcription of the promoter is not required for repression. Further evidence for the lack of role for the sense promoter comes from the ChIP assays. These showed that even when *RME2* is expressed, and *IME4* repressed, Abf1 and TBP remain bound to the *IME4* sense promoter. It was previously observed that incomplete *IME4* transcripts are expressed even in haploid cells (Hongay, et al., 2006). These results indicate that *RME2* is not repressing *IME4* through promoter interference but through another mechanism.

The *ime4::URA3* construct was expected to create a cell-type specific marker system to identify what cofactors are required to allow antisense transcription to repress sense transcription. However, it was a surprise to find that the *IME4* and *RME2* promoters failed to regulate the *URA3* ORF in the same manner as *IME4*; while antisense-*URA3* was expressed in haploids, this did not repress sense-*URA3* expression. This suggested that possibly the *IME4* ORF itself had sequence or spacing specificity that permitted *RME2* to regulate it. The internal deletion analysis of the *IME4* ORF confirmed this finding, showing that a region of the *IME4* ORF, bp225-675, appears to be the DNA element that responds to antisense transcription. The flipping experiments show that this region must be correctly oriented to repress sense transcription.

In diploid cells, antisense transcription is repressed by the **a**1-α2 complex. In haploid cells, the *RME2* transcript is expressed. Transcription of *RME2* across the antisense strand of the ORF, through bp225-675 prevents full-length *IME4* expression. This may be through a folded RNA structure of the elongating *RME2* transcript that blocks *IME4* elongation. Alternately, the act of transcribing the antisense strand may open the local chromatin structure to permit the binding of proteins that prevent *IME4* expression. These findings show an example of a novel mechanism of transcriptional regulation, which may repress other genes in a similar manner.

IV. ANTISENSE TRANSCRIPTION REGULATES *ZIP2*

Introduction

The identification of the cell-type specific regulation of *IME4* by the antisense transcript *RME2* prompted a major question: How widespread is this form of transcriptional regulation? To identify if other diploid-specific genes were regulated by haploid-specific antisense transcription, the data from the previous work on the identification of **a**1-α2 target sites was revisited, as many sites were found with no apparent cell-type specific regulation of the genes downstream of these sites (Galgoczy, et al., 2004; Nagaraj, et al., 2004). We specifically looked for potential **a**1-α2 targets that were downstream of genes that are regulated in a cell-type specific manner.

To better understand the roles of these sites, data from the Yeast Transcriptome Database was employed (David, et al., 2006). This study utilized a novel stranded, tiled microarray technology to assay for the expression of not only coding genes, but also the antisense strands and intergenic regions of the genome. By reapplying the binding site preference data and the expression values of the antisense strand of the upstream gene, other cases of cell-type specific antisense regulation similar to *IME4* could be identified.

Results

1. Bioinformatic Search for Candidate Genes

Since *IME4* is regulated by transcription of a ncRNA in the antisense direction, we wanted identify other genes regulated by a similar mechanism. We therefore re-evaluated our previous data identifying **a**1-α2 sites in the yeast genome by relaxing the sequence requirements for **a**1-α2 sites and specifically searching for sites that are

downstream of cell-type specific genes (Nagaraj, et al., 2004). We then used the Yeast

Transcriptome Database to search for the presence of antisense transcription for the gene

upstream of the **a**1-α2 site (David, et al., 2006). The *RME2* transcript is detected because

the database assays expression in *MAT***a** cells (Fig. 22A).   Using this approach, we

identified **a**1-α2 sites downstream of antisense-expressing genes *HSP26, YFL012W*,

MCK1 and *ZIP2* as possible targets for antisense-mediated regulation (Fig. 22B-22E).

2. The *ZIP2* locus also expresses a haploid-specific antisense transcript, *RME3*

To test for expression of sense and antisense transcripts expressed in a cell-type-

specific manner similar to *IME4,* RT-PCR assays of both strands were performed on

ZIP2, YFL012W, HSP26, and *MCK1,* in haploid and diploid cells grown in both

vegetative and sporulation media. Antisense transcripts of *ZIP2, YFL012W,* and *HSP26*

were detected by the assay (Fig. 23C, D, E). However, the sense and antisense transcripts

of *HSP26* and *YFL012W* are not differentially regulated in a cell-type specific manner

(Fig. 23C, E). In contrast, analysis of *ZIP2*, a meiosis-specific component of the

synaptonemal complex, suggested the sense and antisense transcripts are regulated in a

cell-type dependant manner. In agreement with gene expression profiling experiments

during meiosis, the *ZIP2* (sense) transcript was expressed only in diploid cells under

conditions of sporulation (Fig. 23B, lane 4) (Chu, et al., 1998; Chua and Roeder, 1998).

In contrast, haploid cells under the same conditions express the non-coding antisense

ZIP2 transcript, which we will refer to as *RME3* (Regulator of Meiosis 3) (Fig. 23B, lane

3). Unlike *IME4,* it appeared that the *RME3* transcript was smaller than the *ZIP2* ORF,

Figure 22. Candidate genes for antisense-mediated repression. Images of sense and antisense transcription from the Yeast Transcriptome Database (http://www.ebi.ac.uk/huber-srv/cgi-bin/viewYeastTilingArray) for the *ZIP2*, *YFL012W*, *HSP26,* and *MCK1* genes, which were identified as potential targets of **a**1-α2-mediated antisense regulation. Numbers are log scale expression of the Watson (top) and Crick (bottom) strands.   Green boxes highlight the antisense ncRNAs
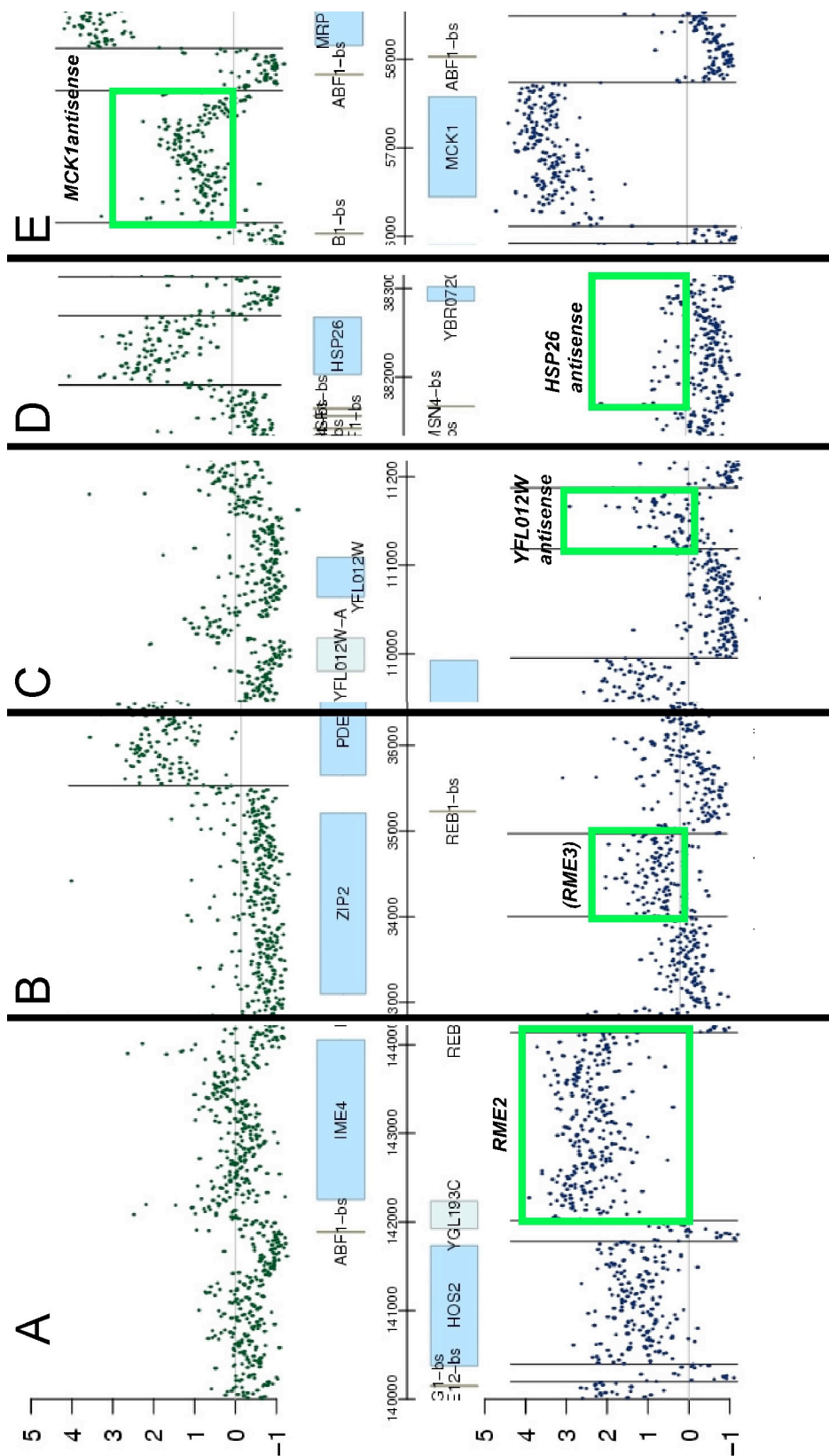
Figure 23. Expression of sense and antisense transcripts of candidate genes. RT-PCR assays of haploid (LNY392) and diploid (YBG144) cells, grown in Vegetative (YEPD) or Sporulation-inducing (SPO) media were performed to detect sense and antisense transcripts of A) *IME4* as a control, B) *ZIP2*, C) *YFL012W*, D) *HSP26*, E) *MCK1*. −RT reactions were run for each primer set. *ACT1* is a loading control for samples.
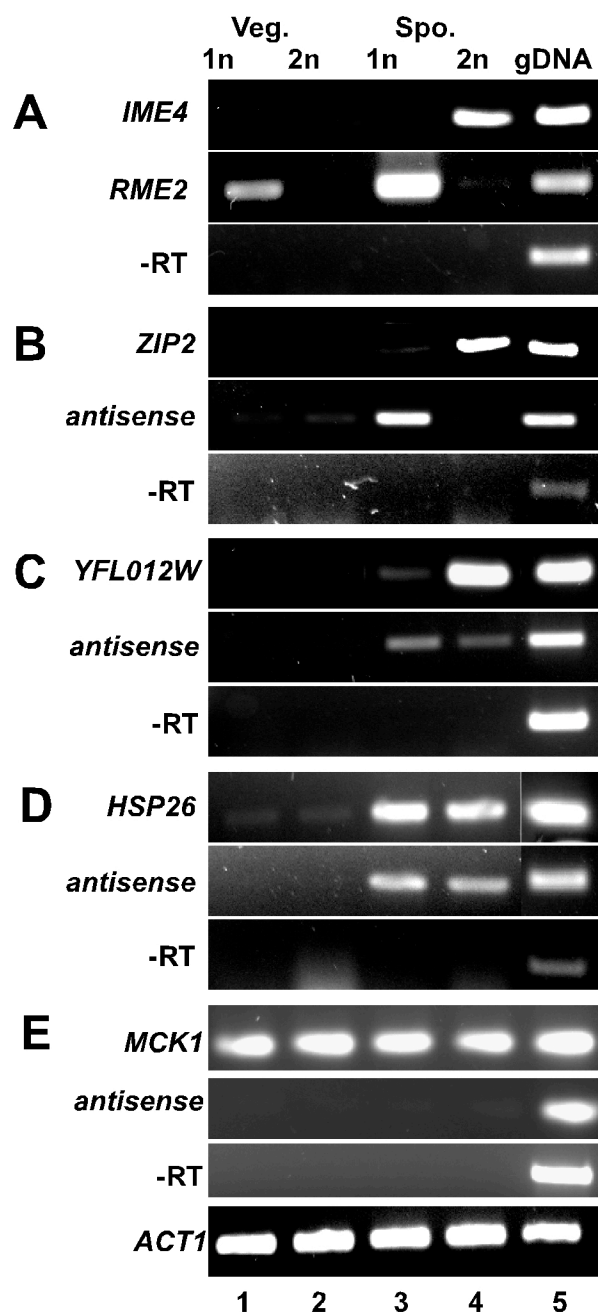
extending only through the last 1200 bp of the 2200 bp *ZIP2* ORF (Fig. 22B)(David, et al., 2006; Miura, et al., 2006).

3. a1-α2 binding represses *RME3,* and *RME3* represses *ZIP2*

The expression pattern of *ZIP2* and *RME3* is similar to that observed for *IME4* and *RME2*. In addition to cell-type specific regulation of the *ZIP2* and *RME3*, we found that **a**1-α2 binds to the predicted site i*n vivo* (Fig. 24A). We were therefore interested if the antisense transcript *RME3* regulated expression of *ZIP2,* similar to the regulation of *IME4* by *RME2*. I first tested if *RME3* is repressed by **a**1-α2. Mutation of four bases in the **a**1-α2 site downstream of *ZIP2* (as described above for *IME4*) allowed expression of *RME3* in diploid cells (Fig. 24B, lane 7 vs. lane 8). This indicated that *RME3* is actively repressed by **a**1-α2 in a cell-type specific manner, similar to *RME2*.

As was previously observed in the mutational analysis of the **a**1-α2 site downstream of *IME4,* expression of the *RME3* transcript in diploid cells repressed transcription of *ZIP2* (Fig. 24B, lane 8). This suggests that a mechanism of antisense-mediated transcriptional regulation, similar to the one regulating *IME4,* represses the *ZIP2* gene in haploid cells.

4. Regulation of *ZIP2* occurs in *cis-*

To test if *RME3* represses *ZIP2* transcription in *cis-* or *tran*s- dependant manner, we constructed a diploid strain heterozygous for the *ZIP2* locus (Fig. 25A). One of the *ZIP2* alleles in this strain is wild-type, while the other contains the 4 bp mutation in the **a**1-α2 site that allows *RME3* expression in diploid cells. If *ZIP2* is repressed by *RME3* in *trans*, we would expect to see a significant drop in the level of detectable *ZIP2* in the

Figure 24. *ZIP2* and *RME3* **are expressed in a cell-type specific manner, regulated by a1-α2 binding.** A) ChIP assay of the *ZIP2-PDE1* intergenic region in *MATα/matΔ* pseudohaploids (JRY118) and **a**/α diploids (JRY103), using antibody to α2 (Nagaraj, et al., 2004). The cartoon illustrates the relative position of this site in relation to *ZIP2* and *RME3*. B) RT-PCR assays of *ZIP2* and *RME3* from haploid and diploid cells grown under vegetative (lanes 1-4) or in sporulation-inducing conditions for 5 hours (lanes 5-8), with either a wild-type (WT) haploid (JMY108, lane 1, 5) or diploid (JMY110, lane 3, 7) strains or **a**1-α2 binding site mutant (Mu) haploid (JMY108 lane 2, 6) or diploid (JMY110 lane 4, 8) strains.

A

	a/α		Δ/α	
	TC	IP	IP	TC
ZIP2-PDE1				
STE6				
YDL223C				

ZIP2 mRNA →   [a1] [α2]   PDE1 mRNA →

ZIP2   PDE1

← RME3 ncRNA

B

Glucose ———————— Sporulation————

	1n WT	1n Mu	2n WT	2n Mu	1n WT	1n Mu	2n WT	2n Mu	gDNA
ZIP2									
RME3									
-RT									
ACT1									
	1	2	3	4	5	6	7	8	9
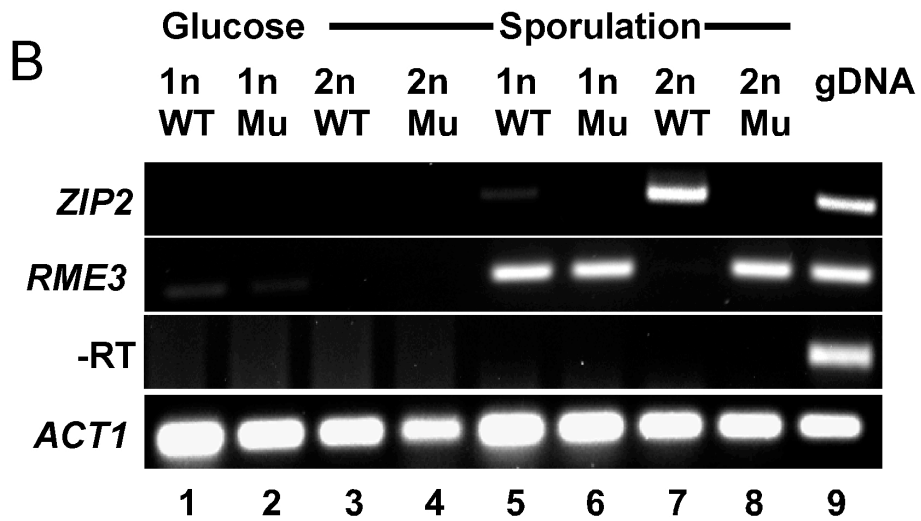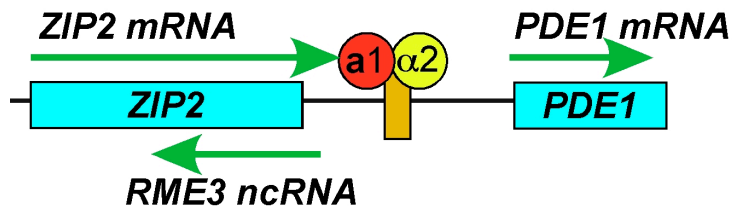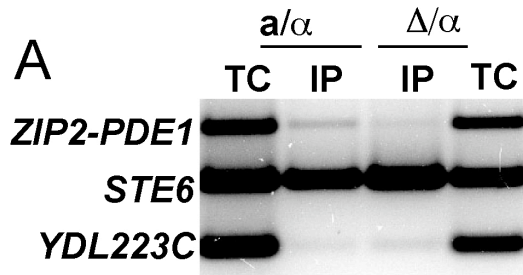
heterozygote compared to wild-type, as previously described for RNAi systems in other eukaryotes (Couzin, 2002; He and Hannon, 2004). However, expression of the *RME3* transcript in *trans* did not repress expression of *ZIP2* in the heterozygote strain in comparison to the wild-type homozygous diploid (Fig. 25B, lane 2 vs. lane 4). This indicates that *RME3* regulates *ZIP2* in a *cis*-dependant configuration, similar to the *IME4/RME2* regulatory system.

5. ChIP for TBP at *ZIP2/RME3* promoters

My previous research on *IME4* had shown that unlike *SRG1* or *ZRR1, RME2* does not appear to block factors binding to the promoter of *IME4*. There is no difference in the amount of TBP bound at the *IME4* promoter, in haploid or diploid cells (Fig. 17A). To determine if the mechanism of antisense mediated repression is similar for *ZIP2/RME3,* I assayed the binding of TBP at the *ZIP2* and *RME3* promoters from wild-type haploid and diploid cells. The *RME3* promoter, which is repressed by the **a**1-α2 complex, is bound by TBP in a haploid but not diploid cells, much like the *RME2* promoter (Fig. 26, *ZIP2* 3' lane 3 vs. 4). Similar to *IME4*, TBP binding was found to be non-cell-type dependant at the *ZIP2* promoter, with the transcriptionally repressed haploid cell showing equivalent levels of binding to diploid cells (Fig. 26, *ZIP2* 3' lane 3, 4).

In contrast, the *HOP1* gene, which is activated at a similar time point in meiosis, only has TBP bound to its promoter in diploid cells. This shows that other cell-type specific promoters are not constitutively bound by TBP (Fig. 26, *HOP1* lane 3 vs. 4). The RT-PCR assays indicate that both *RME3* and *ZIP2* are repressed in vegetative media,

Figure 25. *ZIP2* is repressed in *–cis* by *RME3*. A) Cartoon illustrating the potential

mechanism for dsRNA mediated repression of *ZIP2 in trans*. If repression can function

in *trans,* antisense RNA and mRNA will anneal to form dsRNA leading to repression of

the mRNA. (B) RT-PCR assays of *ZIP2* and *RME3* from wild-type (JMY110 lanes 1,2)

or heterozygous diploid cells (YBG158, lanes 3,4) grown in Veg. or Spo media, as above.

Assays and controls were performed as described in Fig 23.
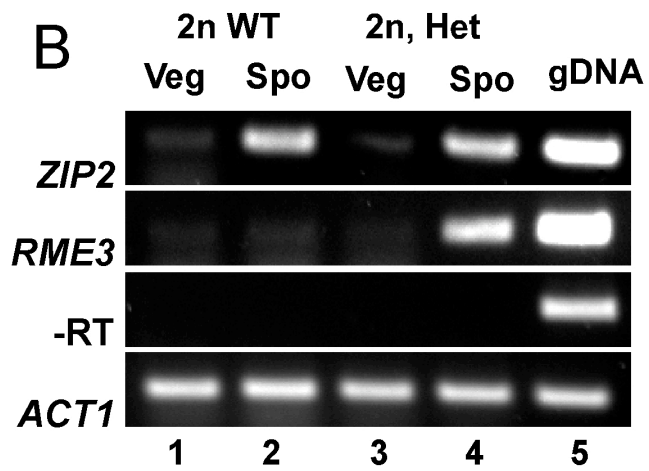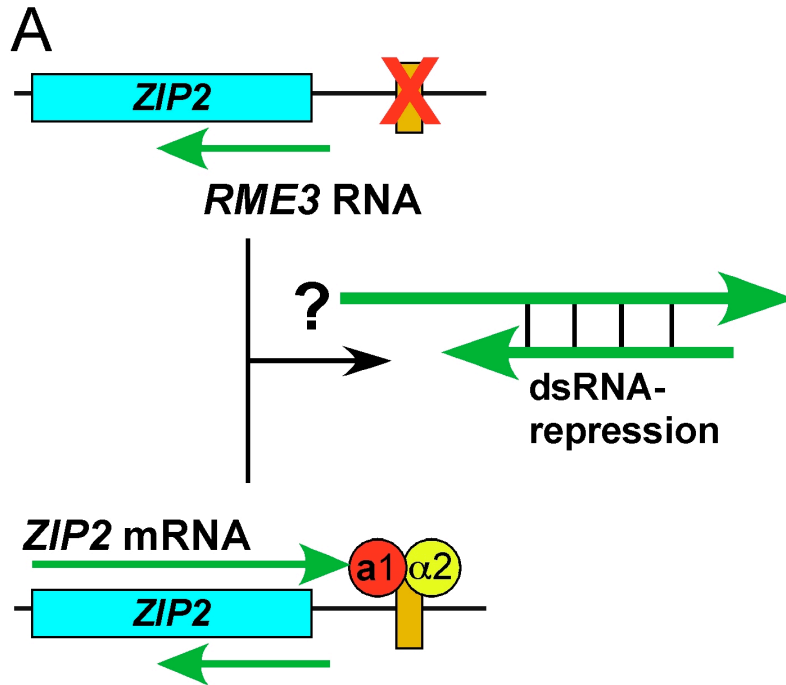
Figure 26. TBP binding at the *ZIP2* and *RME3* promoters. Amplification of the *ZIP2*

and *RME3* promoters from haploid (LNY392; lanes 1, 3) and diploid cells (YBG144;

lanes 2,4). TBP binding was assayed with the 5', Mid-, and 3' amplicons illustrated in the

cartoon. The native *HOP1* promoter was tested as a meiosis-specific, cell-type dependant

promoter. ChIP assays were also performed on the *ZIP2* and *RME3* promoters in

vegetative cultures of both cell types, to confirm that these regions are not constitutively

bound by TBP. Numbers represent the fold enrichment of TBP binding in haploid vs.

diploid, measured and normalized by ImageJ quantitation.
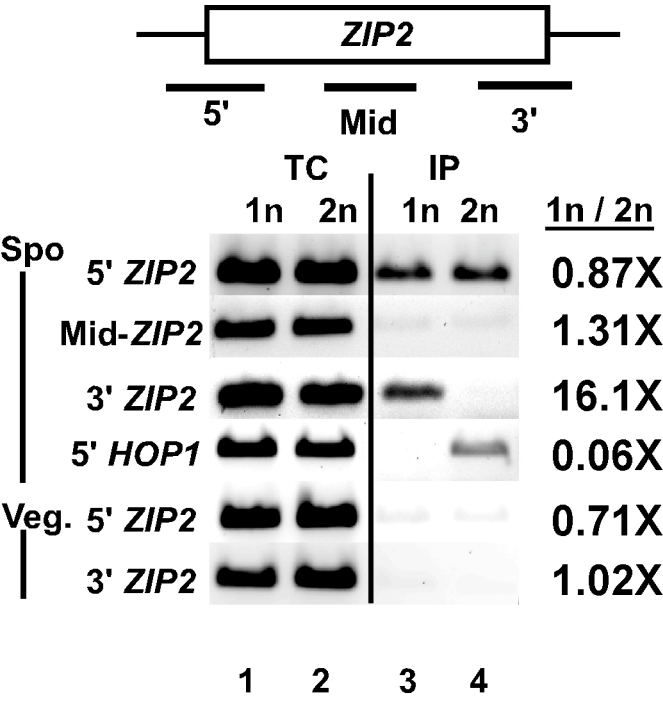
suggesting that TBP binding may also be conditionally regulated. The ChIP assays

directed against the *ZIP2* and *RME3* promoters in vegetatively grown cultures of wild-

type haploid and diploid cells were performed to confirm that these regions are not

constitutively bound by TBP (Fig. 26, Veg. *ZIP2* 5' and *ZIP2* 3' lanes 3, 4).   These

results taken together suggest that like *RME2*-mediated repression of *IME4, RME3*

transcription does not repress *ZIP2* through mechanisms of interference with factors

binding to the promoter.

6. *ZIP2* transcription initiates, but does not extend full-length, in haploid cells

        The ChIP assay showed that TBP is recruited to the *ZIP2* promoter in meiosis-

inducing conditions in both haploid and diploid cells. This suggested that transcription of

ZIP2 may be initiated, but elongation of the full-length transcript may be blocked by

RME3 expression.   *IME4* has been shown to produce short transcripts even when

repressed by *RME2* (Hongay, et al., 2006). To test if this is true for *ZIP2*, I used a series

of RT-PCR assays distributed along the ORF to detect the extent of *ZIP2* expression in

haploid cells (Fig. 27A). Each of these different primer sets detected *ZIP2* from diploid

cells (Fig. 27B, lanes 2, 5, 8, 11).   In haploid cells, while the 5' end primer sets detected

ZIP2, the two 3'-most primer sets failed to detect *ZIP2* mRNA. This showed that

transcription initiation from the *ZIP2* promoter is not regulated by *RME3*. It suggests that

RME3 blocks full-length elongation of the *ZIP2* transcript.

Figure 27. *ZIP2* transcription is initiated, but not extended full-length in haploid cells. A) Primer sets used for tiled RT-PCR of *ZIP2* to assay for aborted sense transcription. B) RT-PCR of *ZIP2* expressed from haploid (LNY392) and diploid (YBG144) strains grown in sporulation media for 5 hours, using tiled primer sets described above. The –RT controls were run for each detection primer pair. *ACT1* control is quantitation of the individual cell extracts. *RME3* was assayed with the 1600-1850 bp primer set as a control for correct cell-type regulation.

A

PCR   1-200   325-650   1100-1350   1600-1850

ZIP2 mRNA

ZIP2

RME3 nc-RNA

B   1-200   325-650   1100-1350   1600-1850

1n   2n   G   1n   2n   G   1n   2n   G   1n   2n   G

ZIP2

-RT

1   2   3   4   5   6   7   8   9   10   11   12

1n   2n   G

RME3

ACT1
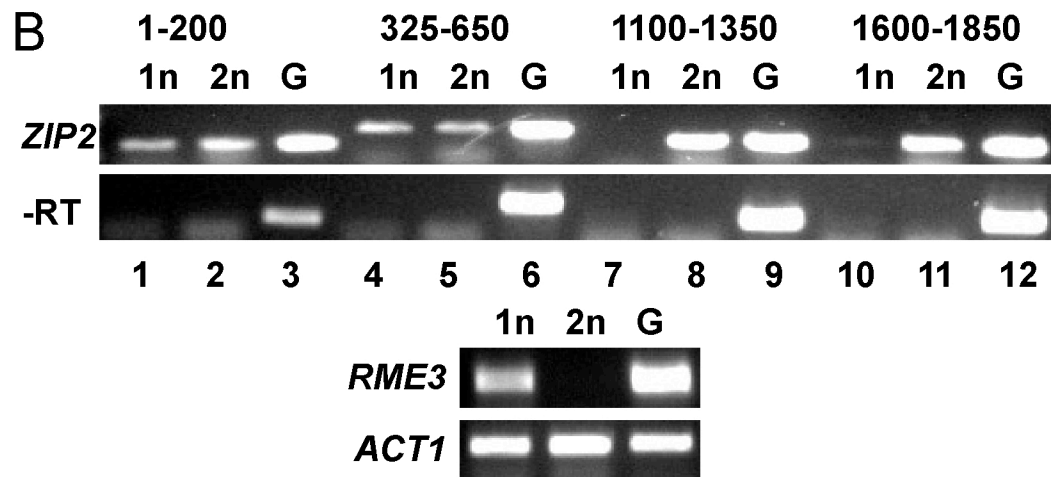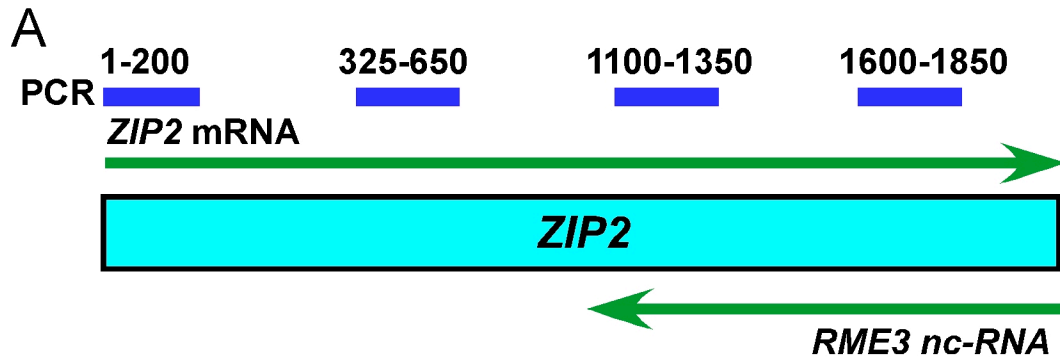
Discussion

The finding of a second cell-type specific gene that is regulated by antisense transcription showed that this is a conserved form of regulation. Like *IME4/RME2*, the diploid-specific gene *ZIP2* is "activated" by **a**1-α2-mediated repression of the haploid-specific antisense ncRNA *RME3*. *RME3* appears to prevent full-length sense transcription by acting in a *cis*- dependant manner at the same ORF, though it does not appear to affect either the transcription factor binding or initiation of *ZIP2*. This raises the question as to how *RME3* regulates *ZIP2*. It is possible that like *IME4*, a region in *ZIP2* creates a chromatin structure that acts as a physical barrier to full-length transcription. This region could also serve as a recruitment site for cofactors that prevent *ZIP2* elongation. It is possible that if repression by *RME2* and *RME3* work by the same mechanism then there would be similar sites or sequences within the coding regions of *IME4* and *ZIP2* that block elongation. However, sequence alignment of the *IME4* bp225-675 region with *ZIP2* did not show any significant homology matches. Therefore, it may be that the repressor element of *ZIP2* is distinct from that in *IME4*, and it is context-specific to *ZIP2*.

It may also be possible that there is no specific element in *ZIP2*, and that repression is related to the act of antisense transcription in the vicinity of the *ZIP2* 3' end. It has been shown that *isw2* mutants express transcripts from cryptic promoters, which do not span the entire coding region (Whitehouse, et al., 2007). These antisense transcripts can reduce the coding gene expression. Therefore, *RME3* may be an example of a non-cryptic antisense transcript that regulates in a similar manner. While the specific mechanisms and cofactors of *IME4* and *ZIP2* regulation have not been completely

discerned, the conservation of cell-type regulatory transcripts indicates that there may be

many more examples of this form of regulation.

V. TRANSCRIPTOME ANALYSIS BY SOLiD SEQUENCING

Introduction

The identification of *ZIP2* as a second case of a gene regulated by a haploid-specific antisense transcript suggested that this mechanism of repression may regulate additional genes in *S. cerevisiae*. In support of this model, antisense ncRNAs were identified by other groups, which regulate *PHO84* and *TY1,* through somewhat different mechanisms (Berretta, et al., 2008; Camblong, et al., 2009). The previous search that identified *RME3* was limited by several factors. One, the search was based on finding specific protein binding sites for **a**1-α2 that were close to the downstream regions of diploid-specific genes. If there is no **a**1-α2 site regulating the antisense RNA, or if its P-value for identification was below threshold, the target will be missed. Second, the identification of antisense RNAs was based on a single tiling array, performed in a *MAT***a** background, under a single growth condition. If a particular ncRNA is not expressed at a high enough level in that condition, such as an **a**/α− specific ncRNA, it would also be missed. Therefore, to truly explore the non-coding transcriptome of yeast by cell-type and growth condition, I used Strand-Specific RNA-sequencing (RNA-seq) Deep-Sequencing technology to assay expression levels of all transcripts expressed in *S. cerevisiae*. This technique was used to detect ncRNAs, so that an analysis to correlate antisense expression to coding repression could reliably be performed.

The main principle of RNA-seq is that the total RNA, depleted of ribosomal RNA, is fragmented to a short segments (<100 bp) and then ligated to adaptors for sequencing. In the ABI SOLiD system used for this study, the adaptors are strand-specific, creating a library that preserves the strandedness of the original RNA (Fig. 28A**)**. Hundreds of millions of short sequences, known as reads, are produced by this

Figure 28. Strand-specific library assembly and barcoding A) Strand-specific library

assembly using the SOLiD transcriptome protocol. 1) Ribosomal RNA is removed, and

2) remaining sample is fragmented with RNAse III. 3) RNA fragments 100-300 bp in

size (green) are ligated with directional RNA-DNA hybrid adaptors (Red/Blue). A

degenerate sequence (N) hybridizes to the RNA and maintains directionality. 4) Reverse

transcriptase fills in the second strand of cDNA from the 3' end. 5) Library is PCR

amplified with a common primer (blue arrow) and a second barcoded library primer

(Pink). 6) Sequencing is performed only through a specific primer, to maintain strand-

specificity. B) Barcoding protocol for parallel running of replicate, randomized, and

blocked samples.   7) Library PCR is performed with unique primers (Pink or Light blue)

for different libraries. Quantitated samples are mixed prior to ePCR and sequencing by

ligation. 8) Reads generated in a common pool have a unique barcode associated with

sequenced DNA.   9) Sequences are bioinformatically deconvoluted into their respective

libraries and data is mapped to the genomic sequence and counted.

A

1) Remove rRNA

2) Fragment RNA

3) Ligate directional adaptors

4) Fill-in with RT

5) PCR with barcode primer (pink)

6) Sequence directionally (red)

B

7) Sequence barcode (pink; blue) and insert

8) Generate pool of barcoded reads

9) Assign to libraries, map, and count reads
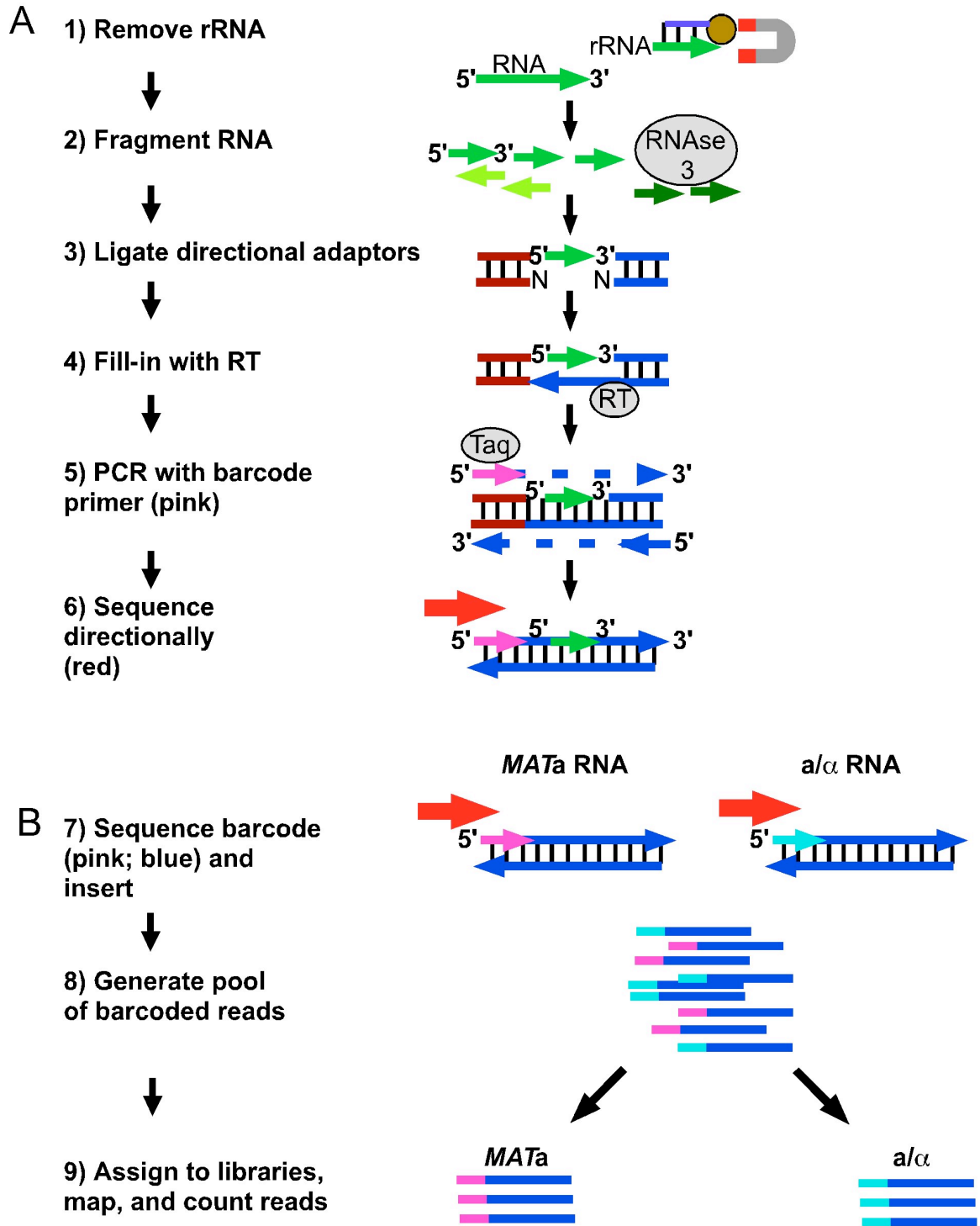
process; they are then bioinformatically reassembled and mapped to the genome, to recreate the transcriptome.

RNA-Seq offers several advantages over the microarray methods. First, barcoding technology allows different replicates, both biological and technical, to be run in a single reaction (Fig. 28B). Second, because RNA-seq reads sequences, instead of using a less stringent hybridization approach, one can assign where each read came from with a greater degree of specificity. This allows genes with high sequence homology to be individually assayed with less problems of cross-reactivity. Third, because the sequences are discrete units, instead of a threshold-based assay like microarrays, RNA-seq has a much greater dynamic range and sensitivity can be detected, allowing rare species to be reliably identified.

To identify antisense ncRNAs which may be differentially regulated in different cell-types, *MATa/MATα* and **a**/α diploid cells were used. These strains preserve the genomic copy number and will only identify expression changes related specifically to the *MAT* locus, and not ploidy (Galitski, et al., 1999; Strathern, et al., 1981). These strains also contain an *ash1::LEU2* mutation, to normalize gene expression between mother and daughter cells (Sil and Herskowitz, 1996). To examine the differences in expression between cells in mitotic growth and those that have entered the early steps of meiosis or starvation, each strain was grown in both vegetative and in sporulation-inducing media for three hours. Triple biological replicates of the four permutations of cell-type/growth were used to reliably detect rare transcripts that are differentially expressed.

Results

Mapping, Transcriptomic Coverage, and Gene Expression Validation

To saturate the yeast transcriptome, and detect low expression species reliably, stranded fragment library RNA-seq with 50 base pair reads using SOLiD 4.0 was performed, using triple biological replicates of *MATa* –Glu, **a**/α–Glu , *MATa*-SPM, and **a**/α -SPM. The number of reads, mappability, and transcriptomic coverage for each of the four biological triplicates of is shown in Table 3. Mappability is defined as the percentage of reads which are mapped correctly by Bioscope to the yeast genome (version SacCer2) permitting 2 mismatches per 50 bp read length (4% of read length). Previous RNA-seq studies have used a standard of 1 mismatch per 10 bp (10% read length) (Auer and Doerge, 2010; Mortazavi, et al., 2008; Nagalakshmi, et al., 2008). However, due to the degree of coverage obtained in our experiments, we have used a more stringent criteria to enhance the reliability of mapping. The *MAT***a**-Glu replicate 3 library was found to have a significant variance in the number of raw and unmappable reads, but the total number of mapped reads was not significantly different from other replicates. This has consistently been observed in other sequencing runs (David Sidote, personal communication). Transcriptome coverage is the total bp length of the reads mapped within the boundaries of the coding strands of ORFs (number mapped x 50 bp) divided by the total bp length of the same coding regions. Previous RNA-seq analysis of *S. cerevisiae*, which was performed with the non-strand specific Illumina protocol had an average coverage of 22X (Nagalakshmi, et al., 2008). The range for individual replicates in my experiment was from 120X-206X coverage (Table 3). Pearson's correlation test

Table 3. Readcount Totals, Mappability, and Transcriptome Coverage of RNA-seq

Total reads is the number of reads generated RNA-seq of a single biological replicate of the listed strain (YBG200, *MATa/matΔ*; JRY103, **a**/α) under the listed growth condition (2%Glu = log phase in YEPD; 3HSpm = 3 hours grown in SPM media). Each library was individually barcoded. Mapped reads are the number mapped to the yeast genome, version SacSer2, using Bioscope 1.2 and allowing for 2 mismatches/read. Mappability is the percent of total reads mapped. Transcriptomic coverage is the calculation of total number of reads mapping to coding strand sequences times read length (50 bp) divided by the total length of coding sequences.

Table 3. Readcount Totals, Mappability, and Transcriptome Coverage of RNA-seq

Cell-type/ Condition	Total	Mapped	Mappability	Transcriptome Coverage
*MAT***a**_2%Glu -1	43988587	30131122	69%	166x
*MAT***a**_2%Glu -2	44301859	31200697	70%	172x
*MAT***a**_2%Glu - 3	113200477	37292719	33%	205x
a/α_2%Glu - 1	36672651	26442372	72%	145x
a/α_2%Glu - 2	35837153	24777275	69%	136x
a/α_2%Glu - 3	37236313	26286950	71%	144x
*MAT***a**_3HSpm - 1	42118488	28439292	68%	156x
*MAT***a**_3HSpm - 2	36759253	24633263	67%	135x
*MAT***a**_3HSpm - 3	34008174	22311885	66%	122x
a/α_3HSpm - 1	31110443	21715294	70%	119x
a/α_3HSpm - 2	31256641	22011841	70%	121x
a/α_3HSpm - 3	54483064	36898931	68%	203x

Figure 29. Correlation of biological replicates of RNA-seq data Log-scale plots of the Pearson's correlation between each pair of biological replicates in each triplet.   The r-value denotes significance of variability between two replicates.
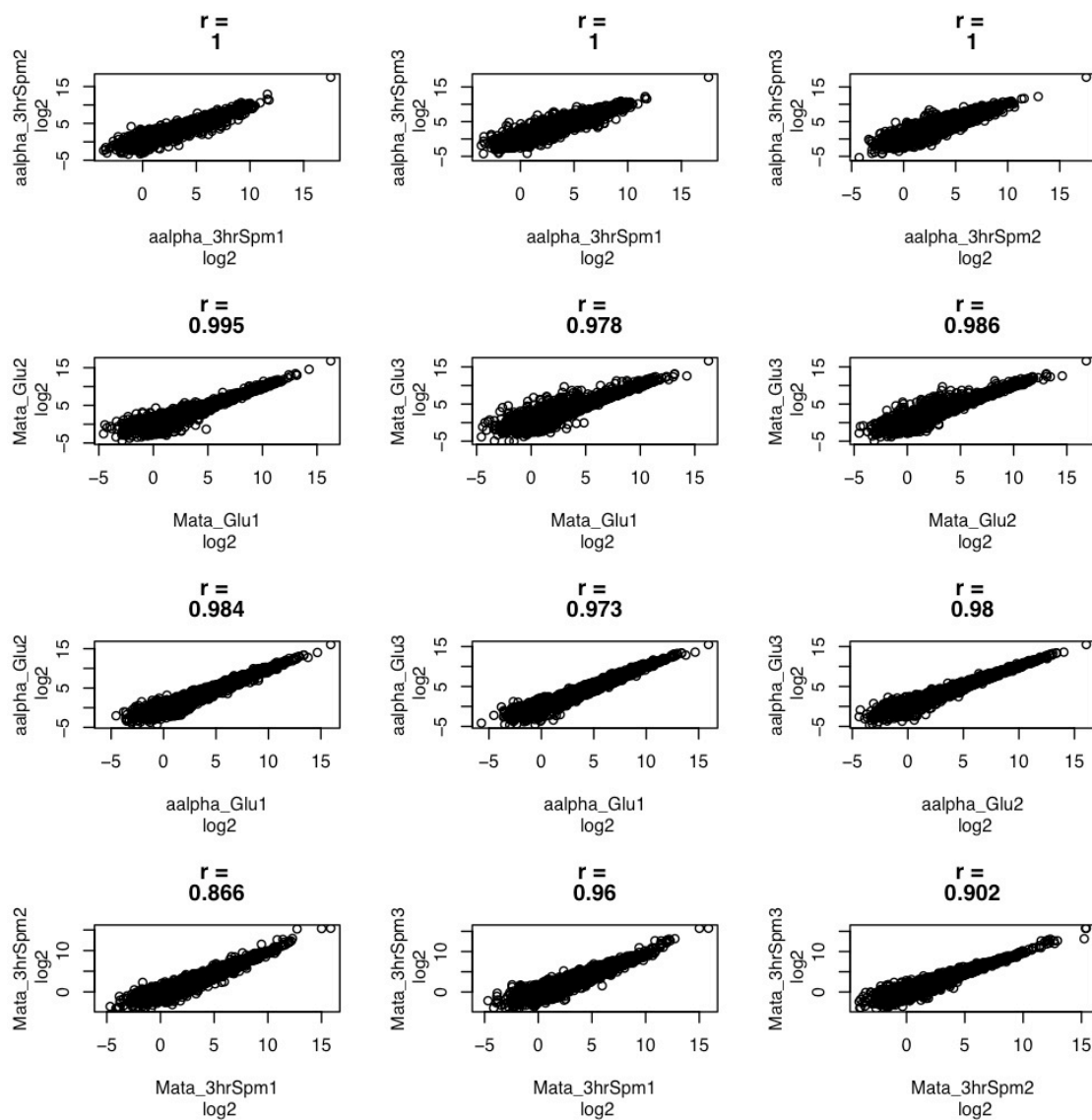
of the biological replicates showed that there was non-significant statistical variance in gene expression from sample to sample (Fig. 29).

To quantify gene expression, Reads-Per KB per Million reads collected (RPKM) values were generated for all ORFs, including both verified and dubious ORFs (Mortazavi, et al., 2008; Pepke, et al., 2009; Wilhelm and Landry, 2009). To verify that the SOLiD protocol detected the correct transcripts for the two different cell-types, I compared the expression level in *MAT*a and a/α cells, by RPKM, to identify cell-type specific genes that are repressed in diploid cells. I then compared this list of genes to with the cell-type specific expression from microarray data sets (Galitski, et al., 1999). The microarray analysis defined differential expression between two samples as a change of greater than 3-fold and a difference of 100 units in the signal values between samples. Haploid-specific expression in the RNA-seq dataset was similarly defined as those genes that show a greater than 4-fold change of the averaged RPKM (*MATa* – Glu divided a/α–Glu). Low expression genes in the RNA-seq dataset (RPKM < 1.0) were filtered against by selecting only genes with a numeric difference of 5.0 in the RPKMs between *MAT*a and a/α.

The SOLiD RNA-seq detected the majority of a1-α2 regulated genes idenfied by microarray (Table 4). *DDR2*, which was shown to be a1-α2 regulated in the microarray experiments, was excluded due to our more stringent filter for expression change (Galgoczy, et al., 2004). This data verifies that the cell-type specific coding transcriptome is being detected properly by SOLiD.

The SOLiD RNA-seq analysis also identified 30 genes that are repressed by greater than 4-fold in the a/α diploid and that had not been previously identified as a-

specific or haploid-specific genes by microarray studies. Thirteen of these genes have no known function. It is possible that they may have functions in haploid-specific pathways such as mating and invasive growth. Although the *MAT***a** and **a**/α strains used in the preparation of the RNA-seq libraries were grown in the same batch of media under the same conditions, *MAT***a** cells interestingly showed a significant enrichment in the expression of genes involved in arginine metabolism and vitamin B6-biosynthesis,. It is possible that these changes in expression are related to the differences in growth and nutrient usage in *MAT***a** and **a**/α cells (Gimeno and Fink, 1992; Gimeno and Fink, 1994; Guo, et al., 2000). This data shows that RNA-seq's increased sensitivity may be useful for detecting new examples of differential expression that were undetectable in microarray studies.

2. *MATa* and a/α cells express different antisense ncRNAs, in both vegetative and sporulation-inducing media

The RPKM analysis method was able to properly identify genes that are regulated in a cell-type specific manner, a modified version of this standard method was used to measure the non-coding strand expression. The value ncRPKM (non-coding Reads Per Kilobase per Million) was generated, using the same formula, but counting reads which mapped to the antisense strand of the ORF (Mortazavi, et al., 2008). Many of the dubious ORFs, which do not have evidence of *in vivo* expression, are arranged in an antisense configuration relative to known coding genes. Expression of the coding gene can therefore appear as false positive for a transcript antisense to the dubious ORF. We therefore excluded all ORFs annotated as dubious from the analysis.

Table 4. RNA-seq expression of haploid-specific genes.   Comparison of averaged
RPKM values of *MAT***a** Glu (Column 3) and **a**/α-Glu libraries (Column 4). Haploid-
specific expression data from the SOLiD RNA-seq was filtered by the standard (RPKM-
*MAT***a**Glu / RPKM **a**/α-Glu) > 4.0 (Column 5), and RPKM-*MAT***a**Glu > RPKM **a**/α-Glu
by five units (Column 6). Column 7 lists what known cell-type specific regulator
represses the listed genes according to ChIP-Chip and microarray data (Galgoczy, et al.,
2004). Comparable haploid-specific genes from a microarray analysis were those
expressing a minimum 100 expression units in *MAT***a**/*MAT***a** cells and a 3-fold increase in
expression in *MAT***a**/*MAT***a** units compared to **a**/α cells (Column 8)(Galitski, et al., 1999).
(*) An extra copy of *TRP1* integrated at the mating type locus is present in YBG200
(*MAT***a**/*MAT*Δ). A (-) indicates these genes were not part of this dataset (Galitski, et al.,
1999; Kessler, et al., 2003).

Table 4. RNA-seq expression values of haploid-specific genes

ORF	Gene	RPKM MATaG	RPKM a/αG	Ratio: MATa / a/α	ΔRPKM	ChIP-ChIP Repression target	Micro-array Haploid-specific	Function
YNL145W	MFA2	121.9	1.1	115.2	120.9	a2-Mcm1	Yes	Mating Factor a
YDL227C	HO	10.4	0.1	80.5	10.3	a1-α2	Yes	Mating type switching
YJL157C	FAR1	19.6	0.4	50.1	19.2	a1-α2	Yes	CDK Inhibitor
YBL016W	FUS3	23.2	0.5	47.8	22.7	a1-α2	Yes	Mitogen activated mating kinase
YMR095C	SNO1	57.8	1.7	33.4	56.0			Pyroxidine metabolism
YLR031W		43.4	1.5	29.7	42.0			Unknown
YGL032C	AGA2	12.9	0.6	21.8	12.3	α2-Mcm1	Yes	Agglutination
YIL015W	BAR1	29.4	1.6	18.7	27.8	α2-Mcm1	Yes	Aspartyl protease of MATa cells
YJR086W	STE18	12.0	0.7	17.5	11.3	a1-α2	Yes	G protein dimer with STE4
YFL026W	STE2	50.1	2.9	17.1	47.2	α2-Mcm1	Yes	Receptor for alpha-factor
YDR461W	MFA1	25.0	1.7	14.5	23.3	α2-Mcm1	Yes	Mating pheromone a-factor
YNL036W	NCE103	308.5	24.5	12.6	284.0			Protein export
YLR265C	NEJ1	13.9	1.2	11.5	12.7	a1-α2		Non-homologous recombination
YDR114C		35.3	3.2	11.1	32.1			Unknown
YMR096W	SNZ1	70.1	6.6	10.7	63.5			Sno1 coregulated/complex
YPL056C	LCL1	144.9	14.4	10.1	130.5			Unknown; deletion is long lifespan
YER138W-A		101.8	10.8	9.4	91.0		(-)	Unknown
YGL170C	SPO74	6.8	0.8	8.7	6.0			Prospore membrane
YHR007C-A		23.0	2.7	8.5	20.3			Unknown
YFL059W	SNZ3	9.3	1.2	8.0	8.2			Sno3 coregulated/complex
YOL058W	ARG1	314.6	39.6	7.9	275.0			arginosuccinate synthestase
YNL333W	SNZ2	7.7	1.0	7.9	6.7			Sno2 coregulated/complex
YMR175W-A		61.1	8.0	7.6	53.1		(-)	Unknown
YDR103W	STE5	11.1	1.5	7.4	9.6	a1-α2		Mating MAPK
YLR030W		9.2	1.3	7.3	7.9			Unknown
YDR366C		6.3	0.9	6.9	5.3			Unknown
YHR005C	GPA1	21.3	3.2	6.7	18.1	a1-α2		alpha subunit, pheromone receptor
YGR240C-A		10.0	1.6	6.2	8.4		(-)	Unknown
YKL209C	STE6	40.9	6.6	6.2	34.3	α2-Mcm1	Yes	Export of a-factor
YFL021W	GAT1	34.6	5.9	5.9	28.8			Regulates nitrogen catabolism
YLR307C-A		6.3	1.1	5.8	5.2		(-)	Unknown
YBR073W	RDH54	80.9	14.2	5.7	66.7	a1-α2	Yes	DNA-dependent ATPase
YOR212W	STE4	52.4	9.4	5.6	43.0	a1-α2	Yes	G protein dimer with Ste18
YDR007W	TRP1*	24.6	4.4	5.6	20.2			Extra copy present in YBG200
YDR019C	GCV1	75.2	13.9	5.4	61.4			Glycine catabolism
YJR109C	CPA2	241.2	44.6	5.4	196.6			Carbamoyl phosphatease synthetase
YGL117W		103.6	20.2	5.1	83.4			Unknown
YLR318W	EST2	15.0	2.9	5.1	12.0			Telomerase RT
YDL181W	INH1	25.4	5.3	4.8	20.1			ATP hydrolysis inhibitor
YGR044C	RME1	89.3	19.1	4.7	70.3	a1-α2	Yes	Repressor of Meiosis
YCL056C		58.0	13.1	4.4	45.0			Unknown; thermotolerance
YGR153W		54.2	12.6	4.3	41.6			Unknown
YJL088W	ARG3	37.3	8.7	4.3	28.5			Ornithine carbamoyltransferase
YPR122W	AXL1	7.7	1.9	4.1	5.8	a1-α2		Haploid specific endoprotease
YDR380W	ARO10	21.0	5.1	4.1	15.9			Phenylpyruvate decarboxylase
YDL210W	UGA4	9.6	2.4	4.0	7.2			Permease that GABA transport
YPL256C	CLN2	41.2	10.3	4.0	30.9			G1 cyclin

Tiled array analysis had previously detected 402 significant antisense transcripts, detected in *MAT*a cells from both poly(A) and total RNA (David, et al., 2006). Strand-specific RNA-seq using the Illumina protocol detected 1103 antisense units expressed in *MAT*a BY4741 background cells (Yassour, et al., 2010). This analysis did not exclude dubious ORFs, and included small transcripts that may only overlap UTRs.

Statistically significant antisense transcription in the SOLiD RNA-seq was defined by a minimum threshold ncRPKM for an expressed transcript being greater than 1.0 (Auer and Doerge; Li, et al.; Wilhelm and Landry, 2009). Based on this criteria, RNA-seq identified a total of 1421 unique antisense transcripts across all four libraries, 1020 of which were expressed (ncRPKM > 1.0) in *MATa* -Glu cells. This contrasts with the antisense expression of *MATa*-SPM (603 transcripts); **a**/α–Glu (732 transcripts); and **a**/α -SPM (742 transcripts). The changes in total number of expressed antisense ncRNAs suggests that cell-type and growth condition induce changes in ncRNA expression, similar to changes in coding gene expression.

Although many genes appeared to have antisense transcripts, this does not necessarily imply that they have a regulatory function in controlling the sense gene. For example, *YGR031W* completely overlaps *NAG1* in the antisense direction (Ma, et al., 2008). However, both *YGR031W* and *NAG1* are expressed. Therefore, to identify the ncRNAs that are regulatory, the relationship between antisense expression and sense repression needs to be determined.

3. Differential shifts between sense and antisense transcripts similar to *IME4*

As shown in Chapter 3, *IME4* is repressed by the haploid-specific antisense transcript *RME2* in *MAT***a** and *MAT*α haploid, cells while the *IME4* sense transcript is normally expressed only in **a**/α diploids (Gelfand, et al., 2011; Hongay, et al., 2006; Shah and Clancy, 1992). Therefore, this locus can be used as a model for stranded differential gene expression. The expression data from the SOLiD transcriptome experiments shows that *MAT***a** cells express *IME4* at an average RPKM value of .061 in the *MAT***a** library, while the *RME2* transcript (measured by the ncRPKM of *IME4*) is expressed at a value of 5.2 (Fig. 30). In contrast, RPKM and ncRPKM values for *IME4* in **a**/α diploid cells change to 9.2 and 0.21, respectively, indicating that *IME4* was expressed and that *RME2* was repressed (Fig 30). Representative data of *IME4* and *RME2* reads mapping to the *IME4* locus visualized with the Integrated Genome Viewer is shown (Fig. 30) (Robinson, et al., 2011). These results suggested that difference in the RPKM and ncRPKM value may be used to bioinformatically identify other genes that are regulated in a similar manner to *IME4*.

With statistically correlated biological replicates of the expression data, we can confidently make comparisons about the changes in expression of both coding and noncoding genes, between two libraries (Auer and Doerge). To identify additional cases of antisense-mediated regulation, I looked for instances where the non-coding transcript is expressed (ncRPKM > 1.0) at a level at least two-fold greater than the expression level of the sense transcript. For all genes with an expressed antisense RNA, stranded differential expression was compared by the ratio of RPKM to ncRPKM value using the formula (coding RPKM + K / noncoding RPKM + K; where K is a constant lower than the lowest RPKM value to correct for a zero-value, 0.0001). This formula for

Figure 30. Expression of *IME4* in SOLiD Dataset. Expression of *IME4* and

RME2 in *MAT***a**-SPM **(**YBG200; library 8) and **a**/α−SPM (JRY103; library 12) cells

assayed by SOLiD RNA-seq. The noncoding transcript (*RME2*) is represented by a green

arrow. Mapped reads were visualized with Integrated Genomics Viewer (Robinson, et

al., 2011). Genes encoded on the Watson strand (*IME4*, *COX13*) are colored red. Genes

encoded on the Crick strand (*HOS2, CDC55, RPS26*) are colored in blue. Individual

reads mapped to *S. cerevisiae* genome are colored according to strand in the same

manner. Reads mapped to *HOS2*, *COX13, CDC55,* and *RPS26* in both cell types, as well

as *IME4* in **a**/α cells, map in the direction of the ORF (black arrow). In a *MAT***a** cell, the

reads mapping to the *IME4* locus are blue, indicating *RME2* expression (green arrow).

The table below lists averaged expression values of both strands in the *MAT***a**-SPM and

a/α−SPM libraries.   Coding genes are measured by Reads Per Kilobase per Million reads

collected (RPKM) and non-coding strand Reads Per Kilobase per Million reads collected

(ncRPKM) values(Mortazavi, et al., 2008).   The expression levels of *RME2* are

measured by ncRPKM-*IME4*. Values under 1.0 are considered to be non-expressed

transcripts.
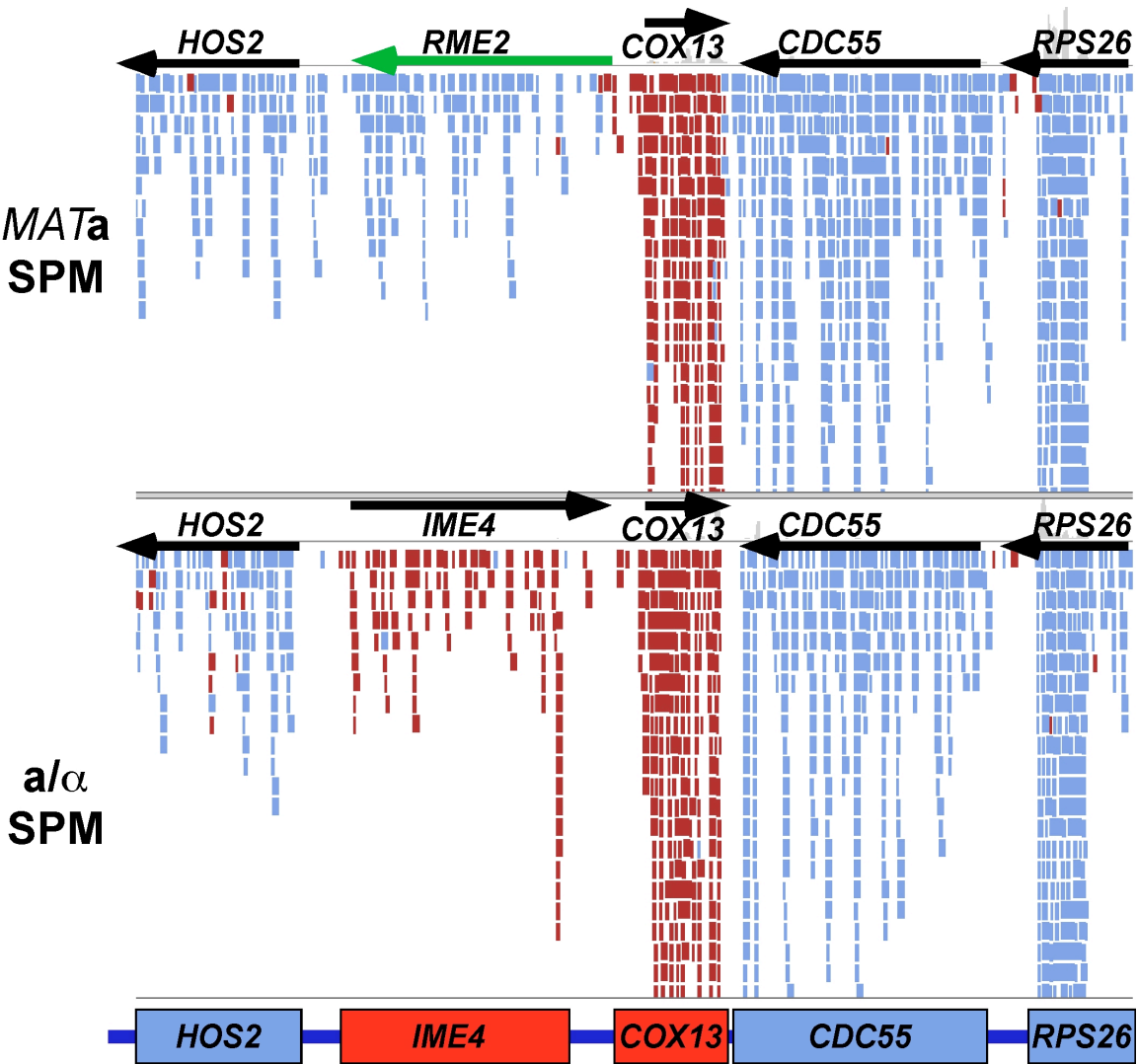
	HOS2	*IME4*	*COX13*	*CDC55*	*RPS26A*
*MAT***a**SPM RPKM	6.3	0.1	46.8	22.3	156.7
a/α SPM RPKM	8.7	9.2	71.5	22.5	117.3
*MAT***a**SPM ncRPKM	0.11	5.2	0.19	0.15	0.2
a/α SPM ncRPKM	0.63	0.21	0	0.08	0.39

expression change, which we termed Stranded RPKM Ratio (SRR) was used to classify

genes into these three categories. Category 1 is comprised of genes where antisense

expression is greater than sense expression by a minimum of 2- fold, like *IME4* in *MAT***a**

cells (Fig. 30, 31). Category 2 is the subset of genes where sense and antisense

expression are within 2-fold of each other. Category 3 is comprised of genes that have

sense expression over 2-fold greater than antisense expression, like *IME4* in diploid cells

(Fig. 30). All genes in each library had SRR calculated based on the averaged RPKM

and ncRPKM expression values. Genes expressing antisense transcripts at an ncRPKM >

1.0 in each library were categorized by SRR value (Table 5).

The majority of expressed antisense transcripts were in SRR category 3,

indicating that they are expressed, but at a level lower than the coding transcript.

Pervasive transcription across non-coding regions of the yeast genome, without respect to

strandedness, has previously been observed (Nagalakshmi, et al., 2008). It is possible

that these transcripts are expressed from cryptic promoters exposed during transcription

of the sense gene. Alternately, SRR category 3 may represent antisense transcripts that

are induced in conditions other than those observed here to repress coding transcription.

There were also a significant number of genes in each library that are in SRR Category 2,

where both strands are transcribed simultaneously. These genes may be similar to the

IME4 deletion mutants and *URA3* replacement, which are not sensitive to repression by

antisense transcription (Chapter 3). Alternately, antisense transcription may partially

downregulate, as opposed to completely repressing sense expression. The genes in SRR

Category 1 are those most like *IME4,* where antisense expression levels of *RME2* are

Figure 31. Stranded RPKM Ratio (SRR) measures relative amounts of sense-antisense transcription.   ORFs are represented by the red boxes, transcribed to the right. Red arrows represent expression of the sense transcript, measured by RPKM. Blue arrows represent antisense transcripts, measured by ncRPKM. Expression values for each non-dubious ORF were compared by the formula ((RPKM+.0001)/(ncRPKM+.0001)) to generate SRR values which fall into three categories illustrated above. (1) ncRPKM greater than RPKM by 2-fold; (2) both RPKM and ncRPKM within 2-fold; (3) RPKM over 2-fold greater than ncRPKM.

Category 1

SRR < 0.5

ORF

Category 2

0.5 < SRR < 2.0

ORF

Category 3

SRR > 2.0

ORF
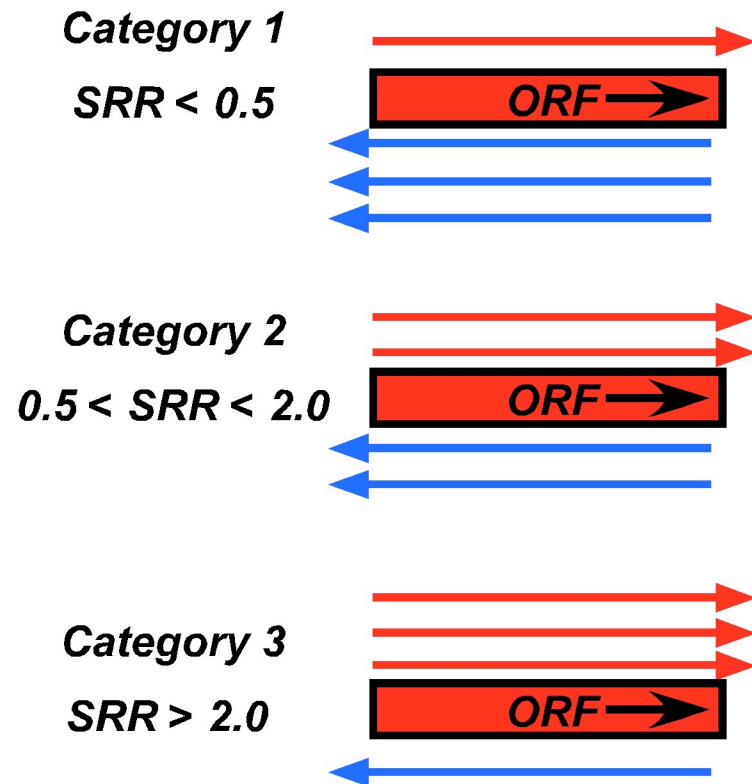
Table 5. SRR Analysis of Expressed (ncRPKM > 1.00) Antisense RNAs

	*MAT*a-Glu	a/α Glu	*MAT*a-SPM	a/α SPM
Total # expressed antisense	1020	732	603	742
Category 1: SRR < 0.5	183	158	145	175
Category 2: 0.5 < SRR <2.0	180	109	148	155
Category 3: SRR > 2.0	657	465	310	412

Totals are the number of genes, out of 5910 non-dubious ORFs, which have an expressed antisense RNA, defined by an ncRPKM >1.00.

significantly higher than sense. If the SRR Category 1 genes switch to SRR category 2 or 3 in other conditions then these may be candidates for antisense mediated regulation.

The first comparison I made was between the *MATa* -SPM and **a**/α–SPM libraries since the regulation of *IME4* by *RME2* is observed under this condition. In this case, the expression levels of *IME4* (RPKM) and *RME2* (ncRPKM) shown in Figure 30 produce SRR values of .012 (Category 1) in *MAT***a** cells, and 43.8 (Catgegory 3) in **a**/α cells.   In *MATa*-SPM, there are 145 genes which have are SRR Category 1, indicating that a higher level of antisense transcripts are expressed than sense (Table 6, Col 3, Total row). When compared to **a**/α–SPM, there were 14 genes, including *IME4,* which change from SRR Category 1 to SRR Category 3 (Table 6, Col 3, Row 1 to 3).   An additional 23 genes change to SRR Category 2. This group may include genes where the antisense transcript is downregulating the sense gene, rather than completely repressing it (Table 6, Col 3, Row 1 to 2). It also includes genes where bidirectional transcription occurs that is independently regulated. Interestingly, there are 108 genes that are in SRR Category 1 in both libraries. These antisense transcripts are expressed at a higher level than sense transcripts in both libraries, and are not differentially regulated by cell-type or under sporulation-inducing conditions (Table 6, Col 3, Row 1 to 1).

A similar comparison was used to identify **a**/α–specific antisense ncRNAs that may repress coding transcription specifically during sporulation. These are genes that are in SRR Category 3 in the *MATa*-SPM library, and change to SRR Category 2 or 3 in the **a**/α–SPM library. Of this subset of 32 genes, 20 change to SRR Category 2 and 12 change to SRR Category 3 in *MATa*-SPM (Table 6, Col 3, Row 2 to 1 and Row 3 to 1, respectively).   Most Category 3 genes remain the same in both libraries.

Table 6: Comparative analysis of antisense expressing genes

	SRR category change between libraries	*MAT***a** SPM vs. **a**/α SPM	*MAT***a** Glu vs. **a**/α Glu	**a**/α Glu vs. **a**/α SPM	*MAT***a** Glu vs. *MAT***a** SPM
Category 1	Total	145	183	158	183
SRR < 0.50	1 to 1	108	134	108	121
	1 to 2	23	29	34	32
	1 to 3	14	20	16	30
Category 2	Total	148	180	109	180
0.50 <= SRR <2.0	2 to 2	64	61	42	51
	2 to 1	30	18	29	32
	2 to 3	54	101	38	97
Category 3	Total	309	657	465	657
SRR > 2.0	3 to 3	277	622	333	598
	3 to 2	20	30	105	43
	3 to 1	12	5	29	16

The remaining comparisons of the four RNA-seq libraries were made using the same breakdown of SRR change as described above for the *MATa*-SPM vs. **a**/α−SPM analysis. The results of these comparisons (**a**/α−Glu vs. **a**/α−SPM; *MATa* **-**Glu vs. *MATa* −SPM; *MATa* **-** Glu vs. **a**/α−Glu) are summarized in Table 6. As stated above, the genes in SRR Category 1 (antisense transcripts) in one library that change to SRR Category 3, or vice versa are of primary interest, as they follow the expression pattern of *IME4* (Table 6, Rows 1 to 3 and 3 to 1).

To select for the genes where sense and antisense expression levels changed in an inversely proportional manner, the list of genes generated by the SRR comparison was filtered by additional standards. In the above analysis only genes with ncRPKM > 1.0 in the SRR Category 1 condition were considered. To look for differential expression, I screened for genes in this list that had an RPKM > 1.0 in the SRR Category 3 condition, and which had at least a 2-fold change in RPKM and ncRPKM. Genes that passed one RPKM-fold filter by this standard, and passed the second RPKM fold filter by a more lenient standard of 1.25-fold were also considered as lower-ranked candidate genes. The specific genes that shift between SRR Categories 1 and 3 in each comparison are listed in Table 7. The same set of filters was applied to genes that shift from SRR category 1 to Category 2, which may represent downregulation by antisense transcription. These genes are listed in Table 8.

These comparisons uncovered many interesting examples of differential expression. *FAR1* was previously shown to be a haploid specific gene (Galgoczy, et al., 2004). RNA-seq showed that there is an **a**/α− specific *FAR1* antisense transcript, which

may regulate *FAR1*. *PRM4*, which is involved in mating, is upregulated in starved *MAT***a** cells (Heiman and Walter, 2000). However, during sporulation, *PRM4*-antisense is expressed instead. *GAL4*, which is required for activation of the other *GAL* genes, expressed an antisense transcript in both cell-types during vegetative growth. Starvation reduced *GAL4*-antisense levels, possibly allowing *GAL4* activation only when glucose is depleted.

I also observed that there are 65 genes that express primarily antisense transcripts across all four libraries (Table 9). These genes may represent cases where antisense-mediated repression is relieved in a condition other than those assayed here. For example, *HXT11, HTX12, HXT15,* and *HXT16* expressed antisense transcripts in all conditions, and did not express sense. These genes may be induced in the presence of specific carbon sources when glucose is also depleted. *AQY1, PRM3, YFL051C, SMK1, AZR1, REC114* and *SPO75* are all induced in mid-meiosis, but expressed antisense transcripts in the conditions analyzed here. (Chu, et al., 1998; Vershon and Pierce, 2000). It is possible that entry into mid-meiosis represses these antisense transcripts, allowing coding expression.

To verify that the bioinformatic comparisons were performed correctly, the mapped RNA-seq data was visualized with the Integrated Genome Viewer (IGV) (Robinson, et al., 2011). Examples of genes that shift between SRR categories 1 and 3 from each comparison were examined. Like *IME4* and *RME2,* these genes show a distinct switch between sense transcription and antisense transcription (Fig. 30, 33). This shows that differential strand-specific expression is a conserved phenomenon.

Table 7. Genes shifting from SRR Category 1 to Category 3 RNA-seq libraries were compared to identify genes which are in SRR Category 1, in one library and SRR Category 3 in the other library. Target genes have RPKM value > 1.0 in the SRR Category 3 condition, and ncRPKM >1.0 in the SRR Category 1 condition, and have greater than 2-fold RPKM and ncRPKM changes. RPKM Fold and ncRPKM fold are the ratios of the respective expression values for a given gene in the first library divided by the expression value of the second library, in a given comparison. Genes with either fold change greater than 2X, and no less than 1.25-fold change in the other comparison were included. A) *MAT***a**-Glu compared to **a**/α-Glu. B) **a**/α Glu-compared with to **a**/α-SPM. C) *MAT***a**-SPM compared to **a**/α-SPM. D) *MAT***a**-Glu compared to *MAT***a**-SPM.

Table 7. Genes shifting from SRR Category 1 to Category 3

A) *MAT*a Glu vs. a/α Glu

gene_id	ORF	*MAT*a Glu				a/α Glu					
		RPKM	ncRPKM	SRR	SRR Cat.	RPKM	ncRPKM	SRR	SRR Cat.	RPKM Fold	ncRPKM Fold
SRR Category 1 *MAT*a Glu; SRR Category 3 a/α Glu											
YGL192W	IME4	0.5	4.5	0.1	1	7.5	0.1	72.6	3	15.1	43.5
YNL146W		0.9	3.2	0.3	1	2.5	0.2	13.6	3	2.9	17.0
YOR316C-A		0.6	1.1	0.5	1	1.8	0.1	15.2	3	3.1	9.5
YGR212W	SLI1	1.9	6.5	0.3	1	5.4	1.6	3.3	3	2.8	4.0
YMR281W	GPI12	7.8	15.8	0.5	1	22.8	7.4	3.1	3	2.9	2.1
YNR014W		0.5	2.3	0.2	1	3.8	1.6	2.3	3	7.1	1.4
YOR049C	RSB1	1.1	10.0	0.1	1	38.8	8.5	4.6	3	34.0	1.2
YDR078C	SHU2	3.5	6.7	0.5	1	6.2	1.3	4.9	3	1.8	5.2
YMR201C	RAD14	2.9	21.4	0.1	1	4.8	1.7	2.8	3	1.6	12.3
YAL063C-A		1.9	5.8	0.3	1	2.7	0.3	7.7	3	1.4	16.9
YLL052C	AQY2	2.1	4.2	0.5	1	2.7	0.5	5.1	3	1.3	8.0
YMR094W	CTF13	4.9	15.8	0.3	1	5.9	0.3	18.2	3	1.2	48.6
YJL149W	DAS1	2.5	6.4	0.4	1	3.0	0.4	7.9	3	1.2	16.9
SRR Category 1 a/α Glu; SRR Category 3 *MAT*a Glu											
YIL160C	POT1	1.5	0.5	2.8	3	0.5	1.3	0.3	1	0.3	0.4
YJL157C	FAR1	19.6	2.9	6.8	3	0.4	5.9	0.1	1	0.0	0.5
YMR095C	SNO1	57.8	3.2	18.0	3	1.7	4.4	0.4	1	0.0	0.7
YFL053W	DAK2	1.8	0.3	6.9	3	1.2	2.4	0.5	1	0.7	0.1

B) a/α Glu vs. a/α SPM

gene_id	ORF	a/α Glu				a/α SPM					
		RPKM	ncRPKM	SRR	SRR Cat.	RPKM	ncRPKM	SRR	SRR Cat.	RPKM Fold	ncRPKM Fold
SRR Category 1 a/α Glu; SRR Category 3 a/α SPM											
YKL103C	LAP4	3.0	21.6	0.3	1	14.2	2.2	6.5	3	4.7	9.9
YKL093W	MBR1	2.8	7.4	0.4	1	51.7	0.8	67.1	3	18.4	9.6
YMR206W		0.5	1.6	0.3	1	33.4	0.3	127.5	3	62.8	6.3
YIL160C	POT1	0.5	1.3	0.3	1	6.0	0.2	24.1	3	13.1	5.4
YOR032W-A		0.4	4.1	0.1	1	2.7	0.9	2.9	3	7.4	4.4
YLR053C		0.2	4.2	0.1	1	18.7	1.1	17.7	3	78.2	4.0
YLL017W		0.9	4.5	0.2	1	4.1	1.5	2.7	3	4.6	2.9
YMR095C	SNO1	1.7	4.4	0.4	1	27.6	1.5	18.2	3	16.0	2.9
YEL035C	UTR5	0.3	1.0	0.3	1	10.9	0.4	26.0	3	41.8	2.4
YPL248C	GAL4	1.6	3.7	0.4	1	6.5	1.8	3.7	3	4.0	2.1
YEL069C	HXT13	0.3	2.7	0.1	1	14.9	1.3	11.1	3	44.9	2.0
YLR446W		2.6	8.5	0.3	1	4.9	0.7	7.1	3	1.9	12.5
YEL067C		0.5	5.6	0.1	1	16.4	3.4	4.8	3	34.7	1.7
YPL017C	IRC15	1.9	5.6	0.3	1	15.5	3.6	4.4	3	8.2	1.6
YHR160C	PEX18	0.8	1.4	0.5	1	4.3	1.2	3.5	3	5.7	1.2
SRR Category 1 a/α SPM; SRR Category 3 a/α Glu											
YNL145W	MFA2	1.1	0.0	105837.8	3	0.0	1.8	0.0	1	0.0	0.0
YIL161W		11.5	0.4	28.5	3	4.0	16.4	0.2	1	0.4	0.0
YMR094W	CTF13	5.9	0.3	18.2	3	2.1	5.1	0.4	1	0.3	0.1
YHR061C	GIC1	4.8	0.6	8.2	3	1.7	5.7	0.3	1	0.4	0.1
YGL209W	MIG2	12.6	0.4	29.2	3	1.5	3.5	0.4	1	0.1	0.1
YOR032C	HMS1	15.1	1.1	13.6	3	2.1	8.1	0.3	1	0.1	0.1
YER060W	FCY21	36.4	2.0	18.1	3	3.5	12.5	0.3	1	0.1	0.2
YLR145W	RMP1	17.6	2.0	8.8	3	3.5	10.9	0.3	1	0.2	0.2
YIL158W	AIM20	21.9	3.0	7.3	3	2.5	14.4	0.2	1	0.1	0.2
YML005W	TRM12	12.0	3.4	3.5	3	5.6	15.4	0.4	1	0.5	0.2
YKL051W	SFK1	25.2	3.1	8.1	3	5.7	12.3	0.5	1	0.2	0.3
YDL037C	BSC1	11.9	0.6	19.5	3	0.4	2.4	0.2	1	0.0	0.3
YNL246W	VPS75	17.6	6.6	2.7	3	3.4	21.4	0.2	1	0.2	0.3
YNL072W	RNH201	5.7	0.6	9.0	3	1.0	1.9	0.5	1	0.2	0.3
YOR339C	UBC11	4.7	0.8	6.0	3	0.2	2.3	0.1	1	0.0	0.3
YOR190W	SPR1	1.6	0.5	3.1	3	0.4	1.1	0.4	1	0.2	0.5
YDR254W	CHL4	9.6	3.9	2.5	3	5.8	11.2	0.5	1	0.6	0.3
YGR041W	BUD9	12.4	2.6	4.8	3	2.3	4.7	0.5	1	0.2	0.5
YJR147W	HMS2	6.2	3.0	2.1	3	2.6	4.9	0.5	1	0.4	0.6
YMR006C	PLB2	12.1	2.4	5.1	3	0.9	3.7	0.2	1	0.1	0.6
YGR213C	RTA1	8.9	2.6	3.4	3	0.8	4.0	0.2	1	0.1	0.7
YDR414C	ERD1	10.0	3.3	3.0	3	1.4	4.6	0.3	1	0.1	0.7
YIL073C	SPO22	4.1	1.9	2.1	3	0.6	2.7	0.2	1	0.1	0.7
YOR373W	NUD1	2.7	0.4	6.2	3	2.0	4.4	0.5	1	0.7	0.1

C) *MAT*a SPM vs. a/α SPM

		*MAT*a SPM				a/α SPM					
gene id	ORF	RPKM	ncRPKM	SRR	SRR Cat.	RPKM	ncRPKM	SRR	SRR Cat.	RPKM Fold	ncRPKM Fold
SRR Category 1 *MAT*a SPM; SRR Category 3 a/α SPM											
YAL055W	PEX22	2.6	14.9	0.2	1	6.8	0.5	14.7	3	2.7	32.2
YGL192W	IME4	0.1	5.2	0.0	1	9.2	0.2	43.1	3	150.0	24.6
YNR014W		0.3	1.2	0.2	1	21.3	0.1	208.3	3	74.1	12.0
YOR010C	TIR2	0.4	1.9	0.2	1	1.1	0.4	2.5	3	2.9	4.6
YKL087C	CYT2	1.5	2.9	0.5	1	28.5	1.5	18.5	3	19.1	1.9
YOL155C	HPF1	11.0	60.7	0.2	1	137.5	35.7	2.9	3	12.5	1.7
YPL124W	SPC29	0.8	2.0	0.4	1	1.1	0.2	4.9	3	1.4	9.1
SRR Category 1 a/α SPM; SRR Category 3 *MAT*a SPM											
YPL276W		4.0	0.2	17.1	3	0.0	8.4	0.0	1	0.0	0.0
YDR536W	STL1	8.8	0.3	25.4	3	0.6	3.5	0.2	1	0.1	0.1
YLR092W	SUL2	6.8	0.4	18.7	3	0.6	3.0	0.2	1	0.1	0.1
YLR004C	THI73	3.1	0.7	4.2	3	1.1	2.6	0.4	1	0.4	0.3
YGR213C	RTA1	6.1	1.2	5.2	3	0.8	4.0	0.2	1	0.1	0.3
YNL145W	MFA2	24.4	0.6	42.1	3	0.0	1.8	0.0	1	0.0	0.3
YNR056C	BIO5	1.7	0.5	3.1	3	0.6	1.4	0.4	1	0.4	0.4
YPL156C	PRM4	7.7	2.4	3.2	3	1.3	6.0	0.2	1	0.2	0.4
YDL037C	BSC1	2.2	1.0	2.3	3	0.4	2.4	0.2	1	0.2	0.4
YIL060W		17.8	4.4	4.0	3	4.3	9.9	0.4	1	0.2	0.4
YMR006C	PLB2	4.0	1.8	2.2	3	0.9	3.7	0.2	1	0.2	0.5
YJL157C	FAR1	3.5	1.4	2.6	3	0.3	1.8	0.2	1	0.1	0.8
YKL096C-B		17.3	6.7	2.6	3	1.1	2.6	0.4	1	0.1	2.6
YDR383C	NKP1	11.8	5.2	2.3	3	7.6	62.5	0.1	1	0.6	0.1
YPL096C-A	ERI1	15.2	0.3	46.4	3	10.6	49.7	0.2	1	0.7	0.0

D) *MAT*a Glu v s. *MAT*a SPM

		*MAT*a Glu				*MAT*a SPM					
gene id	ORF	RPKM	ncRPKM	SRR	SRR Cat.	RPKM	ncRPKM	SRR	SRR Cat.	RPKM Fold	ncRPKM Fold
SRR Category 1 *MAT*a Glu; SRR Category 3 *MAT*a SPM											
YJL149W	DAS1	2.5	6.4	0.4	1	5.1	0.1	38.1	3	2.0	48.0
YMR034C		2.7	4.9	0.5	1	9.8	0.1	71.3	3	3.7	35.8
YOR293C-A		1.1	24.4	0.0	1	8.2	1.2	6.8	3	7.3	20.1
YOR032W-A		0.3	4.8	0.1	1	1.6	0.5	3.1	3	4.7	9.1
YDR536W	STL1	0.4	2.8	0.1	1	8.8	0.3	25.4	3	24.8	8.1
YMR194C-B	CMC4	5.2	9.7	0.5	1	28.0	1.2	23.0	3	5.3	7.9
YLR446W		1.6	4.8	0.3	1	5.4	0.6	8.9	3	3.4	7.9
YMR081C	ISF1	2.6	6.8	0.4	1	18.6	0.9	21.1	3	7.0	7.8
YEL067C		1.9	7.2	0.3	1	6.0	1.6	3.7	3	3.1	4.4
YDR253C	MET32	2.7	5.6	0.5	1	8.4	1.3	6.3	3	3.1	4.3
YML087C	AIM33	0.9	2.5	0.4	1	3.9	0.7	5.6	3	4.3	3.6
YNR002C	ATO2	1.0	4.3	0.2	1	6.1	1.3	4.6	3	6.2	3.2
YCR061W		1.2	2.7	0.4	1	5.3	0.9	5.8	3	4.5	3.0
YKL096C-B		3.0	20.0	0.2	1	17.3	6.7	2.6	3	5.7	3.0
YEL075C		0.5	1.1	0.5	1	1.4	0.4	3.5	3	2.6	3.0
YBR294W	SUL1	1.8	4.3	0.4	1	71.1	1.5	47.5	3	40.6	2.8
YPL017C	IRC15	1.8	5.0	0.4	1	5.2	2.1	2.5	3	2.8	2.4
YDR223W	CRF1	0.7	2.0	0.3	1	11.6	1.0	11.9	3	17.7	2.0
YJL100W	LSB6	4.6	22.8	0.2	1	7.7	0.4	19.6	3	1.7	58.3
YGL154C	LYS5	5.3	13.1	0.4	1	8.7	4.2	2.1	3	1.6	3.1
YKL103C	LAP4	11.6	28.3	0.4	1	15.1	5.8	2.6	3	1.3	4.9
SRR Category 1 *MAT*a SPM; SRR Category 3 *MAT*a Glu											
YGR153W		54.2	1.1	51.3	3	5.8	12.2	0.5	1	0.1	0.1
YIL161W		10.2	1.2	8.6	3	2.4	11.5	0.2	1	0.2	0.1
YGL209W	MIG2	5.4	0.4	13.8	3	1.0	3.5	0.3	1	0.2	0.1
YHR061C	GIC1	3.0	0.4	7.1	3	1.3	3.6	0.4	1	0.5	0.1
YOR032C	HMS1	8.9	0.8	11.4	3	0.9	4.9	0.2	1	0.1	0.2
YML027W	YOX1	2.4	0.5	5.0	3	0.4	2.5	0.1	1	0.1	0.2
YER060W	FCY21	38.2	1.8	21.3	3	1.8	4.2	0.4	1	0.0	0.4
YCR105W	ADH7	2.6	0.5	4.9	3	0.3	1.1	0.3	1	0.1	0.5
YDR317W		3.2	0.8	4.3	3	0.5	1.5	0.4	1	0.2	0.5
YKL087C	CYT2	12.9	1.6	8.0	3	1.5	2.9	0.5	1	0.1	0.6

Table 8. Genes shifting from SRR Category 1 to Category 2 RNA-seq libraries were compared to identify genes which are in SRR Category 1, in one library and SRR Category 2 in the other library.   The highest scoring target genes have RPKM value > 1.0 in the SRR Category 2 condition, and ncRPKM >1.0 in the SRR Category 1 condition. RPKM Fold and ncRPKM fold are the ratios of the respective expression values for a given gene in the first library divided by the expression value of the second library, in a given comparison.   Genes with either fold change greater than 2X, and no less than 1.25-fold change in the other comparison were included. A) *MAT***a**-Glu compared to **a**/α-Glu.   B) **a**/α Glu-compared with to **a**/α-SPM.   C) *MAT***a**-SPM compared to **a**/α-SPM.   D) *MAT***a**-Glu compared to *MAT***a**-SPM.

Table 8. Genes shifting from SRR Category 1 to Category 2

A) *MAT***a** Glu vs. a/α Glu

		*MAT***a** Glu				a/α Glu				RPKM	ncRPKM
gene id	ORF	RPKM	ncRPKM	SRR	SRR Cat.	RPKM	ncRPKM	SRR	SRR Cat.	Fold	Fold
SRR Category 1 *MAT***a** Glu; SRR Category 2 a/α Glu											
YHL015W-A		0.8	5.1	0.2	1	1.0	1.2	0.8	2	1.3	4.3
YLR405W	DUS4	15.1	104.1	0.1	1	21.3	27.6	0.8	2	1.4	3.8
YJL100W	LSB6	4.6	22.8	0.2	1	10.1	6.2	1.6	2	2.2	3.7
YOL144W	NOP8	6.3	18.0	0.4	1	9.9	5.4	1.8	2	1.6	3.4
YOL114C		1.1	4.8	0.2	1	2.2	1.7	1.3	2	2.0	2.8
YJL133W	MRS3	8.7	30.9	0.3	1	14.9	14.3	1.0	2	1.7	2.2
YOL155C	HPF1	1.4	46.5	0.0	1	14.2	22.0	0.6	2	10.5	2.1
SRR Category 1 a/α Glu; SRR Category 2 *MAT***a** Glu											
YHL012W		1.2	1.6	0.8	2	0.7	6.2	0.1	1	0.6	0.3
YKL037W	AIM26	1.1	1.0	1.1	2	0.1	2.4	0.1	1	0.1	0.4
YJR086W	STE18	12.0	9.9	1.2	2	0.7	19.5	0.0	1	0.1	0.5
YKL093W	MBR1	7.6	4.0	1.9	2	2.8	7.4	0.4	1	0.4	0.5
YIL013C	PDR11	2.9	1.6	1.8	2	1.3	3.0	0.4	1	0.4	0.6
YJL170C	ASG7	1.2	1.1	1.1	2	0.2	1.6	0.2	1	0.2	0.7
YLR053C		2.4	3.3	0.7	2	0.2	4.2	0.1	1	0.1	0.8

B) a/α Glu vs. a/α SPM

		a/α Glu				a/α SPM				RPKM	ncRPKM
gene id	ORF	RPKM	ncRPKM	SRR	SRR Cat.	RPKM	ncRPKM	SRR	SRR Cat.	Fold	Fold
SRR Category 1 a/α Glu; SRR Category 2 a/α SPM											
YOR242C	SSP2	0.1	4.7	0.0	1	1.0	0.9	1.1	2	7.4	5.1
YBR294W	SUL1	0.4	7.1	0.1	1	5.8	4.3	1.4	2	15.2	1.7
YGR109W-B		0.9	3.1	0.3	1	2.7	2.1	1.3	2	2.9	1.4
YIL082W-A		1.2	5.2	0.2	1	2.8	3.6	0.8	2	2.4	1.4
YIL080W		1.3	4.9	0.3	1	2.8	3.5	0.8	2	2.3	1.4
SRR Category 1 a/α SPM; SRR Category 2 a/α Glu											
YMR105W-A		2.3	1.7	1.4	2	0.6	7.2	0.1	1	0.2	0.2
YBL100W-C		1.8	3.3	0.6	2	1.3	13.3	0.1	1	0.7	0.2
YHL015W-A		1.0	1.2	0.8	2	0.0	2.9	0.0	1	0.0	0.4
YGR109C	CLB6	2.0	2.4	0.8	2	0.3	5.6	0.0	1	0.1	0.4
YFL064C		1.2	1.0	1.2	2	0.9	2.3	0.4	1	0.7	0.4
YIL072W		2.1	1.2	1.8	2	0.6	2.5	0.3	1	0.3	0.5
YDR317W		1.6	1.9	0.9	2	0.9	3.7	0.2	1	0.6	0.5

C) *MAT***a** SPM vs. a/α SPM

		*MAT***a** SPM				a/α SPM				RPKM	ncRPKM
gene id	ORF	RPKM	ncRPKM	SRR	SRR Cat.	RPKM	ncRPKM	SRR	SRR Cat.	Fold	Fold
SRR Category 1 *MAT***a** SPM; SRR Category 2 a/α SPM											
YJL133W	MRS3	1.9	7.4	0.3	1	4.3	2.5	1.7	2	2.2	2.9
YBR296C-A		0.3	5.9	0.0	1	1.8	2.1	0.9	2	6.2	2.9
YKL071W		2.1	12.3	0.2	1	4.2	4.5	0.9	2	2.0	2.7
YOR242C	SSP2	0.2	1.5	0.1	1	1.0	0.9	1.1	2	6.0	1.6
YDR478W	SNM1	0.8	3.5	0.2	1	3.2	2.9	1.1	2	4.0	1.2
SRR Category 1 a/α SPM; SRR Category 2 *MAT***a** SPM											
YIL072W		2.4	1.9	1.3	2	0.6	2.5	0.3	1	0.3	0.8
YIL073C	SPO22	1.6	1.6	1.0	2	0.6	2.7	0.2	1	0.4	0.6
YJR079W		6.8	6.6	1.0	2	2.6	24.1	0.1	1	0.4	0.3
YAR066W		1.2	1.4	0.9	2	0.5	4.0	0.1	1	0.4	0.4
YOR192C-C		8.3	8.9	0.9	2	3.6	15.3	0.2	1	0.4	0.6
YMR105W-A		1.3	1.8	0.7	2	0.6	7.2	0.1	1	0.5	0.3
YOL047C		1.1	1.5	0.7	2	0.5	4.0	0.1	1	0.5	0.4
YDR114C		10.5	9.8	1.1	2	7.0	25.3	0.3	1	0.7	0.4
YFL064C		1.3	1.1	1.2	2	0.9	2.3	0.4	1	0.7	0.5

D) *MAT***a** Glu vs. *MAT***a** SPM

		*MAT***a** Glu				*MAT***a** SPM				RPKM	ncRPKM
gene id	ORF	RPKM	ncRPKM	SRR	SRR Cat.	RPKM	ncRPKM	SRR	SRR Cat.	Fold	Fold
SRR Category 1 *MAT***a** Glu; SRR Category 2 *MAT***a** SPM											
YAR068W		0.5	3.9	0.1	1	2.4	1.4	1.7	2	5.2	2.8
YNL196C	SLZ1	0.9	4.5	0.2	1	3.4	2.2	1.5	2	3.7	2.0
YIL082W-A		0.6	3.4	0.2	1	1.9	1.7	1.1	2	3.1	2.0
YIL080W		0.7	3.2	0.2	1	1.9	1.6	1.2	2	2.9	2.0
YGR109W-B		0.6	2.1	0.3	1	1.5	0.8	1.8	2	2.6	2.5
YNL146C-A		2.9	11.4	0.3	1	6.3	3.2	2.0	2	2.1	3.6
YPL248C	GAL4	0.9	3.9	0.2	1	1.8	1.2	1.5	2	2.0	3.3
YJR079W		5.1	36.0	0.1	1	6.8	6.6	1.0	2	1.3	5.5
SRR Category 1 *MAT***a** SPM; SRR Category 2 *MAT***a** Glu											
YAL055W	PEX22	3.5	2.0	1.8	2	2.6	14.9	0.2	1	0.7	0.1
YAL034W-A	MTW1	2.2	1.4	1.6	2	1.3	7.1	0.2	1	0.6	0.2
YOR298W	MUM3	1.7	1.9	0.9	2	0.5	9.8	0.0	1	0.3	0.2
YHL026C		2.0	1.5	1.3	2	0.5	3.8	0.1	1	0.3	0.4
YMR279C		1.3	2.3	0.6	2	0.3	5.5	0.0	1	0.2	0.4

Table 9. Genes with constitutively expressed antisense transcripts. Genes in this list are in SRR Category 1 in all four libraries. The expression value of the average ncRPKM of all four libraries must be greater than 1.00, with not more than one individual library average less than 1.0. Genes are ranked by average RPKM value. Genes antisense to ribosomal RNA and UTRs were excluded. RNA (Column 8) indicates whether the antisense transcript is a ncRNA (nc) or coding RNA (cod).

Table 9. Genes with constitutively expressed antisense transcripts

| gene_id	ORF	*MATa*G ncRPKM	a/α G ncRPKM	*MATa*SPM ncRPKM	a/α SPM ncRPKM	Avg ncRPKM	RNA
YLR264C-A		4090.71	1006.98	225.54	391.48	1428.68	nc
YDR524W-C		820.27	727.60	484.89	538.47	642.81	nc
YFL010W-A	AUA1	124.93	126.61	64.74	30.28	86.64	cod.
YAR035C-A		16.12	7.44	144.54	48.35	54.11	nc
YGL168W	HUR1	72.22	44.70	55.65	28.39	50.24	cod.
YGR035W-A		66.10	30.50	9.86	15.58	30.51	nc
YIL169C		20.01	19.13	33.25	49.50	30.47	nc
YOL155W-A		20.36	11.02	4.08	49.66	21.28	nc
YGR240C-A		29.86	15.25	12.64	20.66	19.60	nc
YOR068C	VAM10	25.26	36.70	4.54	4.91	17.85	cod.
YIR018C-A		16.69	13.07	19.67	11.12	15.14	nc
YDL118W		24.41	18.70	2.95	3.97	12.51	cod.
YCR050C		11.84	14.73	10.00	12.45	12.25	cod.
YOL164W-A		11.29	6.47	12.33	17.95	12.01	nc
YER014C-A	BUD25	21.40	10.62	7.50	7.59	11.78	cod.
YHR021W-A	ECM12	13.55	23.16	3.26	2.08	10.51	cod.
YJL077C	ICS3	8.97	10.02	7.38	12.76	9.78	nc
YKL070W		7.60	12.53	4.70	7.88	8.18	nc
YFL051C		1.14	7.56	1.48	20.49	7.67	nc
YNR075C-A		11.33	13.14	1.82	1.70	7.00	nc
YFR035C		8.83	8.89	4.14	5.42	6.82	nc
YMR247W-A		5.43	9.81	2.93	7.74	6.48	nc
YGR109W-A		6.38	9.25	2.97	6.38	6.24	nc
YGL041C-B		5.22	7.67	1.46	8.70	5.76	nc
YLL005C	SPO75	2.60	4.73	7.07	5.95	5.09	nc
YGR221C	TOS2	3.88	5.77	3.29	6.97	4.98	nc
YPL192C	PRM3	4.07	8.17	5.13	2.39	4.94	nc
YMR317W		3.64	4.00	3.86	8.14	4.91	nc
YPR192W	AQY1	2.95	4.46	6.46	4.16	4.51	nc
YGL258W	VEL1	7.99	4.17	2.24	3.62	4.50	nc
YOR384W	FRE5	1.82	2.57	3.78	9.72	4.47	nc
YGR224W	AZR1	4.08	7.41	2.93	2.75	4.29	nc
YOR008C-A		4.88	6.01	3.30	2.08	4.07	nc
YAR050W	FLO1	3.46	3.72	4.54	4.41	4.03	nc
YPR054W	SMK1	3.25	4.52	2.27	5.94	3.99	nc
YPL165C	SET6	3.07	3.60	3.72	5.01	3.85	nc
YOR387C		2.30	5.38	2.78	2.85	3.33	nc
YNL033W		2.83	2.85	2.98	4.34	3.25	nc
YPR027C		6.28	2.12	2.19	2.15	3.19	nc
YHR213W-A		3.12	2.73	1.76	4.59	3.05	nc
YNL019C		2.84	2.47	2.60	3.29	2.80	nc
YHR213W-B		2.69	2.05	2.86	2.67	2.57	nc
YHR214W		2.35	2.20	2.67	2.77	2.50	nc
YJR160C	MPH3	1.81	3.04	2.12	2.86	2.46	nc
YNL210W	MER1	1.61	2.96	2.21	2.94	2.43	nc
YFL067W		1.36	3.25	1.39	3.59	2.40	nc
YOR378W		1.64	3.73	0.96	3.17	2.38	nc
YDL247W	MPH2	1.77	2.55	1.94	2.92	2.30	nc
YDR246W-A		2.63	3.46	1.11	1.82	2.26	nc
YIL171W	HXT12	5.45	1.03	0.75	1.69	2.23	nc
YMR133W	REC114	2.08	1.10	2.43	2.74	2.09	nc
YHR212W-A		3.20	2.03	1.81	1.14	2.04	nc
YDL245C	HXT15	2.22	1.81	1.31	1.95	1.82	nc
YKR105C	VBA5	1.72	1.93	1.32	2.13	1.78	nc
YGL089C	MF(ALPHA)2	2.90	1.71	0.61	1.79	1.75	nc
YAR064W		1.46	1.56	1.76	2.06	1.71	nc
YHR213W		2.43	1.01	1.76	1.37	1.64	nc
YJR158W	HXT16	1.75	1.53	0.84	1.75	1.46	nc
YOL156W	HXT11	2.16	1.44	0.62	1.12	1.33	nc
YIL170W	HXT12	2.50	1.45	0.43	0.93	1.33	nc
YCL069W	VBA3	1.04	1.15	0.85	1.58	1.15	nc
YBR019C	GAL10	1.05	0.62	1.18	1.75	1.15	nc
YJR153W	PGU1	1.48	1.18	0.79	1.02	1.12	nc
YOR268C		1.22	1.05	1.29	0.63	1.05	nc

Figure 32. Comparison of genes which differentially express sense and antisense

transcripts. Venn diagrams of the number of genes that express primarily antisense

RNA (SRR category 1) in one library, and shift to increased levels of sense mRNA (SRR

category 2 or 3) in the other library (non-overlapping portions). This dataset was filtered

based on averaged RPKM and ncRPKM of three biological replicates, where ncRPKM >

1.0 in the antisense expressing (SRR Category 1) condition, and RPKM > 1.0 in the

sense expressing (SRR Category 2 or 3) condition, and the previously listed fold changes

in both RPKM and ncRPKM. The overlap is the subset of genes expressing the same

sense:antisense ratio (SRR Category 1) in both of the tested conditions.

MAT*a* Glucose **vs.** **a/α** Glucose
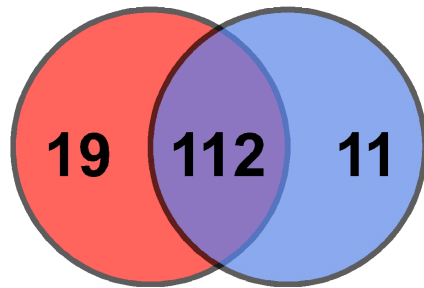
a/α Glucose **vs.** **a/α** SPM

19   112   11

20   85   31
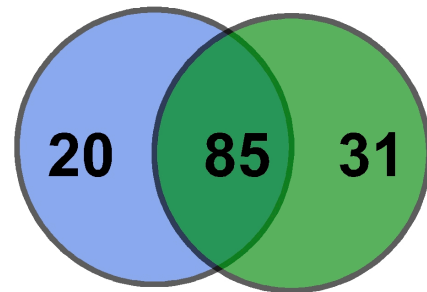
MAT*a* SPM **vs.** **a/α** SPM
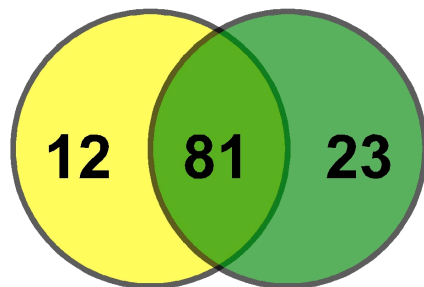
MAT*a* Glucose **vs.** **MAT*a*** SPM

12   81   23

29   99   15
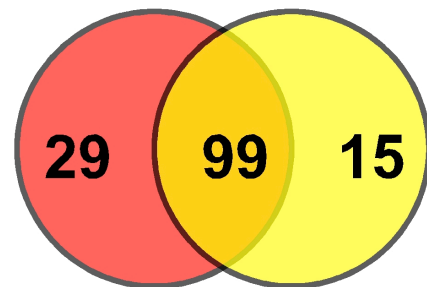
Figure 33. Visualization of sample genes that change between sense/antisense expression, as determined by SRR comparison. IGV was used to visualize representative data from the indicated libraries (Robinson, et al., 2011). Two examples from each library are shown. Watson-strand features (reads and gene annotations) are in Red; Crick strand features are Blue. Novel antisense transcripts, correlated with lowered coding expression are indicated by the arrows in green.

MATa Glu

a/α Glu

GPI12

FAR1

MATa SPM

a/α SPM

HPF1

SUL2

MATa Glu

MATa SPM

SUL1

HMS1

a/α Glu

a/α SPM

LAP4

BSC1
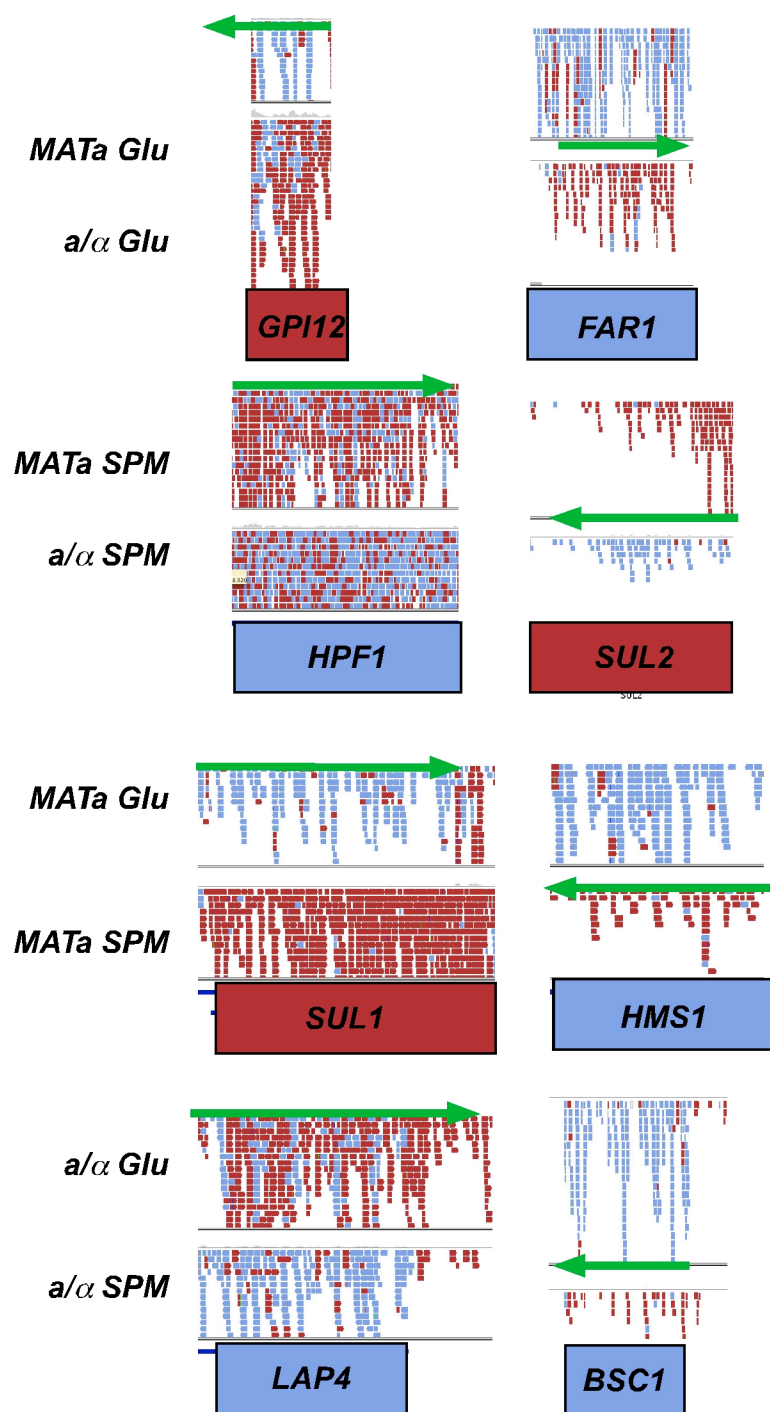
4. Verification of SOLiD findings

i. RNA-seq identified *HPF1*, a new gene regulated by a haploid-specific antisense RNA

While a significant number of genes showed differential expression patterns similar to *IME4*, the presence of an antisense RNA does not necessarily mean that the overlapping coding is regulated by the antisense transcript (Jenks, et al., 2008; Xu, et al., 2009). For example, as shown by the *IME4* deletion and *URA3* experiments shown in Chapter 3 (Fig. 17, 18), not all antisense transcripts that do not affect coding transcription. To investigate if antisense ncRNAs discovered by the RNA-seq experiments have functional roles in the cell, I selected several genes that showed changes from SRR category 1 to 3 in at least one comparison (Fig. 33).

One gene of interest was *HPF1,*which is known to have a role in breaking down mannoprotein aggregates in wine (Brown, et al., 2007). In *MAT**a**-SPM libraries, this ORF expresses the highest level of antisense transcript of any verified ORF (ncRPKM = 60.6,), with an SRR value of 0.217. *HPF1* is therefore in SRR Category 1 in *MAT**a* cells, much like *IME4*. In contrast, the *HPF1* transcript is highly upregulated and the antisense downregulated in **a**/α cells (SRR = 2.90) much like *IME4* and *RME2* (Fig. 33B). RT-PCR of wild-type cells verified there was differential expression of both transcripts (Fig. 34A, lane 1 vs. 2). The antisense transcript, termed H̲az̲e R̲egulator 1̲, is only expressed in *MATa* cells, while *HPF1* is the predominate species in **a**/α cells.

To test if *HZR1* expression directly regulates *HPF1*, I deleted the *HZR1* promoter. If *HZR1* regulates *HPF1,* preventing *HZR1* expression would permit haploid cells to express high levels of *HPF1*. This would be similar to the way the *ime4Δ3'* mutant fails

Figure 34. ***HPF1* is repressed in a cell-type specific manner by the antisense**

transcript *HZR1* A) RT-PCR assays of *HPF1* and its antisense transcript *HZR1* in *MAT***a**

(YBG200) and **a**/α cells (JRY103). Strand-specific cDNA of both transcripts was

generated, and PCR amplified by a common nested primer pair.   *ACT1* is a loading

control for the RNA samples, and –RT reactions were performed for each primer set

used. B) Deletion of the *HZR1* promoter permits *HPF1* expression.   Cartoon illustrates

the deletion of 250 bp downstream of *HPF1*, containing the presumptive *HZR1* promoter.

RT-PCR assays show the expression of *HPF1* and *HZR1* in *MAT***a** (YBG200, lane 1),

*MAT***a** *–hzr1Δpr* (YBG228, lane 2), and **a**/α cells (JRY103, lane 3).   Controls and assays

were performed as above.
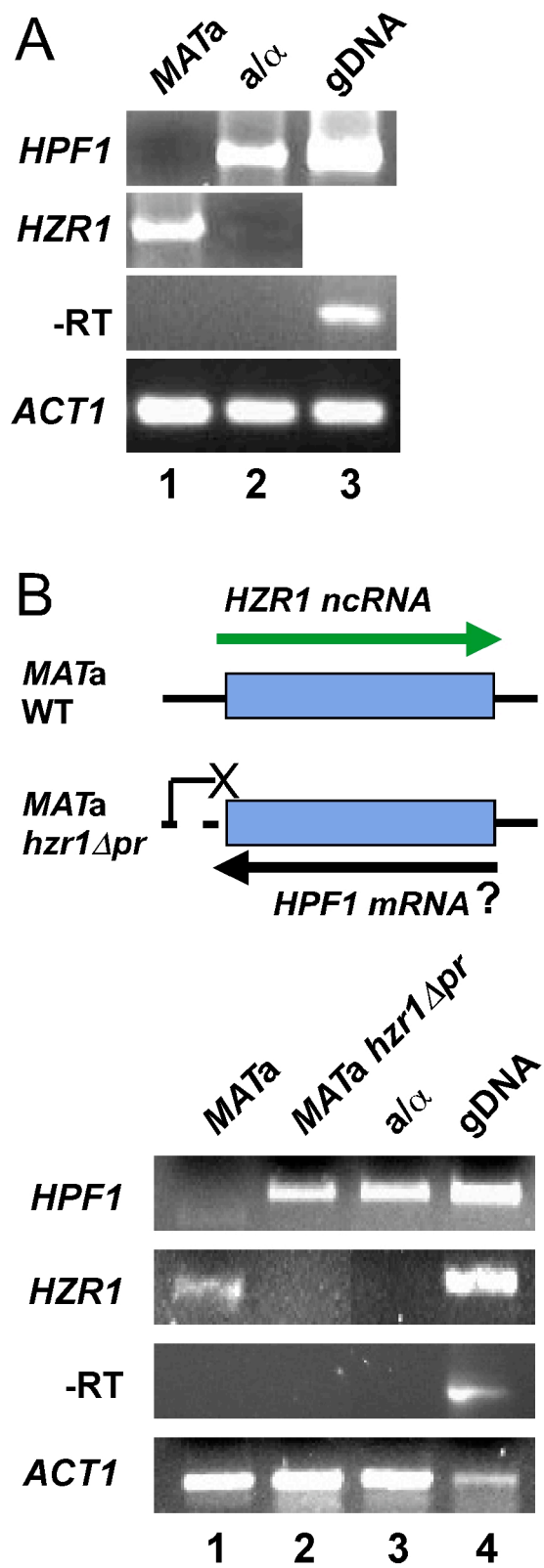
to express *RME2*, allowing expression of *IME4* (Fig. 10C, Fig. 11). Deletion of a 250 bp

region downstream of *HPF1,* containing the presumptive *HZR1* promoter, prevented

HZR1 expression in *MAT***a** cells (Fig. 34B, lane 1 vs. lane 2-*HZR1*). More importantly,

the loss of *HZR1* resulted in diploid-like expression of *HPF1* (Fig. 34B, lane 2 vs. lane 3-

HPF1). This showed that *HPF1* is regulated in a cell-type specific manner, and

suggested that genes which shift SRR ratios between Category 1 and 3 could be likely

examples of antisense-mediated regulation.

ii. The *LAP4 l*ocus expresses an antisense RNA that represses *LAP4*

in vegetative cells.

In addition to genes regulated by cell-type specific antisense transcription, the

RNA-seq experiments also looked for genes regulated by growth conditions. *LAP4* is an

aminopeptidase with a role in vacuolar formation (Cueva, et al., 1989; Schu, 2008).

This gene showed significant levels of antisense expression in both cell-types during

vegetative growth (*MAT***a**-Glu ncRPKM = 28.0; **a**/α ncRPKM = 21.5), which gave SRR

values of .408 in *MAT***a** and .258 in **a**/α cells. This antisense RNA, which we call *ALF4*

(Antisense *LAP4*), spans 2.2 KB and completely overlaps the *LAP4* gene, similar to the

arrangement of *IME4/RME2* (Hongay, et al., 2006).   Verification of the RNA-seq

dataset via RT-PCR showed that *ALF4* is expressed in both cell-types in vegetative media

(Fig. 35A.**)** *LAP4* is not expressed under the same conditions, suggesting regulation of

ALF4 and *LAP4* is not cell-type specific.   Previous studies showed that *LAP4* RNA and

protein expression are both highly induced when cells are starved for fermentable carbon

sources, or in the presence of metal ions such as Zinc (Adamis, et al., 2009; Cueva, et al.,

1989; Schu, 2008). If *ALF4* expression functions to repress *LAP4* in vegetative media, we would expect to see down-regulation of the antisense transcript under the conditions that induce *LAP4* expression. In our transcriptome data, we observed that transfer to sporulation media, which also lacks a fermentable carbon source, lowered expression of *ALF4* by an average of 7.2X and *LAP4* expression was increased by an average of 3.0X (Fig. 33D). When *MATa* cells were grown without glucose, during stationary phase, and when exposed to 1 mM Zn^{2+}, which are conditions known to induce *LAP4* expression, we observed that *ALF4* is no longer expressed and *LAP4* is expressed (Fig. 35B, lanes 2-4). This result indicates that there is a conditional switch between sense and antisense expression, similar to that observed for *IME4*.

To verify that *ALF4* is the direct repressor of *LAP4* in vegetatively grown cells, I made a mutant, *alf4Δpr*, which deletes a 250 bp sequence downstream of *LAP4* that likely contains the presumptive *ALF4* promoter. If *ALF4* is the direct repressor of *LAP4*, then we would expect to see derepression of *LAP4* in the *alf4Δpr* strain due to the loss of *ALF4*. Conditions that induce *LAP4* would show no change in expression. As expected, both wild-type and *alf4Δpr* cells express *LAP4* when induced by the presence of 0.1 mM Zn2+, indicating that this mutation does not effect the ability to induce *LAP4* (Fig. 35C, lanes 3, 4). However, in *alf4Δpr* mutants, the expression of *ALF4* is lost under vegetative conditions (Fig. 35C, lane 2). This loss also causes *LAP4* expression under conditions where it should be repressed. Therefore, we can conclude that the *ALF4* ncRNA has a direct role in vegetative-state repression of *LAP4*, similar to *IME4* and *ZIP2*.

Figure 35. *LAP4* is a gene that is repressed by antisense transcription in vegetative cells. A) A) RT-PCR assays of *LAP4* and its antisense transcript *ALF4* in *MAT***a** (YBG200, lane 1) and **a**/α cells (JRY103, lane 2) grown in YEPD media. B) RT-PCR assays of *LAP4* and *ALF4* from *MAT***a** (LNY433) cells grown to log phase in vegetative media (lane 1); vegetative media to stationary phase (lane 2); shifted to glucose-free media (lane 3); or induced by 15 min exposure to 1 mM Zn^{2+} (lane 4).   C) RT-PCR assays of *LAP4* and *ALF4* in Wild type (LNY433, lanes 1,3) and *alf4Δpr* (YBG227, lanes 2,4) strains, either in vegetative media (lanes 1,2) or induced by 15 min exposure to 1 mM Zn^{2+} (lanes 3,4).
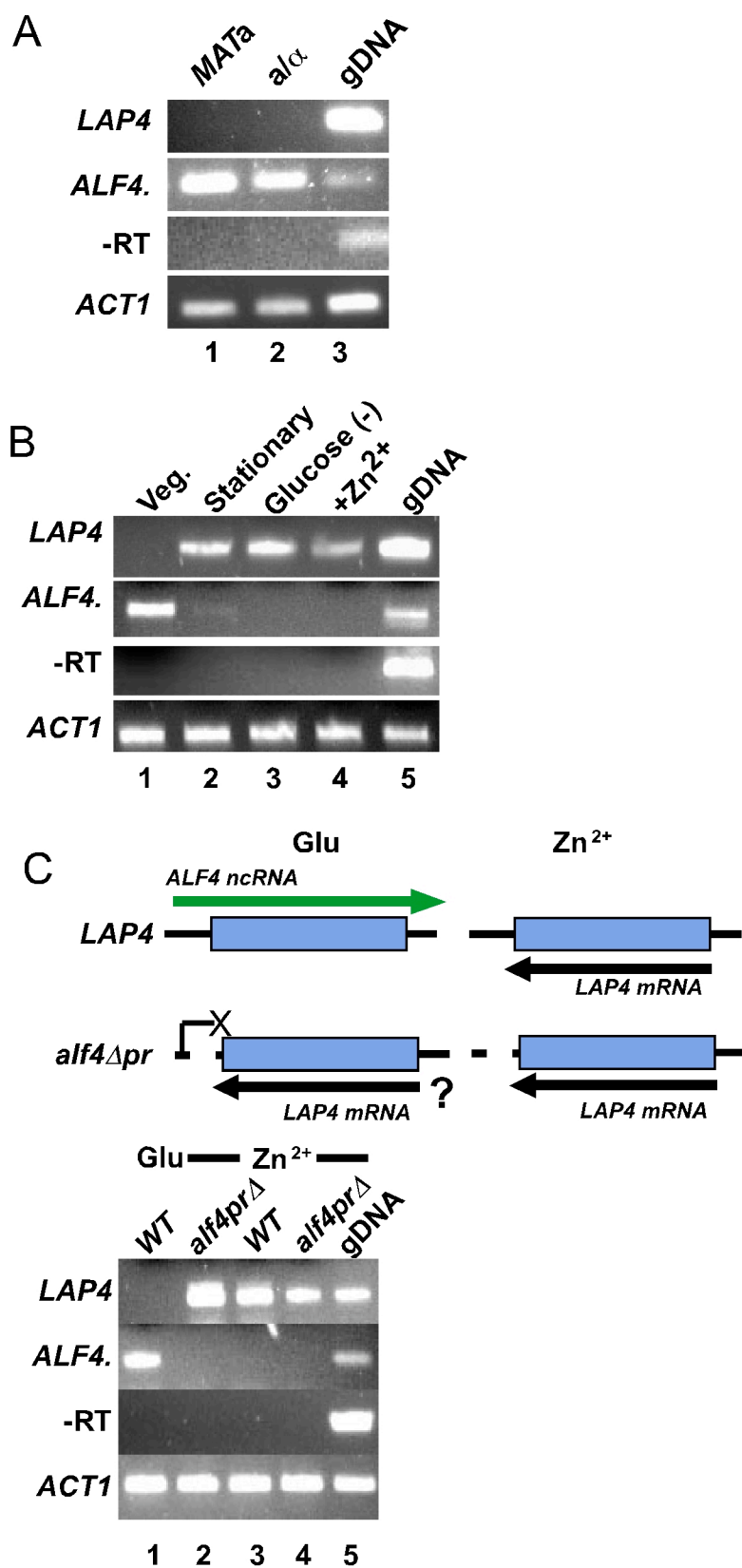
iii. *HIM1* is repressed in non-*MAT*a cells by *cis*-acting antisense transcript, *RHI1*

In the bioinformatics analysis of the RNA-seq data, changes from SRR Category 1 and SRR Category 3 appear to be the most likely candidates for antisense-mediated repression. However, there are also many genes in each comparison that show less dramatic shifts in the ratio of sense: antisense transcription. As an example, a shift from Category 3 to 2 is indicative of a reduction in sense transcript in the presence of an antisense transcript.   An example of this is the *HIM1* gene, which encodes a protein involved in DNA mutation repair (Kelberg, et al., 2005).   In the *MAT*a-Glu library the *HIM1* sense transcript is expressed. In the a/α-Glu library, *HIM1* sense transcription was reduced, and an antisense transcript that does not apparently encode a protein, was detected (Fig. 36A, bottom).   I wanted to see if this transcript may have a role in regulating *HIM1* in a cell-type specific manner, like the *HZR1* and *ALF4* transcripts do.

RT-PCR assays showed that the *HIM1* sense transcript is expressed in *MAT*a cells when grown vegetatively (Fig. 36B. *HIM1*-lane 2).   Neither *MAT*α nor a/α mating-types express detectable levels of *HIM1* under the same conditions (Fig. 36B. *HIM1*-lane 1, 3). Interestingly, both *MAT*α and a/α cells express the non-coding *HIM1* antisense transcript in this assay, which is consistent with the SOLiD data from a/α cells. I have named this transcript <u>R</u>egulator of <u>Him</u>1 (*RHI1*). This indicated that there was differential cell-type regulation of the *HIM1* and *RHI1* transcripts.

To further understand the interaction between *HIM1* and *RHI1,* I wanted to identify if there is a regulation, between the two transcripts and if so, what transcript is regulating the other.   If it is similar to *IME4/RME2*, the non-coding *RHI1* transcript

would repress *HIM1* transcription in *MATα* and **a**/α cells. If this model is correct, the

loss of *RHI1* transcription would likely result in expression of *HIM1* in *MATα* cells. On

the other hand, the loss of *HIM1* would likely have no effect on *RHI1* transcription in

*MAT***a** cells. If the *HIM1* coding transcript instead represses the ncRNA *RHI1*, it would

be expected that preventing *HIM1* expression in *MAT***a** cells would allow *RHI1*

expression. In this case, *MATα* cells would not express either transcript in an *RHI1*

promoter mutant.

To inactivate *HIM1* sense transcription, 250 bp upstream of the *HIM1* start codon

was deleted (*him1Δpr;* Fig. 36C). This mutation deleted the TATA box and any

upstream activation elements of *HIM1*. *RHI1* transcription was inactivated by deletion of

250 bp downstream of the *HIM1* ORF (r*hi1Δpr;* Fig. 36C). Both mutants were

reintegrated in *MAT***a** and *MATα him1Δ* cells. In *MAT***a** cells, *HIM1* RNA is expressed

when integrated at the *LEU2* locus at the same level as in wild-type cells (Fig. 36D lane

1). The *rhi1Δpr* mutation prevented the expression of *HIM1* as predicted (Fig. 36 lane

2). However, *RHI1* was not expressed in this strain, which indicates that *RHI1* is not

expressed as a consequence of *HIM1* repression.   The *rhi1Δpr* mutation did not affect

HIM1 expression in *MAT***a** cells.

In *MATα* cells, there was no *HIM1* expression in either wild-type or *him1Δpr*

strains (Fig. 36E, lanes 2,3). *RHI1* was present in both strains. In contrast, *rhi1Δpr* not

only failed to express *RHI1,* it also allowed expression of *HIM1* (Fig. 36E lane 3). This

directly showed that like *IME4*, loss of the *RHI1* ncRNA resulted in derepression of the

coding transcript in the inappropriate cell-type.

Figure 36. *HIM1* **is an a-specific gene repressed by antisense transcription in**

MATα **cells.** A) Reads mapping to the *HIM1* locus in *MAT***a**-Glu and **a**/α cells,

visualized with IGV. B) Expression of *HIM1* and *RHI1* in *MAT*α (JRY118, lane 1) *MAT***a**

(YBG200, lane 2), and **a**/α (JRY103, lane 3). C) Cartoons illustrating the *him1Δpr* and

rhi1Δpr constructs and presumptive effect on transcription of *HIM1* and *RHI1*. D)

Assays of *MAT***a-** *him1Δ* (YBG205) cells with reintegrated *HIM1* (YBG207, lane 1),

him1Δpr (YBG211, lane 2)*, rhi1Δpr* (YB212, lane 3)*,* or *him1Δpr* and *rhi1Δpr* (YBG213,

lane 4). E) Assays of *MAT*α*- him1Δ* cells (YBG204, lane 1) with reintegrated *HIM1*

(YBG206, lane 2), *him1Δpr* (YBG208, lane 3)*, rhi1Δpr* (YB209, lane 4)*,* or *him1Δpr* and

rhi1Δpr (YBG210, lane 5).
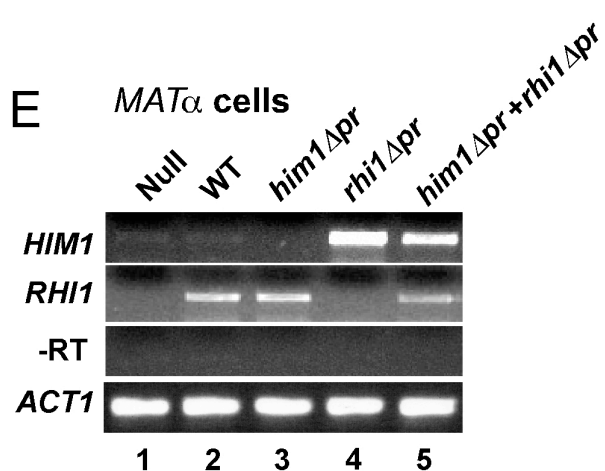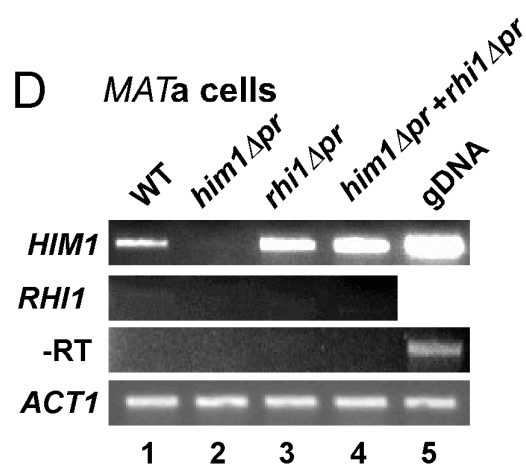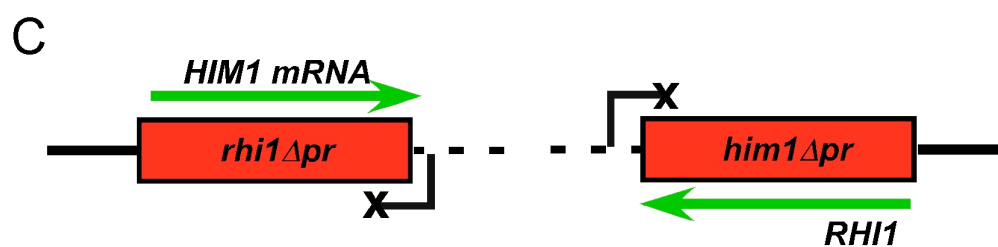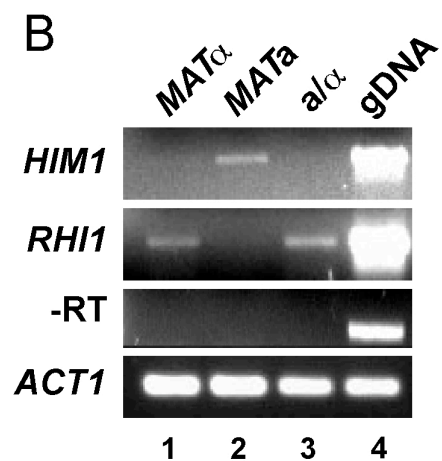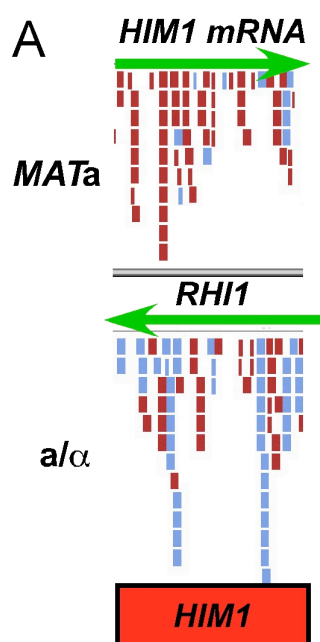
*MAT***a** and *MAT*α strains carrying both mutants in a *trans-* configuration were also constructed to determine if the relationship between *HIM1* and *RHI1* occurs in a *cis-* mechanism like *IME4* and *RME2,* or if it can function in *trans-* like miRNA (He and Hannon, 2004; Martens, et al., 2004). *MAT*α cells with *him1Δpr* and *rhi1Δpr* integrated at a distant locus showed expression of both transcripts similar to the single constructs (Fig. 36E, lane 5). This indicated that *HIM1* regulation by *RHI1* likely occurs in *–cis.*

Discussion

RNA-seq analyses of the cell-type and early meiotic transcriptome uncovered a significant number of novel antisense transcripts. Previous studies of antisense transcription in *S. cerevisiae* using Tiled Microarrays and SAGE Tagging suggested that almost 400 genes have reliably expressed antisense transcripts (David, et al., 2006; Miura, et al., 2006). From a combined analysis of all four replicated libraries, using the minimal criteria of averaged ncRPKM > 1.0, my work has identified over 1400 examples of transcripts antisense to coding RNAs. Interestingly, in any given cell-type under either vegetative or meiosis-inducing condition, there appear to be about 600 to 1000 antisense ncRNAs expressed. This clearly indicates that like the coding transcriptome, many of the ncRNAs are differentially regulated by both cell-type and growth condition. This is consistent with previous studies of specific ncRNAs in yeast. For example, *RME2* and *RME3* are cell-type regulated; *SRG1* is activated by low serine; *PHO84*-antisense is activated by long-term stasis (Camblong, et al., 2009; Gelfand, et al., 2011; Hongay, et al., 2006; Martens, et al., 2005).

In addition to the number of antisense transcripts discovered, comparing sense and antisense expression across cell-types and growth condition was able to directly identify many specific cases where differential stranded gene expression occurs. A change between SRR Category 1 and 3, like is observed for *IME4/RME2*, appears to be a good detector of regulated antisense expression. In each comparison performed in this analysis, between thirty-one and fifty-one genes are differentially expressed in this manner. I have shown here that *HPF1*, *LAP4,* and *HIM1* have differentially regulated expression of sense and antisense transcripts similar to *IME4* and *ZIP2*. In each of these cases, in conditions or cell-types where the antisense transcript is expressed, there is no detectable sense expression. When the antisense transcription was prevented by deletion of the antisense promoter, sense expression was concurrently induced.

Some genes expressed antisense transcripts at higher levels than sense under all four conditions. This included the genes *FRE5, AZR1,* and *SMK1*. The coding transcripts for these genes are induced in low iron, resistance to azoles, and middle sporulation, respectively (Pierce, et al., 1998; Shakoury-Elizeh, et al., 2004; Tenreiro, et al., 2000). It is possible that these antisense transcripts are downregulated in these conditions, to permit coding expression.

While differential expression of sense and antisense transcripts may be suggestive of regulation, it does not mean that it is always occurring. For example, non-regulatory, cryptic antisense transcription can be induced in absence of Isw2, without a large effect on the expression of sense genes (Whitehouse, et al., 2007). Furthermore, as described in Chapter 3, there are also deletions within *IME4* and bp225-675 Flip mutants that cause a lack of repression, despite the presence of an antisense transcript. Therefore, I wanted to

prove that antisense transcripts identified by RNA-seq directly regulate gene expression. The promoter deletions of the non-coding antisense genes *HZR1, ALF4* and *RHI1* proved that these ncRNAs directly regulate expression of the overlapping coding genes. In each case, loss of the antisense transcript permitted sense expression in the inappropriate cell-type or growth condition. Of the over 1400 antisense ncRNAs identified by SOLiD, there are likely many more cases where antisense transcription regulates sense transcription. Additionally, if the cases of low-level antisense expression are also considered, (0.00 < ncRPKM < 1.00), there exists the potential for over 2000 additional antisense transcripts. These low-level antisense transcripts are not highly induced in the four conditions assayed here; however, it is distinctly possible that other conditions may induce their expression, leading to conditional downregulation of the coding gene. *S. cerevisiae* has previously been shown to express ncRNAs that regulate genes during depletion of serine, depletion of Zinc, and extended exposure to low temperature (Berretta, et al., 2008; Bird, et al., 2006; Martens, et al., 2004). The data I have collected for the cell-type and early meiotic transcriptomes suggests 30-50 genes differentially express sense and antisense transcripts in response to a single change. Taken together, this suggests that regulation by antisense transcription may be a widespread, conserved process in the cellular response and differentiation.

VI. CONCLUSIONS

The study presented here had two primary goals: to determine the mechanism by which the antisense transcript *RME2* represses *IME4*, and if this mechanism of repression regulates the expression of other genes in *S. cerevisiae*. Identifying a novel widespread mechanism of gene regulation would represent a significant change in how we think about cells function. If this mechanism were conserved in other eukaryotes, it would likely have a significant impact on the study of many human genetic disorders and diseases caused by eukaryotic parasites.

I have shown in this work that *IME4,* a diploid-specific gene required for sporulation, is differentially regulated by cell-type. In haploid cells, the antisense transcript *RME2* is expressed from the same locus, which prevents transcription of full length *IME4*. In diploid cells, the **a**1-α2 repressor complex binds downstream of *IME4,* in the *RME2* promoter. This prevents *RME2* expression, which in turn permits *IME4* expression only in the appropriate **a**/α cell type. This mechanism is distinctly different from miRNA-mediated regulation, in that it only functions in a *cis-* configuration. This mechanism of regulation is similar in some aspects to other regulatory ncRNAs present in *S. cerevisiae*, such as *SRG1* and *ZRR1*.

I hypothesize that gene regulation by ncRNAs benefits the cell by reducing the number of different regulatory proteins expressed. For example, in the case of *SER3*, the Swi/Snf activator complex functions as a repressor by inducing *SRG1*, which blocks transcription of *SER3*. *ADH1* and *ADH3* similarly are repressed by Zap1, which activates transcription of *ZRR1* and *ZRR3,* respectively, which block access to the *ADH* promoters. Similarly, the **a**1-α2 complex, which normally functions as a repressor of haploid-

specific genes such as *GPA1* and *STE5,* also functions as an activator by preventing

expression of the *RME2* and *RME3* transcripts. By repressing these antisense ncRNAs,

a1-α2 activates the diploid-specific genes *IME4* and *ZIP2.* It has long been known that

the mating-type locus in diploid yeast functions in the repression of both **a**-specific and

haploid-specific genes (Fig. 2). The discovery of this mechanism may provide insight

into how diploid-specific processes, including meiosis and sporulation are controlled.

While the initial finding that *IME4* is antisense-regulated has been observed

elsewhere, there was little explanation of the mechanism of repression (Hongay et al.

2006; Shah and Clancy 1992). In this work, the mechanism by which antisense

transcription regulates *IME4* has been further examined (Gelfand, et al., 2011; Hongay, et

al., 2006; Shah and Clancy, 1992). Regulation by *SRG1* and *ZRR1* in *S. cerevisiae*

suggested a model where transcription across their respective coding promoters blocks

transcription factor binding and gene expression. In the case of *SER3,* this involves

nucleosome remodeling by Spt2, Spt6, Spt16 to occlude promoter access (Hainer, et al.,

2011; Thebault, et al., 2011). However, *IME4* does not appear to be regulated by *RME2*

in the same manner. The *spt6-1004* and *spt16-197* mutants that prevent *SRG1* from

repressing *SER3* did not affect *IME4* repression by *RME2*. Truncation of *RME2,* so that

it is not transcribed across the *IME4* promoter region did not affect regulation of *IME4*.

A similar mutant in *SRG1*, that terminates before crossing the *SER3* TATA box was

defective in repression (Martens, et al., 2004). The ChIP assays of both *IME4* and *ZIP2*

showed no appreciable differences in TBP binding difference between haploid and

diploid cells. It has been shown that repression of *PHO84* by *PHO84*-antisense does not

disrupt TBP, either (Camblong, et al., 2009). This is unlike repression of *SER3,* where

transcription of *SRG1* prevents TBP binding. These findings show that regulation by ncRNAs may utilize several different molecular mechanisms.

Interestingly, the mechanism of repression by antisense transcription for *IME4* appears to be dependant on sequences within the ORF itself, rather than the promoter region. A copy of the *URA3* gene cloned in place of *IME4* was not cell-type regulated by growth assays. RT-PCR showed the presence of both sense and antisense transcripts in haploid cells. This indicated that antisense transcription alone was not sufficient for regulation. This was verified by the deletion analysis of the ORF, which identified a direction-specific DNA element at bp225-675 that is critical for *RME2*-mediated repression.

The exact function of this region within *IME4* is not completely clear. It is possible that this region is a recruitment site for chromatin remodeling complexes. Histone H2A.Z, and chromatin modifiers Isw2, and Set2 have been shown to repress cryptic antisense transcription (Whitehouse, et al., 2007, Carrozza, 2005 #195; Zofall, et al., 2009).   If these complexes are involved in antisense mediated regulation, then deletion mutants of these genes should derepress *IME4* transcription in haploid cells. If a particular mutant shows a defect in *IME4* repression, then ChIP assays could be used to determine if the complex binds directly to the bp225-675 element when *RME2* is expressed, or if there is an indirect interaction between this region and chromatin remodeling complexes.

Alternately, the RNA structure of *RME2* in this region may play a role in the inhibition of full-length transcription. I believe that further deletion analysis of this region would be able to provide clues to which mechanism may be true.   If smaller sub-

regions of bp225-675 are required for regulation, then these regions could be examined more specifically for role they may play in regulation. One possibility is that sequences within this region of the *RME2* are binding sites for RNA-binding proteins, such as *THO1* and *THO2*, which have been shown to regulate transcription and chromatin remodeling (Piruat and Aguilera, 1998). RNA-binding proteins recruited to this region may act alone or as part of a complex to physically block full-length *RME2* expression.

RNA folding and base pairing of the required regions within bp225-675 could also be analyzed. One possibility is that sequences within *RME2* form stem-loop structures, which are recognized by the polymerase transcribing *IME4* on the opposite strand. Binding of the RNA stem-loops to PolII may then forcibly terminate transcription of *IME4*. Similar RNA folding and base pairing has been shown to have a role in Rho-independent termination in bacterial transcription (Wilson and von Hippel, 1995). If there are sequences that suggest the formation of such structures, then mutants that prevent folding could be examined for failure to repress *IME4*.

It is also possible that the act of transcription of *RME2* across this region of DNA, rather than the RNA molecule itself remodels the chromatin structure in a manner that exposes a premature polyadenylation or termination site for *IME4*. The factors involved these processes would then be able to bind and prevent the full length *IME4* RNA from being expressed. If there is a polyadenylation site for *IME4,* which is exposed by transcription of *RME2*, then it may be possible to use oligo-d(T) 3' RACE to detect the prematurely polyadenylated *IME4* transcript (Hague, et al., 2008; Lutz, 2008). It is also possible *IME4* is terminated by the Nab-Nrd pathway that terminates non-polyadenylated RNA PolII transcripts (Carroll, et al., 2007). Nab1 and Nrd3 bind directly to the RNA

transcript to prevent elongation and recruit other proteins involved in termination. If this pathway is involved in the termination of *IME4* RNA-protein analysis (RIP-ChIP) for the Nab-Nrd complex bound to *IME4* near the possible termination site could be used to detect the interaction (Baroni, et al., 2008). Similar experiments would also be performed on the other antisense-regulated genes identified in this study to determine if there is a conserved mechanism of antisense-mediated termination, or if they are regulated in a different manner.

RNA PolII ChIP-on-Chip or ChIP analysis by Deep-sequencing (ChIP-Seq) showed that RNA Polymerase II frequently pauses and stalls during transcription (Churchman and Weissman, 2011). It is possible that sense-strand RNAPolII pausing at specific sites, such as within the bp225-675 region of *IME4*, is susceptible to disruption by polymerase moving in the opposite direction. *IME4* and the other genes regulated by antisense transcription could be examined for pause sites. If pausing is involved in regulation, then deletion of pause sites would be expected to prevent antisense-mediated regulation.

The initial finding of the similar mechanism regulating *ZIP2* suggested the possibility that this mechanism of regulation was not limited to a single gene, but may be employed as a widespread mechanism of transcriptional regulation. The RNA-seq experiments performed here show that differential antisense transcription is a widespread phenomenon, suggestive of regulation. With a switch in cell-type or growth condition, an average of 30 different genes are differentially expressed in a strand-specific manner. It is likely that other changes in growth conditions show similar numbers of genes changing stranded expression.

The RNA-seq described here examined the expression of *MAT***a** and **a**/α cells. This identified changes in haploid vs. diploid expression, but also includes **a**-specific genes which are repressed by α2-Mcm1. Therefore, RNA-seq of the *MAT*α cell-type should be performed to determine which of the haploid specific antisense transcripts are **a**1-α2 or α2-Mcm1 regulated. This would also identify α-specific ncRNAs which may be induced by the α1-Mcm1 activator complex (Bender and Sprague, 1987).

RNA-seq was used to identifiy target genes that differentially express antisense ncRNAs, which may regulate coding gene expression. The cell-type specific antisense ncRNAs *HZR1* and *RHI1* were shown to repress *HPF1* in haploid cells, and *HIM1* in *MAT*α and **a**/α cells, respectively. While deletion of the antisense promoters showed that expression of these transcripts regulated coding genes, the specific activators of *RHI1* and *HZR1* are not known. There are no binding sites for mating-type regulators in their promoter regions. *ALF4* represses *LAP4* during vegetative growth, and repression is relieved by glucose depletion or addition of Zinc, independent of cell-type. This implies that *ALF4* is regulated by conditional transcription factors. To identify the factors that regulate these ncRNAs, the RNA-seq data could be cross-referenced to ChIP on Chip or ChIP by Deep-Sequencing experiments (ChIP-seq) datasets to try and correlate regulated antisense transcription with specific factors. It is highly likely that other cell-type and conditional regulatory proteins and complexes regulate ncRNAs. I would be particularly interested in knowing how many different transcription factors have dual functions as shown for Swi/Snf, Zap1, and **a**1-α2.

Additionally, the use of strand-specific RNA-seq to examine the full scope of transcriptional regulation during meiosis and sporulation could possibly explain the

regulation of many genes. Previous studies of the yeast transcriptome during this process were performed using non-stranded microarray technology (Chu, et al., 1998). Therefore, antisense and sense transcription would not have been distinguished. Based on the results observed by RNA-seq here, it is distinctly possible that some of the apparently upregulated genes are in fact downregulated through antisense transcription. Additionally, I observed that several mid to late meiotic genes, including *SMK1, AZR1,* and *SPO75* express antisense transcripts in vegetative and early meiotic conditions. RNA-seq of the full program of sporulation would show if the sense transcript is differentially expressed during mid-meiosis, and determine if the antisense transcripts are re-activated in late meiosis to repress these genes again.

As stated above, activator and repressor proteins can apparently have dual roles, acting through an ncRNA intermediate. I think that it is highly likely that some of the coordinate gene activation and repression that occurs during this process depends on this property. For example many middle-sporulation genes are activated by Ndt80 (Chu, et al., 1998; Pierce, et al., 2003). During middle sporulation, early meiotic and vegetative stage genes are downregulated (Vershon and Pierce, 2000).   It is possible that this is done through Ndt80-activated ncRNAs which act as repressors. RNA-seq would be able to identify the ncRNAs ,which may be repressing coding expression. To further examine these regulatory circuits, cross reference of the RNA-seq data of the sporulation process would be cross referenced to global binding data of Sum1, Ndt80, and **a**1-α2 to determine if these or other transcription factors regulate the sporulation-specific antisense ncRNAs.

The final question that this research provokes is how common this system of regulation occurs in other organisms. It has been shown that *S. cerevisiae* lacks the endogenous ability to produce and utilize miRNA (Drinnenberg, et al., 2009). Therefore, any endogenous antisense RNAs expressed in *S. cerevisiae* must be repressing through alternative mechanisms. This study has shown that *IME4, ZIP2,* and *HIM1* are regulated through *cis-* acting pathways. In higher eukaryotes, the RNAi –response can also be activated by larger antisense RNAs annealing to target species in *trans-* (Drinnenberg, et al., 2009; He and Hannon, 2004; Mallory, et al., 2004). However, there are no logical reasons why other eukaryotes cannot also use *cis-* acting antisense transcription to regulate gene expression. To examine this, stranded RNA-seq of related organisms utilize RNAi, such as *Saccharomyces castelli* and *Schizosaccharomyces pombe,* would identify the ncRNAs present in these species (Drinnenberg, et al., 2009; Kato, et al., 2005; Martienssen, et al., 2005). If antisense RNAs expressed in *cis-* to coding ORFs are identified in these species, promoter knockouts like the *ime4Δ3', alf4Δpr, rhi1Δpr,* and *hzr1Δpr* mutants assayed here would indicate if the sense genes are regulated by their respective antisense transcripts. Antisense promoter mutants that derepress sense transcription would then have the antisense transcript supplied in *trans-*, as I have done for *IME4, ZIP2,* and *HIM1*. If the mutant is rescued by *trans-* antisense RNA, it would suggest that gene may also be regulated through the RNAi pathway. Mutants which fail to be repressed by *trans-* supplied antisense RNA would prove that similar mechanisms of *cis-* acting RNA are present in species which can also use RNAi. This would show that this mechanism is conserved in other eukaryotes, which would potentially have a dramatic impact on studies of gene regulation of genetic disorders and cancer cells.

REFERENCES

Abraham, D.S. and Vershon, A.K. (2005) N-terminal arm of Mcm1 is required for transcription of a subset of genes involved in maintenance of the cell wall, *Eukaryotic cell*, **4**, 1808-1819.

Adamis, P.D., Mannarino, S.C., Riger, C.J., Duarte, G., Cruz, A., Pereira, M.D. and Eleutherio, E.C. (2009) Lap4, a vacuolar aminopeptidase I, is involved in cadmium-glutathione metabolism, *Biometals*, **22**, 243-249.

Auer, P.L. and Doerge, R.W. (2010) Statistical design and analysis of RNA sequencing data, *Genetics*, **185**, 405-416.

Ausubel, F.M. (1987), *Current Protocols in Molecular Biology*, **1**.

Baroni, T.E., Chittur, S.V., George, A.D. and Tenenbaum, S.A. (2008) Advances in RIP-chip analysis : RNA-binding protein immunoprecipitation-microarray profiling, *Methods Mol Biol*, **419**, 93-108.

Bender, A. and Sprague, G.F., Jr. (1987) MAT alpha 1 protein, a yeast transcription activator, binds synergistically with a second protein to a set of cell-type-specific genes, *Cell*, **50**, 681-691.

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

Bird, A.J., Gordon, M., Eide, D.J. and Winge, D.R. (2006) Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts, *Embo J*, **25**, 5726-5734.

Brown, S.L., Stockdale, V.J., Pettolino, F., Pocock, K.F., de Barros Lopes, M., Williams, P.J., Bacic, A., Fincher, G.B., Hoj, P.B. and Waters, E.J. (2007) Reducing haziness in white wine by overexpression of Saccharomyces cerevisiae genes YOL155c and YDR055w, *Appl Microbiol Biotechnol*, **73**, 1363-1376.

Buchman, A.R., Kimmerly, W.J., Rine, J. and Kornberg, R.D. (1988) Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in Saccharomyces cerevisiae, *Mol Cell Biol*, **8**, 210-225.

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

Carroll, K.L., Ghirlando, R., Ames, J.M. and Corden, J.L. (2007) Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements, *RNA*, **13**, 361-373.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O. and Herskowitz, I. (1998) The transcriptional program of sporulation in budding yeast, *Science*, **282**, 699-705.

Chua, P.R. and Roeder, G.S. (1998) Zip2, a meiosis-specific protein required for the initiation of chromosome synapsis, *Cell*, **93**, 349-359.

Churchman, L.S. and Weissman, J.S. (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution, *Nature*, **469**, 368-373.

Clancy, M.J., Shambaugh, M.E., Timpte, C.S. and Bokar, J.A. (2002) Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N6-methyladenosine in

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

Cogoni, C. and Macino, G. (1997) Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in Neurospora crassa, *Proc Natl Acad Sci U S A*, **94**, 10233-10238.

Couzin, J. (2002) Breakthrough of the year. Small RNAs make big splash, *Science*, **298**, 2296-2297.

Covitz, P.A., Herskowitz, I. and Mitchell, A.P. (1991) The yeast RME1 gene encodes a putative zinc finger protein that is directly repressed by a1-alpha 2, *Genes Dev*, **5**, 1982-1989.

Cueva, R., Garcia-Alvarez, N. and Suarez-Rendueles, P. (1989) Yeast vacuolar aminopeptidase yscI. Isolation and regulation of the APE1 (LAP4) structural gene, *FEBS Lett*, **259**, 125-129.

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

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R. and Bartel, D.P. (2009) RNAi in budding yeast, *Science*, **326**, 544-550.

Duina, A.A., Rufiange, A., Bracey, J., Hall, J., Nourani, A. and Winston, F. (2007) Evidence that the localization of the elongation factor Spt16 across transcribed genes is dependent upon histone H3 integrity in Saccharomyces cerevisiae, *Genetics*, **177**, 101-112.

Esposito, M.S. and Esposito, R.E. (1974) Genes controlling meiosis and spore formation in yeast, *Genetics*, **78**, 215-225.

Esposito, M.S. and Esposito, R.E. (1975) Mutants of meiosis and ascospore formation, *Methods Cell Biol*, **11**, 303-326.

Gagneur, J., Sinha, H., Perocchi, F., Bourgon, R., Huber, W. and Steinmetz, L.M. (2009) Genome-wide allele- and strand-specific expression profiling, *Molecular systems biology*, **5**, 274.

Gailus-Durner, V., Xie, J., Chintamaneni, C. and Vershon, A.K. (1996) Participation of the yeast activator Abf1 in meiosis-specific expression of the HOP1 gene, *Mol Cell Biol*, **16**, 2777-2786.

Galgoczy, D.J., Cassidy-Stone, A., Llinas, M., O'Rourke, S.M., Herskowitz, I., DeRisi, J.L. and Johnson, A.D. (2004) Genomic dissection of the cell-type-specification circuit in Saccharomyces cerevisiae, *Proc Natl Acad Sci U S A*, **101**, 18069-18074.

Galitski, T., Saldanha, A.J., Styles, C.A., Lander, E.S. and Fink, G.R. (1999) Ploidy regulation of gene expression, *Science*, **285**, 251-254.

Ge, D., Lamontagne, B. and Elela, S.A. (2005) RNase III-mediated silencing of a glucose-dependent repressor in yeast, *Curr Biol*, **15**, 140-145.

Gelfand, B., Mead, J., Bruning, A., Apostolopoulos, N., Tadigotla, V., Nagaraj, V., Sengupta, A.M. and Vershon, A.K. (2011) Regulated antisense transcription controls expression of cell-type specific genes in yeast, *Mol Cell Biol*.

Gimeno, C.J. and Fink, G.R. (1992) The logic of cell division in the life cycle of yeast, *Science*, **257**, 626.

Gimeno, C.J. and Fink, G.R. (1994) Induction of pseudohyphal growth by overexpression of PHD1, a Saccharomyces cerevisiae gene related to transcriptional regulators of fungal development, *Mol Cell Biol*, **14**, 2100-2112.

Goutte, C. and Johnson, A.D. (1988) a1 protein alters the DNA binding specificity of alpha 2 repressor, *Cell*, **52**, 875-882.

Guldener, U., Heck, S., Fielder, T., Beinhauer, J. and Hegemann, J.H. (1996) A new efficient gene disruption cassette for repeated use in budding yeast, *Nucleic Acids Res*, **24**, 2519-2524.

Guo, B., Styles, C.A., Feng, Q. and Fink, G.R. (2000) A Saccharomyces gene family involved in invasive growth, cell-cell adhesion, and mating, *Proc Natl Acad Sci U S A*, **97**, 12158-12163.

Hague, L.K., Hall-Pogar, T. and Lutz, C.S. (2008) In vivo methods to assess polyadenylation efficiency, *Methods Mol Biol*, **419**, 171-185.

Hainer, S.J., Pruneski, J.A., Mitchell, R.D., Monteverde, R.M. and Martens, J.A. (2011) Intergenic transcription causes repression by directing nucleosome assembly, *Genes Dev*, **25**, 29-40.

He, L. and Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation, *Nat Rev Genet*, **5**, 522-531.

Heiman, M.G. and Walter, P. (2000) Prm1p, a pheromone-regulated multispanning membrane protein, facilitates plasma membrane fusion during yeast mating, *J Cell Biol*, **151**, 719-730.

Hongay, C.F., Grisafi, P.L., Galitski, T. and Fink, G.R. (2006) Antisense transcription controls cell fate in Saccharomyces cerevisiae, *Cell*, **127**, 735-745.

Jenks, M.H., O'Rourke, T.W. and Reines, D. (2008) Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast, *Mol Cell Biol*, **28**, 3883-3893.

Jin, Y., Mead, J., Li, T., Wolberger, C. and Vershon, A.K. (1995) Altered DNA recognition and bending by insertions in the alpha 2 tail of the yeast a1/alpha 2 homeodomain heterodimer, *Science*, **270**, 290-293.

Kanduri, C., Thakur, N. and Pandey, R.R. (2006) The length of the transcript encoded from the Kcnq1ot1 antisense promoter determines the degree of silencing, *Embo J*, **25**, 2096-2106.

Kassir, Y., Granot, D. and Simchen, G. (1988) IME1, a positive regulator gene of meiosis in S. cerevisiae, *Cell*, **52**, 853-862.

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K. and Murakami, Y. (2005) RNA polymerase II is required for RNAi-dependent heterochromatin assembly, *Science*, **309**, 467-469.

Kelberg, E.P., Kovaltsova, S.V., Alekseev, S.Y., Fedorova, I.V., Gracheva, L.M., Evstukhina, T.A. and Korolev, V.G. (2005) HIM1, a new yeast Saccharomyces cerevisiae gene playing a role in control of spontaneous and induced mutagenesis, *Mutat Res*, **578**, 64-78.

Kessler, M.M., Zeng, Q., Hogan, S., Cook, R., Morales, A.J. and Cottarel, G. (2003) Systematic discovery of new genes in the Saccharomyces cerevisiae genome, *Genome Res*, **13**, 264-271.

Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A. and Dewey, C.N. RNA-Seq gene expression estimation with read mapping uncertainty, *Bioinformatics*, **26**, 493-500.

Lutz, C.S. (2008) Alternative polyadenylation: a twist on mRNA 3' end formation, *ACS Chem Biol*, **3**, 609-617.

Ma, J., Dobry, C.J., Krysan, D.J. and Kumar, A. (2008) Unconventional genomic architecture in the budding yeast saccharomyces cerevisiae masks the nested antisense gene NAG1, *Eukaryotic cell*, **7**, 1289-1298.

Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K. and Bartel, D.P. (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region, *Embo J*, **23**, 3356-3364.

Martens, J.A., Laprade, L. and Winston, F. (2004) Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene, *Nature*, **429**, 571-574.

Martens, J.A., Wu, P.Y. and Winston, F. (2005) Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae, *Genes Dev*, **19**, 2695-2704.

Martienssen, R.A., Zaratiegui, M. and Goto, D.B. (2005) RNA interference and heterochromatin in the fission yeast Schizosaccharomyces pombe, *Trends Genet*, **21**, 450-456.

Mathias, J.R., Hanlon, S.E., O'Flanagan, R.A., Sengupta, A.M. and Vershon, A.K. (2004) Repression of the yeast HO gene by the MATalpha2 and MATa1 homeodomain proteins, *Nucleic Acids Res*, **32**, 6469-6478.

Matzke, M.A. and Birchler, J.A. (2005) RNAi-mediated pathways in the nucleus, *Nat Rev Genet*, **6**, 24-35.

Mitchell, A.P., Driscoll, S.E. and Smith, H.E. (1990) Positive control of sporulation-specific genes by the IME1 and IME2 products in Saccharomyces cerevisiae, *Mol Cell Biol*, **10**, 2104-2110.

Mitchell, A.P. and Herskowitz, I. (1986) Activation of meiosis and sporulation by repression of the RME1 product in yeast, *Nature*, **319**, 738-742.

Miura, F., Kawaguchi, N., Sese, J., Toyoda, A., Hattori, M., Morishita, S. and Ito, T. (2006) A large-scale full-length cDNA analysis to explore the budding yeast transcriptome, *Proc Natl Acad Sci U S A*, **103**, 17846-17851.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq, *Nat Methods*, **5**, 621-628.

Muhlrad, D., Hunter, R. and Parker, R. (1992) A rapid method for localized mutagenesis of yeast genes, *Yeast*, **8**, 79-82.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing, *Science*, **320**, 1344-1349.

Nagaraj, V.H., O'Flanagan, R.A., Bruning, A.R., Mathias, J.R., Vershon, A.K. and Sengupta, A.M. (2004) Combined analysis of expression data and transcription factor binding sites in the yeast genome, *BMC Genomics*, **5**, 59.

Nasr, F., Becam, A.M., Brown, S.C., De Nay, D., Slonimski, P.P. and Herbert, C.J. (1995) Artificial antisense RNA regulation of YBR1012 (YBR136w), an essential gene

from Saccharomyces cerevisiae which is important for progression through G1/S, *Mol Gen Genet*, **249**, 51-57.

Nasr, F., Becam, A.M., Slonimski, P.P. and Herbert, C.J. (1994) YBR1012 an essential gene from S. cerevisiae: construction of an RNA antisense conditional allele and isolation of a multicopy suppressor, *C R Acad Sci III*, **317**, 607-613.

Olsson, L., Larsen, M.E., Ronnow, B., Mikkelsen, J.D. and Nielsen, J. (1997) Silencing MIG1 in Saccharomyces cerevisiae: effects of antisense MIG1 expression and MIG1 gene disruption, *Appl Environ Microbiol*, **63**, 2366-2371.

Oshima, J. and Campisi, J. (1991) Fundamentals of cell proliferation: control of the cell cycle, *Journal of dairy science*, **74**, 2778-2787.

Park, H., Shin, M. and Woo, I. (2001) Antisense-mediated inhibition of arginase (CAR1) gene expression in Saccharomyces cerevisiae, *J Biosci Bioeng*, **92**, 481-484.

Pepke, S., Wold, B. and Mortazavi, A. (2009) Computation for ChIP-seq and RNA-seq studies, *Nat Methods*, **6**, S22-32.

Pickford, A.S., Catalanotto, C., Cogoni, C. and Macino, G. (2002) Quelling in Neurospora crassa, *Adv Genet*, **46**, 277-303.

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

Pierce, M., Wagner, M., Xie, J., Gailus-Durner, V., Six, J., Vershon, A.K. and Winter, E. (1998) Transcriptional regulation of the SMK1 mitogen-activated protein kinase gene during meiotic development in Saccharomyces cerevisiae, *Mol Cell Biol*, **18**, 5970-5980.

Piruat, J.I. and Aguilera, A. (1998) A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination, *EMBO J*, **17**, 4859-4872.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer, *Nature biotechnology*, **29**, 24-26.

Rump, P., Zeegers, M.P. and van Essen, A.J. (2005) Tumor risk in Beckwith-Wiedemann syndrome: A review and meta-analysis, *Am J Med Genet A*, **136**, 95-104.

Rusche, L.N. and Rine, J. (2001) Conversion of a gene-specific repressor to a regional silencer, *Genes Dev*, **15**, 955-967.

Schu, P. (2008) Aminopeptidase I enzymatic activity, *Methods in enzymology*, **451**, 67-78.

Shah, J.C. and Clancy, M.J. (1992) IME4, a gene that mediates MAT and nutritional control of meiosis in Saccharomyces cerevisiae, *Mol Cell Biol*, **12**, 1078-1086.

Shakoury-Elizeh, M., Tiedeman, J., Rashford, J., Ferea, T., Demeter, J., Garcia, E., Rolfes, R., Brown, P.O., Botstein, D. and Philpott, C.C. (2004) Transcriptional remodeling in response to iron deprivation in Saccharomyces cerevisiae, *Mol Biol Cell*, **15**, 1233-1243.

Shang, Y., Hu, X., DiRenzo, J., Lazar, M.A. and Brown, M. (2000) Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription, *Cell*, **103**, 843-852.

Sikorski, R.S. and Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae, *Genetics*, **122**, 19-27.

Sil, A. and Herskowitz, I. (1996) Identification of asymmetrically localized determinant, Ash1p, required for lineage-specific transcription of the yeast HO gene, *Cell*, **84**, 711-722.

Steigele, S. and Nieselt, K. (2005) Open reading frames provide a rich pool of potential natural antisense transcripts in fungal genomes, *Nucleic Acids Res*, **33**, 5034-5044.

Strathern, J., Hicks, J. and Herskowitz, I. (1981) Control of cell type in yeast by the mating type locus. The alpha 1-alpha 2 hypothesis, *J Mol Biol*, **147**, 357-372.

Strich, R., Surosky, R.T., Steber, C., Dubois, E., Messenguy, F. and Esposito, R.E. (1994) UME6 is a key regulator of nitrogen repression and meiotic development, *Genes Dev*, **8**, 796-810.

Tenreiro, S., Rosa, P.C., Viegas, C.A. and Sa-Correia, I. (2000) Expression of the AZR1 gene (ORF YGR224w), encoding a plasma membrane transporter of the major facilitator superfamily, is required for adaptation to acetic acid and resistance to azoles in Saccharomyces cerevisiae, *Yeast*, **16**, 1469-1481.

Thakur, N., Tiwari, V.K., Thomassin, H., Pandey, R.R., Kanduri, M., Gondor, A., Grange, T., Ohlsson, R. and Kanduri, C. (2004) An antisense RNA regulates the bidirectional silencing property of the Kcnq1 imprinting control region, *Mol Cell Biol*, **24**, 7855-7862.

Thebault, P., Boutin, G., Bhat, W., Rufiange, A., Martens, J. and Nourani, A. (2011) Transcription Regulation by the non-coding RNA SRG1 Requires Spt2-dependent Chromatin Deposition in the Wake of RNAP II, *Mol Cell Biol*.

Vershon, A.K., Hollingsworth, N.M. and Johnson, A.D. (1992) Meiotic induction of the yeast HOP1 gene is controlled by positive and negative regulatory sites, *Mol Cell Biol*, **12**, 3706-3714.

Vershon, A.K., Jin, Y. and Johnson, A.D. (1995) A homeo domain protein lacking specific side chains of helix 3 can still bind DNA and direct transcriptional repression, *Genes Dev*, **9**, 182-192.

Vershon, A.K. and Pierce, M. (2000) Transcriptional regulation of meiosis in yeast, *Curr Opin Cell Biol*, **12**, 334-339.

Wach, A., Brachat, A., Pohlmann, R. and Philippsen, P. (1994) New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae, *Yeast*, **10**, 1793-1808.

Whitehouse, I., Rando, O.J., Delrow, J. and Tsukiyama, T. (2007) Chromatin remodelling at promoters suppresses antisense transcription, *Nature*, **450**, 1031-1035.

Wilhelm, B.T. and Landry, J.R. (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing, *Methods*, **48**, 249-257.

Wilson, K.S. and von Hippel, P.H. (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin, *Proc Natl Acad Sci U S A*, **92**, 8793-8797.

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

Yassour, M., Pfiffner, J., Levin, J.Z., Adiconis, X., Gnirke, A., Nusbaum, C., Thompson, D.A., Friedman, N. and Regev, A. (2010) Strand-specific RNA sequencing reveals

extensive regulated long antisense transcripts that are conserved across yeast species, *Genome biology*, **11**, R87.

Yin, Z., Wilson, S., Hauser, N.C., Tournu, H., Hoheisel, J.D. and Brown, A.J. (2003) Glucose triggers different global responses in yeast, depending on the strength of the signal, and transiently stabilizes ribosomal protein mRNAs, *Mol Microbiol*, **48**, 713-724.

Zhao, H., Butler, E., Rodgers, J., Spizzo, T., Duesterhoeft, S. and Eide, D. (1998) Regulation of zinc homeostasis in yeast by binding of the ZAP1 transcriptional activator to zinc-responsive promoter elements, *J Biol Chem*, **273**, 28713-28720.

Zofall, M., Fischer, T., Zhang, K., Zhou, M., Cui, B., Veenstra, T.D. and Grewal, S.I. (2009) Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs, *Nature*, **461**, 419-422.

<center>**VITA**</center>

<center>Brian Nathan Gelfand</center>

EDUCATION

Rutgers University, New Brunswick, NJ   August 2005 –PresentCurrent
Studies: PhD, Microbiology and Molecular Genetics
Graduation: Spring 2011

Rutgers University, New Brunswick, NJ   August 2003 – July 2005
M.S. Degree, Microbiology
Transferred to PhD program

Purdue University, West Lafayette, IN   August 1997 – May 2001
B.S. Degree, Microbiology
Minors: Psychology, Spanish

Publications
Differential antisense transcription is a widespread mechanism of regulation in yeast, Gelfand, B, Golhar, R, Apostolopoulos, N, Rajaghopal, G, Vershon, A. In Prep.

Gelfand, B, Meade, J, Nagaraj, V, Tadigolta, V, Sengupta A, Vershon A.K. Antisense transcription regulates cell-type specific genes in yeast. Molecular and Cellular Biology 2011 Apr;31(8):1701-9