ANALYSIS OF NON-PROTEIN CODING DNA TRANSCRIPTION FOR ROLES IN REGULATING GENE EXPRESSION

by

Elizabeth Anne Raupach

Bachelor of Science, University of Wisconsin—Eau Claire, 2009

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Elizabeth Raupach

It was defended on

August 21, 2015

and approved by

Mark Rebeiz, PhD, Assistant Professor, Department of Biological Sciences

Deborah Chapman, PhD, Associate Professor, Department of Biological Sciences

Jeffrey Hildebrand, PhD, Associate Professor, Department of Biological Sciences

Martin Schmidt, PhD, Associate Professor, Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine

Dissertation Advisor: Karen Arndt, PhD, Professor, Department of Biological Sciences

ANALYSIS OF NON-PROTEIN CODING DNA TRANSCRIPTION FOR ROLES IN REGULATING GENE EXPRESSION

Elizabeth Raupach, PhD

University of Pittsburgh, 2016

ANALYSIS OF NON-PROTEIN CODING DNA TRANSCRIPTION FOR ROLES IN REGULATING GENE EXPRESSION

Elizabeth Raupach, PhD

University of Pittsburgh, 2016

During transcription, the DNA sequence encoding a gene is "read" and copied into an RNA molecule. Many RNA molecules convey instructions for synthesizing proteins. Traditionally, proteins are thought of as molecular workhorses, which perform cellular functions. However, recent work has discovered that transcription of non-protein coding DNA is widespread in eukaryotes and plays important regulatory roles for many genes including genes mis-regulated in cancers. For example, in *S. cerevisiae*, the act of transcribing *SRG1*, a non-coding RNA (ncRNA), across the *SER3* promoter positions nucleosomes over the *SER3* upstream activating sequences. This creates a physical barrier preventing transcription of *SER3*. The pervasiveness of non-coding transcription suggests that regulatory roles for non-coding transcription may exist throughout the genome. To explore this possibility, we selected six

candidate yeast genes expressing ncRNAs over their promoters and analyzed the effects of disrupting intergenic transcription on neighboring protein-coding transcript expression. Through this unbiased approach, we identified a previously unknown mechanism of transcription regulation at the *ECM3* gene. Intergenic transcription seems to activate *ECM3* expression. Further analyses identified roles for the Paf1 complex in *ECM3* activation through methylation of histone H3 at lysine 4. Additionally, the NuA3 and SAGA chromatin modifying complexes are also required to activate *ECM3* expression. These data support a model where co-transcriptional methylation of histone H3 at lysine 4 is required to recruit the NuA3 complex and other downstream modifiers to activate transcription of *ECM3*. Other cases of regulation by intergenic transcription had previously identified this modification in negative regulation of neighboring gene expression. Thus, *ECM3* is an interesting model gene for elucidation of a novel regulatory mechanism mediated by non-coding transcription. The results presented here add to the growing number of cases where noncoding transcription has important roles in regulating gene expression. This work also indicates that the molecular mechanisms by which transcription of noncoding DNA exerts regulatory effects depend on the local chromatin environment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 CHAPTER ONE: INTRODUCTION

1.1 GENE EXPRESSION

1.1.1 Transcription

Genes can be defined as units of cellular information carrying instructions for performing cellular functions that are transmitted from one generation to the next. Genetic information is stored and transmitted in the form of deoxyribonucleic acid (DNA) molecules. DNA is a polymer of nucleotides covalently linked together to form molecules called chromosomes. The nucleotides that form DNA consist of a deoxy-ribose sugar, a phosphate group, and a nucleobase, adenine (A), thymine (T), guanine (G), or cytosine (C). Eukaryotic organisms have a set of linear chromosomes consisting of double-stranded DNA where the two strands wrap around each other forming a helical structure. In this helical structure, adenine bases will pair with thymine bases and guanine bases will pair with cytosine bases on the opposite strand (WATSON AND CRICK 1953). In this way, the two strands of DNA carry complementary sequences. The potentially limitless order of these four nucleotides into chromosomes encodes the information of genes.

Each cell in the human body contains the entire complement of genes encoded in its DNA (aside from notable exceptions such as cells involved in immune functions). Thus, what makes an eye cell different from a skin cell is not that they contain different genes, but that they express different subsets of those genes. The proper control of gene expression is critical to regulate all cellular processes in order to form a fully functional organism. Mis-regulation of genes can result in a wide variety of human diseases, including cancer.

The first step of gene expression is to read the genes stored in DNA. This is accomplished by transcription, the process by which the sequence for a particular gene is copied from DNA into a ribonucleic acid (RNA) transcript. Much like DNA, RNA is a polymer of nucleotides. RNA nucleotides consist of a phosphate group, a ribose sugar, and a nucleobase, either adenine (A), uracil (U), guanine (G), or cytosine (C). Although RNA is synthesized as a single strand, adenine bases can pair with uracil bases and guanine bases can pair with cytosine bases to form three-dimensional RNA structures or as a heterodimer pairing with DNA. Cellular machines called RNA polymerases synthesize RNA using DNA as a template via complementary base-pairing interactions. All eukaryotic cells have three nuclear multisubunit DNA-dependent RNA polymerases, RNA polymerases I, II, and III (ROEDER AND RUTTER 1969). RNA polymerase I transcribes ribosomal RNA (rRNA), RNA polymerase II transcribes messenger RNA (mRNA) and a variety of regulatory RNAs, and RNA polymerase III transcribes the 5S rRNA and transfer RNA (tRNA) (REEDER AND ROEDER 1972; WEINMANN *et al.* 1974; WEINMANN AND ROEDER 1974). Plants have two additional DNA-dependent RNA polymerases, polymerases IV and V, which are involved in the production of regulatory RNAs (reviewed in (HE *et al.* 2014)). It was also recently discovered that mammals have an alternatively spliced variant of the mitochondrial RNA polymerase gene (spRNAP-IV), which is

localized to the nucleus and transcribes some mRNA genes (KRAVCHENKO *et al.* 2005). The genes discussed in this work are all transcribed by RNA polymerase II (Pol II).

Pol II is highly conserved among all eukaryotes. Pol II consists of 12 subunits that form a catalytic core of the enzyme where DNA threads through the active site and ribonucleotides are added to an elongating RNA molecule as well as peripheral structures that interact with a host of factors that regulate the activities of Pol II throughout the transcription cycle (reviewed in (CRAMER *et al.* 2008)). A distinguishing feature of Pol II is the extended carboxy-terminal domain (CTD) of a core subunit, Rpb1, which is flexible and contains a repeated heptapeptide consensus sequence of tyrosine-serine-proline-threonine-serine-proline-serine. Each of these seven residues can be post-translationally modified and these modifications are important for regulation of interactions with and activity of polymerase associated factors throughout the transcription cycle (reviewed in (BURATOWSKI 2009)). Several of these modifications will be discussed further as they relate to following information. The process of transcription can be classified into three general stages, initiation, elongation, and termination (depicted in Figure 1).

1.1.1.1 Initiation

During initiation, Pol II must find the start of a gene in order to begin transcription. The beginning of a gene is marked by a promoter, which consists of DNA sequences that directly or indirectly interact with Pol II to facilitate its recruitment to an active gene. Transcription factors mediate indirect interactions between promoter sequences and Pol II. Transcription factors are proteins that bind to regulatory sequences of a gene that influence whether Pol II will transcribe that gene. These factors may be either transcriptional activators or transcriptional repressors. Some transcription factors are specific to individual genes or sets of genes that are co-regulated while other transcription factors are necessary for initiating transcription at most genes. The

transcription factors acting at most genes are termed basal transcription factors. The basal factors include TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. There are multiple modes of assembly for these factors and the way these methods are employed *in vivo* likely varies between individual genes and is an active area of research (discussed in (LUSE 2014)). A step-wise mode of assembly has been characterized *in vitro*. In this classical model system, TFIID is the factor that nucleates assembly (MATSUI *et al.* 1980; SAWADOGO AND ROEDER 1985). This nucleation event is guided by the interaction between a subunit of TFIID called TATA binding protein (TBP), which binds to a consensus TATA(A/T)A(A/T)(A/G) DNA sequence in *S. cerevisiae* referred to a TATA box (BASEHOAR *et al.* 2004). Although a TATA consensus sequence is common among all eukaryotes, TBP is capable of binding a variety of sequences and many genes do not contain a recognizable TATA element (SINGER *et al.* 1990). At some loci, TBP is capable of binding to consensus TATA sequences in the absence of other TFIID subunits while the other TFIID subunits co-occupy non-consensus TATA sequences with TBP (RHEE AND PUGH 2012). In the classical model of assembly following nucleation by TFIID, TFIIA then TFIIB bind to the promoter DNA. At this stage, both Pol II and TFIIF enter the complex. Recruitment of TFIIE and subsequent association of TFIIH follow this (BURATOWSKI *et al.* 1989; FLORES *et al.* 1991; GERARD *et al.* 1991).

In addition to these factors, occupancy of gene-specific transcription factors may precede recruitment of basal factors. Binding of basal transcription factors often requires interaction with other protein complexes referred to as co-activators in order to recruit Pol II to a promoter. These co-activators include the Mediator, SAGA, and Swi/Snf complexes. Initial studies identified roles for co-activators serving as molecular bridges between gene-specific transcription factors and the basal transcription machinery. These complexes also have

important roles in regulating the accessibility and architecture of promoter DNA and transitioning between Pol II binding to a promoter and engaging in productive elongation (reviewed in (NAAR *et al.* 2001)). The assembly of transcription factors, co-activators, and Pol II at a promoter is called a pre-initiation complex (PIC).

Once assembled, a PIC must denature the DNA at the promoter in order to read the sequence of the gene and synthesize an RNA transcript complementary to the gene. To denature the DNA, the double stranded helix is separated into two strands. This critical function is mediated by TFIIH and requires the input of cellular energy (SCHAEFFER *et al.* 1993; SCHAEFFER *et al.* 1994). Pol II uses the template strand to determine the order of nucleotides to be incorporated into the RNA transcript by matching incoming RNA nucleotides to the template strand by complementary base-pairing. Incoming nucleotides are added to the 3' hydroxyl of the previous nucleotide, thus providing the chemical orientation of an RNA being synthesized beginning at its 5' end and progressing toward its 3' end.

Pol II enters a PIC with its CTD lacking phosphorylation. Mediator stimulates the kinase activity of TFIIH, which phosphorylates the Pol II CTD at serine 5 (Ser5). The phosphorylation level of Ser5 of the Pol II CTD is high at the 5' ends of genes. This modification facilitates binding and activity of factors important for early stages of transcription and RNA processing (reviewed in (BURATOWSKI 2009)). Quickly following initiation, 7-methylguanosine is covalently attached to the 5' end of the elongating RNA. This cap protects the 5' end from degradation and is added in a three-step enzymatic process that requires phosphorylation of Ser5 (reviewed in (SHUMAN 2001; HOCINE *et al.* 2010)).

1.1.1.2 Elongation

During the elongation phase, Pol II will add nucleotides to the growing RNA transcript until a termination signal is reached. Many factors influence the processivity of Pol II elongation, including RNA processing, regulation of Pol II enzymatic activity, and the local chromatin structure. The accessibility of DNA being hindered by chromatin structure is a major contribution to processivity. The Pol II machinery has a host of tools known as elongation factors, such as the Paf1 complex, histone chaperones, and histone modifiers, to regulate chromatin during elongation. These elongation factors will be discussed in more detail below.

The modifications of the Pol II CTD change throughout elongation. As mentioned above, TFIIH phosphorylates Ser5 and this mark is enriched at the 5' ends of genes. As elongation progresses, Ser5 phosphorylation decreases and phosphorylation of serine 2 (Ser2) gradually increases. The *S. cerevisiae* Ctk1 and Bur1 (or Cdk12 and Cdk9 in mammals, respectively) kinases are responsible for phosphorylation of Ser2 (BARTKOWIAK *et al.* 2010). The phosphorylation state of these residues is important for interactions of Pol II with other factors at specific stages of the transcription cycle. For example, phosphorylation of Ser5 facilitates the dissociation of Pol II from Mediator, an important step to transition from initiation to elongation. Like Ser5, serine 7 (Ser7) is also phosphorylated by TFIIH at the 5' ends of genes, but its role is less understood. Other modifications of the CTD residues likely display specific localization patterns and functions that have yet to be fully characterized (reviewed in (BURATOWSKI 2009)).

The Spt4/5 complex is a highly conserved elongation complex, being shared between Pol I and Pol II and having related components in bacteria and archaea. Spt4/5 associates with Pol II very early after initiation and remains associated with Pol II to the 3' ends of genes. This

complex is important for Pol II progression through paused sites, which can be introduced by chromatin structure and also occurs when transitioning from initiation to elongation. Spt4/5 also facilitate interactions with other elongation factors, such as the Paf1 complex. Analogous to the CTD of Pol II, Spt5 has a disordered carboxy-terminal region (CTR) that contains a repeated sequence that is also subject to post-translational modifications. Much like the Pol II CTD, modifications of the Spt5 CTR regulate interactions with chromatin and RNA processing regulators such as the RNA capping machinery (reviewed in (HARTZOG AND FU 2013)).

1.1.1.3 Termination

At the 3' end of a gene, Pol II will encounter termination signals, which will cause Pol II to stop transcription and facilitate the dissociation of Pol II from the DNA at that locus. There are at least two major modes of Pol II termination in yeast cells. The cleavage and poly-adenylation factor (CPF) pathway is responsible for termination of longer transcripts, often including mRNAs. The Nrd1-Nab3-Sen1 pathway is responsible for termination of short transcripts including small nucleolar RNAs (snoRNAs) and other non-protein-coding RNAs.

The CPF complex is recruited to the 3' ends of genes by sequence-specific interactions with the nascent RNA transcript and direct interactions with Pol II. Recruitment of CPF depends on phosphorylation of Ser2 of the Pol II CTD (LICATALOSI *et al.* 2002; AHN *et al.* 2004; KIM *et al.* 2004b; PEARSON AND MOORE 2014). The RNA transcript is then cleaved by an endolytic subunit of CPF, Ysh1, at the poly-adenylation (poly-A) site. Following cleavage, the free 3' end of the RNA is poly-adenylated by the poly-A polymerase Pap1, which associates with CPF. The poly-A tail is protected from degradation by the poly-A binding protein, which also plays a role in mRNA export and translation. The cleavage of RNA transcripts appears to be necessary for subsequent dissociation of Pol II (SADOWSKI *et al.* 2003; SCHAUGHENCY *et al.*

2014). The exact mechanism to dislodge Pol II from the DNA template beyond the poly-A site is a focus of current research. Several non-mutually exclusive models are possible. In one model, the association of Pcf11 destabilizes an elongating Pol II to promote dissociation (ZHANG *et al.* 2005). Another model proposes that cleavage of the nascent RNA transcript provides a substrate for exonuclease activity carried out by Rat1 in yeast (XRN2 in humans) and this degradation machinery will approach Pol II and subsequently destabilize Pol II elongation. It has previously been thought that cleavage and degradation of the remaining RNA are necessary for poly-A dependent termination (KIM *et al.* 2004b; WEST *et al.* 2004). However, more recent research has elucidated that the isolated termination complex alone is sufficient to carry out poly-A dependent termination, which suggests that a conformational change in the complex is sufficient for termination (ZHANG *et al.* 2015). It is possible that all of these factors contribute to termination and that each might play more or less prominent roles at individual genes *in vivo* (reviewed in (PORRUA AND LIBRI 2015)).

The Nrd1, Nab3, and Sen1 proteins mediate the termination pathway for short RNAs in yeast. Recruitment of the Nrd1-Nab3-Sen1 (NNS) complex to transcribing polymerases occurs through interactions with the nascent RNA transcript and the Pol II CTD. The Nrd1 subunit preferentially binds to phosphorylated Ser5 of the CTD, which provides a molecular basis for the NNS complex selectively terminating shorter RNAs as Ser5 phosphorylation is higher near the 5' ends of genes (GUDIPATI *et al.* 2008; VASILJEVA *et al.* 2008; TUDEK *et al.* 2014). Nrd1 and Nab3 each have RNA recognition motifs (RRMs) that recognize GUAA/G and UCUUG sequence motifs in the nascent RNA (CARROLL *et al.* 2007; PORRUA *et al.* 2012). The association of Nrd1 and Nab3 with the nascent RNA and Pol II CTD aids in the recruitment of Sen1. Current models suggest that Sen1 has helicase activity that generates a force to dislodge

the RNA from the transcription bubble, which destabilizes Pol II elongation. After dissociating from Pol II, the Nrd1 CTD interacting domain (CID) is available to interact with the Trf4 subunit of the Trf4-Air2-Mtr4 (TRAMP) complex. This complex poly-adenylates RNA transcripts and facilitates their processing or degradation by interactions with the nuclear exosome which will be discussed further with non-protein-coding transcripts in a later section. Interestingly, these factors appear to have a very broad RNA surveillance role, as they are also responsible for turnover of Pol I and Pol III transcripts (WYERS *et al.* 2005; WLOTZKA *et al.* 2011). Termination by the NNS pathway has recently been reviewed in (ARNDT AND REINES 2015; PORRUA AND LIBRI 2015).

Although these two modes of transcription termination are thought to be the predominant mechanisms of yeast Pol II termination, at least two other mechanisms have been described and Pol I and III employ other strategies to terminate transcription. Pol II termination can also be terminated by a "roadblock" mechanism where a DNA binding protein, Reb1, physically blocks Pol II elongation. Subsequently, Pol II is ubiquitylated and likely degraded to be removed from the DNA template (WILSON *et al.* 2013; COLIN *et al.* 2014). A similar termination strategy to the CPF pathway depends on Rnt1, which can cleave a nascent RNA and provide a substrate for Rat1 (GHAZAL *et al.* 2009; RONDON *et al.* 2009).

Completion of transcription generates RNA molecules, which carry out many cellular functions including serving as a messenger of genetic information. In eukaryotic organisms, genes encoded in DNA are highly organized into chromatin structures that can facilitate or impede transcriptional activity. The important regulatory roles of chromatin in transcription are discussed below.

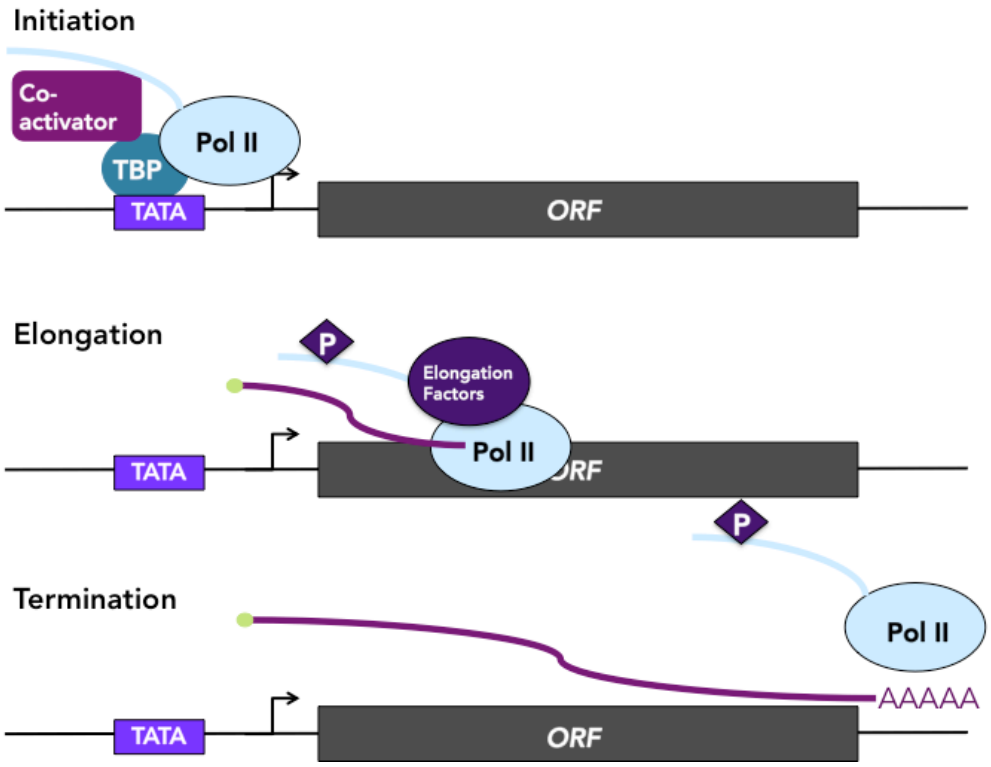

Figure 1. Schematic of the transcription cycle.

The process of transcription occurs in three basic stages, initiation, elongation, and termination. The open reading frame (ORF) of a protein-coding gene is depicted as a gray box. The transcription start site is depicted as an arrow. The TATA box is depicted as a purple box. During initiation, transcription factors, such as TBP, and co-activators recruit Pol II to the promoter of a gene. During elongation, RNA nucleotides are added to a polymerizing RNA transcript (depicted in maroon). After elongation begins, a 5' methyl-guanosine is covalently attached to the 5' end of the transcript (depicted in green). Elongation factors, such as the Paf1 complex, travel with polymerase and regulate activities, such as phosphorylation of the Pol II carboxy-terminal domain (CTD, depicted as a curved blue line), to facilitate processive transcription. During termination, transcription is halted, Pol II is released from the DNA, and the 3' end of the newly synthesized RNA molecule is poly-adenylated.

1.1.2 Organization of genes into chromatin

Inside nuclei of cells, eukaryotic genomes are highly organized and packaged into chromatin, consisting of DNA and its associated proteins. The basic unit of chromatin is the nucleosome, which consists of about 147 base pairs of DNA wrapped twice around a core octamer of two copies of each of four histone proteins, H2A, H2B, H3, and H4 (LUGER *et al.* 1997). Nucleosomes are assembled by loading of a heterotetramer consisting of two H3 proteins and two H4 proteins onto a DNA template followed by assembly of two H2A/H2B dimers (reviewed in (CUTTER AND HAYES 2015; RAMACHANDRAN AND HENIKOFF 2015). In an extended conformation, DNA wraps around histones connected by linker DNA and this repeating structure is referred to as "beads on a string." Higher levels of compaction involve forming loops of the beads on a string structure and forming contacts to scaffolding molecules (reviewed in (LUGER *et al.* 2012)). Although this packaging is necessary for storing and transmitting a vast amount of genetic material, it presents a barrier to Pol II machinery at genes that need to be expressed. Cells have multiple mechanisms to manipulate and alter chromatin to favor a more accessible or less accessible state. These include chromatin remodelers, histone chaperones, variants of the histone proteins, and histone modifiers.

1.1.2.1 Chromatin remodelers

Chromatin remodelers, such as Swi/Snf and Chd1, are enzyme complexes that use the energy of ATP hydrolysis to alter nucleosome structure in a variety of ways that are necessary to regulate DNA-templated processes such as transcription and DNA replication. These functions include sliding the positions of nucleosomes relative to the underlying DNA, loosening histone-DNA contacts, removing histone dimers or entire octamers from DNA, and

facilitating the exchange of histones with variants (reviewed in (CLAPIER AND CAIRNS 2009)). Importantly for facilitation of gene expression, remodelers can slide nucleosomes away from the promoter of a gene, enabling Pol II binding to promoter sequences to initiate transcription.

1.1.2.2 Histone chaperones

Histone chaperones, such as Spt6 and the FACT complex, are proteins capable of binding histones to remove them from or place them onto DNA (BORTVIN AND WINSTON 1996; LEROY *et al.* 1998; ORPHANIDES *et al.* 1998). Chaperones are necessary for deposition of newly synthesized histones onto DNA and also facilitate transcription elongation by removing histones from DNA ahead of a transcribing polymerase and replacing histones onto DNA in the wake of polymerase. The FACT complex (facilitates chromatin transcription) consists of Spt16, Pob3, and Nhp6 in yeast. It was first named for its roles in transcription, but further investigations have revealed roles for FACT in DNA replication, DNA repair, and centromere integrity. For this reason, it has been proposed that the acronym FACT should instead represent "facilitates chromatin transactions (FORMOSA 2012)." Pol II has some inherent ability to begin transcribing through a nucleosome, but becomes paused partway through. *In vivo*, passage of one Pol II through a nucleosome displaces one H2A/H2B dimer leaving an intact hexasome in the same location of the original nucleosome (KIREEVA *et al.* 2002). Through a mechanism reviewed in (KULAEVA *et al.* 2013), Pol II is able to transcribe through hexasomes by sequential destabilization of histone-DNA contacts. Formation of a hexasomal template is facilitated by FACT, possibly by removal of an H2A/H2B dimer or by destabilization of intra-nucleosomal contacts or a combination of these roles (reviewed in (FORMOSA 2012)). FACT is subsequently required to maintain nucleosome occupancy at highly transcribed regions (HAINER *et al.* 2011). Spt6 has been less studied and traditionally viewed as an H3-H4 chaperone. Recent genetic

evidence indicates that although Spt6 and FACT likely have distinct molecular roles, they function together in nucleosome reorganization (MCCULLOUGH *et al.* 2015). In line with this view, Spt6 is also required to maintain nucleosome occupancy at highly transcribed regions (HAINER *et al.* 2011).

1.1.2.3 Histone variants

The bulk of histone proteins deposited onto newly synthesized DNA are the canonical H2A, H2B, H3, and H4 proteins mentioned above (MARZLUFF AND DURONIO 2002). Cells also utilize variant histones throughout the cell cycle that have specialized roles in regulation of gene expression, DNA repair, or chromatin structure (reviewed in (TALBERT AND HENIKOFF 2010)). These variants typically have small changes in the amino acid sequence compared to their canonical counterparts. Several histone variants are found in all eukaryotes, while many more variants are lineage-specific. One universal variant is CenH3 (Cse4 in yeast), which is a centromeric form of H3 that is required for proper assembly of kinetochores during mitosis (reviewed in (SANTAGUIDA AND MUSACCHIO 2009)). Another universal variant is H2A.Z, which is incorporated into nucleosomes on either side of a transcription start site, demarcating the boundaries of nucleosome free regions and assisting in the recruitment of Pol II to promoters (ADAM *et al.* 2001; HARDY *et al.* 2009). H2A.X is another universal variant of H2A, which becomes phosphorylated and signals the presence of DNA double-strand breaks for repair (reviewed in (VAN ATTIKUM AND GASSER 2009)). Histone variants provide a crucial layer of complexity in genome regulation that have vital impacts on human development and diseases (reviewed in (MAZE *et al.* 2014)).

1.1.2.4 Histone modifications

Histone modifiers are enzymes that catalyze chemical reactions resulting in covalent attachment of small chemical moieties or proteins to specific residues of the histone proteins. Many of these modifications take place at histone residues located on flexible tails of the histone proteins that extend outside of the nucleosome core structure, but several of these modifications are placed onto the core of a nucleosome. These modifications include acetylation, methylation, and ubiquitylation of histone proteins (reviewed in (ZENTNER AND HENIKOFF 2013)). Histone modifications have a wide variety of effects. Some are repressive to transcription, including the methylation of H3 at lysine 27 (K27) (in mammals and flies), and others positively regulate transcription, such as acetylation of H3 and H4. These relationships are further complicated by the fact that the same chemical modification can have opposite effects if placed on different residues of the histones or at different genetic loci. At a basic level, these modifications are thought to exert their effects either by promoting or preventing the compaction of chromatin or by serving as a signal that marks a specific locus to be recognized by downstream effectors.

Actively transcribed loci display a characteristic array of histone modifications. Histone acetylation is high over the promoters and 5' ends of transcription units. Trimethylation of H3 at lysine 4 (K4) is highest at the 5' ends of transcription units and the level of methylation decreases gradually throughout the bodies of transcription units following with enrichment of dimethylation of H3 K4 and further downstream enrichment of mono-methylation of H3 K4 (NG *et al.* 2003b; LIU *et al.* 2005; POKHOLOK *et al.* 2005). The methyltransferase enzyme, Set1, catalyzes methylation of H3 K4 (SANTOS-ROSA *et al.* 2002). Mono-ubiquitylation of H2B at lysine 123 (K123, K120 in humans) is found throughout the bodies of transcription units. The

S. cerevisiae ubiquitin conjugating and ligase enzymes, Rad6 and Bre1, catalyze the ubiquitylation of H2B K123 (ROBZYK *et al.* 2000; HWANG *et al.* 2003; WOOD *et al.* 2003a). Methylation of H3 at lysine 36 (K36) is enriched at the 3' ends of transcription units (KROGAN *et al.* 2003b; LI *et al.* 2003; SCHAFT *et al.* 2003; XIAO *et al.* 2003). The methyltransferase, Set2, catalyzes the methylation of H3 K36 (STRAHL *et al.* 2002). Methylation of H3 at lysine 79 is broadly distributed, occurring throughout the bodies of transcription units. The methyltransferase, Dot1, catalyzes the methylation of H3 K79 (VAN LEEUWEN *et al.* 2002).

Histone acetylation alters the electrostatic properties of lysine residues and is thought to exert is regulatory effects by loosening histone-DNA contacts, and thus, this promoter-associated modification may make promoter DNA more accessible to RNA polymerases. Histone methylation can have both positive and negative regulatory roles. This can be explained by what effect proteins that recognize or "read" this mark have on gene expression (reviewed in (LI *et al.* 2007)). Many "readers" of methylated lysine residues have been identified and several domains of these proteins have been characterized for their ability to bind to methylated lysine residues. These domains include chromodomains and tudor domains. For example, the chromodomain of Eaf3, a subunit of the Rpd3S histone de-acetylase complex (HDAC), recognizes methylated H3 K36 residues and localizes Rpd3S to gene bodies (KEOGH *et al.* 2005). This results in histone de-acetylation toward the 3' ends of genes and is important for maintaining chromatin structure at the 3' ends of genes to prevent spurious transcription from within gene bodies (CARROZZA *et al.* 2005). Acetylated histone residues can be recognized by bromodomains contained in histone acetyltransferase (HAT) complexes as well as chromatin remodelers. The Gcn5 subunit of the SAGA co-activator complex contains a

bromodomain that may contribute to recruitment of this complex to the 5' ends of genes (OWEN *et al.* 2000).

The alterations in chromatin discussed here that facilitate transcription are regulated by transcription elongation factors. An important regulatory elongation factor that has many roles related to chromatin during transcription is the Paf1 complex. This complex is a critical regulator of the candidate gene that is the subject of much of this work.

1.1.3 The Paf1 complex

The evolutionarily conserved Paf1 complex consists of five subunits in *S. cerevisiae*, Paf1, Cdc73, Ctr9, Rtf1, and Leo1, and one additional subunit, Ski8, in humans (SHI *et al.* 1997; KROGAN *et al.* 2002; MUELLER AND JAEHNING 2002; SQUAZZO *et al.* 2002; ZHU *et al.* 2005). The Paf1 complex is a transcription elongation factor that associates with Pol II at all actively transcribed regions (WADE *et al.* 1996; KROGAN *et al.* 2002; POKHOLOK *et al.* 2002; KIM *et al.* 2004a; MAYER *et al.* 2010). The Paf1 complex plays a number of important regulatory roles throughout the transcription cycle that help to facilitate transcription. These roles include regulation of post-translational histone modifications, maintenance of nucleosome occupancy, recruitment of chromatin remodelers, regulation of the phosphorylation state of the Pol II CTD, and regulation of termination and 3' end formation of RNA transcripts (CRISUCCI AND ARNDT 2011).

The Rtf1 subunit is responsible for the regulation of a set of histone modifications as well as the recruitment of the chromatin remodeler, Chd1 (NG *et al.* 2003a; SIMIC *et al.* 2003). A small domain of Rtf1 called the histone modification domain (HMD) is required for the

ubiquitylation of H2B K123 by the enzymes Rad6 and Bre1 (WARNER *et al.* 2007). This ubiquitylation mark is an upstream pre-requisite for the downstream methylation of H3 at K4 and at K79 catalyzed by the methyltransferases Set1 and Dot1 (BRIGGS *et al.* 2002; KROGAN *et al.* 2003a; NG *et al.* 2003a; NG *et al.* 2003b). The Paf1 complex also regulates methylation of H3 at K36. This modification does not require Rtf1 or ubiquitylation of H2B. Instead, methylation of K36 by Set2 is dependent on the Paf1 and Ctr9 subunits (CHU *et al.* 2007). The histone modifications regulated by the Paf1 complex are depicted in Figure 2.

The Paf1 complex is required for normal phosphorylation of Ser2 of the Pol II CTD (NORDICK *et al.* 2008). This may be due to the roles of the Paf1 complex in regulating ubiquitylation of HB2 K123, as de-ubiquitylation of H2B is required for proper recruitment of Ctk1, the enzyme that phosphorylates Ser2 of the Pol II CTD (CHO *et al.* 2001; WYCE *et al.* 2007). The Paf1 complex is also required to maintain nucleosome occupancy at highly transcribed regions (PRUNESKI *et al.* 2011). This is likely mediated by interactions between the Paf1 complex and a histone chaperone, the FACT complex (SQUAZZO *et al.* 2002).

The Paf1 complex is also required for proper termination of transcription and 3' end formation of RNA transcripts as loss of the Paf1 complex results in transcriptional read-through of many genes. The Paf1 complex is required for recruitment of the CPF complex, which mediates termination of longer transcripts including the majority of mRNAs in *S. cerevisiae,* which is the model system for these studies (MUELLER *et al.* 2004; NORDICK *et al.* 2008). This association is likely to be at least partly dependent on the Paf1 complex's role in promoting the phosphorylation of Ser2 on the Pol II CTD (AHN *et al.* 2004). Paf1 complex-dependent histone modifications are also required for proper termination by the *S. cerevisiae* Nrd1-Nab3-Sen1 (NNS) complex, which is responsible for the termination of short, ncRNAs (SHELDON *et al.*

2005; TOMSON *et al.* 2011; TOMSON *et al.* 2013). The Paf1 complex also plays important post-transcriptional roles. Loss of the Paf1 complex results in reduced poly-A tail length and altered poly-A site usage (MUELLER *et al.* 2004; PENHEITER *et al.* 2005).

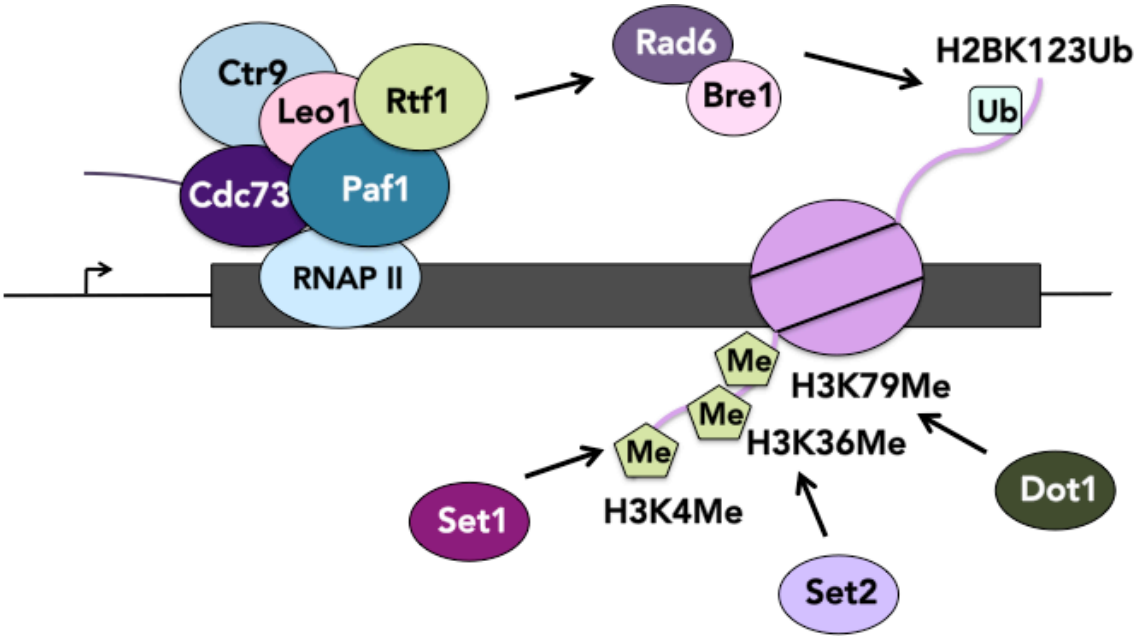

Figure 2. Schematic of histone modifications regulated by the Paf1 complex.

The Rtf1 subunit of the Paf1 complex is required for the monoubiquitylation of H2B K123 catalyzed by the ubiquitin conjugase and ligase enzymes, Rad6 and Bre1. The ubiquitylation of H2B K123 is required for methylation of H3 K4, catalyzed by Set1, and H3 K79, catalyzed by Dot1. The Paf1 subunit is required for methylation of H3 K36, catalyzed by Set2. Arrows to histone residues indicate which enzymes catalyze the modification of that residue. The arrow between Rtf1 and Rad6/Bre1 indicates that Rtf1 is required for the ubiquitylation of H3K123 by Rad6/Bre1 through a direct interaction (M. Shirra and K. Arndt, unpublished data).

1.2 PERVASIVENESS OF NON-CODING TRANSCRIPTION

The sequencing of the human genome and subsequent genome-wide expression analyses revealed that although about 99% of the human genome does not contain protein-coding sequences, the majority of the genome is transcribed into RNA. At least 80% of the human genome displays biochemical activities associated with transcription in at least one cell type (CONSORTIUM 2012). These non-coding transcripts vary widely in their length as well as their stability. The locations of the DNA sequences containing non-coding transcripts also vary. Noncoding transcription units occur in all possible orientations with respect to protein-coding genes and their promoters. Longer noncoding transcripts may span more than one open reading frame (ORF) of multiple protein-coding genes while others are entirely intergenic (reviewed in (GUTTMAN AND RINN 2012)). It has been proposed that not all of these non-coding transcripts have biological roles and that many ncRNAs might arise from spurious transcription from somewhat leaky transcriptional machinery (STRUHL 2007; SEILA *et al.* 2009). However, it is becoming increasingly clear that many ncRNAs do have important biological roles. Furthermore, this pervasiveness of transcription is a feature common to all eukaryotic genomes, suggesting that non-coding transcription has been heavily conserved throughout evolution of eukaryotic organisms. It is possible that increased non-coding transcription could be correlated with greater complexity of an organism. An emerging theme is that many ncRNAs have important regulatory roles. This is evidenced by the fact that many ncRNAs have been found to play a role in cancer progression (GUPTA *et al.* 2010; GUTSCHNER AND DIEDERICHS 2012). As more functions for ncRNAs are elucidated, the mechanisms employed by ncRNAs become more widely varied, indicating a previously under-explored wealth of regulatory mechanisms.

1.3 NON-CODING TRANSCRIPTS IN *S. CEREVISIAE*

Several classes of non-coding transcripts have been characterized genome-wide in *S. cerevisiae*. These classes vary in their length and stability. Small, unstable ncRNAs consisting of about 200-400 nucleotides are called cryptic unstable transcripts (CUTs). CUTs are terminated by the NNS complex, polyadenylated by the TRAMP complex, and quickly degraded by the nuclear exosome (WYERS *et al.* 2005; ARIGO *et al.* 2006; THIEBAUT *et al.* 2006). During this process, CUTs are handed from NNS to the TRAMP complex, which polyadenylates and unwinds CUTs. Polyadenylation and unwinding stimulates degradation of these transcripts by the nuclear exosome, which interacts with the TRAMP and NNS complexes (LACAVA *et al.* 2005; VANACOVA *et al.* 2005; VASILJEVA AND BURATOWSKI 2006). As a result, CUTs are so quickly degraded that they are not normally detected in wild-type cells. CUTs can be detected by disrupting their degradation, which can be achieved by deletion of the *TRF4* gene, encoding a subunit of the TRAMP complex, or by deletion of the *RRP6* gene, encoding a catalytic subunit of the nuclear exosome (DAVIS AND ARES 2006). RNA turnover by the nuclear exosome has been reviewed in (SCHNEIDER AND TOLLERVEY 2013). A distinct class of ncRNAs from CUTs is referred to as stable unannotated transcripts (SUTs) (XU *et al.* 2009). SUTs are larger in size and are terminated by the CPF complex. SUTs can be detected without altering RNA degradation.

Other classes of ncRNAs are named for how they are terminated or for the biological functions with which they are associated. These classes are not distinct and can include SUTs and CUTs as well as transcripts that are only expressed under specific cellular conditions. For example, meiotic unannotated transcripts (MUTs) are a class of ncRNAs that are normally degraded by the nuclear exosome in mitotic cells, but upon transitioning to meiosis are

stabilized (LARDENOIS *et al.* 2011). Transcripts that accumulate in the absence of the Xrn1 endonuclease are called Xrn1-sensitive unstable transcripts (XUTs). Some of these transcripts have also been annotated as CUTs and SUTs while the majority of XUTs are a distinct set of transcripts. Most XUTs are transcribed antisense to a protein-coding gene and could provide a broadly distributed mechanism to regulate expression of protein coding-genes (VAN DIJK *et al.* 2011). Transcripts dependent on the NNS complex are called Nrd1 unterminated transcripts (NUTs), as loss of the NNS complex results in extended, read-through transcripts. NUTs share a great degree of overlap with CUTs, SUTs, and XUTs and also include most snoRNAs. Proper termination of these transcripts has been proposed to be necessary for prevention of transcription interference that may otherwise result from extended read-through transcription (SCHULZ *et al.* 2013).

1.4 GENE REGULATION BY NON-CODING RNAS

Non-coding RNAs can regulate gene expression in many ways. Some of these mechanisms are dependent on the RNA molecule to interact with proteins to perform their regulatory functions while other mechanisms are only dependent on the act of transcription to regulate transcription of neighboring genes. There are several well-studied cases of regulation by intergenic transcription, which are discussed here.

An important distinction in thinking about the ways that ncRNAs can regulate gene expression is whether the regulation works *in cis* or *in trans*. If a regulator works *in cis*, the regulator must be encoded in the DNA at the locus of a target gene in order to perform its function. If a regulator works *in trans*, the regulator can be encoded in a different region of the

genome than its target and still be able to perform its function. The distinction between *cis* and *trans* is clear when considering protein transcription factors. Protein transcription factors regulate *in trans* as they can exert their function at any locus regardless of the genomic location of the gene encoding the transcription factor itself. This is attributed to the ability of proteins to diffuse throughout the nucleus. The DNA sequences bound by transcription factors are *cis*-regulatory elements and must be physically present at the target gene locus. The distinction is less obvious when considering RNA regulators as Pol II effectively tethers an individual RNA to a specific locus while it is being transcribed but the RNA molecule is able to diffuse after transcription has been terminated. This makes it possible for RNA molecules to be either *cis*- or *trans*- regulators. Thus, determining whether RNA regulators work *in cis* or *in trans* can be insightful when postulating models for mechanisms of regulation by ncRNAs.

1.4.1 Regulation of gene expression by non-coding RNA interactions with protein complexes

Many ncRNAs mediate their functions through interactions with proteins or protein-complexes. Among the earliest characterized ncRNAs were ribosomal RNAs (rRNAs), which together with ribosomal proteins form the structure of the catalytic machinery necessary for translating mRNAs into proteins. The later discovery of micro-RNAs (miRNAs) was an eye-opening landmark in our understanding of RNA biology. Micro-RNAs are very short (about 20 nucleotides) RNAs that associate with Argonaute protein complexes and serve as a guide to regulate the binding of Argonaute complexes to specific sequences of another RNA. The functions of miRNAs have been reviewed in (BARTEL 2004). In the cytoplasm, miRNAs can initiate mRNA degradation or block translation of specific mRNAs in order to regulate gene

expression at a post-transcriptional level. Micro-RNAs that block translation have the ability to regulate *in trans*. In the nucleus, miRNAs can guide Argonaute complexes to RNAs tethered to specific genetic loci resulting in downstream chromatin modifications that regulate transcription of protein-coding genes at those loci. Nuclear-acting miRNAs have both features of *cis* and *trans* regulation. The miRNA guide can be encoded anywhere in the genome, working *in trans*. These miRNAs can be recruited to specific loci by the transcription of ncRNAs *in cis* at the target locus. The discovery of miRNAs was very influential because it brought to our attention that non-protein-coding molecules so small that they had previously escaped detection can have a huge impact on regulation of gene expression and are critical regulators for normal development as well as prevention of human disorders.

A more recently characterized class of much larger ncRNAs that plays important regulatory roles is called long non-coding RNAs (lncRNAs). The GENCODE database has currently annotated 19,815 lncRNAs in the human genome, which have been identified by chromatin landmarks, RNA-seq expression data, and computational analyses (HARROW *et al.* 2012). This number is likely to increase as more expression data become available. Furthermore, the potential for alternative isoforms of these lncRNAs presents a rich pool of biological functions to be characterized. Currently, only a few lncRNAs have been well studied, including *Xist* and *HOTAIR*. Both of these lncRNAs are critical developmental regulators. *Xist* is involved in X chromosome inactivation, while *HOTAIR* regulates expression of the Hox genes (RINN *et al.* 2007; LEE 2009; SIMON AND KINGSTON 2013). Although the mechanistic details of how these lncRNAs function are not completely understood, both of these lncRNAs result in recruitment of the polycomb repressive complex 2 (PRC2) which alters the local chromatin landscape by promoting methylation of H3 K27 at these loci (TSAI *et al.* 2010).

In order to regulate the Hox genes, *HOTAIR* must be transcribed *in cis* at the Hox locus in order to recruit PRC2. *Xist* regulates X chromosome inactivation *in cis*, at least in initial stages of X chromosome inactivation, as only the X chromosome expressing the *Xist* RNA will become inactive. However, the *Xist* RNA is able to spread across the entire inactive X chromosome and may display some features of *trans* regulation in later stages of inactivation. The mechanisms that allow this spreading are an interesting area of active research. Current models suggest that the Xist RNA spreads to loci on the X chromosome by proximity transfer between sequences that are close together in the folded three-dimensional structure of the X chromosome rather than by spreading in linear order of the DNA molecule (ENGREITZ *et al.* 2013).

In addition to interacting with chromatin modifiers, lncRNAs can interact with transcription factors to regulate gene expression. Several lncRNAs have been implicated in cancer progression (SPIZZO *et al.* 2012). *PCGEM1* is a lncRNA that is specifically expressed in prostate cancer (SRIKANTAN *et al.* 2000). Recent work by the Kung lab revealed that one of the ways *PCGEM1* leads to cancer progression is through its interaction with the oncogenic transcription factor c-Myc. *PCGEM1* directly binds to the c-Myc protein, and results in enhanced DNA binding and enhanced transactivation activity of c-Myc, allowing the *trans*-regulation of many target genes by *PCGEM1*. Transcriptional targets include genes involved in the tricarboxylic acid cycle and nucleotide biosynthesis, explaining part of *PCGEM1*'s contribution to cancer progression (HUNG *et al.* 2014).

1.4.2 Regulation of *SER3* expression by intergenic transcription

The act of transcribing ncRNAs can also have regulatory roles. Previous work in the Martens lab described one such regulatory mechanism at the *S. cerevisiae SER3* gene. *SER3*

encodes an enzyme involved in serine biosynthesis. When serine is not present in the environment, the *SER3* gene is expressed, enabling cells to synthesize their own serine. When serine is present in the environment, cells do not need to make their own serine and the *SER3* gene is repressed. The repression of *SER3* is mediated by transcription of a ncRNA, called *SRG1*, over the promoter of the *SER3* gene *in cis* (MARTENS *et al.* 2004; MARTENS *et al.* 2005). The repression by *SRG1* is dependent on the histone chaperones, Spt6 and the FACT complex, which travel along with Pol II during transcription (HAINER *et al.* 2011). These histone chaperones facilitate transcription through a chromatin template by removing histones from the DNA ahead of an elongating Pol II enzyme and replacing the histones in the wake of Pol II. Histone chaperones position nucleosomes in a characteristic array with specific spacing relative to the transcription start site. During transcription of *SRG1*, the placement of nucleosomes coincides with the upstream activating sequences for the *SER3* promoter (HAINER *et al.* 2011). This creates a physical barrier to potential activating proteins, as the DNA sequences they recognize are inaccessible. Further work in the Martens lab has identified specific amino acid residues of the histones and of the histone chaperone, Spt16, that are necessary for *SER3* repression and for maintaining nucleosome occupancy at other highly transcribed loci (HAINER AND MARTENS 2011a; HAINER *et al.* 2012). A model of *SER3* repression by intergenic transcription is depicted in Figure 3.

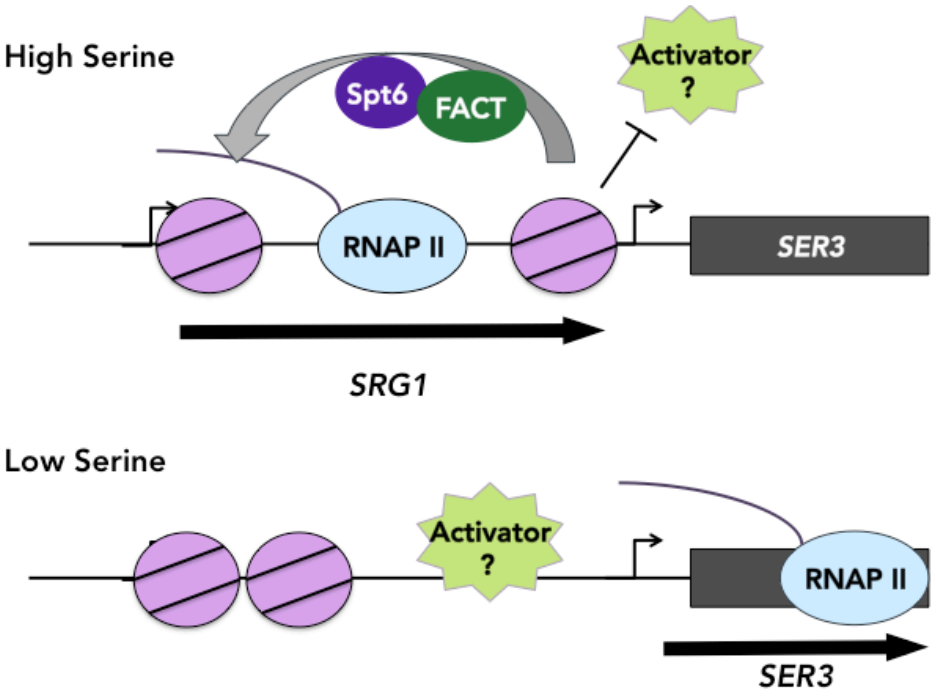

Figure 3. Schematic of *SER3* repression by intergenic transcription.

In high serine conditions, *SER3* is repressed by transcription of a ncRNA, *SRG1*, over its promoter. During transcription of *SRG1*, histone chaperones, Spt6 and the FACT complex, place nucleosomes over the upstream activating sequences for the *SER3* gene. These nucleosomes provide a physical barrier to transcription factors accessing the *SER3* promoter. In low serine conditions, *SRG1* is not transcribed and the nucleosomes at this locus are shifted farther upstream, allowing transcriptional machinery to access the *SER3* promoter. This model is based on previous discoveries in the Martens lab (MARTENS *et al.* 2004; MARTENS *et al.* 2005; HAINER *et al.* 2011).

1.4.3 Regulation by intergenic transcription at other loci

Several other cases of regulation by the act of transcribing ncRNAs have been reported and more cases continue to be elucidated. The models proposed to mediate regulation by ncRNA transcription include transcriptional interference, alterations in histone occupancy, and alterations in post-translational histone modifications. A common feature of these mechanisms is that they work *in cis* and do not necessarily require the ncRNA product for regulation. Instead, the act of transcribing these ncRNAs at specific loci is required for these regulatory mechanisms.

1.4.3.1 Transcriptional interference

Transcriptional interference is most simply described as the physical blockage of one elongating RNA polymerase due to the presence of another polymerase complex transcribing through the same region on the opposite strand of DNA. The expression of many genes is thought to be negatively regulated by transcription of a ncRNA antisense to the mRNA gene. It has been shown *in vivo* and *in vitro* that two converging polymerases cannot pass each other and require other factors to remove one of the polymerases from the DNA in order to resume transcription (HOBSON *et al.* 2012). Although many antisense transcripts occur throughout the genome, an estimate of how frequently these events are regulatory is difficult to measure as individual genes are regulated differently in varying environmental conditions. One study that probed for correlation patterns between sense and antisense transcript pairs across several environmental conditions found the potential for up to 5% of the *S. cerevisiae* protein-coding genes to be regulated by antisense transcription (XU *et al.* 2011). This figure is likely to be an

under-estimate, as more regulatory pairs would be revealed in environmental conditions that have not yet been explored.

A classic example of transcriptional interference has been described at the *S. cerevisiae* *IME4* gene. *IME4* is a gene necessary for diploid yeast cells to undergo meiosis. Diploid cells not undergoing meiosis transcribe a ncRNA antisense to *IME4*. Transcription of the ncRNA does not interfere with transcription factor binding or initiation at the *IME4* promoter and likely acts by preventing elongation of *IME4* transcription (HONGAY *et al.* 2006; GELFAND *et al.* 2011).

1.4.3.2 Noncoding DNA transcription alters nucleosome occupancy

Similar to *SER3*, at least two other loci are regulated by transcription of noncoding DNA (ncDNA) by a mechanism that involves alterations in nucleosome occupancy. These occur at the *Schizosaccharomyces pombe* locus, *fbp1*$^+$ locus and the mouse Igl loci. The *fbp1*$^+$ locus is activated by transcription of ncRNAs over its promoter in a stepwise fashion. This series of non-coding transcription events displace nucleosomes over the *fbp1*$^+$ promoter and lead to induction of *fbp1*$^+$ expression (HIROTA *et al.* 2008). During B-cell development, recombination of the immunoglobulin genes allows the immune system to generate a vast array of antibodies to recognize a wide variety of antigens. This process is facilitated by non-coding transcription across the IgL loci, which evicts H2A/H2B dimers from the DNA. The eviction of H2A/H2B dimers allows the recombination machinery to access the DNA in order to perform recombination (BEVINGTON AND BOYES 2013). This example is of particular interest as it demonstrates that the regulatory potential of non-coding transcription is not limited to gene expression, but has the potential to regulate all DNA-templated processes.

1.4.3.3 Noncoding DNA transcription alters post-translational histone modifications

As mentioned before, histones can be post-translationally modified with small chemical moieties and proteins that result in specific regulatory events. The modifications regulated by the Paf1 complex occur co-transcriptionally at both protein-coding and non-coding loci. Several cases have been reported in which noncoding transcription results in the placement of a histone modification that is necessary for downstream regulation of a neighboring protein-coding gene. One example occurs at the *S. cerevisiae GAL1-10* locus. The *GAL* genes are involved in catabolism of the sugar galactose as an energy source. If glucose is available, glucose will be catabolized first. The *GAL1-10* genes are repressed in the presence of glucose, and only become activated when galactose is present and glucose is absent. A ncRNA transcribed at the *GAL1-10* locus is necessary for the repression of these genes in the presence of glucose. This noncoding transcription results in methylation of H3 K4 that is required for downstream de-acetylation of histones over the divergent *GAL1-10* promoter by the Rpd3S complex. This de-acetylation is necessary to maintain a repressed state (HOUSELEY *et al.* 2008; PINSKAYA *et al.* 2009).

The *S. cerevisiae PHO84* gene is repressed by an extended antisense transcript, which results in recruitment of the histone deacetylase, Hda1, to the *PHO84* promoter (CAMBLONG *et al.* 2007). The recruitment of Hda1 deacetylates histones at the *PHO84* promoter and is necessary for full repression of *PHO84* mRNA transcription. The antisense ncRNA exists as either a short isoform that is terminated by the NNS complex or a long isoform that is terminated by the CPF complex. Termination of the antisense RNA by the NNS complex does not result in transcription across the *PHO84* promoter and allows *PHO84* to be expressed.

Termination of the antisense RNA by the CPF complex results in a much longer transcript that spans the promoter of *PHO84*, leading to the recruitment of Hda1 and resulting in repression of *PHO84* (CASTELNUOVO *et al.* 2013).

The *S. cerevisiae* genes *DCI1* and *DUR3* are also repressed by transcription of ncRNAs across their promoters. *DCI1* and *DUR3* each have a short ncRNA transcribed over their promoters in a tandem orientation to the mRNA transcript. Transcription of these short ncRNAs results in methylation of H3 K4 across the promoters of these genes, which recruits the histone deacetylase Set3. Set3 leads to repression of the *DCI1* and *DUR3* protein-coding transcripts, much resembling the function of Set3 in repressing cryptic intragenic transcripts (KIM *et al.* 2012).

Taken together, these examples clearly highlight the importance of non-coding transcription in regulating gene expression. A variety of mechanisms for regulation by non-coding transcription have been identified and many have opened up new questions for further investigation. As more examples of regulatory non-coding transcriptional events are characterized, these mechanisms will become more widely varied and will likely give insights into how these types of mechanisms contribute to human development and disease.

1.5 THESIS SIGNIFICANCE AND AIMS

The goal of the work described in this dissertation is to further our understanding of regulation by intergenic transcription. While only a few examples of regulation by intergenic transcription have been well described, the huge number and pervasiveness of non-coding transcripts presents the possibility that these mechanisms are widespread and diverse. Work in

the Martens lab (and others) describing mechanisms of regulation by intergenic transcription has left two important questions, which have been the focus of my thesis: (1) is regulation by intergenic transcription widespread? and (2) do regulatory non-coding transcripts share common regulatory mechanisms? To address the first question, I was part of a collaborative genome-wide project with the goal of identifying how many genes in the *S. cerevisiae* genome are likely to be regulated by maintenance of transcription-coupled nucleosome occupancy, similar to the *SER3* gene. These data are discussed in the Appendix. These interesting results give insight into how frequently individual regulatory mechanisms might be repeated throughout the genome and add depth to our perspective of regulatory mechanisms elucidated at individual loci. To address the second question, I selected candidate protein-coding genes that have CUTs transcribed over their promoters to determine whether these genes are regulated by CUT transcription over their promoters for further mechanistic studies. This strategy allows for the characterization of novel regulatory mechanisms as well as determining if these genes share regulatory mechanisms with other genes regulated by noncoding transcription. These data are discussed in Chapters Two, Three, and Four. These results identified several interesting model genes for further studies that have previously unknown regulatory roles for noncoding transcripts. I investigated one of these regulatory roles in more detail, which revealed a positive regulatory role for histone H3 K4 methylation. Interestingly, this modification has a negative regulatory role at other loci associated with noncoding transcription. This work highlights the potential diversity of regulatory mechanisms employed by noncoding transcription and the importance of studying individual loci as one regulatory factor might have different effects in different local chromatin environments. Together, these studies advance our understanding of gene regulation by noncoding transcription. Enhancing our understanding of gene regulation is

essential to future progress in developing treatments for disorders caused by misregulation of genes, including developmental disorders and cancer.

2.0 CHAPTER TWO: SELECTION AND INITIAL CHARACTERIZATION OF CANDIDATE GENES FOR REGULATION BY INTERGENIC TRANSCRIPTION

2.1 INTRODUCTION

The goal of the work presented in this chapter was to survey a set of *S. cerevisiae* genes expressing non-coding transcripts over their promoters to determine if transcription of intergenic DNA at the promoters of these genes contributes to regulating expression of the neighboring protein-coding gene. We selected six candidate genes expressing CUTs over their promoters and disrupted transcription of these CUTs initially by insertion of a transcription termination sequence to prematurely stop CUT transcription. We then observed the effect that disruption of CUT transcription had on expression of the neighboring protein-coding genes. Also summarized in this chapter is initial characterization of several of these candidates which ultimately led us to select *ECM3* as the candidate gene we would investigate further with the goal of elucidating the mechanism of how intergenic transcription regulates this gene.

2.2 MATERIALS AND METHODS

2.2.1 *S. cerevisiae* strains and media

S. cerevisiae strains used in Chapter Two are listed in Table 1. All strains used in this chapter are *GAL2+* derivatives of S288C generated by standard genetic crosses and transformations (WINSTON *et al.* 1995). Insertion of a 180 base pair transcription termination sequence (TTS) from the 3' end of the *HIS3* gene was done by a two-step integration method, using the pDW1 plasmid as a PCR template to generate integration products using the primers listed in Table 2. The PCR products were transformed into diploid strains (YEAR001 and YEAR002). URA3+ diploids were isolated then sporulated. After isolation of spores with the desired genotypes, recombination was allowed to occur between the two tandem copies of the TTS, and non-recombinants were selected against on media containing 5FOA. This left behind a single copy of the TTS at the desired locations, which was confirmed by PCR. The TTS sequence was integrated at 190 base pairs upstream of the *ECM3* start codon, 400 bases upstream of the *KNH1* start codon, 300 bases upstream of the *ARO8* start codon, 520 bases upstream of the *CLN3* start codon, or 400 bases upstream of the *FET4* start codon. For the experiments described in this chapter, all strains were grown to early log phase at 30°C in rich medium (1% yeast extract, 2% peptone, 2% glucose) for RNA isolation.

Strain	Genotype	Reference or Source
FY4	*MATa*	Winston, *et al*., 1995
FY5	*MATα*	Winston, *et al*., 1995
YEAR001	*MATa/α ura3Δ0/ura3Δ0 leu2Δ0/LEU2 his3Δ200/HIS3 rrp6Δ0::KanMX/RRP6*	This study
YEAR002	*MATa/α ura3Δ0/ura3Δ0 leu2Δ0/LEU2 his3Δ200/HIS3 rrp6Δ0::KanMX/RRP6*	This study
YEAR034	*MATα rrp6Δ0::KanMX ura3Δ0 KNH1::HIS3 TTS at -300*	This study
YEAR037	*MATα rrp6Δ0::KanMX ura3Δ0 KNH1::HIS3 TTS at -300*	This study
YEAR038	*MATa ura3Δ0 KNH1::HIS3 TTS at -300*	This study
YEAR040	*MATα ura3Δ0 leu2Δ0 KNH1::HIS3 TTS at -300*	This study
YEAR041	*MATα rrp6Δ0::KanMX ura3Δ0 his3Δ200 KNH1::HIS3 TTS at -300*	This study
YEAR043	*MATα ura3Δ0 his3Δ200 leu2Δ0 KNH1:: HIS3 TTS at -300*	This study
YEAR045	*MATα ura3Δ0 his3Δ200 leu2Δ0 ECM3:: HIS3 TTS at -190*	This study
YEAR048	*MATa rrp6Δ0::KanMX ura3Δ0 leu2Δ0 ECM3::HIS3 TTS at -190*	This study
YEAR049	*MATa ura3Δ0 his3Δ200 ECM3::HIS3 TTS at -190*	This study
YEAR051	*MATa rrp6Δ0::KanMX ura3Δ0 his3Δ200 ECM3::HIS3 TTS at -190*	This study
YEAR053	*MATa ura3Δ0 leu2Δ0 ECM3::HIS3 TTS at -190*	This study
YEAR055	*MATα rrp6Δ0::KanMX ura3Δ0 his3Δ200 leu2Δ0 ECM3::HIS3 TTS at -190*	This study
YEAR058	*MATα rrp6Δ0::KanMX ura3Δ0 leu2Δ0 ECM3::HIS3 TTS at -190*	This study
YEAR059	*MATa ura3Δ0 his3Δ200 leu2Δ0 ECM3:: HIS3 TTS at -190*	This study
YEAR061	*MATα ura3Δ0 his3Δ200 KNH1:: HIS3 TTS at -300*	This study
YEAR062	*MATa rrp6Δ0::KanMX ura3Δ0 leu2Δ0 KNH1::HIS3 TTS at -300*	This study
YEAR071	*MATα ura3Δ0 leu2Δ0 ARO8::TTS at -300*	This study
YEAR072	*MATa ura3Δ0 leu2Δ0 ARO8::TTS at -300*	This study
YEAR073	*MATα ura3Δ0 ARO8::TTS at -300*	This study
YEAR075	*MATa ura3Δ0 leu2Δ0 ARO8::TTS at -300*	This study
YEAR077	*MATa ura3Δ0 leu2Δ0 trf4Δ0::NatMX ARO8::TTS at -300*	This study
YEAR078	*MATα ura3Δ0 leu2Δ0 trf4Δ0::NatMX ARO8::TTS at -300*	This study
YEAR079	*MATa ura3Δ0 leu2Δ0 trf4Δ0::NatMX ARO8::TTS at -300*	This study
YEAR080	*MATa ura3Δ0 leu2Δ0 trf4Δ0::NatMX ARO8::TTS at -300*	This study

YEAR083	*MATα ura3Δ0 leu2Δ0 CLN3::TTS at -520*	This study
YEAR085	*MATa ura3Δ0 leu2Δ0 CLN3::TTS at -520*	This study
YEAR086	*MATα ura3Δ0 leu2Δ0 CLN3::TTS at -520*	This study
YEAR089	*MATα ura3Δ0 leu2Δ0 trf4Δ0::NatMX CLN3::TTS at -520*	This study
YEAR091	*MATa ura3Δ0 leu2Δ0 trf4Δ0::NatMX CLN3::TTS at -520*	This study
YEAR092	*MATa ura3Δ0 leu2Δ0 trf4Δ0::NatMX CLN3::TTS at -520*	This study
YEAR120	*MATα ura3Δ0 leu2Δ0 rox1Δ0::kanMX*	This study
YEAR122	*MATa ura3Δ0 leu2Δ0 rox1Δ0::kanMX*	This study
YEAR123	*MATa ura3Δ0 leu2Δ0 rox1Δ0::kanMX*	This study
YEAR133	*MATa ura3Δ0 his3Δ200 leu2Δ0 FET4::HIS3 TTS at -400*	This study
YEAR134	*MATa ura3Δ0 his3Δ200 leu2Δ0 FET4::HIS3 TTS at -400*	This study
YEAR135	*MATa ura3Δ0 FET4::HIS3 TTS at -400*	This study
YEAR137	*MATα rrp6Δ0::KanMX ura3Δ0 leu2Δ0 FET4::HIS3 TTS at -400*	This study
YEAR138	*MATa rrp6Δ0::KanMX ura3Δ0 FET4::HIS3 TTS at -400*	This study
YEAR139	*MATa rrp6Δ0::KanMX ura3Δ0 leu2Δ0 FET4::HIS3 TTS at -400*	This study
YEAR140	*MATa rrp6Δ0::KanMX ura3Δ0 his3Δ200 FET4::HIS3 TTS at -400*	This study
YEAR141	*MATα rrp6Δ0::kanMX rox1Δ0::kanMX ura3Δ0*	This study
YEAR142	*MATa rrp6Δ0::kanMX rox1Δ0::kanMX ura3Δ0 leu2Δ0*	This study
YEAR144	*MATα rrp6Δ0::kanMX rox1Δ0::kanMX ura3Δ0 leu2Δ0*	This study
YJ1091	*MATα ura3Δ0 leu2Δ0 lys2-128δ trp1Δ63 spt16Δ::KanMX <pSPT16- URA3>*	Hainer, *et al*. 2012
YJ1125	*MATa ura3Δ0 his3Δ200 leu2Δ0*	This study
YJ1126	*MATa rrp6Δ0::KanMX ura3Δ0 his3Δ200 leu2Δ0*	This study
YJ744	*MATa rrp6Δ0::KanMX*	J. Pruneski and J. Martens, Tomson *et al*., 2013
YJ746	*MATα rrp6Δ0::KanMX*	J. Pruneski and J. Martens, Tomson *et al*., 2013
YS404	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 met15Δ0 (hht1-hhf1)Δ::hhts -K122A/HHFS-Hygro (hht2-hhf2)Δ::hhts-K122A/HHFS- URA3 can1Δ::MFApr -HIS3*	S. Hainer
YS417	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 met15Δ0 (hht1-hhf1)Δ::HHTS/HHFS-Hygro (hht2-hhf2)Δ::HHTS/HHFS-URA3 can1Δ::MFApr-HIS3*	S. Hainer
YS418	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 met15Δ0 (hht1-hhf1)Δ::HHTS/HHFS-Hygro (hht2-hhf2)Δ::HHTS/HHFS-URA3 can1Δ::MFApr-HIS3*	S. Hainer
YS419	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 met15Δ0 (hht1-hhf1)Δ::HHTS/HHFS-Hygro (hht2-hhf2)Δ::HHTS/HHFS-URA3 can1Δ::MFApr-HIS3*	S. Hainer
YS420	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 met15Δ0 (hht1-hhf1)Δ::hhts -K122A/HHFS-Hygro (hht2-hhf2)Δ::hhts-K122A/HHFS- URA3 can1Δ::MFApr -HIS3*	S. Hainer
YS421	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 met15Δ0 (hht1-hhf1)Δ::hhts-K122A/HHFS-Hygro (hht2-hhf2)Δ::hhts-K122A/HHFS- URA3 can1Δ::MFApr -HIS3*	S. Hainer

Table 1. *S. cerevisiae* strains used in Chapter Two.

2.2.2 Northern blot analysis

Total RNA was isolated from yeast cells as previously described using acid phenol (buffered to a pH of 5) at 65°C for one hour followed by a second acid phenol extraction at room temperature and a final extraction using chloroform at room temperature (AUSUBEL 1987). Ethanol precipitated RNA was re-suspended in water and stored at -80°C

Northern blot analyses were performed using 20 μg total RNA samples resolved in gels containing 2% agarose, 6.5% formaldehyde, and 1X MOPS as previously described (AUSUBEL 1987). Most of the northern blots in this work were run for 500 volt hours, which separates total RNA over a gel spanning about 10 cm. The high-resolution northern blot (Figure 8) was run for 2200 volt hours, after which low molecular weight RNAs will have run off the gel leaving only high molecular weight RNAs that are separated over about 20 cm of gel. Double-stranded probes were generated by random-primed labeling and single-stranded probes were generated by asymmetric PCR with α-^{32}P-dATP and purified PCR products as templates (RIO 2011). Probe templates were amplified from genomic DNA to contain the following sequences relative to the +1 ATG of the protein-coding gene at each locus: *EUC1* (-541 to -100), *ECM3* (+545 to +976), *ARO2*-CUT (-494 to -49), *ARO8*-CUT (-429 to -28), *ARO8*-ORF (+556 to +1051), *CLN3*-CUT (-671 to -321), *CLN3*-ORF (+420 to +896), *FET4*-CUT (-479 to -114), *FET4*-ORF (+261 to +651), *KNH1*-CUT (-496 to -235), *KNH1*-ORF (+250 to +670), *SCR1* (-182 to +284), and *ACT1* (+277 to +845). Oligonucleotides used to generate these probe templates are listed in Table 2. *SCR1* and *ACT1* RNA levels serve as internal loading controls. Images were generated by phosphorimaging and quantified using ImageJ software. For each experiment, the data from at least three biological replicates were averaged.

2.2.3 5’ RACE

5’ RACE was performed on total RNA isolated from an *rrp6Δ* strain (YJ744) using the RLM-RACE kit (Ambion) following the manufacturer’s guidelines. Briefly, the 5’ phosphate of any non-capped nucleic acid is removed by CIP, intact 5’ caps are cleaved by tobacco acid pyrophosphatase (TAP), leaving behind a 5’ phosphate specifically from capped RNAs, and an adapter sequence is ligated to these remaining phosphates. Nested PCR reactions amplify the ligated region spanning the adapter sequence at the 5’ end to a gene-specific primer at the 3’ end. Gene-specific PCR products are then cloned into a topoisomerase vector (TOPO TA cloning kit, Invitrogen) and sequenced. The nucleotide directly adjacent to the adapter sequence was identified as a start site for each clone. I sequenced 40 clones for each transcript to determine major and minor start sites as single clones are not indicative of relative abundance of each species.

2.2.4 Primer extension analysis

Primer extension assays were performed as previously described using 20 μg total RNA samples (AUSUBEL 1987). Sequencing reactions were performed using the Sequenase kit following manufacturer’s guidelines (Affymetrix USB) using a purified PCR product as a template. Oligonucleotides were gel-purified and end-labeled with α-^{32}P-ATP and T4 polynucleotide kinase using standard protocols (AUSUBEL 1987). The major *ECM3* start sites were each mapped with two different oligonucleotides (OJ1173, OJ1270, and OJ1362) listed in Table 2.

Purpose	Primer set	Oligo name	Orientation	Location relative to +1 ATG	Sequence
Northern probe template	***ARO2* CUT**	OJ1160	F	-494	5'- CTACGTTGGGCACGTCTAAG
		OJ1161	R	-49	5'- ATAACGCTTAGATGATGCCGT
Northern probe template	***ARO2* ORF**	OJ1162	F	+565	5'- CCTATCAGATGTCCAGACGC
		OJ1163	R	+1039	5'- TAGCTCTTGGAGTGACAGCA
Northern probe template	***ARO8* CUT**	OJ1156	F	-429	5'- CATGGCTCATATACACCATCC
		OJ1157	R	-28	5'- TGTCTGTATCAACTGCAGGG
Northern probe template	***ARO8* ORF**	OJ1158	F	+556	5'- GACGCTGATGGTATCATTCC
		OJ1159	R	+1051	5'- TGGATAGCGTAGCGTTGACC
Northern probe template	***CLN3* CUT**	OJ1164	F	-671	5'- GGAAGTGTCGCAACCAAACG
		OJ1166	R	-321	5'- GGCAGACTCAGTAGTAGAAG
ChIP qPCR	***CLN3* ORF**	OJ1165	F	+420	5'- CTCGCGGTTCATTATCAAGAG
		OJ1167	R	+896	5'- CAGCGCAATTAGTGAACGATC
Northern probe template	***FET4* CUT**	OJ1152	F	-497	5'- GCGTAAATCACACAGGTGTTG
		OJ1153	R	-114	5'- CAATTAATTCATGCCGTGTGAAG
Northern probe template	***FET4* ORF**	OJ1154	F	+261	5'- GGATTTCCTGGTACGAGTGG
		OJ1155	R	+265	5'- CGTTAGATAAACGGTCGTACC
Northern probe template	***KNH1* CUT**	OJ1168	F	-496	5'- TCAGCTGTACAAGCCTAGGC
		OJ1169	R	-235	5'- CTGTTGGAGTTGGTCAACAAT
Northern probe template	***KNH1* ORF**	OJ1170	F	+250	5'- AGTGCGAGTGATCTGACAGA
		OJ1171	R	+670	5'- CCTGGTGTTACGGTAGTATGT
Northern probe template	***ECM3 CUT***	OJ1172	F	-541	5'- CCATGCTTATCTGCCGTCTT
		OJ1173	R	-100	5'-GGTAATGGTCAACAATACGC
Northern probe template	***ECM3***	OJ1174	F	+545	5'- TGACCAATGATGATTCTGCCC
		OJ1175	R	+976	5'- GTAGTTCACGCATATCGATGG
Northern probe template	***SER3***	OJ174	F	+111	5'-CGTTCCACAGCGCTTGAATGCTG
		OJ244	R	+1342	5'-CGCTTTGGTCAACAGAAGAG
Northern	***SCR1***	OJ459	F	-182	5'-CAACTTAGCCAGGACATCCA

probe template		OJ460	R	+284	5’-AGAGAGACGGATTCCTCACG
Northern probe template	***ACT1***	OJ257	F	+277	5’-ATCGATTGCTTCATTCTTTTTGTT
		OJ258	R	+845	5’-ATCGATTCTCAAAATGGCGTGAGG
TTS integration	***ARO2* TTS**	OJ1180	F	-320	5’-GCCGGCTGTGACGCTGGCGCGCGATCCAAAAAGAAGTGTCTTTGATGACTCGCGGTGGCGGCCGCGGAAA
		OJ1181	R	-320	5’-CAAGAGAATGATGCTGAGTTACGTCTGTTCTGAAGCTGAATGGTCTGTATCCCCTCGAGGCGCGCCTCGT
TTS integration	***ARO8* TTS**	OJ1178	F	-300	5’-ACCTCATTCAAGAATCTGGCTTCTGAATTGCCATTGATAGAAGAACAGTACGCGGTGGCGGCCGCGGAAA
		OJ1179	R	-300	5’-AAGACATCTAAGGTAACCCGGATGTTCTTCATTATTCCGGCGCAATTTAGCCCCTCGAGGCGCGCCTCGT
TTS integration	***CLN3* TTS**	OJ1182	F	-520	5’-ACCAAGCCTGCTCTCACTGTAATGATCAAGTTACATAAATTTACTATCGGCGCGGTGGCGGCCGCGGAAA
		OJ1183	R	-520	5’-GGCAAAAACCCAAGGCCAAATATGGAAATGTGGCAGAGGGACACACTAATCCCCTCGAGGCGCGCCTCGT
TTS integration	***FET4* TTS**	OJ1176	F	-400	5’-GGAGCTTGTGCGGTTATGTATTAGATATGGGCAGTTTCCTTTAACGTTGGCGCGGTGGCGGCCGCGGAAA
		OJ1177	R	-400	5’-TGACCTAATCAGTTTCGAGAGCAACCCCACGGGTAGGAAAGAAAGAGGCGCCCCTCGAGGCGCGCCTCGT
TTS integration	***KNH1* TTS**	OJ1184	F	-400	5’-TCAAGTCATGGACTCATTATCGATTCGTCTTTTTTTAGCTGCCCCCACGTCGCGGTGGCGGCCGCGGAAA
		OJ1185	R	-400	5’-ACCCTTTCAAGGGACTCGCCGCGATGCAGAGAAAAAAAAAAGAAGAGATACCCCTCGAGGCGCGCCTCGT
TTS integration	***ECM3* TTS**	OJ1305	F	-190	5’-ATAGGCCTTTCATTGTTTTTAATATAGACTTTCATCATAGGGCATCCGGACGCGGTGGCGGCCGCGGAAA
		OJ1306	R	-190	5’-GTCTGAAAAACCTAATATAAAGAAAAATTGCGAGGCTTCTGAGAAGAAATCCCCTCGAGGCGCGCCTCGT
***ECM3* primer**		OJ1173	R	-100	5’-GGTAATGGTCAACAATACGC

extension					
***ECM3* primer extension**		OJ1270	R	-1	5'-TGTCTACTTGTCTTGAACTTAC
***ECM3* primer extension**		OJ1362	R	+1	5'- CCCAGTGTGATGTGTGTCAT

Table 2. Oligonucleotides used in Chapter Two.

2.3 RESULTS

2.3.1 Selection of candidate genes

We decided to look at cases where the ncRNA is transcribed over the promoter of a protein-coding gene in a tandem orientation, as this resembles the structure of the *SRG1-SER3* locus. Although the orientation of neighboring transcript pairs is not likely to limit their regulatory potential, selecting this class increased the likelihood that any regulatory effects would be due to transcription elongation rather than collision of convergent polymerase complexes as has been characterized at loci with sense and antisense transcript pairs. We selected six candidate genes showing expression of CUTs over their promoters in a tandem orientation as identified in microarray expression data (NEIL *et al.* 2009). We were especially interested in the CUT class of non-coding transcripts as they are quickly degraded by the nuclear exosome. Because of this, we reasoned that any regulatory effects we would find would likely be due to the act of transcription rather than the ncRNA molecule itself.

We intentionally wanted to leave open the possibility of finding a wide variety of potential regulatory mechanisms and not direct our studies toward re-discovering any previously characterized mechanisms of regulation by intergenic transcription. For this reason, the only criterion that we required of a candidate gene is that it expresses a CUT over its promoter in a tandem orientation. There were approximately 200 CUTs expressed in this orientation relative to protein-coding genes under steady-state rich medium growth conditions (NEIL *et al.* 2009). We also confirmed that these CUTs showed evidence of Pol II occupancy over their promoter regions in available genome-wide chromatin immunoprecipitation (ChIP) data (STEINMETZ *et al.* 2006). We considered other factors including the level of expression of

the neighboring protein-coding gene (NEIL *et al.* 2009), whether the genes are regulated by Swi/Snf (SUDARSANAM *et al.* 2000) or by depletion of histones (WYRICK *et al.* 1999), and whether the promoters of these genes contained nucleosomes or were free of nucleosomes (LEE *et al.* 2007). For these last criteria, although we considered each factor, we selected some genes that displayed these characteristics and some genes that displayed the opposite characteristics with the goal of diversifying the potential regulatory mechanisms we could investigate. The genes we selected as candidates are *ARO2*, *ARO8*, *CLN3*, *ECM3*, *FET4*, and *KNH1*. Snapshots of the microarray expression data identifying CUTs over the promoters of these genes are shown in Figure 4. A summary of the other characteristics of each candidate gene is listed in Table 3.

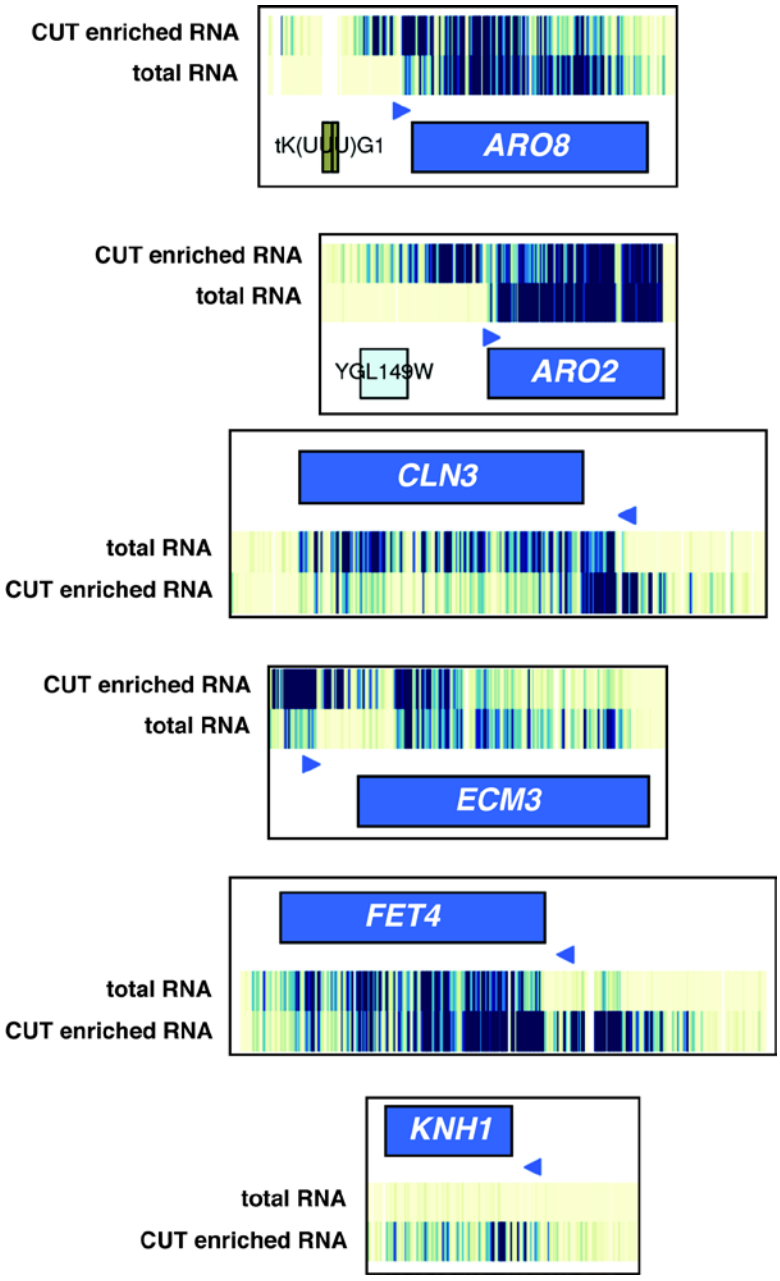

Figure 4. Microarray expression detection of CUT species upstream of candidate protein-coding genes. (Legend continued on next page.)

The images above are snapshots of microarray expression data of total RNA isolated from wild-type cells compared to CUT-enriched RNA isolated from *rrp6Δ* cells (NEIL *et al.* 2009). Darker blue indicates higher RNA levels and lighter yellow indicates lower RNA levels in raw microarray intensity measurements. Whether the expression data is above or below the

diagram of the ORF indicates whether that gene is transcribed from the Watson or Crick strand of DNA at that locus. Blue arrows indicate the positions of previously identified transcription start sites (NAGALAKSHMI *et al.* 2008).

Gene	Pol II density	Swi/Snf affected	Histone H4 depletion affected	Nucleosome occupancy	Gene function
SER3	**+**	**+**	**+**	**+**	Serine biosynthesis
KNH1	**+**	**+**	**+**	**-**	Cell wall synthesis
ECM3	**+**	**-**	**-**	**-**	Cell wall synthesis
FET4	**+**	**-**	**+**	**-**	Plasma membrane iron transporter
ARO8	**+**	**-**	**-**	**+**	Aromatic amino acid biosynthesis
CLN3	**+**	**-**	**-**	**+**	G1 cyclin
ARO2	**+**	**-**	**-**	**-**	Aromatic amino acid biosynthesis

Table 3. Summary of candidate gene characteristics compared to the *SER3* gene.

Each of the six candidate genes shows CUT expression (NEIL *et al.* 2009) and Pol II occupancy over their promoters (STEINMETZ *et al.* 2006). *KNH1* expression is affected by mutation of Swi/Snf, while the other genes are not (SUDARSANAM *et al.* 2000). *KNH1* and *FET4* expression are affected by depletion of H4 in genome-wide expression data (WYRICK *et al.* 1999). *ARO8* and *CLN3* have nucleosome occupancy over their promoters while the other genes have nucleosome free regions across their promoters; however, the positions of these nucleosome free regions may not be directly over the promoters of these protein coding genes, as we did not have clear boundaries for the CUT and ORF promoters (LEE *et al.* 2007). The functions of each gene are as listed on the *Saccharomyces cerevisiae* genome database.

I first confirmed that we could detect expression of the CUTs over the promoters of these six candidate genes by northern blot analysis. For each of the six candidate genes, I was able to detect expression of a CUT specifically in either *rrp6Δ* or *trf4Δ* using northern probe templates designed to be upstream of the protein-coding gene at each locus (Figure 5).

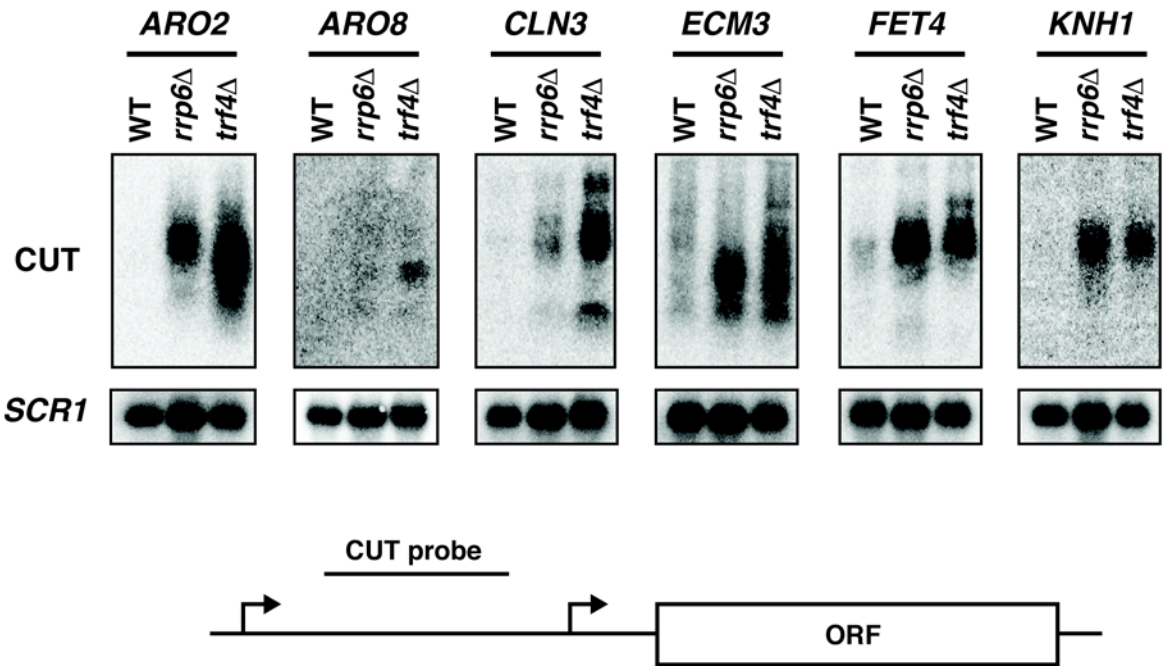

Figure 5. Detection of CUT transcripts across the promoters of six candidate genes.

Northern analysis was performed on RNA isolated from a wild-type strain (FY4) or strains where CUTs are stabilized (*rrp6Δ* YJ744; *trf4Δ* KY1975). Probes were designed to detect transcripts produced upstream of the neighboring protein-coding gene as diagramed below. *SCR1* serves as a loading control. Arrows on the diagram below indicate the relative positions of transcription start sites for the CUT and ORF transcripts.

It should be noted that *trf4Δ* strains show higher CUT levels than *rrp6Δ* strains. This holds true in most cases and stems from the fact that Rrp6 is only one of several catalytically active subunits of the nuclear exosome, leaving a partially active exosome complex. Additionally, the TRAMP complex acts at an earlier step in 3' end processing of CUTs, which may contribute to its stronger phenotype (CALLAHAN AND BUTLER 2010; SCHNEIDER AND TOLLERVEY 2013). An even stronger stabilization of CUTs can be observed in *rrp6Δ trf4Δ* double mutant strains; however, the double mutant strains are extremely slow-growing and can accumulate suppressor mutations. The *rrp6Δ* strains grow nearly as well as wild-type strains and much faster than *trf4Δ* strains. For ease of manipulation as well as to avoid suppressor mutations, I used *rrp6Δ* strains in subsequent studies whenever possible. I used *trf4Δ* strains to study CUTs that were undetectable in *rrp6Δ* strains. Another consideration with strains that stabilize CUTs is that these mutations are not inert. Genome-wide studies have indicated that deletion of *RRP6* does alter the level of expression as well as the site of transcription termination for a subset of genes (FOX *et al.* 2015). Consistent with these data, I found that deletion of *RRP6* increases levels of *KNH1* mRNA while expression of the other candidate genes seems to be unaffected by this mutation (Figure 6D). In order to ensure that any regulatory effects I observed were not due to the deletion of either *RRP6* or *TRF4*, I have performed the experiments presented here in strains where CUTs are stabilized as well as in strains with wild-type alleles of *RRP6* and *TRF4*.

2.3.2 Disruption of CUT transcription alters expression of downstream protein-coding genes

We initially inserted a transcription termination sequence (TTS) to prematurely stop transcription of these CUTs as this strategy does not require any prior knowledge of promoter sequence elements. The TTS element was integrated at the endogenous locus for each gene using a two-step integration method. The plasmid used for integration (pDW1) contains two copies of 180 base-pairs of the 3' end of the *HIS3* gene flanking a *URA3* cassette, which was made by Danielle Wagner. This strategy was piloted by Robin Monteverde and shown to be effective to terminate transcription of *SRG1* and resulted in *SER3* de-repression. Interestingly, at each of the 5 candidate genes I examined in this way, I saw a change in expression of the downstream protein-coding gene (Figure 6).

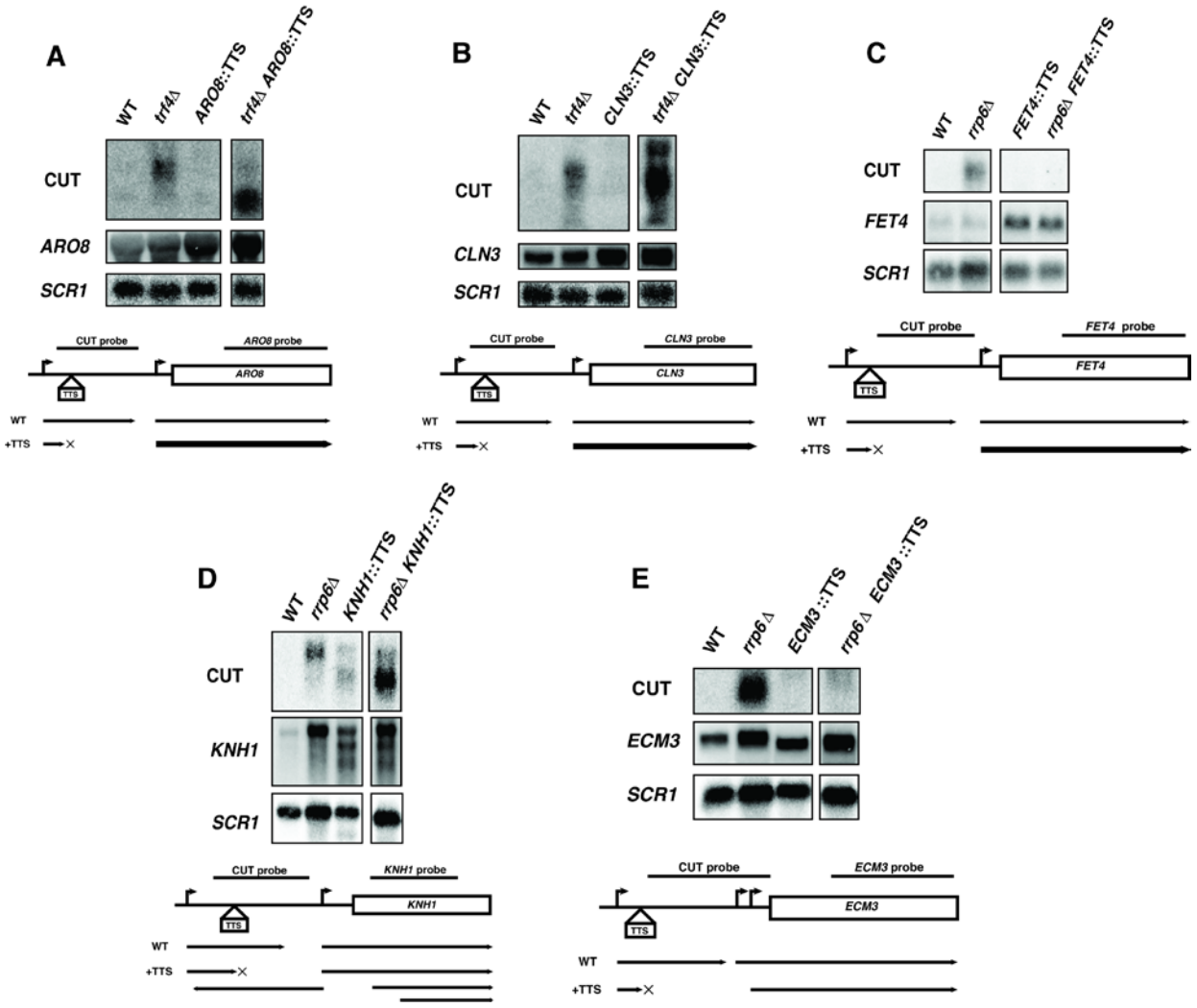

Figure 6. Termination of CUT transcription upstream of candidate protein-coding genes alters expression of the neighboring protein-coding genes.

(A) Northern analysis of *ARO8* and its promoter associated CUT in the presence and absence of the *HIS3* TTS inserted at -300 relative to the *ARO8* open reading frame (+1 ATG) in both wild-type (YJ713 and YEAR071) and *trf4Δ* (KY1976 and YEAR079) strains. (B) Northern analysis of *CLN3* and its promoter associated CUT in the presence and absence of the *HIS3* TTS inserted at -520 relative to the *CLN3* open reading frame (+1 ATG) in both wild-type (YJ713 and YEAR083) and *trf4Δ* (KY1976 and YEAR089) strains. (C) Northern analysis of

FET4 and its promoter associated CUT in the presence and absence of the *HIS3* TTS inserted at -400 relative to the *FET4* open reading frame (+1 ATG) in both wild-type (YEAR030 and YEAR137) and *rrp6Δ* (YJ744 and YEAR138) strains. (D) Northern analysis of *KNH1* and its promoter associated CUT in the presence and absence of the *HIS3* TTS inserted at -400 relative to the *KNH1* open reading frame (+1 ATG) in both wild-type (YJ713 and YEAR038) and *rrp6Δ* (YJ744 and YEAR062) strains. *SCR1* serves as a loading control. (E) Northern analysis of *ECM3* and its promoter associated CUT in the presence and absence of the *HIS3* TTS inserted at -190 relative to the *ECM3* open reading frame (+1 ATG) in both wild-type (YJ713 and YEAR045) and *rrp6Δ* strains (YJ744 and YEAR048). *SCR1* serves as a loading control. Schematics below each panel indicate relative positions of Northern probes as well as *HIS3* TTS insertion positions.

First, we wanted to ensure that the insertion of the *HIS3* TTS sequence had terminated CUT transcription. Supporting this, there is a clear loss of CUT expression upon TTS insertion at the *FET4* and *ECM3* loci in this northern analysis (Figure 6C and 6E). This result is most clear at these loci as the northern probe to detect CUT expression was downstream of where the TTS had been integrated. At other loci, you might expect to observe a shortened CUT isoform if the CUT was prematurely terminated and the northern probe spans the TTS insertion site. This is what I observed at the *ARO8* and *KNH1* loci (Figure 6A and 6D). At the *CLN3* locus, I was not able to confirm that CUT transcription had been prematurely terminated, as a CUT of the same size is abundant in the presence of the TTS (Figure 6B). This may be expected if the TTS were integrated at a site that is normally a termination sequence for CUT transcription. Another complication at *CLN3* is that there are at least three detectable CUT isoforms. We anticipated that we could encounter many technical challenges in attempting to terminate all CUT isoforms at the *CLN3* locus and we may not be able to distinguish separate regulatory effects for the individual isoforms. For these reasons, we decided to focus our attention on other loci.

FET4, *ARO8*, and *CLN3* all appear to be up-regulated upon termination of CUT transcription (Figure 6A-C). With no other obvious phenotypes by northern analysis, these genes appear to be repressed by intergenic transcription over their promoters, potentially in a similar manner to *SER3*, which we investigated further below. Interestingly, *KNH1* and *ECM3* also appeared to be up-regulated upon CUT termination; however, I observed different transcript isoforms in these conditions. The *ECM3* ORF transcript appears to be shifted to a smaller isoform upon CUT termination (Figure 6E). At *KNH1*, CUT termination results in the appearance of three new transcripts (Figure 6D). Single stranded northern blot analysis showed

that one of these transcripts is transcribed in the antisense direction relative to the ORF and is located across the promoter (Figure 7A-B). This transcript could be a product of the *KNH1* promoter initiating bidirectional transcription. The other two previously undetected transcripts are transcribed in the sense direction relative to the ORF and arise from within the coding region of the *KNH1* gene, resembling cryptic transcripts (Figure 7C).

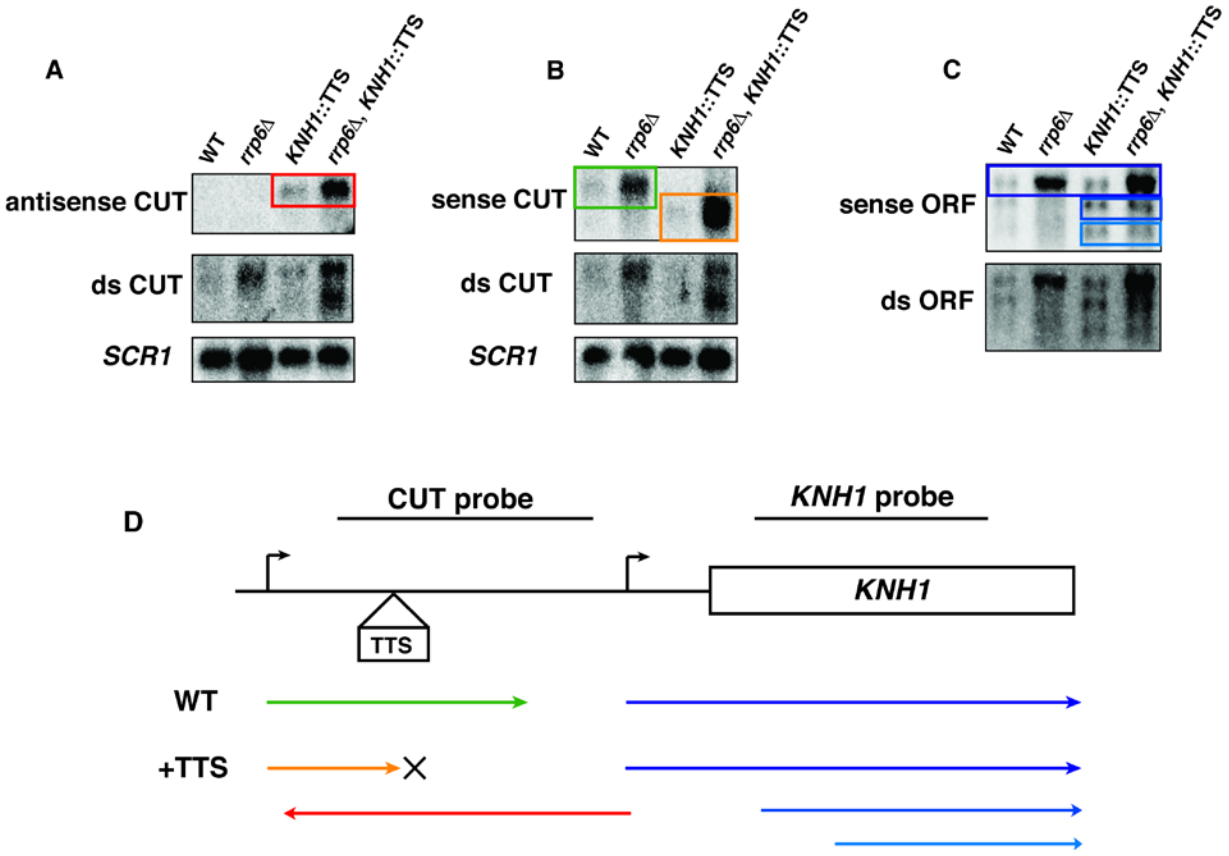

Figure 7. Transcripts arising at the *KNH1* locus upon CUT termination include an antisense transcript over the promoter region and sense transcripts arising from within the *KNH1* ORF.

(A) Northern analysis performed with single-stranded DNA probes detecting antisense transcripts over the *KNH1* promoter region on RNA isolated from control strains (FY4 and YJ744) and strains where the CUT upstream of *KNH1* has been terminated by insertion of the *HIS3* TTS 400 bases upstream of the *KNH1* start codon (YEAR038 and YEAR034). dsCUT indicates a double stranded CUT probe. (B) Northern analysis performed with single-stranded DNA probes detecting sense transcripts over the *KNH1* promoter region using strains described

in (A). dsCUT indicates a double stranded CUT probe. (C) Northern analysis performed with single-stranded DNA probes detecting sense transcripts over the *KNH1* open reading frame using strains described in (A). dsORF indicates a double stranded open reading frame probe. *SCR1* serves as a loading control. (D) Model for relative positions of transcript isoforms at the *KNH1* locus based on data in panels A-C. Positions of Northern probes and the *HIS3* TTS integration site are indicated. The arrows indicate the directionality and relative position of transcripts observed at the *KNH1* locus. The color of the transcript isoforms in this schematic correspond to the boxed northern analyses in panels A-C that support this model for their relative positions at the *KNH1* locus.

We investigated the apparent change in size of the *ECM3* ORF transcript upon CUT termination by higher resolution northern blot analysis. Standard northern blot analysis in our lab involves running RNA samples in an agarose gel for 500 Volt hours, which separates total RNA over about 10 centimeters of gel. To further separate higher molecular weight RNA molecules, I ran RNA samples in agarose gels for 2200 Volt hours over twice the length of gel. This resulted in gels where smaller molecular weight RNAs were run off the end of the gel and higher molecular weight RNAs were separated over about 20 centimeters of gel. This higher resolution northern analysis revealed that *ECM3* has two RNA isoforms. These two isoforms appeared to be expressed at equal levels in wild-type and *rrp6Δ* strains; however, upon CUT termination, only the short isoform is produced (Figure 8). We hypothesized that this could be due differential transcription start site selection.

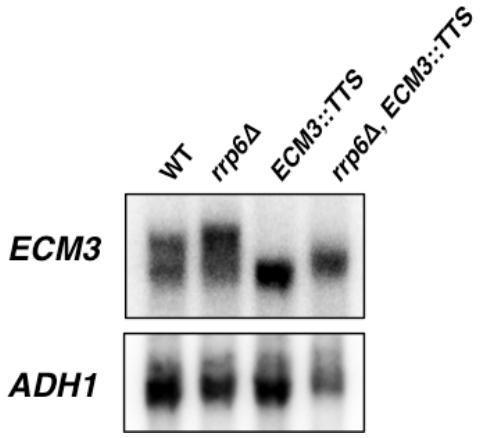

Figure 8. Termination of the CUT upstream of *ECM3* alters *ECM3* expression in an isoform-specific manner.

Representative high resolution northern analysis of *ECM3* expression on RNA isolated from control strains (WT, YJ713; *rrp6Δ*, YJ744) and strains where the CUT upstream of *ECM3* has been terminated by insertion of the *HIS3* TTS at 190 base pairs upstream of the *ECM3* start codon (*ECM3::TTS*, YEAR045; *rrp6Δ ECM3::TTS*, YEAR048). *ADH1* serves as a loading control.

We explored the potential for differential start site selection by mapping the 5' ends of the *ECM3* ORF transcripts by primer extension analysis. I successfully mapped two prominent start sites for *ECM3*, one at 80 base pairs upstream of the *ECM3* start codon and the other 190 base pairs upstream of the *ECM3* start codon (Figure 9). I had integrated the TTS at 190 bases upstream of the start codon, which led us to conclude that the differences in start site selection we observed may be an artifact of having integrated the TTS directly at one of these start sites. I made several attempts to integrate the TTS at different locations and these new mutations were all unsuccessful. Some did not effectively terminate CUT transcription while others produced non-specific changes in isoforms as these changes were also observed when the TTS was integrated in the antisense orientation as a control. This led us to abandon the strategy of disrupting CUT transcription by insertion of a premature TTS at the *ECM3* locus. As discussed in Chapter Three, we were successful with an alternative approach to disrupt the *ECM3* CUT promoter.

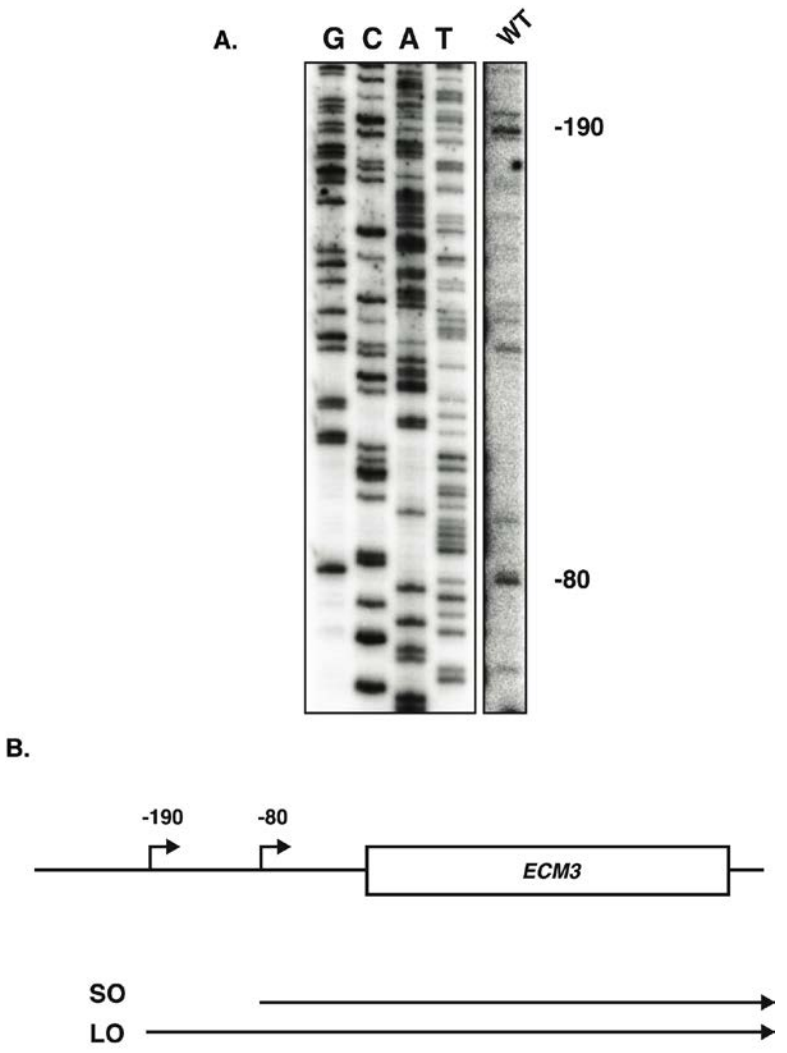

Figure 9. Identification of two major transcription start sites for *ECM3*.

(A) Primer extension analysis was performed using RNA isolated from a wild-type strain (FY5). The oligonucleotides used to map these start sites are listed in Table 2. These start sites were identified using two different oligonucleotides and RNA isolated from more than one strain. (B) Diagram of the results shown in (A). *ECM3* shows at least two major start sites initiating 80 bases and 190 bases upstream of the +1 ATG of the *ECM3* ORF. The isoform initiating at -80 is referred to as the short ORF (SO) isoform and the isoform initiating at -190 is referred to as the long ORF (LO) isoform.

2.3.3 Mapping of CUT 5' ends by RACE and primer extension analysis

Concurrently with experiments integrating the TTS to stop transcription of candidate CUTs, I also mapped putative transcription start sites (TSSs) for each CUT by 5' rapid amplification of cDNA ends (RACE) analysis with the thought that these TSS data would be informative for further experiments. I performed RACE analysis to map 5' ends for the CUTs upstream of *ARO2*, *ECM3*, *FET4*, and *KNH1*. The 5' ends mapped by RACE analysis are summarized graphically in Figure 10.

The 5' RACE data are difficult to interpret due to limitations of the assay. Many sequence reads were non-specific, the majority of which corresponded to rRNA transcripts. For this reason, I would have depleted my RNA samples of rRNA if I had continued 5' RACE analysis. Aside from this technical difficulty, the assay requires a set of plasmids to be sequenced for each transcript to be mapped. Each of these plasmids presumably contains cDNA from a single RNA molecule ligated to an adapter sequence. Sequencing these plasmids generates a library of TSSs for each transcript. A challenge of this technique is determining an appropriate number of plasmids to be sequenced to successfully map the 5' ends of all possible isoforms with sufficient reproducibility to confidently consider it a verified start site. Furthermore, this technique involves synthesis of cDNA and two amplifications in a nested PCR reaction, each of which can introduce error. These issues led us to use primer extension as an alternative approach in later experiments.

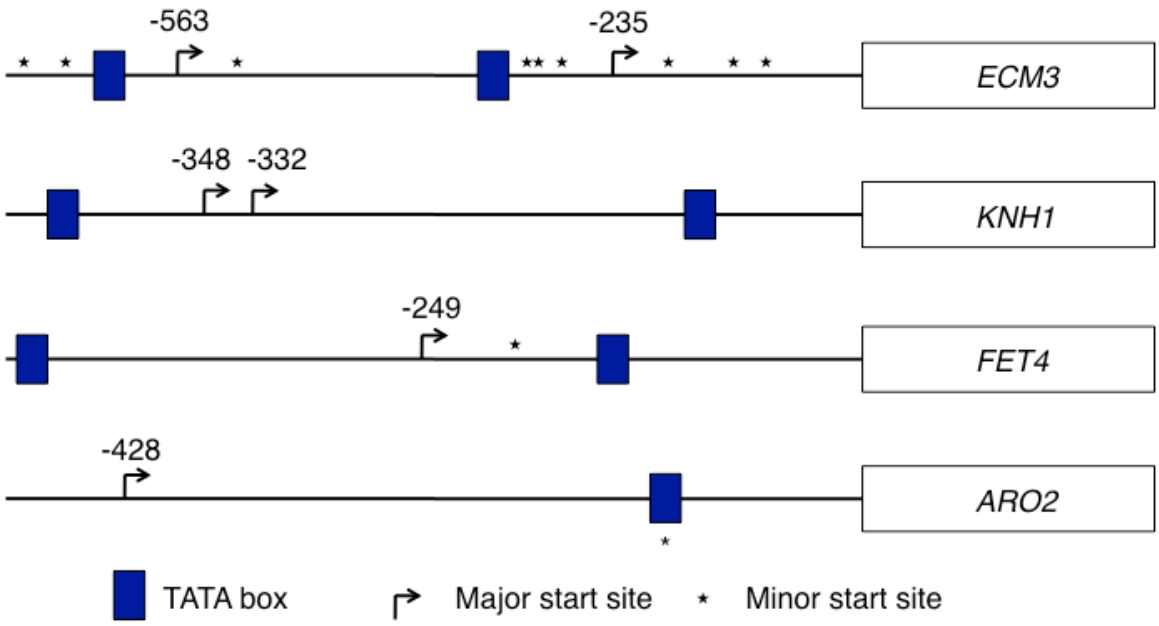

Figure 10. Graphical summary of putative transcription start sites of CUT transcripts upstream of *ECM3, KNH1, FET4,* and *ARO2*.

Transcription start sites for CUTs upstream of *ECM3*, *KNH1*, *FET4*, and *ARO2* were mapped by 5' RACE analysis of cDNA synthesized from RNA isolated from an *rrp6Δ* strain (YJ744). Putative TATA sequences identified by multiple sequence alignment of four related yeast species (*S. cerevisiae, S. mikatae*, *S. bayanus*, and *S. paradoxus*) are shown as blue boxes. An arrow marks major start sites. Major start sites are numbered in bases upstream of the +1 start codon of each ORF. An asterisk marks minor start sites.

2.3.4 Investigation of regulation of candidate genes by transcription-coupled nucleosome occupancy

The *ARO8*, *CLN3*, and *FET4* genes appeared to be simply upregulated upon CUT termination, resembling *SER3* de-repression upon *SRG1* termination. To determine if these three genes might utilize the same mechanism of repression by intergenic transcription as observed at the *SER3* locus, I performed northern blot analysis of these genes in strains carrying an *spt16-E857K* or *H3 K122A* mutation as these mutations have a very strong defect in transcription-coupled nucleosome occupancy and result in very strong de-repression of *SER3* (HAINER AND MARTENS 2011a; HAINER *et al.* 2012). These analyses show that although the expression of these candidates may increase slightly in these strains (Figure 11A and 11B), this does not account for the level of de-repression observed upon CUT termination (Figure 11C). This suggests that these three genes are repressed via intergenic transcription by a mechanism that is distinct from the mechanism observed at *SER3*. It is interesting to note that *FET4* expression appears to increase in strains where one copy of the H3 and H4 genes has been deleted (Figure 11, lanes 1 and 3). This suggests that *FET4* expression is sensitive to histone dosage and this could be interesting for future studies of *FET4* regulation.

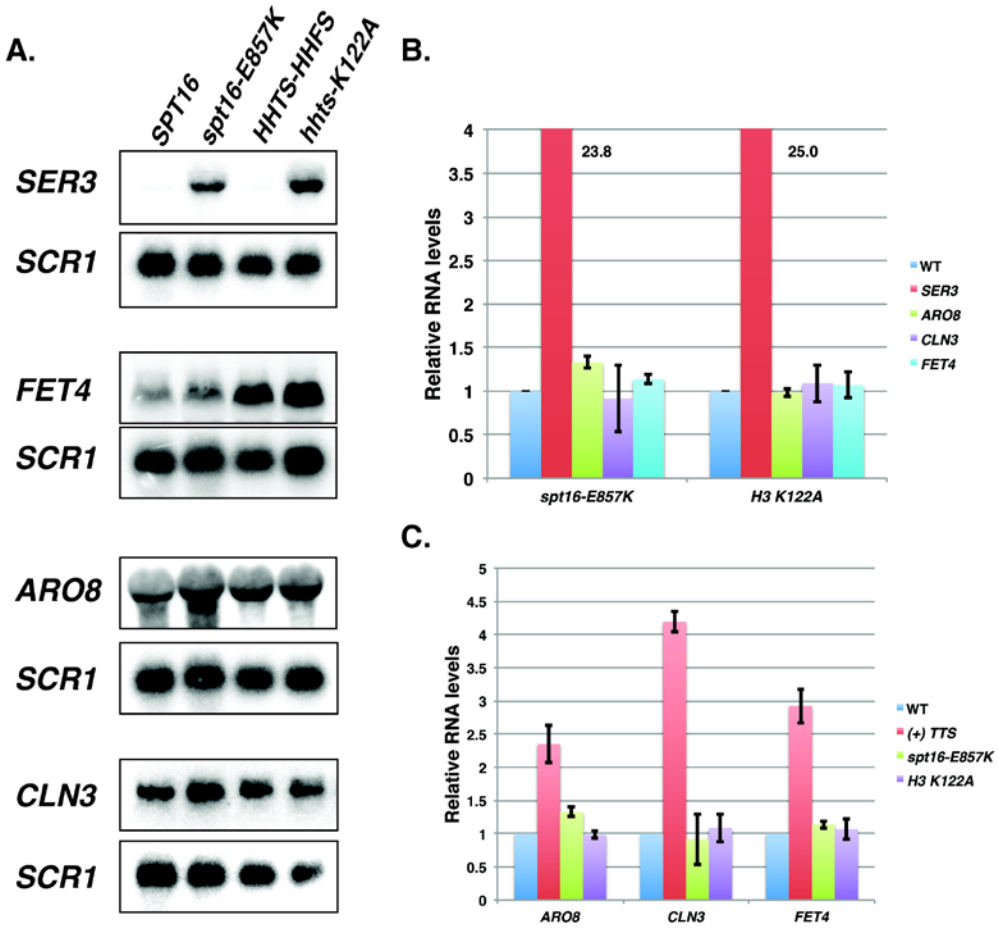

Figure 11. Mutations in a gene encoding a member of the FACT complex or H3 that decrease transcription coupled nucleosome occupancy do not account for the level of de-repression observed by CUT termination upstream of *ARO8, CLN3*, and *FET4*.

(A) Representative northern analysis of *SER3*, *FET4*, *ARO8*, and *CLN3* expression on RNA isolated from wild-type strains (*SPT16*, YJ1091 <p*SPT16-LEU2*>; *HHTS-HHFS*, YS417, YS418, YS419) and strains where transcription-coupled nucleosome occupancy is decreased by mutation of the FACT histone chaperone, *spt16-E857K* (YJ1091 <p*spt16-E857K*>), or by mutation of histone H3 K122A (YS404, YS420, YS421). *SCR1* serves as a loading control. (B) Quantitation of northern analyses shown in (A) comparing the level of de-repression of *FET4*, *ARO8*, and *CLN3* compared to *SER3* in *spt16-E857K* and H3 K122A strains (C)

Quantitation of northern analysis shown in (A) comparing the level of de-repression of *FET4*, *ARO8*, and *CLN3* compared to the level of de-repression observed by termination of CUT transcription upstream of each gene (shown in Figure 6). Values show the average of three biological replicates where wild-type expression of each gene was set to 1. RNA levels for each transcript are normalized to the respective control (lane 2 to lane 1 and lane 4 to lane 3). Error bars represent the standard error of the mean.

2.3.5 Investigation of CUT transcription and known transcription factor regulation of *FET4*

The Rox1 transcription factor represses expression of *FET4* in aerobic conditions (JENSEN AND CULOTTA 2002; WATERS AND EIDE 2002). With this knowledge, we were interested to explore how the Rox1 transcription factor and CUT transcription across this promoter may interplay to regulate expression of *FET4*. To analyze this, I deleted the *ROX1* gene in our strain background by PCR amplification of the *rox1Δ::KanMX* allele from the deletion collection, which I then integrated into wild-type and *rrp6Δ* strains. One hypothesis that we had for this analysis was that perhaps Rox1 repressed expression of *FET4* by activating CUT transcription across its promoter. Northern analysis of these strains revealed that the de-repression of *FET4* in the absence of Rox1 is dependent on having a wild-type *RRP6* allele (Figure 12). This suggests that the CUT RNA product itself has a role in *FET4* regulation. Alternatively, termination of the *FET4* CUT may be altered in the absence of Rrp6 and perhaps this plays a role in preventing normal regulation by Rox1. We were specifically interested in characterizing mechanisms of regulation by intergenic transcription that were dependent on events associated with the act of transcription rather than the ncRNA product or defects associated with mutations that would not be present in environmental regulation of the gene, and for this reason, we did not continue investigating *FET4* regulation.

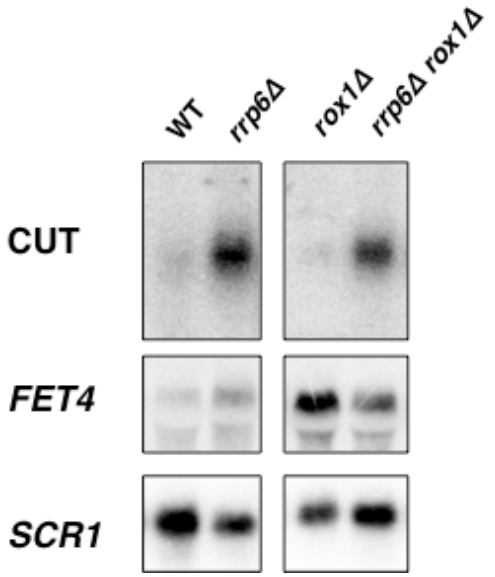

Figure 12. De-repression of *FET4* expression in the absence of the repressive transcription factor, Rox1, is dependent on *RRP6*.

Representative northern analysis of strains lacking the Rox1 repressive transcription factor *(rox1Δ*, YEAR123; *rrp6Δ rox1Δ*, YEAR141) compared to wild-type (YJ1125) and *rrp6Δ* (YJ744) strains. *SCR1* serves as a loading control.

2.4 DISCUSSION

The interesting findings of gene regulation by intergenic transcription at the *SER3* locus led us to explore the potential for intergenic transcription to regulate other loci. We took a candidate gene approach for this aim. We first selected genes expressing CUTs over their promoters in a tandem orientation, resembling the structure of the *SER3* locus. We then integrated a TTS to prematurely stop transcription of the CUTs over five of the candidate gene promoters. Interestingly, each of these cases resulted in a change in expression of the downstream protein-coding gene. The results discussed in this chapter should be considered with the caveat that I cannot exclude the possibility that insertion of the TTS sequence may have other effects in addition to termination of CUT transcription. It is perhaps not surprising that this mutational strategy was not ideal for the *ECM3* locus as this locus is very compact and the TTS sequence introduces more than a nucleosome's worth of DNA. Importantly, the regulatory effects I have observed here could be different in physiological conditions or in the context of more targeted disruption of CUT transcription by other strategies.

I investigated several known mechanisms involved in repressing cryptic transcription, such as histone chaperones, Spt6 and Spt16 (BELOTSERKOVSKAYA *et al.* 2003; KAPLAN *et al.* 2003; MASON AND STRUHL 2003; CARROZZA *et al.* 2005; CHU *et al.* 2007; KIM *et al.* 2012); however, these factors do not appear to be necessary for repression of the cryptic transcripts that I observed at *KNH1*. Therefore, the *KNH1* gene provides an interesting model for elucidation of a novel pathway to repress cryptic transcripts. This locus is particularly interesting, as CUT

transcription across the *KNH1* promoter also appears to influence *KNH1* promoter directionality. Although this has been questioned recently, several lines of evidence indicate that promoters are inherently bidirectional (NEIL *et al.* 2009; XU *et al.* 2009). In the case of *KNH1*, it would be interesting to study how CUT transcription may promote sense transcription of *KNH1*.

At *FET4*, the observation that normal repression by Rox1 requires a functional *RRP6* allele presents the possibility that stabilization of a CUT may impact regulation of a neighboring protein-coding gene. It is possible that this could occur under relevant biological conditions. Alternatively, loss of Rrp6 may alter where CUT termination occurs (FOX *et al.* 2015), and at the *FET4* locus, the effect I have observed may be due to abnormal transcriptional read-through of the upstream CUT. Although these hypotheses would not explain the de-repression of *FET4* that I have observed upon CUT termination, they would provide an interesting model to study how the nuclear exosome regulates gene expression.

The isoform-specific regulation of *ECM3* upon CUT termination was particularly intriguing. Although this turned out to be due to a technical difficulty with insertion of a large exogenous DNA sequence into a compact locus, we were enticed by the potential for differential regulation of specific *ECM3* isoforms. This led us to further analyze the effect of intergenic transcription on *ECM3* regulation as discussed in Chapter Three.

3.0 CHAPTER THREE: EVIDENCE FOR POSITIVE REGULATION OF THE *ECM3* GENE BY INTERGENIC TRANSCRIPTION

3.1 INTRODUCTION

As discussed in Chapter Two, initial analyses of candidate genes expressing CUTs over their promoters revealed several interesting hypotheses to explain gene regulation by intergenic transcription. Intrigued by the potential for isoform-specific regulation by intergenic transcription at *ECM3*, the goal of the work presented in Chapter Three was to test if transcription of the CUT across the *ECM3* promoter regulates expression of *ECM3*. Disrupting intergenic transcription by mutating the promoter of this CUT revealed a positive correlation between CUT transcription and *ECM3* expression. This chapter also summarizes data investigating whether CUT transcription regulates *ECM3* expression in physiological conditions where cells are naturally responding to changes in their environment.

Prior to the present study, very little was known about the *ECM3* gene. *ECM3* was first identified in a screen for sensitivity to the general cell wall stressor, calcofluor white (LUSSIER *et al.* 1997). For this reason, *ECM3* is thought to be involved in cell wall maintenance. *ECM3* is lowly expressed in rich medium at 30°C and induced in response to genotoxic stress (DARDALHON *et al.* 2007). Computational analysis predicts the Ecm3 protein to be a transmembrane protein localized to the endoplasmic reticulum or the Golgi.

3.2 MATERIALS AND METHODS

3.2.1 *S. cerevisiae* strains and media

S. cerevisiae strains used in this study are listed in Table 4. All strains used in Chapter Three are derived from a *GAL2*$^{+}$ S288C isolate using standard genetic crosses and transformations (WINSTON *et al.* 1995). The *EUC1* promoter deletions were made by two-step integration of an HA-*URA3*-HA cassette that was PCR-amplified from the plasmid pMPY-3XHA (SCHNEIDER *et al.* 1995). This resulted in strains where a portion of the *EUC1* promoter has been replaced with a DNA sequence encoding one copy of the 3XHA tag, which is serving as spacer DNA. Cells were grown at 30°C in YPD medium (1% yeast extract, 2% peptone, 2% dextrose) until cultures reached a density of 1 X 10^7 to 2 X 10^7 cells per mL for isolation of RNA for use in northern blotting analyses.

Strain	Genotype	Reference or Source
FY4	*MATa*	Winston, *et al*., 1995
FY5	*MATα*	Winston, *et al*., 1995
YJ744	*MATa rrp6Δ0::KanMX*	J. Pruneski and J. Martens, Tomson *et al*., 2013
YJ746	*MATα rrp6Δ0::KanMX*	J. Pruneski and J. Martens, Tomson *et al*., 2013
YJ1125	*MATa ura3Δ0 his3Δ200 leu2Δ0*	This study
YJ1126	*MATa rrp6Δ0::KanMX ura3Δ0 his3Δ200 leu2Δ0*	This study
YJ1127	*MATα rrp6Δ0::KanMX ura3Δ0 his3Δ200 leu2Δ0 ecm3-pEUC1Δ1::HA (-400 to -350)*	This study
YJ1128	*MATa rrp6Δ0::KanMX ura3Δ0 ecm3-pEUC1Δ1::HA (-400 to -350)*	This study
YJ1129	*MATα rrp6Δ0::KanMX ura3Δ0 ecm3-pEUC1Δ1::HA (-400 to -350)*	This study
YJ1130	*MATα rrp6Δ0::KanMX ura3Δ0 leu2Δ0 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1131	*MATa rrp6Δ0::KanMX ura3Δ0 his3Δ200 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1132	*MATα rrp6Δ0::KanMX ura3Δ0 leu2Δ0 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1133	*MATa ura3Δ0 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1134	*MATα ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1135	*MATa leu2Δ0 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YEAR217	*MATa ura3Δ0 his3Δ200 leu2Δ0 ecm3-1 (TATA sequence at -344 mutated to AvrII site)*	This study
YEAR218	*MATa ura3Δ0 his3Δ200 ecm3-1*	This study
YEAR219	*MATa ura3Δ0 his3Δ200 ecm3-1*	This study
YEAR221	*MATa rrp6Δ0::kanMX, ura3Δ0, his3Δ200, leu2Δ0, ecm3-1*	This study
YEAR222	*MATα rrp6Δ0::kanMX, ura3Δ0, ecm3-1*	This study
YEAR223	*MATα rrp6Δ0::kanMX, ura3Δ0, ecm3-1*	This study

Table 4. *S. cerevisiae* strains used in Chapter Three.

3.2.2 Northern blot analysis

Northern blot analyses were performed using 20 μg total RNA samples resolved in gels containing 2% agarose, 6.5% formaldehyde, and 1X MOPS as previously described (AUSUBEL 1987). Double-stranded probes were generated by random-primed labeling (RIO 2011). Probe templates were amplified from genomic DNA to contain the following sequences relative to the +1 ATG of the protein-coding gene at each locus: *EUC1* (-541 to -100), *ECM3* (+545 to +976), and *SCR1* (-182 to +284). *SCR1* RNA levels serve as an internal loading control. Images were generated by phosphorimaging and quantified using ImageJ software.

Purpose	Primer set	Oligo name	Orientation	Location relative to +1 ATG	Sequence
Generation of *pEUC1Δ1*		OJ1722	F	-400	5'-GGATTAAAAGTGCTCGAGTTTTGCTCTCTATTTCACAATCagggaacaaaagctgg
		OJ1723	R	-350	5'-GCGTCAAACATATATTCAGGCCCTCTAAGTATATAGCCTCctatagggcgaattgg
Generation of *pEUC1Δ2*		OJ1722	F	-400	5'-GGATTAAAAGTGCTCGAGTTTTGCTCTCTATTTCACAATCagggaacaaaagctgg
		OJ1724	R	-300	5'-CATATAATAACGGTGCAAGAGTAACAGATGGTAGCGTACGctatagggcgaattgg
Northern probe template	***EUC1***	OJ1172	F	-541	5'- CCATGCTTATCTGCCGTCTT
		OJ1173	R	-100	5'-GGTAATGGTCAACAATACGC
Northern probe template	***ECM3***	OJ1174	F	+545	5'- TGACCAATGATGATTCTGCCC
		OJ1175	R	+976	5'- GTAGTTCACGCATATCGATGG
Northern probe template	***SCR1***	OJ459	F	-182	5'-CAACTTAGCCAGGACATCCA
		OJ460	R	+284	5'-AGAGAGACGGATTCCTCACG

Table 5. Oligonucleotides used in Chapter Three.

3.3 RESULTS

3.3.1 Intergenic transcription positively correlates with *ECM3* expression

As discussed in Chapter Two, *ECM3* presented an interesting candidate as termination of CUT transcription appeared to have isoform specific regulatory effects on *ECM3* expression. For subsequent experiments, we termed this CUT *EUC1* for *ECM3* upstream CUT. Premature termination of *EUC1* transcription showed effects on *ECM3* start site selection that were not specific to the termination as start site selection was also affected when the termination sequence was integrated in the antisense direction. Consequently, we turned our attention to disrupting *EUC1* transcription by disrupting the *EUC1* promoter. To do this, we needed to first identify putative promoter sequence elements for *EUC1*. To search for potentially conserved TATA box sequences, I performed a sequence alignment of four related yeast species (*S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. bayanus*). This sequence alignment identified a conserved putative TATA sequence 344 base pairs upstream of the *ECM3* start codon (Figure 14). This putative TATA sequence was near the location where we estimated the promoter might be based on the size of the *EUC1* transcript in northern blot analyses as well as the location of *EUC1* expression detected in genome-wide tiling microarray expression data (NEIL *et al.* 2009). A site of pre-initiation complex (PIC) assembly was also detected nearby at 382 bases upstream of the *ECM3* start codon in genome-wide ChIP-exo analyses (RHEE AND PUGH 2012). The sequence at the detected PIC assembly site did not resemble a traditional TATA sequence, and was not an A/T rich sequence. Therefore, we directed a targeted mutagenesis to the nearby putative TATA sequence. I mutated the putative TATA sequence to CCTAGG, the sequence recognized by the AvrII restriction enzyme to aid in identification of successful

mutagenesis, by site directed mutagenesis. I then integrated this mutation at the endogenous *EUC1* promoter. This mutation had no effect on *EUC1* transcription (Figure 13). As this locus is very A/T rich, we thought that it may be necessary to use multiple directed mutations to eliminate PIC assembly at the *EUC1* promoter and turned to promoter deletion mutations as an alternative approach.

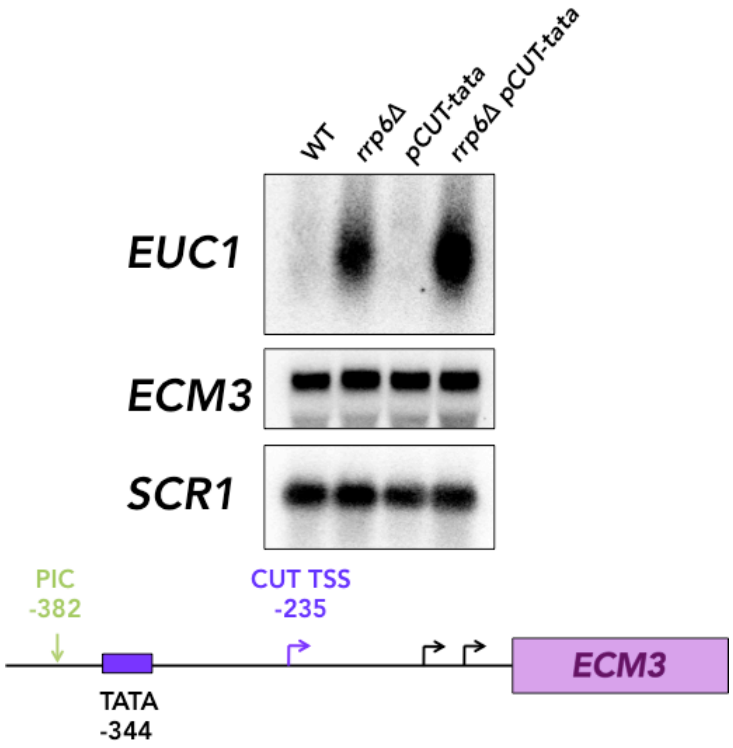

Figure 13. Mutation of a putative *EUC1* TATA element does not disrupt *EUC1* transcription.

Representative northern analysis of strains where a putative TATA sequence for *EUC1* was mutated to the AvrII restriction site sequence (*pCUT-tata*, YEAR219; *rrp6Δ pCUT-tata*, YEAR223) compared to wild-type (YJ1125) and *rrp6Δ* (YJ1126) strains. *SCR1* serves as a loading control. Schematic below indicates the relative positions of a detected site of PIC assembly (RHEE AND PUGH 2012), the putative TATA sequence that was mutated, and a putative transcription start site for *EUC1* mapped by 5' RACE analysis (Figure 10).

FIGURE S2

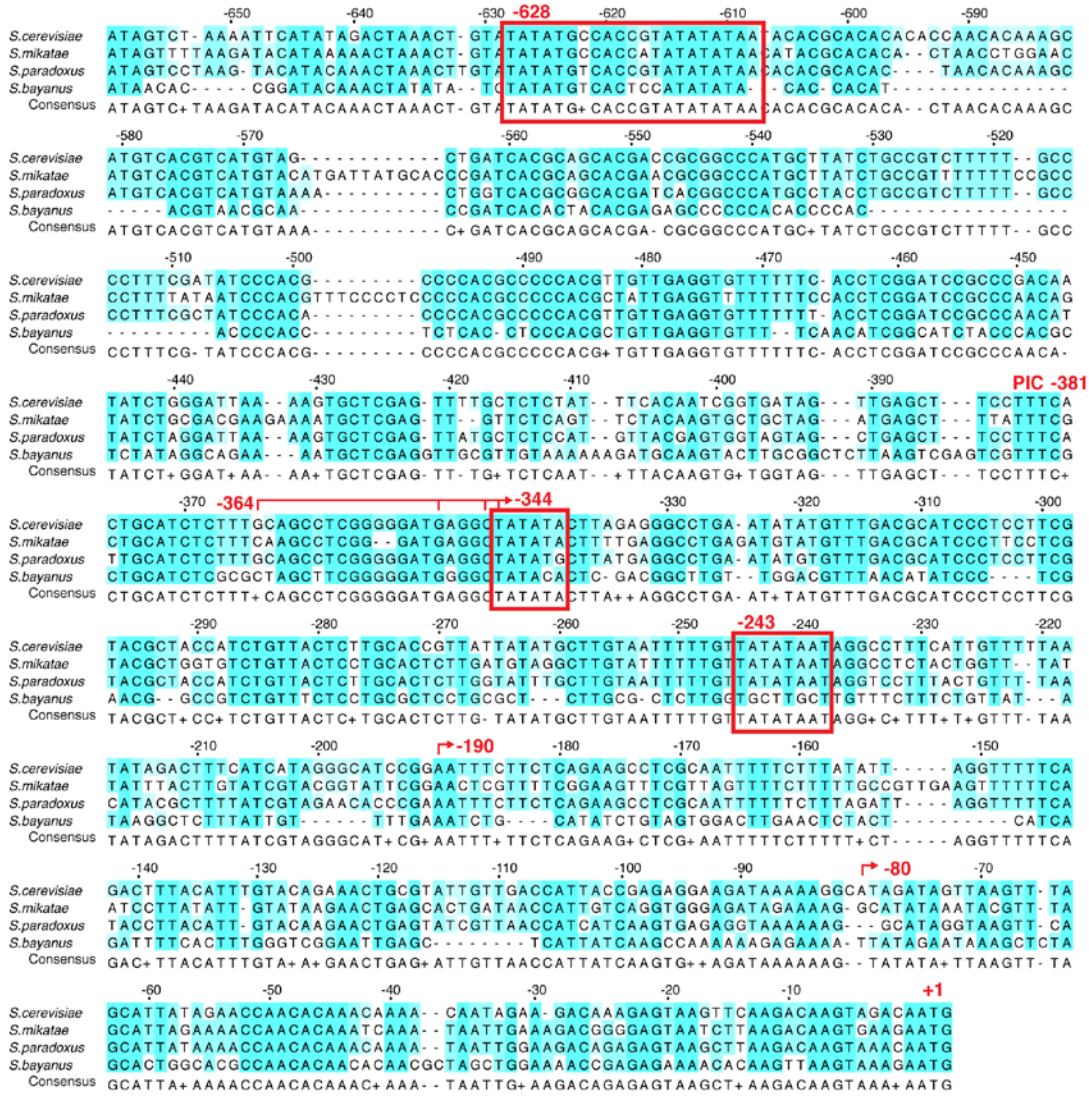

Figure 14. Multiple sequence alignment of the intergenic region upstream of the *ECM3* ORF in four related yeast species.

Sequences in this alignment were obtained from the *Saccharomyces cerevisiae* genome database. Alignment of these sequences was performed using Clustal X software and this image was generated using Jalview and Adobe Illustrator software. The alignment was performed including sequence from the upstream tRNA gene extending into the *ECM3* ORF. The first three nucleotides in this image are the 3' end of the upstream tRNA gene. The last

three nucleotides in this image are the +1 start codon of the *ECM3* ORF. Darker cyan indicates a higher degree of sequence identity. The locations of the putative TATA sequences are highlighted in red boxes. The location of the PIC identified by Rhee and Pugh (2012) is indicated above the sequence. The locations of transcription start sites for both *EUC1* and *ECM3* as mapped by primer extension (Figures 9 and 23) are shown as red arrows.

I designed two promoter deletion mutations that I integrated at the putative *EUC1* promoter. The deleted sequences are replaced with the 3XHA tag sequence, which serves as a DNA spacer. Deletion 1 replaced 50 base pairs (bps) (*pEUC1Δ1*, -400 to -350, relative to +1 start codon of *ECM3*) and deletion 2 replaced 100 bps (*pEUC1Δ2*, -400 to -300) at the predicted *EUC1* promoter. Both of these deletions result in dramatically reduced *EUC1* levels and *ECM3* levels (Figure 15). This suggests that *EUC1* transcription has a positive role in *ECM3* expression. Another plausible explanation for these data is that the deleted sequences may contain a positive regulatory element for the *ECM3* promoter. Although we are currently unable to distinguish between these models, other data presented in Chapters 3 and 4 suggest that regulation by intergenic transcription may be the more likely model.

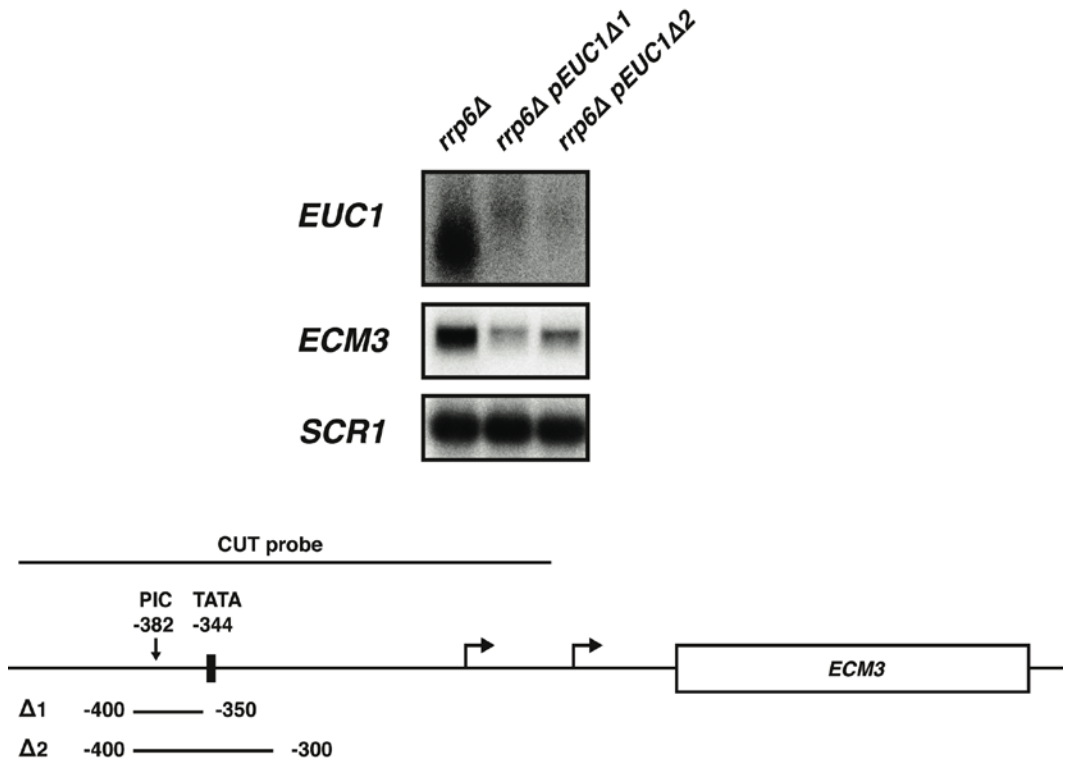

Figure 15. Deletion of a putative *EUC1* promoter upstream of *ECM3* results in reduced *ECM3* expression.

Representative northern blot analysis of RNA isolated from strains carrying an *RRP6* deletion and containing either a wild type *ECM3* locus (YJ1126) or the indicated *EUC1* promoter deletion mutations. Promoter deletion mutations were introduced at the endogenous *ECM3* locus and replaced either 50 bp (*pEUC1*Δ*1*; region -400 to -350 deleted; YJ1128) or 100 bp (*pEUC1*Δ*2*; region -400 to -300 deleted; YJ1131) upstream of the +1 ATG of *ECM3*, as diagrammed below. The locations of a pre-initiation complex (PIC) identified by Rhee and Pugh (2012) and a putative TATA sequence are indicated on the diagram below. *SCR1* serves as a loading control.

3.3.2 Investigation of physiological regulation of *ECM3*

Although we had identified a positive correlation between *EUC1* and *ECM3* expression using a mutational approach, we wanted to determine if *ECM3* is regulated by *EUC1* transcription in response to environmental stimuli. To do this, I performed northern blot analysis of *EUC1* and *ECM3* expression under various growth conditions. I did not observe any significant changes in *EUC1* or *ECM3* expression when cells were grown in rich medium containing either ethanol or galactose as a carbon source compared to glucose in steady state conditions. The expression of many genes in response to environmental conditions can be transient. To determine if there was a transient change in *ECM3* expression, I performed time course experiments taking samples for RNA isolation at timed intervals after shifting cells from one growth condition to another.

One way to stress the *S. cerevisiae* cell well is to expose cells to heat shock. As Ecm3 is predicted to be involved in cell wall maintenance, I first analyzed *EUC1* and *ECM3* expression when shifting cells from 30°C (normal growth temperature for *S. cerevisiae*) to 37°C (heat shock temperature for *S. cerevisiae*). This analysis was done using exponentially growing cultures in rich medium containing glucose under aerobic conditions. For each experiment, a log phase culture was grown at 30°C until cells reached a density of approximately 2×10^7 cells per milliliter. This culture was then divided in half. One half was diluted 1:1 with pre-warmed media at 30°C and continued growing at 30°C. The other half was diluted 1:1 with media pre-warmed to 42°C and continued growing at 37°C. I then removed samples of each culture at timed intervals for RNA isolation and analysis by northern blotting. Initially after shifting cells from 30°C to 37°C, I observed a decrease in *ECM3* expression that correlates with a decrease in

levels of the *EUC1* isoform detected in *rrp6Δ* cells and an increase in a larger isoform of *EUC1* (Figure 16). I will refer to the smaller *EUC1* isoform as "SC" for short CUT and the larger *EUC1* isoform as "LC" for long CUT. The LC isoform was detected using a strand-specific northern probe to detect sense transcripts, indicating that like the SC, the LC isoform of *EUC1* is transcribed in the sense direction relative to *ECM3*. Gradually over the time course, *ECM3* levels rise with levels of the SC *EUC1* isoform as levels of the LC *EUC1* isoform decrease. I have observed this trend in four separate experiments. Although this trend is repeatable, the magnitude of the effect is subtle. Furthermore, there is variability in the exact timing of when the changes in expression begin, making these data difficult to quantify.

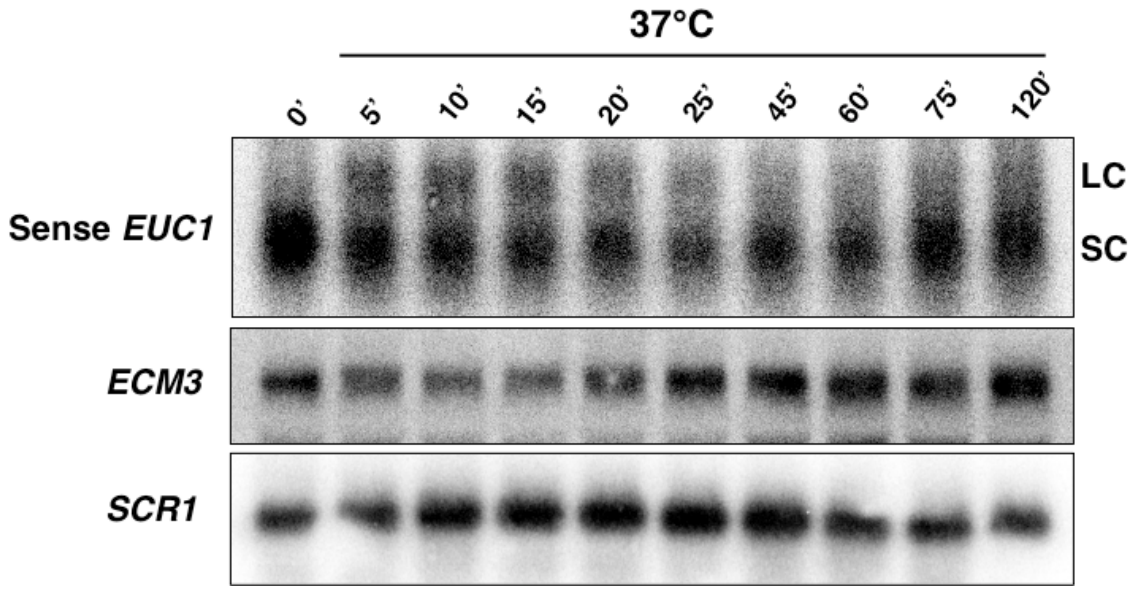

Figure 16. *ECM3* expression decreases immediately after exposure to heat shock and subsequently increases over time.

Representative northern blot analysis of *EUC1* and *ECM3* expression on RNA isolated from an *rrp6Δ* strain (YJ1126) comparing cells grown at 30°C (0') to cells grown at 37°C for the indicated lengths of time. Each sample used in this analysis was taken from a single log-phase culture grown at 30°C and following its progression after shifting to 37°C. *EUC1* was probed with a strand-specific probe designed to detect sense *EUC1* transcripts relative to the *ECM3* ORF. SC indicates the short CUT *EUC1* isoform. LC indicates the long CUT *EUC1* isoform. *SCR1* serves as a loading control.

With the hope of finding a more dramatic change in *ECM3* expression, I explored a change in carbon source as another physiological condition. Modeling my experiment after microarray expression results published by the Buratowski lab (KIM *et al.* 2012), I analyzed *EUC1* and *ECM3* expression when shifting cells from media containing raffinose to media containing galactose as a carbon source. Under these conditions, the Buratowski lab observed a decrease in *ECM3* expression and an increase in *EUC1* transcription, although in these microarray data, the SC and LC isoforms of *EUC1* are indistinguishable. In my northern blot analyses, I observed a similar trend when shifting cells from raffinose to galactose as I observed when shifting cells from 30°C to 37°C. This experiment was only performed once, but contrary to the results from the Buratowski lab, it appears that *ECM3* levels increase slightly in my experiment. Similar to the heat shock conditions, I observe a decrease in levels of the SC *EUC1* isoform at early time points corresponding with an increase in levels of the LC *EUC1* isoform (Figure 17). Although I did not follow up on this result, as observed in heat shock conditions, changes in carbon source appear to have subtle, transient effects on *EUC1* and *ECM3* expression.

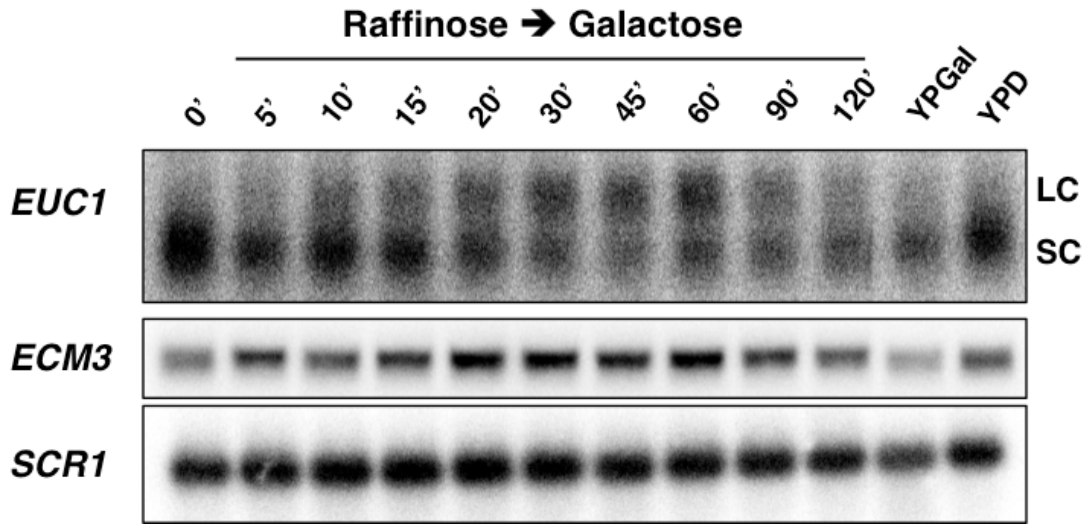

Figure 17. *EUC1* and *ECM3* expression changes in response to galactose after growth in raffinose containing media.

Northern blot analysis of *EUC1* and *ECM3* expression on RNA isolated from an *rrp6Δ* strain (YJ1126) comparing cells grown in media containing raffinose (0') compared to cells grown after addition of galactose to this media for the indicated lengths of time. Each sample used in this analysis was taken from a single log-phase culture grown in raffinose containing media and following its progression after addition of galactose to the media. SC indicates the short CUT *EUC1* isoform. LC indicates the long CUT *EUC1* isoform. *SCR1* serves as a loading control.

As *ecm3* mutant cells are sensitive to the cell wall stressor, calcofluor white (CFW), we reasoned that *ECM3* expression might be induced in the presence of CFW since the Ecm3 protein is beneficial to cell growth in this condition (LUSSIER *et al.* 1997). To test this, I analyzed *EUC1* and *ECM3* expression in the presence of CFW by northern blot analysis. Because the changes in *ECM3* expression might be transient, I performed this experiment as a time course, isolating RNA from samples taken at timed intervals after addition of CFW to a log-phase culture. Contrary to our prediction, *ECM3* levels are lower in the presence of CFW with a corresponding decrease in *EUC1* levels (Figure 18). These results should be considered with the caveats that this experiment was only performed once and an equivalent amount of ethanol used to dissolve the CFW was not added to the negative control culture. In addition to these caveats, CFW is not standardly used in liquid media and this procedure may require optimization of the experimental conditions. For these reasons, it is difficult to draw conclusions from these data. If these data are reflective of *ECM3* expression in response to CFW, the lack of increase in *ECM3* expression in the presence of CFW may indicate that steady state levels of Ecm3 protein are sufficient for the function of Ecm3 under cell wall stress conditions.

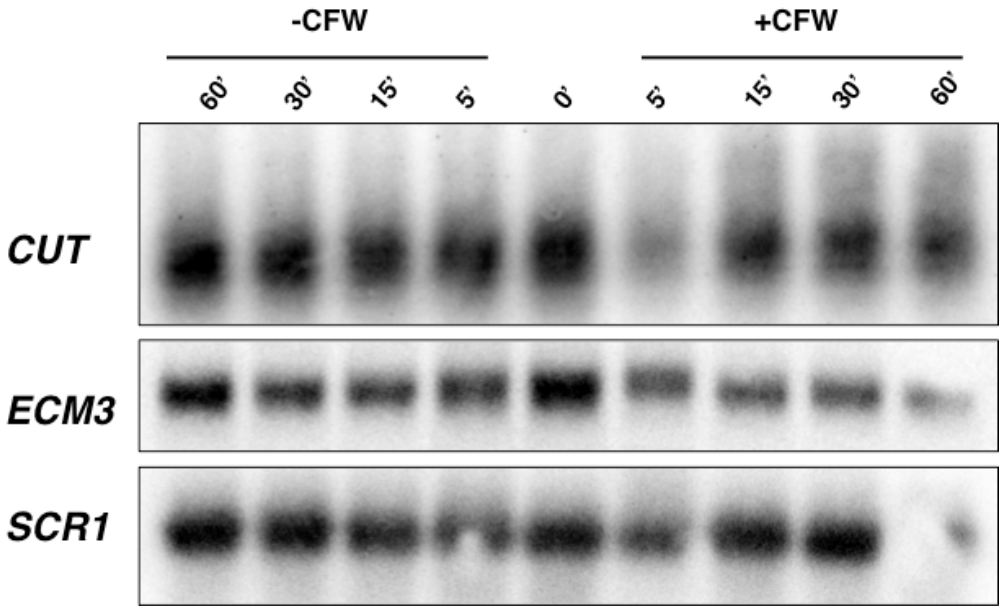

Figure 18. *ECM3* and *EUC1* levels decrease in the presence of calcofluor white.

Northern blot analysis of *EUC1* and *ECM3* expression on RNA isolated from an *rrp6Δ* strain (YJ744) comparing cells grown in rich media (0') compared to cells grown after addition of calcofluor white (CFW) to this media for the indicated lengths of time. Each sample used in this analysis was taken from a single log-phase culture grown without CFW and following its progression after splitting the culture in half and adding CFW to only one half of the culture. *SCR1* serves as a loading control.

Although regulation of start site selection was what drew our attention to the *ECM3* locus, we did not observe any evidence for differential regulation of the two *ECM3* mRNA isoforms in these conditions. In all conditions that we have examined in this chapter, the two *ECM3* isoforms appear to be co-regulated.

3.4 DISCUSSION

Intrigued by the potential of the *ECM3* locus to be a model for isoform-specific regulation by intergenic transcription, we sought to further characterize a role for *EUC1* transcription in *ECM3* regulation. The results discussed in this chapter collectively suggest that *ECM3* may be positively regulated by *EUC1* transcription across its promoter. This is interesting as a positive regulatory role for intergenic transcription is the opposite of what had been previously characterized at the *SER3* locus. We first showed this by disrupting *EUC1* transcription by deletion of the *EUC1* promoter and observing a decrease in both *EUC1* and *ECM3* expression. We then analyzed several physiological conditions where we saw a transient, but repeatable correlation between expression of *ECM3* and the short isoform of *EUC1* with an opposing correlation in expression of the long isoform of *EUC1*. These results present an interesting possibility that the different isoforms of *EUC1* could have opposing effects on *ECM3* expression. Alternatively, one of these isoforms could be a by-product of some transient change in expression of the normal isoform or in the local chromatin landscape under these conditions.

We cannot rule out the possibility that the sequences deleted in the *pEUC1* deletions may contain transcription factor binding sites for the *ECM3* mRNA promoter; however, several

lines of evidence indicate that this is not the most likely explanation for our data. First, we observe two distinct peaks of H3 K4 trimethylation in ChIP data presented in Chapter Four (Figure 25). This indicates that the promoters of *EUC1* and *ECM3* are likely to be separate. The observations presented here that expression of *EUC1* and *ECM3* are correlated in varying environmental conditions suggests that this correlation is not limited to a context where the *EUC1* promoter has been deleted. In genome-wide ChIP analyses, one transcription factor, Phd1, has been detected to occupy the sequences we deleted in the promoter of *EUC1* (RHEE AND PUGH 2011). Overexpression of Phd1 is sufficient to induce pseudohyphal growth in the Σ1278b strain of *S. cerevisiae* (GIMENO AND FINK 1994). However, *ECM3* is not required for pseudohyphal growth (RYAN *et al.* 2012). Although these results are not directly comparable, this leaves no obvious connection between Phd1 and *ECM3* expression. If Phd1 does play a regulatory role at this locus, its binding site being overlapping with the *EUC1* promoter may suggest that it is more likely to regulate *EUC1* than *ECM3*. Additionally, although it is not impossible, it would be an unusually long distance in *S. cerevisiae* for an activator-binding site to be 350 bases upstream of a target gene as average promoters are between 100-200 bases upstream of an ORF for tandemly oriented genes (PELECHANO *et al.* 2006). For these reasons, we favor a model where transcription of *EUC1* positively regulates *ECM3* expression.

4.0 CHAPTER FOUR: THE *ECM3* GENE IS POSITIVELY REGULATED BY METHYLATION OF H3 K4

4.1 INTRODUCTION

Our studies to this point have identified the potential for *ECM3* to be positively regulated by *EUC1* transcription across its promoter. Although the *ECM3* mRNA isoforms appear to be co-regulated, the potential for *EUC1* isoform-specific regulation arose from our investigations of physiological regulation of *ECM3*. The goal of the work presented in this chapter was to characterize the mechanism by which *EUC1* transcription positively regulates *ECM3* expression. As other genes regulated by intergenic transcription exert their effects by altering the local chromatin landscape, we surveyed various chromatin regulators to examine the role of chromatin in *ECM3* regulation. These analyses revealed that the Paf1 complex and methylation of H3 K4 are positive regulators of *ECM3* expression. We confirmed that loss of *EUC1* transcription and loss of *PAF1* both result in loss of H3 K4me3 at the *ECM3* locus by ChIP analyses. Genetic analyses suggest that *EUC1* transcription and the Paf1 complex have separable roles in *ECM3* regulation, but the common link between regulators of *ECM3* identified in this work is that they promote methylation of H3 K4.

4.2 MATERIALS AND METHODS

4.2.1 *S. cerevisiae* strains and media

S. cerevisiae strains used in Chapter Four are listed in Table 6. The strains used to perform the anchor away experiment are W303 derivatives purchased from Euroscarf or generously provided by Patrick Cramer (SCHULZ *et al.* 2013). For anchor away experiments, cells were grown at 30°C in YPD (1% yeast extract, 2% peptone, 2% dextrose) medium until cultures reached a density of 2 X 10^7 cells per mL. Rapamycin was then added to cultures for one hour at a final concentration of 1 μg/mL from a stock of 1 mg/mL rapamycin suspended in ethanol. All other strains used in this study are derived from a *GAL2*$^+$ S288C isolate using standard genetic crosses and transformations (WINSTON *et al.* 1995). Cells were grown at 30°C in YPD medium (1% yeast extract, 2% peptone, 2% dextrose) until cultures reached a density of 1 X 10^7 to 2 X 10^7 cells per mL for isolation of either RNA or chromatin for use in northern blotting, primer extension, and chromatin immunoprecipitation analyses.

Strain	Genotype	Reference or Source
FY4	*MATa*	Winston, *et al*., 1995
FY5	*MATα*	Winston, *et al*., 1995
JDY86	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 met15Δ0 can1::MFA1pr-HIS3 hht1-hhf1::NatMX4 hht2-hhf2::[HHTS-HHFS]-URA3*	Dai *et al*., 2008
JDY86 derivative	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 met15Δ0 can1::MFA1pr-HIS3 hht1-hhf1::NatMX4 hht2-hhf2::[hhts-K4A-HHFS]-URA3*	Dai *et al*., 2008
JDY86 derivative	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 met15Δ0 can1::MFA1pr-HIS3 hht1-hhf1::NatMX4 hht2-hhf2::[hhts-K36A-HHFS]-URA3*	Dai *et al*., 2008
JDY86 derivative	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 met15Δ0 can1::MFA1pr-HIS3 hht1-hhf1::NatMX4 hht2-hhf2::[hhts-K79A-HHFS]-URA3*	Dai *et al*., 2008
KY930	*MATα ura3-52 lys2-128δ rad6Δ0::URA3*	M. Braun and K. Arndt, unpublished
KY1130	*MATα his3Δ200 leu2Δ1 bre1Δ0::KanMX*	M. Braun and K. Arndt, unpublished
KY1700	*MATα paf1Δ0::KanMX*	Crisucci and Arndt, 2011
KY1701	*MATa leu2Δ0 paf1Δ0::KanMX*	Crisucci and Arndt, 2012
KY1703	*MATa rtf1Δ0::KanMX*	Crisucci and Arndt, 2011
KY1704	*MATα rtf1Δ0::KanMX*	Crisucci and Arndt, 2011
KY1705	*MATa ctr9Δ0::KanMX*	Crisucci and Arndt, 2011
KY1706	*MATα cdc73Δ0::KanMX*	Crisucci and Arndt, 2011
KY1711	*MATa rad6Δ0::KanMX*	Crisucci and Arndt, 2011
KY1712	*MATα rad6Δ0::KanMX*	E. Crisucci and K. Arndt, unpublished
KY1713	*MATα bre1Δ0::KanMX*	Crisucci and Arndt, 2011
KY1715	*MATa set1Δ0::KanMX*	Crisucci and Arndt, 2011
KY1716	*MATa set2Δ0::KanMX*	Crisucci and Arndt, 2011
KY1717	*MATa dot1Δ0::KanMX*	Crisucci and Arndt, 2011
KY1755	*MATα set1Δ0::KanMX*	Crisucci and Arndt, 2011
KY1805	*MATα leo1Δ0::KanMX*	Tomson *et al*., 2013
KY1975	*MATa trf4Δ0::NatMX leu2Δ0 ura3Δ0*	E. Crisucci and K. Arndt, unpublished
KY2170	*MATa leu2Δ1 ctr9Δ0::KanMX*	E. Crisucci and K. Arndt, Pruneski *et al*., 2011

KY2171	*MATa cdc73Δ0::KanMX*	E. Crisucci and K. Arndt, unpublished
KY2173	*MATa bre1Δ0::KanMX*	E. Crisucci and K. Arndt, unpublished
KY2377	*MATα rrp6Δ0::kanMX paf1Δ0::kanMX*	Tomson *et al*., 2013
KY2720	*MATa leu2Δ0 set1Δ0::KanMX*	E. Crisucci and K. Arndt, unpublished
KY2721	*MATα ura3Δ0 set1Δ0::KanMX*	E. Crisucci and K. Arndt, unpublished
KY2722	*MATα set1Δ0::KanMX*	E. Crisucci and K. Arndt, unpublished
KY2723	*MATα ura3Δ0 set2Δ0::KanMX*	E. Crisucci and K. Arndt, unpublished
KY2724	*MATa leu2Δ0 set2Δ0::KanMX*	E. Crisucci and K. Arndt, unpublished
KY2725	*MATa leu2Δ0 dot1Δ0::KanMX*	E. Crisucci and K. Arndt, unpublished
KY2726	*MATα ura3Δ0 dot1Δ0::KanMX*	E. Crisucci and K. Arndt, unpublished
KY2727	*MATα rrp6Δ0::kanMX paf1Δ0::kanMX*	E. Crisucci and K. Arndt, unpublished
KY2728	*MATα rrp6Δ0::kanMX paf1Δ0::kanMX*	E. Crisucci and K. Arndt, unpublished
KY2729	*MATα rrp6Δ0::kanMX paf1Δ0::kanMX leu2Δ0*	E. Crisucci and K. Arndt, unpublished
OKA279	*MATα tor1-1 frp1::NAT RPL13A-2XFKBP12::TRP1 his3-11,15 leu2-3,112 ura3 trp1-1 ade2-1 can1-100 GAL psi*$^{+}$	Euroscarf
OKA292	*MATα tor1-1 frp1::NAT RPL13A-2XFKBP12::TRP1 his3-11,15 leu2-3,112 ura3 trp1-1 ade2-1 can1-100 GAL psi*$^{+}$ *NRD1-FRB::KanMX6*	Shulz *et al*., 2013
YEAR205	*MATa rrp6Δ0::kanMX set3Δ0::natMX*	This study
YEAR362	*MATa rrp6Δ0::KanMX gcn5Δ0::NatMX ura3Δ0*	This study
YEAR365	*MATa rrp6Δ0::KanMX sas3Δ0::NatMX ura3Δ0 leu2Δ0*	This study
YJ744	*MATa rrp6Δ0::KanMX*	J. Pruneski and J. Martens, Tomson *et al*., 2013

YJ746	*MATα rrp6Δ0::KanMX*	J. Pruneski and J. Martens, Tomson *et al*., 2013
YJ760	*MATα ura3Δ0 ctr9Δ0::KanMX*	K. Petrov and J. Martens, unpublished
YJ761	*MATα ura3Δ0 leo1Δ0::KanMX*	K. Petrov and J. Martens, unpublished
YJ764	*MATa lys2Δ0 ura3Δ0 his3Δ200 cdc73Δ0::KanMX*	K. Petrov and J. Martens, unpublished
YJ766	*MATa lys2Δ0 ura3Δ0 his3Δ200 leu2Δ0 leo1Δ0::KanMX*	K. Petrov and J. Martens, unpublished
YJ788	*MATα his3Δ200 rtf1Δ0::KanMX*	K. Petrov and J. Martens, unpublished
YJ807	*MATα his3Δ200 paf1Δ0::KanMX*	K. Petrov and J. Martens, unpublished
YJ809	*MATα ura3Δ0 paf1Δ0::KanMX*	K. Petrov and J. Martens, unpublished
YJ1125	*MATa ura3Δ0 his3Δ200 leu2Δ0*	This study
YJ1126	*MATa rrp6Δ0::KanMX ura3Δ0 his3Δ200 leu2Δ0*	This study
YJ1127	*MATα rrp6Δ0::KanMX ura3Δ0 his3Δ200 leu2Δ0 ecm3-pEUC1Δ1::HA (-400 to -350)*	This study
YJ1128	*MATa rrp6Δ0::KanMX ura3Δ0 ecm3-pEUC1Δ1::HA (-400 to -350)*	This study
YJ1129	*MATα rrp6Δ0::KanMX ura3Δ0 ecm3-pEUC1Δ1::HA (-400 to -350)*	This study
YJ1130	*MATα rrp6Δ0::KanMX ura3Δ0 leu2Δ0 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1131	*MATa rrp6Δ0::KanMX ura3Δ0 his3Δ200 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1132	*MATα rrp6Δ0::KanMX ura3Δ0 leu2Δ0 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1133	*MATa ura3Δ0 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1134	*MATα ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1135	*MATa leu2Δ0 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1136	*MATa paf1Δ0::kanMX his3Δ200 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1137	*MATα paf1Δ0::kanMX ura3Δ0 his3Δ200 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study

YJ1138	*MATα paf1Δ0::kanMX ura3Δ0 his3Δ200 ecm3-pEUC1Δ2::HA (-400 to -300)*	This study
YJ1139	*MATα rrp6Δ0::kanMX set1Δ0::kanMX his3Δ200*	This study
YJ1140	*MATa rrp6Δ0::kanMX set1Δ0::kanMX*	This study
YJ1141	*MATa rrp6Δ0::kanMX set1Δ0::kanMX his3Δ200*	This study
YJ1142	*MATa rrp6Δ0::kanMX rtf1Δ0::kanMX*	This study
YJ1143	*MATα rrp6Δ0::kanMX rtf1Δ0::kanMX ura3Δ0 his3Δ200*	This study
YJ1144	*MATa rrp6Δ0::kanMX rtf1Δ0::kanMX leu2Δ0*	This study
YJ1145	*MATα rrp6Δ0::kanMX set2Δ0::kanMX leu2Δ0*	This study
YJ1146	*MATα rrp6Δ0::kanMX set2Δ0::kanMX*	This study
YJ1147	*MATα rrp6Δ0::kanMX set2Δ0::kanMX leu2Δ0 his3Δ200*	This study

Table 6. *S. cerevisiae* strains used in Chapter Four.

4.2.2 Northern blot analysis

Northern blot analyses were performed using 20 μg total RNA samples resolved in gels containing 2% agarose, 6.5% formaldehyde, and 1X MOPS as previously described (AUSUBEL 1987). Most of the northern blots in this work were run for 500 volt hours, which separates total RNA over a gel spanning about 10 cm. The high-resolution northern (Figure 24B) was run for 2200 volt hours, after which low molecular weight RNAs will have run off the gel leaving only high molecular weight RNAs that are separated over about 20 cm of gel. Double-stranded probes were generated by random-primed labeling and single-stranded probes were generated by asymmetric PCR with α-^{32}P-dATP and purified PCR products as templates (RIO 2011). Probe templates were amplified from genomic DNA to contain the following sequences relative to the +1 ATG of the protein-coding gene at each locus: *EUC1* (-541 to -100), *ECM3* (+545 to +976), *SCR1* (-182 to +284), and *ACT1* (+277 to +845). *SCR1* and *ACT1* RNA levels serve as internal loading controls. Images were generated by phosphorimaging and quantified using ImageJ software. For each experiment, the data from at least three biological replicates were averaged.

4.2.3 Primer extension analysis

Primer extension assays were performed as previously described using 20 μg total RNA samples (AUSUBEL 1987). Sequencing reactions were performed using the Sequenase kit following manufacturer's guidelines (Affymetrix USB) using a purified PCR product as a template. Oligonucleotides were gel-purified and end-labeled with α-^{32}P-ATP and T4

polynucleotide kinase using standard protocols (AUSUBEL 1987). The *EUC1* transcription start sites were mapped by primer extension using two different oligonucleotides listed in Table 7.

4.2.4 Chromatin immunoprecipitation analysis

Chromatin was isolated as previously described and sheared by sonication using a Misonix 3000 sonicator (SHIRRA *et al.* 2005). Immunoprecipitations were performed by incubating sheared chromatin with 5 μL antisera to histone H3 (TOMSON *et al.* 2011) or 2.5 μL of antibody to H3 K4me3 (Active Motif, catalogue number 39159) at 4°C overnight followed by precipitation using Protein A sepharose beads (GE Healthcare) for 2 hours at 4°C. All ChIP results were quantified using quantitative PCR amplification of immunoprecipitated DNA compared to input DNA. Real-time PCR reactions were performed using SYBR green reagents (Fermentas) and a Step One Plus instrument (Applied Biosystems). The oligonucleotides used for qPCR amplification are listed in Table 7. Data were analyzed using the Pfaffl relative quantitation method (PFAFFL 2001). H3 K4me3 occupancy values were normalized to total H3 occupancy values.

Purpose	Primer set	Oligo name	Orientation	Location relative to +1 ATG	Sequence
ChIP qPCR	**p*EUC1***	OJ1172	F	-541	5'-CCATGCTTATCTGCCGTCTT
		OJ1257	R	-381	5'-AGGAAGCTCAACTATCACCG
ChIP qPCR	**5' *EUC1***	OJ1730	F	-299	5'-CGTACGCTACCATCTGTTACTCTTGC
		OJ1731	R	-191	5'-CCGGATGCCCTATGATGAAAGTCTATAT
ChIP qPCR	***EUC*1**	OJ1329	F	-186	5'-CTTCTCAGAAGCCTCGCAAT
		OJ1173	R	-100	5'-GGTAATGGTCAACAATACGC
ChIP qPCR	**p*ECM3***	OJ1715	F	-119	5'-GCGTATTGTTGACCATTACC
		OJ1270	R	-1	5'-TGTCTACTTGTCTTGAACTTAC
ChIP qPCR	**5' *ECM3***	OJ1732	F	+3	5'-ACACACATCACACTGGGACAAG
		OJ1733	R	+142	5'-CGATATCAGAGATGGACCTTGTG
ChIP qPCR	**mid *ECM3***	OJ1728	F	+838	5'-AGCGAGCTTAACGATCCTACT
		OJ1175	R	+976	5'-GTAGTTCACGCATATCGATGG
Northern probe template	***EUC1***	OJ1172	F	-541	5'- CCATGCTTATCTGCCGTCTT
		OJ1173	R	-100	5'-GGTAATGGTCAACAATACGC
Northern probe template	***ECM3***	OJ1174	F	+545	5'- TGACCAATGATGATTCTGCCC
		OJ1175	R	+976	5'- GTAGTTCACGCATATCGATGG
Northern probe template	***SCR1***	OJ459	F	-182	5'-CAACTTAGCCAGGACATCCA
		OJ460	R	+284	5'-AGAGAGACGGATTCCTCACG
Northern probe template	***ACT1***	OJ257	F	+277	5'-ATCGATTGCTTCATTCTTTTTGTT
		OJ258	R	+845	5'-ATCGATTCTCAAAATGGCGTGAGG
***EUC1* primer extension**		OJ1258	R	-180	5'- GTAACAGATGGTAGCGTACG
***EUC1* primer extension**		OJ1521	R	-192	5'- CGGATGCCCTATGATGAAAGTC

Table 7. Oligonucleotides used in Chapter Four.

4.3 RESULTS

4.3.1 The Paf1 complex positively regulates *ECM3* expression

Other genes regulated by intergenic transcription exert their regulatory effects through alterations in the local chromatin. Examples have been reviewed in (HAINER AND MARTENS 2011b; FU 2014; JOH *et al.* 2014; RINN AND GUTTMAN 2014). We hypothesized that regulation of *ECM3* expression by *EUC1* transcription might also involve a chromatin-mediated mechanism. To test this, I surveyed *ECM3* expression in a variety of strains with chromatin defects by northern blot analysis. Through this survey, we observed that the Paf1 complex positively regulates *ECM3* expression as loss of any one of the five Paf1 complex members (Paf1, Ctr9, Rtf1, Cdc73, and Leo1) results in reduced *ECM3* expression (Figure 19A).

We also wanted to know if the regulatory effects of the Paf1 complex are dependent on *EUC1* transcription. To address this, I analyzed *ECM3* expression in a *paf1Δ pEUC1Δ2* double mutant compared to *paf1Δ* and *pEUC1Δ2* single mutant strains. This analysis shows a greater reduction in *ECM3* expression in the *paf1Δ pEUC1Δ2* double mutant strain than either single mutant alone (Figure 19B). This result should be considered with the caveat that the *pEUC1Δ2* mutation is not a complete null allele as some small amount of *EUC1* transcript is detected in the presence of this mutation. However, these data are consistent with *EUC1* transcription and the Paf1 complex having separable roles that are additive in regulation of *ECM3* expression. We have identified one shared role of *EUC1* transcription and the Paf1 complex, which will be discussed below, but each of these factors may have other interesting roles remaining to be identified.

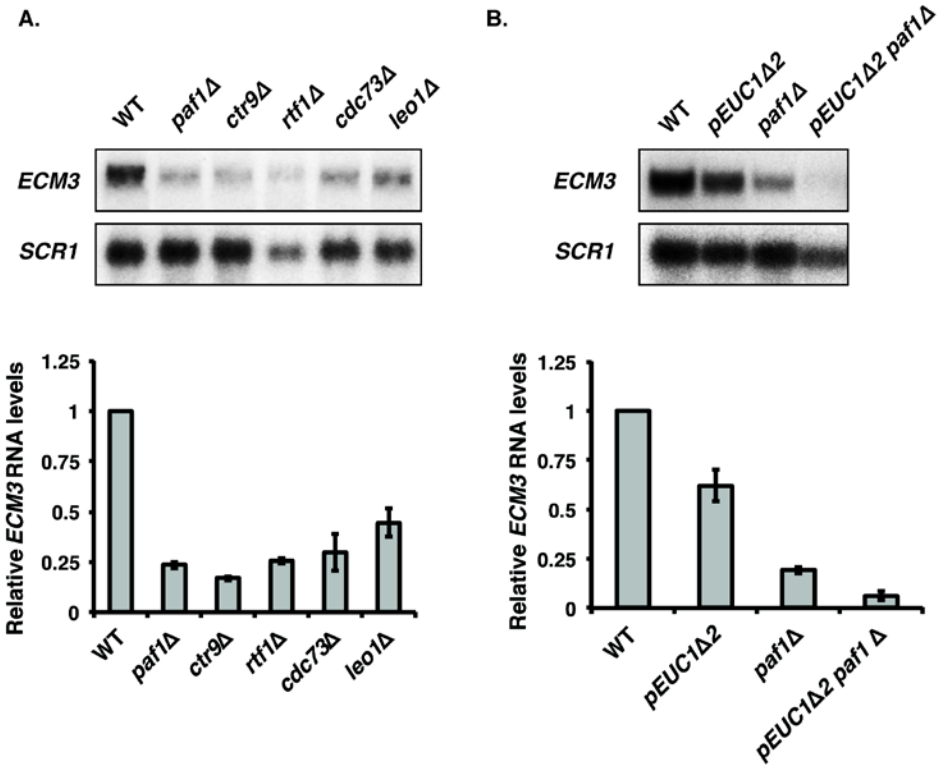

Figure 19. The Paf1 complex positively regulates *ECM3* expression.

(A) Representative northern blot analysis of *ECM3* transcript levels in a wild-type strain (FY4) or strains lacking one of the five members of the Paf1 complex (*paf1*Δ, YJ807; *ctr9*Δ, KY2170; *rtf1*Δ, YJ788; *cdc73*Δ, KY2171; *leo1*Δ, KY1805). (B) Representative northern blot analysis of *ECM3* transcript levels in a *pEUC1*Δ*2* strain (YJ1135), a *paf1*Δ strain (KY1701), and a *pEUC1*Δ*2 paf1*Δ double mutant strain (YJ1138). Bar graphs show the average *ECM3* mRNA levels relative to WT (YJ1125, set to 1) from three biological replicates. Error bars represent the SEM. *SCR1* serves as a loading control.

4.3.2 Methylation of H3 K4 positively regulates *ECM3* expression

We also investigated whether histone modifications regulated by the Paf1 complex might be contributing to *ECM3* regulation. As discussed in Chapter One, the Rtf1 subunit of the Paf1 complex is required for monoubiquitylation of H2B K123 catalyzed by the Rad6 and Bre1 ubiquitin conjugase and ligase enzymes (KROGAN *et al.* 2003a; NG *et al.* 2003a; NG *et al.* 2003b; WOOD *et al.* 2003b). The ubiquitylation of H2B K123 is required for methylation of H3 at K4 and K79, which are catalyzed by the Set1 and Dot1 methyltransferase enzymes, respectively (DOVER *et al.* 2002; SUN AND ALLIS 2002). Tri-methylation of H3 K36 is also dependent on the Paf1 complex; however, this modification is dependent on the Paf1 and Ctr9 subunits but not the Rtf1 subunit (CHU *et al.* 2007). I examined whether these modifications contribute to *ECM3* regulation by northern blot analysis of RNA isolated from strains lacking one of the enzymes that catalyze these modifications. This analysis revealed that ubiquitylation of H2B K123 and methylation of H3 K4 positively regulate *ECM3* expression, as loss of *RAD6*, *BRE1*, or *SET1* results in reduced *ECM3* expression (Figure 20A). This regulation is specific to methylation of H3 K4 as deletion of genes encoding the other methyltransferases regulated by the Paf1 complex, *SET2* or *DOT1*, does not alter *ECM3* expression (Figure 20A). I also tested the effect of H3 K4 methylation on *ECM3* regulation by analyzing *ECM3* expression in strains where H3 K4 has been mutated to an unmodifiable residue, alanine. Although the magnitude of the decrease in *ECM3* levels is slightly different between these two contexts, mutation of H3 K4 to alanine also results in reduced *ECM3* expression (Figure 20B). This may be due to the fact that these histone mutations are integrated into a strain background where one copy of the H3 and H4 histone genes has been deleted. Consequently, this strain background has a lower

histone dosage compared to the *set1Δ* strains, which may have an additional effect on *ECM3* expression. Importantly mutation of other H3 residues, K36 and K79, to unmodifiable residues does not alter *ECM3* levels (Figure 20B).

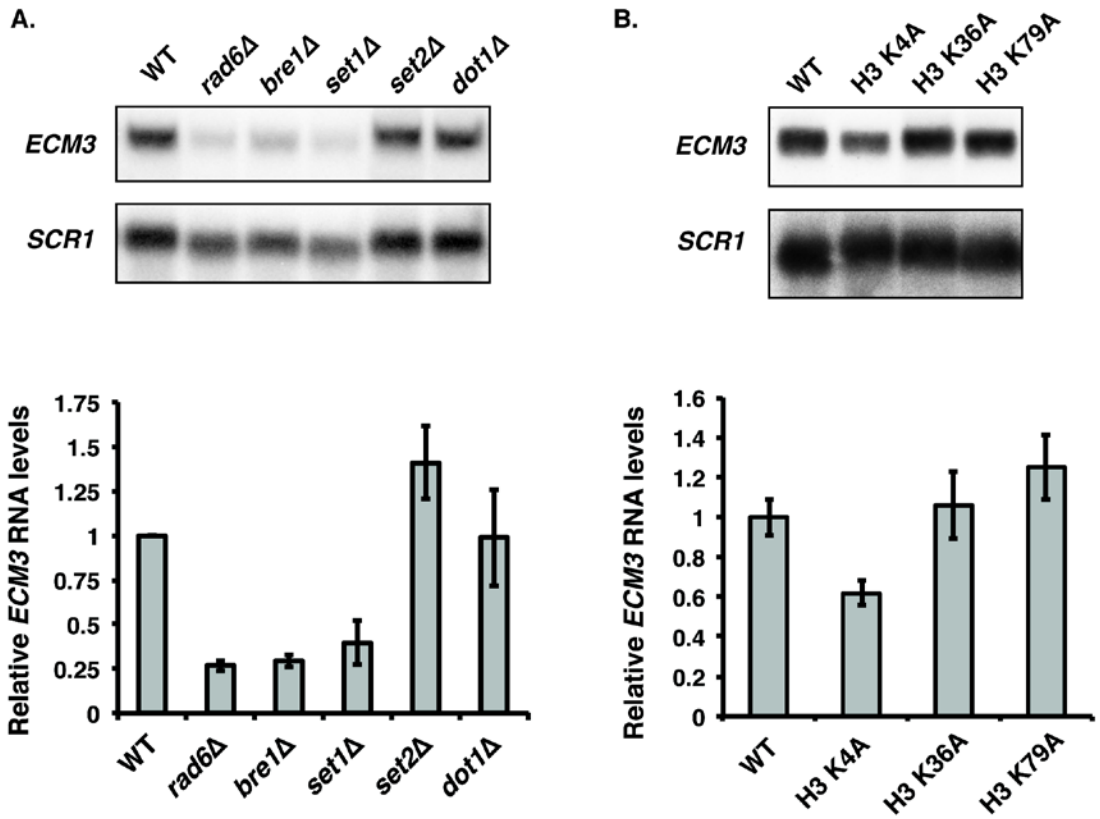

Figure 20. Methylation of H3 K4 positively regulates *ECM3* expression.

(A) Representative northern blot analysis of *ECM3* transcript levels in a wild-type strain (FY4) and strains where the genes encoding histone modifiers that work downstream of the Paf1 complex have been deleted (*rad6*Δ, KY1712; *bre1*Δ, KY1713; *set1*Δ, KY2720; *set2*Δ, KY2723; *dot1*Δ, KY2725). (B) Representative northern blot analysis of *ECM3* transcript levels in a wild-type control strain, lacking one copy of the genes for H3 and H4 (JDY86), and derivatives of JDY86 in which the only copy of the H3-H4 genes encodes the indicated amino acid substitutions in H3. Quantitation below shows the average *ECM3* mRNA levels relative to WT (set to 1) from three biological replicates. Error bars represent the SEM. *SCR1* serves as a loading control.

We next investigated whether the Paf1 complex might be functioning during *EUC1* transcription or if the Paf1 complex works farther upstream to regulate levels of *EUC1* transcription. To do this, I analyzed *EUC1* expression levels in strains lacking *PAF1*, *RTF1*, *SET1*, or *SET2* by northern blot analysis. In the absence of *PAF1*, the levels of the short *EUC1* isoform present in *rrp6Δ* strains do not significantly change, however, a longer, more abundant *EUC1* isoform is detected (Figure 21A and 21C). This long *EUC1* isoform is the same size as the LC isoform detected in response to heat shock and when cells are shifted from media containing raffinose to media containing galactose. Furthermore, this transcript is detected with a strand-specific northern probe that hybridizes to sense transcripts relative to the *ECM3* ORF and is only detected in *paf1Δ* strains lacking *RRP6* (Figure 22), indicating that this larger transcript is also transcribed in the sense direction and is unstable. In the absence of *RTF1*, the levels of the SC *EUC1* isoform decrease slightly. In the absence of *SET1*, I observe no significant change in the levels of either *EUC1* isoform, while levels of *ECM3* are reduced. Importantly, loss of H3 K36 methylation by deletion of *SET2* does not alter *EUC1* or *ECM3* levels (Figure 21A, 21B, 21D).

The mutations that reduce *ECM3* levels (*paf1Δ*, *rtf1Δ*, and *set1Δ*) have variable effects on *EUC1* expression levels in an isoform-specific manner. There appears to be no correlation among these mutants regarding how they influence *EUC1* transcription. Each of these mutants has reduced levels of H3 K4 methylation, which is a likely mechanism underlying reduced *ECM3* expression. Loss of H3 K4 methylation in the absence of *SET1* does not alter *EUC1* expression, but still results in reduced *ECM3* levels. This suggests that synthesis of the *EUC1* RNA or passage of Pol II across the *ECM3* promoter alone is not responsible for the regulatory effects we have observed by transcription of *EUC1*. Instead, it seems that without methylation

of H3 K4, the positive regulation of *ECM3* is lost. As methylation of H3 K4 is catalyzed during transcription elongation, it is possible that transcription of *EUC1* results in methylation of H3 K4 at the *ECM3* promoter and this methylation mark is necessary for downstream activation of *ECM3* transcription. This effect could be direct or indirect and is likely mediated by intermediate proteins that bind to methylated H3 K4 and carry out further positively regulating functions.

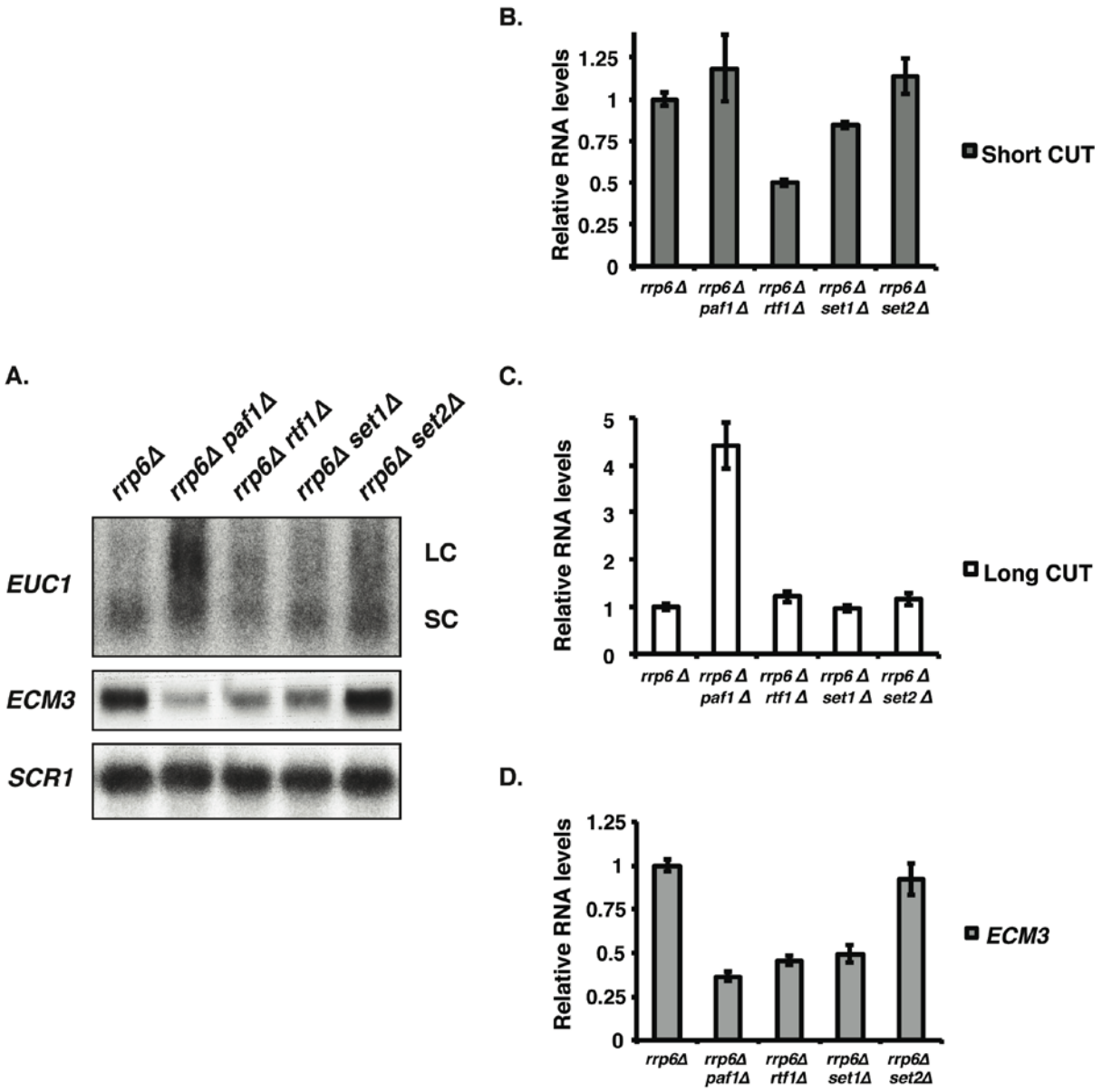

Figure 21. Effect of the Paf1 complex and histone methyltransferases on *EUC1* transcription.

(A) Northern blot analysis of RNA isolated from strains lacking *RRP6* to stabilize CUTs and lacking subunits of the Paf1 complex (*paf1*Δ *rrp6*Δ, KY2727; *rtf1*Δ *rrp6*Δ, YJ1143) or histone methyltransferases (*set1*Δ *rrp6*Δ, YJ1140; *set2*Δ *rrp6*Δ, YJ1146). The *rrp6*Δ control strain was YJ746. (B-D) Quantitation shows average transcript levels relative to those observed

in the *rrp6*Δ strain, which were set to 1. Averaged results from three biological replicates for the *EUC1* short isoform (SC, panel B), the *EUC1* long isoform (LC, panel C) and the *ECM3* ORF transcript (panel D) are shown. Error bars represent the SEM. *SCR1* serves as a loading control.

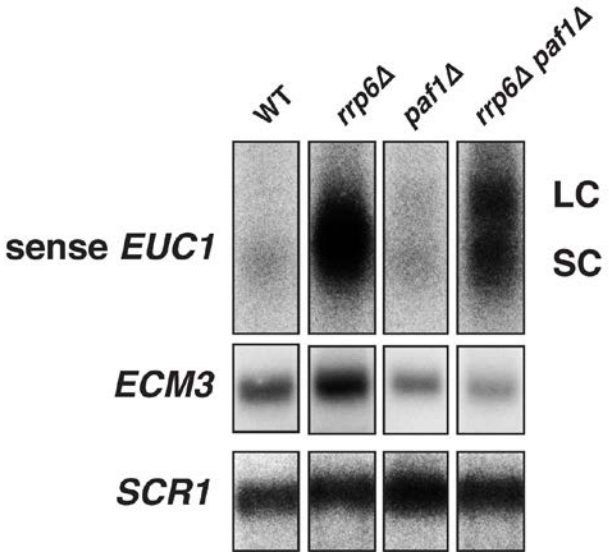

Figure 22. The long *EUC1* isoform is an unstable transcript that is transcribed in the sense direction relative to *ECM3*.

Representative northern analysis of *EUC1* and *ECM3* levels in a wild-type strain compared to strains where CUTs are stabilized (*rrp6*Δ) or lacking Paf1 (*paf1*Δ and *paf1*Δ *rrp6*Δ). *EUC1* was detected using a sense strand-specific probe labeled by asymmetric PCR. *SCR1* serves as a loading control.

4.3.3 Downstream termination of *EUC1* is not sufficient to regulate *ECM3* expression

To this point, we have identified two roles for Paf1 in *ECM3* regulation: Paf1-dependent promoter methylation of H3 K4 and repression of the long isoform of *EUC1*. As Paf1 is required for proper termination by the Nrd1-Nab3-Sen1 complex, we hypothesized that the longer *EUC1* isoform observed in the absence of *PAF1* could be a read-through product arising from a termination defect. I tested this in two ways. First, I mapped the 5' ends of the short and long *EUC1* isoforms by primer extension analysis. Second, I analyzed the effect of disrupting CUT termination on *ECM3* expression. I performed primer extension analysis on RNA isolated from an *rrp6Δ* strain, which expresses the short *EUC1* isoform, an *rrp6Δ paf1Δ* strain, which expresses both short and long *EUC1* isoforms, and a *pEUC1Δ2* strain, which served as a negative control as this strain shows reduced *EUC1* levels. This analysis revealed a cluster of start sites closely spaced around 344 base pairs upstream of the *ECM3* start codon that appear to be utilized equally in *rrp6Δ* and *rrp6Δ paf1Δ* strains (Figure 23). This suggests that the short and long *EUC1* isoforms utilize the same transcription start sites and the difference in size may be attributed to differences in termination.

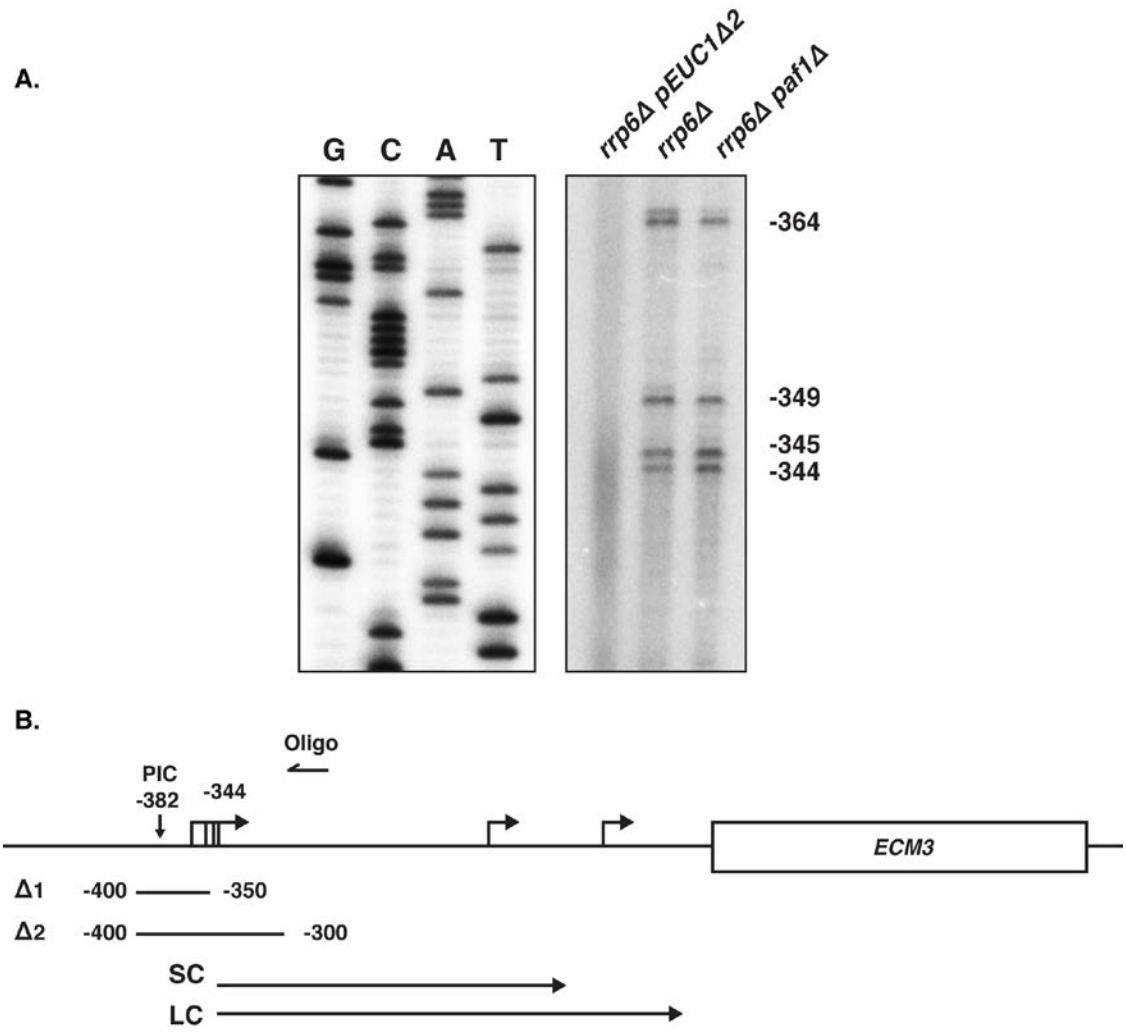

Figure 23. Evidence that the short and long isoforms of *EUC1* initiate from the same transcription start sites.

(A) Primer extension analysis of the 5' ends of *EUC1* transcripts produced in strains that express the *EUC1* SC transcript (*rrp6*Δ, YJ746) or both the *EUC1* SC and LC transcripts (*paf1*Δ *rrp6*Δ, KY2729). The *pEUC1*Δ*2* mutant (YJ1130) was used as a negative control as this strain is severely reduced for *EUC1* transcription. A DNA sequencing ladder is shown on the left. (B) A schematic diagram of the *ECM3* locus with the positions of the upstream *EUC1* CUTs and the *pEUC1*Δ*1* and *pEUC1*Δ*2* mutations indicated. For simplicity, the *EUC1* SC and LC

isoforms are diagrammed as initiating at a single start site to reflect that the closely positioned start sites detected in (A) do not appear as distinct isoforms by northern blot analysis.

To analyze the effects of *EUC1* termination more directly, I used the anchor away technique to deplete Nrd1 from the nucleus. This method allows inducible re-localization a tagged protein of interest and is particularly effective for essential nuclear proteins. This system employs two tags that interact only in the presence of the small molecule, rapamycin. One of these tags serves as the anchor and the other tag is placed on your protein of interest. In this case, the anchor tag is placed on a ribosomal protein and the other tag is placed on the Nrd1 protein. During ribosomal maturation, ribosomal proteins are translated in the cytoplasm, shuttled through the nucleus, and exported back to the cytoplasm where they will exist as part of a mature ribosome. In the presence of rapamycin, as newly synthesized ribosomal proteins are shuttled through the nucleus, the tag on Nrd1 will interact with the tagged ribosomal protein and Nrd1 will be exported to the cytoplasm where it will remain associated with ribosomal proteins. The Cramer lab successfully piloted this method for use with the Nrd1 protein and I was able to validate their results using a serial dilution assay (SCHULZ *et al.* 2013). I performed northern blot analysis on RNA isolated from a tagged *NRD1* strain in the presence and absence of rapamycin compared to an untagged control strain. Upon depletion of Nrd1 from the nucleus, I observed both the short and long isoforms of *EUC1* (Figure 24A). These transcripts are the same size as the isoforms observed in the absence of *PAF1*, which is consistent with the idea that the longer *EUC1* isoform results from read-through transcription of a more upstream termination site. I also observed a third transcript that hybridizes to the *EUC1* northern probe that is the appropriate size to be a full-length read-through (RT) transcript initiating at the *EUC1* promoter and terminating at the 3' end of *ECM3*. This read-through transcript migrates to the same position in the gel as *ECM3* by standard northern blot analysis, which obscured the interpretation of whether production of the long *EUC1* isoform regulates *ECM3* expression. To

separate these transcripts, I performed higher resolution northern blot analysis on these RNA samples as described in Chapter Two. Separation of the full-length read-through transcript from *ECM3* transcripts revealed that synthesis of the long *EUC1* isoform upon depletion of Nrd1 from the nucleus does not significantly alter the levels of *ECM3* expression (Figure 24B). It does appear that depletion of Nrd1 from the nucleus favors expression of the long *ECM3* isoform; however, this result may not be biologically relevant as we do not observe Pol II reading through *ECM3* from the *EUC1* promoter in other conditions. We interpret these results to indicate that the primary regulatory function of Paf1 on *ECM3* expression in these conditions is through its role in promoting H3 K4 methylation.

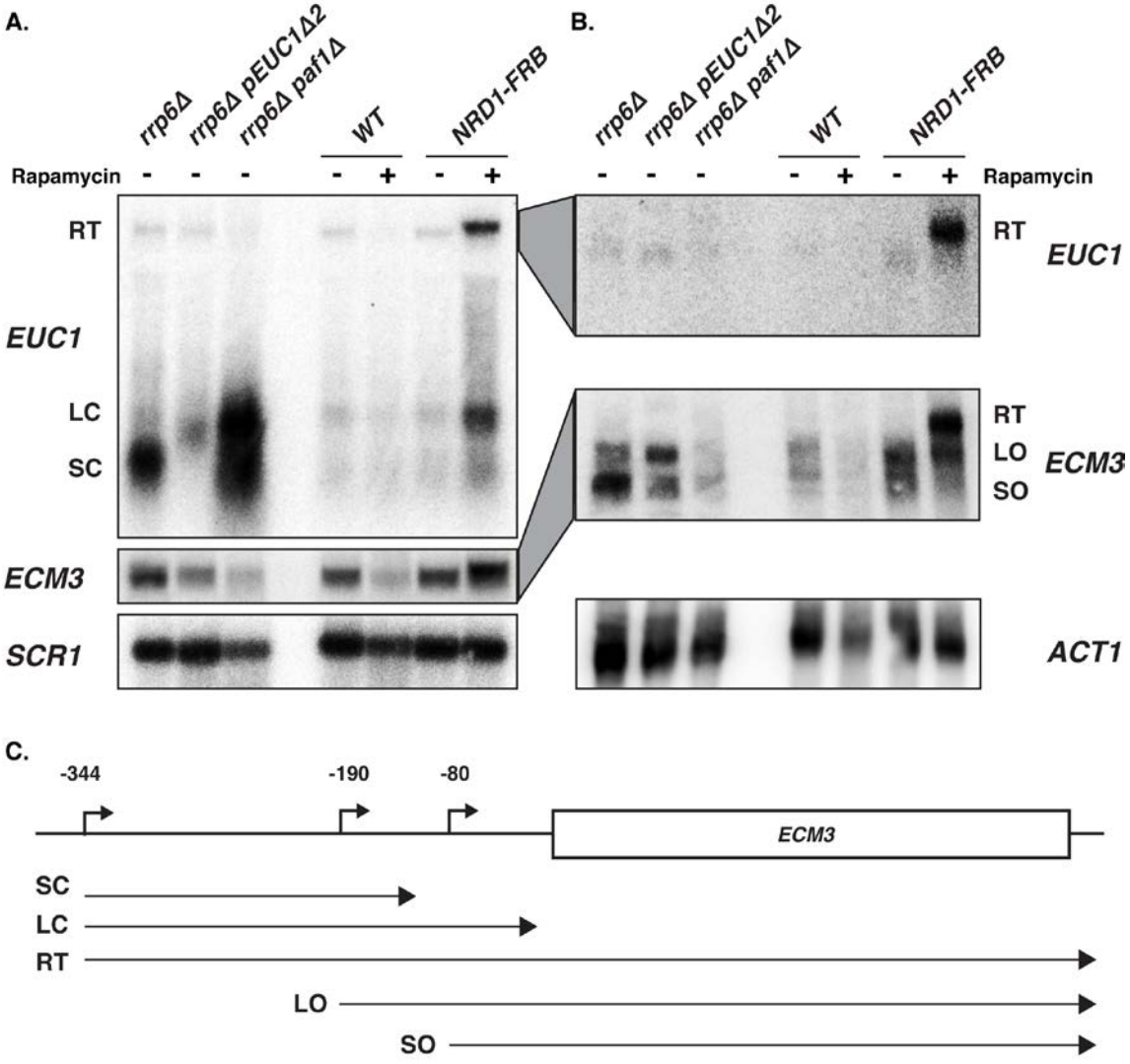

Figure 24. Disruption of CUT termination produces the long *EUC1* isoform.

(A) Representative northern blot analysis comparing *EUC1* and *ECM3* transcript patterns in *rrp6*Δ (YJ746) and *paf1*Δ *rrp6*Δ (KY2729) strains to those of a strain in which Nrd1 (OKA292, Cramer lab) has been depleted from the nucleus by the anchor away method. An untagged anchor away strain was used as a control (OKA279, Euroscarf). *SCR1* serves as a loading control. (B) Representative high-resolution northern blot analysis of strains where Nrd1 has been depleted from the nucleus by the anchor away method using the same strains as in (A) (see Materials and Methods for details). *ACT1* serves as a loading control because the *SCR1*

transcript has run off the gel. (C) The diagram below shows the relative positions of *EUC1* and *ECM3* isoforms at the *ECM3* locus.

4.3.4 *ECM3* expression is not regulated by Mpk1 in response to cell stress

Recent studies have identified non-catalytic roles for the Mpk1 MAPK in regulation of gene expression in response to stress. One of these roles is mediated through a direct interaction with Paf1 that can prevent recruitment of the Nrd1-Nab3-Sen1 termination complex (KIM AND LEVIN 2011). This association can induce expression of target stress responsive genes by alleviating premature termination by the NNS complex. As some of the Mpk1 induced genes have roles in cell wall maintenance and we have observed production of a longer *EUC1* isoform that terminates farther downstream, we analyzed whether Mpk1 may have a role in mediating production of the long *EUC1* isoform. To test this, I performed northern blot analysis on RNA isolated from strains where *mpk1* had been deleted. If Mpk1 is responsible for causing *EUC1* termination to occur farther downstream, you would expect to observe only the short *EUC1* isoform in the absence of *MPK1*. Contrary to this hypothesis, I observed both the short and long isoforms in the absence of *MPK1* (data not shown), even under stress conditions. This suggests that Mpk1 does not mediate differential termination of *EUC1*.

4.3.5 *EUC1* transcription and the Paf1 complex promote methylation of H3 K4 at the *ECM3* locus

The common link between factors that regulate *ECM3* expression is loss of these factors results in reduced H3 K4 methylation. This led us to examine the occupancy of H3 K4me3 directly at the *ECM3* locus by ChIP analysis. I analyzed the occupancy of H3 K4me3 across the *ECM3* locus in strains lacking *EUC1* transcription as well as in strains lacking *PAF1* and *SET1*

as negative controls. The occupancy of H3 K4me3 was made relative to input and normalized to the total H3 occupancy. As expected, the levels of H3 K4me3 in the absence of *PAF1* and *SET1* were nearly undetectable (Figure 25). In the wild-type strain, I observed two peaks of H3 K4me3 occupancy, one over the 5' end of *EUC1* and one over the 5' end of *ECM3*. In the absence of *EUC1* transcription (*pEUC1Δ2*), the levels of H3 K4me3 occupancy at both of these locations are reduced (Figure 25).

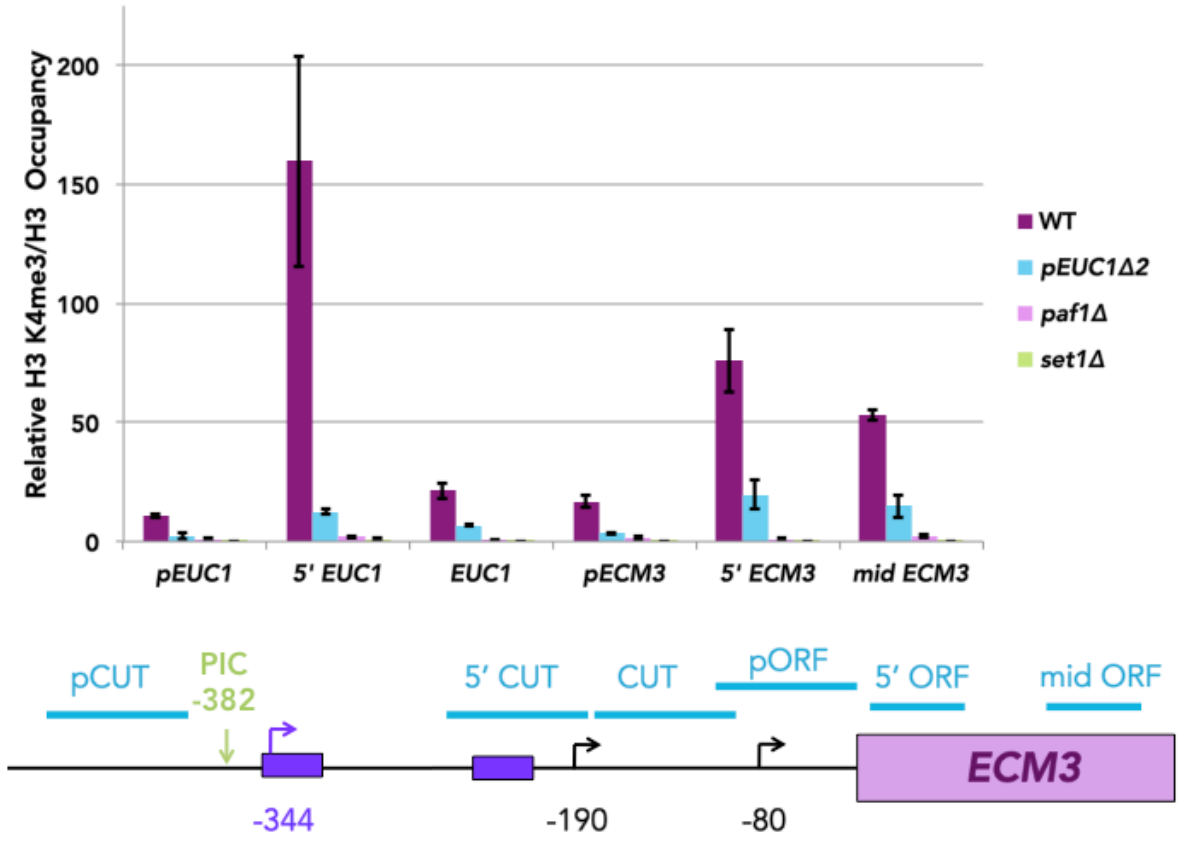

Figure 25. *EUC1* promoter deletion reduces H3K4me3 occupancy across the *ECM3* locus.

ChIP analysis of H3 K4me3 levels at the *ECM3* locus. Immunoprecipitations were performed in biological triplicate using chromatin isolated from wild type (FY4, FY5, YJ1125), *pEUC1Δ2* (YJ1133, YJ1134, YJ1135), *paf1*Δ (YJ807, YJ809, KY1701), and *set1*Δ (KY1755, KY1715, KY2722) strains. Occupancy of H3 K4me3 relative to input DNA was measured by qPCR and normalized to H3 occupancy. Error bars represent the SEM of three biological replicates. The relative locations of qPCR primers are indicated on the diagram below (the mid *ECM3* primer set is not shown to scale).

One possible model is that H3 K4 methylation at the *ECM3* promoter serves as a signal to downstream activators of *ECM3* expression. To test this, I analyzed whether factors known to recognize methylated H3 K4 residues regulate *ECM3* expression. The SAGA, NuA3, and NuA4 histone acetyltransferase (HAT) complexes and the Set3 histone deacetylase (HDAC) complex can bind to methylated H3 K4 in *S. cerevisiae* (MARTIN *et al.* 2006; GINSBURG *et al.* 2009; KIM AND BURATOWSKI 2009; BIAN *et al.* 2011). I examined the effect of SAGA, NuA3, and Set3 as potential "readers" of H3 K4 methylation by northern blot analysis of RNA isolated from strains where one subunit of each of these complexes has been deleted. Consistent with a positive role for H3 K4me in *ECM3* regulation, *ECM3* levels do not change in the absence of the HDAC Set3 (Figure 26). Both SAGA and NuA3 are required for *ECM3* expression as mutation of SAGA (*gcn5Δ*) or NuA3 (*sas3Δ*) results in dramatically reduced *ECM3* levels (Figure 26). It is possible that methylation of H3 K4 during transcription of *EUC1* serves as a signal for SAGA and NuA3 to activate transcription at the *ECM3* promoter. However, these data are complicated by the fact that mutation of SAGA and NuA3 also reduce *EUC1* levels, suggesting that they may act farther upstream to regulate *EUC1* transcription.

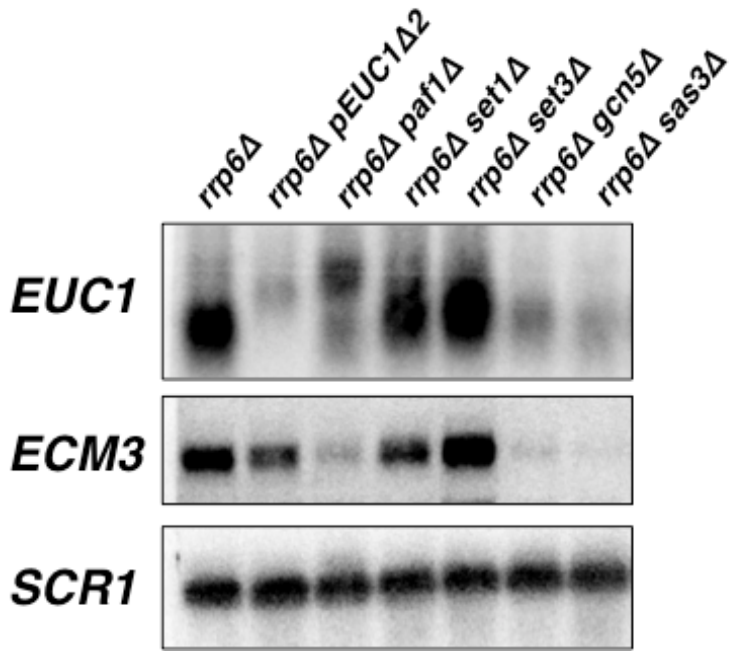

Figure 26. The SAGA and NuA3 histone acetyltransferase complexes positively regulate *ECM3* expression.

Representative northern blot analysis of *EUC1* and *ECM3* expression on RNA isolated from strains where the Set3 HDAC complex (*rrp6Δ set3Δ*, YEAR205), the SAGA HAT (*rrp6Δ gcn5Δ*, YEAR362), or the NuA3 HAT (*rrp6Δ sas3Δ*, YEAR365) have been mutated compared to strains lacking *EUC1* transcription (*rrp6Δ pEUC1Δ2*, YJ1132) or lacking H3 K4 methylation (*rrp6Δ paf1Δ*, KY2729; *rrp6Δ set1Δ*, YJ1141) and a control strain (*rrp6Δ*, YJ1126). *SCR1* serves as a loading control.

4.4 DISCUSSION

With the goal of characterizing how transcription of *EUC1* positively regulates *ECM3*, we surveyed various chromatin regulators to determine how transcription of *EUC1* might alter the chromatin landscape at the *ECM3* locus. Through these analyses, we identified the Paf1 complex as a positive regulator of *ECM3* expression by northern blot analysis. We also identified roles for histone modifications regulated by the Paf1 complex in positive regulation of *ECM3* as deletion of the enzymes that catalyze ubiquitylation of H2B K123 or methylation of H3 K4 results in reduced *ECM3* expression. This effect is specific to methylation of H3 K4 as loss of methylation of other histone residues regulated by the Paf1 complex does not alter *ECM3* expression. We observed an additional role for Paf1 in the regulation of *EUC1* transcription termination by primer extension analysis showing that the long and short *EUC1* isoforms utilize the same 5' ends and by detection of the same *EUC1* isoforms when Nrd1 is depleted from the nucleus by the anchor away method. Although the data presented here suggest that alternative termination of *EUC1* transcription is not sufficient to regulate *ECM3* expression, the fact that we have observed changes in the levels of these isoforms under the physiological conditions discussed in Chapter Three leave open the possibility for the long *EUC1* isoform to have a spatial or temporal role under physiological conditions. Future studies of these roles could be possible in DNA damaging conditions, as *ECM3* expression increases in response to DNA damage (DARDALHON *et al.* 2007). We also showed that loss of *EUC1* transcription results in reduced H3 K4 methylation at the *ECM3* locus by ChIP analysis. The trend that emerges from these data is that loss of H3 K4 methylation results in reduced *ECM3*

expression, indicating a positive role for this modification in *ECM3* regulation. We next analyzed what factors might be recognizing this modification to activate *ECM3* expression and identified that the SAGA and NuA3 HAT complexes are required for proper *ECM3* regulation, which could be related to their ability to recognize methylated H3 K4.

Although positive regulation of *ECM3* expression by promoting methylation of H3 K4 is a function shared by *EUC1* transcription and the Paf1 complex, each of these components must have additional roles in *ECM3* regulation that remain to be identified. This is clearly indicated by the fact that loss of *EUC1* transcription and loss of the *PAF1* have an additive effect on *ECM3* expression levels. This is also evidenced by the fact that we observe a decrease in *ECM3* expression in strains lacking *LEO1*, which is not required for methylation of H3 K4 (NG *et al.* 2003b). It would be interesting to identify the additional roles of *EUC1* transcription and the Paf1 complex in regulation of *ECM3* expression.

Collectively, the results presented in this chapter identify a key role for methylation of H3 K4 as one component necessary for positive regulation of *ECM3* expression. This is interesting, as methylation of H3 K4 by intergenic transcription has been shown to repress expression of two other yeast genes, *DCI1* and *DUR3* (KIM *et al.* 2012). Thus, methylation of H3 K4 can have both positive and negative effects in regulatory mechanisms mediated by intergenic transcription. We know that the different effects of H3 K4 methylation at *ECM3* compared to *DCI1* and *DUR3* is at least partly due to interactions with different readers of H3 K4 methylation. It will be interesting to determine what factors influence the readers of H3 K4 methylation that will act at individual loci. As more cases of regulation by intergenic transcription are characterized, the mechanisms by which intergenic transcription can mediate

regulation of neighboring genes become more widely varied. This may indicate a previously unexplored wealth of regulatory mechanisms.

5.0 CHAPTER FIVE: CONCLUSIONS AND FUTURE DIRECTIONS

Pervasive intergenic transcription is a conserved feature of eukaryotic genomes. Recent studies show clear evidence that intergenic transcription has a wide variety of regulatory roles. In some cases, the RNA products of non-coding genes mediate this regulation. In other cases, the act of transcribing non-coding regions has important regulatory roles. Inspired by the interesting mechanism of regulation by intergenic transcription observed in previous studies at the *S. cerevisiae SER3* gene, the intent of this work was to further investigate the roles of intergenic transcription by identifying and analyzing regulation by intergenic transcription at other loci. Through a candidate gene approach, I have identified *ECM3* as an interesting model gene for elucidation of regulatory mechanisms by intergenic transcription. Interestingly, intergenic transcription positively regulates expression of *ECM3*. Further studies have identified roles for the Paf1 complex and methylation of H3 K4 in positive regulation of *ECM3*, which may be facilitated by transcription of *EUC1* across the *ECM3* promoter. We hypothesized that methylation of H3 K4 may serve as a signal to downstream activators of *ECM3* expression. I found preliminary evidence to suggest that the downstream activators might be the SAGA and NuA3 HAT complexes. Collectively, this work shows a positive role for *EUC1* transcription in *ECM3* regulation, and suggests that this regulatory effect is carried out, at least in part, by post-translational modifications of histone proteins.

5.1 CONCLUSIONS

5.1.1 Premature termination of CUT transcription upstream of protein-coding genes affects expression of the downstream protein-coding gene

In Chapter Two, I selected *ARO2*, *ARO8*, *CLN3*, *ECM3*, *FET4*, and *KNH1* as candidate genes for regulation by intergenic transcription as these six genes all displayed evidence of CUT expression over the promoter of each protein-coding gene in a tandem orientation. I confirmed expression of these six CUTs by northern blot analysis. I then investigated whether premature termination of CUT transcription altered expression of the downstream protein coding genes. Although control experiments would need to be performed by insertion of the TTS in the antisense orientation to confirm these results, I did observe some interesting effects by termination of CUT transcription.

At the *FET4*, *ARO8*, and *CLN3* loci, I observed expression of these protein-coding transcripts to be up-regulated upon termination of upstream CUT transcription. This may suggest that CUT transcription represses expression of these three protein-coding genes. Although we cannot rule out a role for the CUT RNA products in this regulation, we expect that the mechanism of regulation would be less likely to work *in trans* as degradation of CUTs is coupled to their termination. In contrast to what we observed at the *SER3* locus, these genes do not appear to be repressed by maintenance of nucleosome occupancy as the level of de-repression observed in a histone chaperone mutant strain and a histone mutant strain that lose nucleosome occupancy over highly transcribed regions, *spt16-E857K* and H3 K122A, does not account for the level of de-repression observed by termination of CUT transcription. These loci may provide interesting model genes for elucidation of novel mechanisms for repression by

intergenic transcription. I further investigated how CUT transcription interplays with transcription factors known to regulate *FET4* and found that the expected increase in *FET4* RNA levels in the absence of the repressor, *ROX1*, is dependent on a wild-type allele of *RRP6*. This result is interesting and may indicate that CUT stability or perhaps CUT termination is a critical regulatory factor in regulation of *FET4* expression.

At the *ECM3* and *KNH1* loci, I observed changes in transcript isoforms upon termination of CUT transcription. We found the preliminary results at the *ECM3* locus particularly interesting and investigated the mechanism further in Chapters Three and Four. At *KNH1*, disruption of CUT transcription resulted in an antisense transcript that spans the CUT region and two intragenic transcripts arising from within the *KNH1* ORF that are transcribed in the sense direction. *KNH1* could provide an interesting model for regulation of promoter directionality and for repression of cryptic transcripts by intergenic transcription.

5.1.2 Transcription of a CUT, *EUC1*, across the *ECM3* promoter positively correlates with expression of *ECM3*

In Chapter Three, I further investigated the role of CUT transcription in regulation of *ECM3* expression. For subsequent experiments, we named the CUT upstream of *ECM3*, *EUC1*, for *E*CM3 upstream CUT. We discovered that insertion of large, exogenous DNA sequences had effects on which *ECM3* isoforms were expressed and those effects were unrelated to loss of *EUC1* transcription. Therefore, we took an alternative approach to disrupt *EUC1* transcription and were successful in deleting the *EUC1* promoter. Disruption of *EUC1* transcription resulted in lower *ECM3* expression levels by northern blot analysis, suggesting a positive role for *EUC1* transcription in *ECM3* regulation.

I explored several environmental conditions for potential correlations between *EUC1* and *ECM3* expression in a biological setting. In two of these conditions, I observed a longer *EUC1* isoform that appears to be negatively correlated with expression of the short *EUC1* isoform and *ECM3* expression in these conditions. Importantly, fitting with the idea that *EUC1* transcription is positively regulating *ECM3*, the levels of the short *EUC1* isoform do appear to be correlated with *ECM3* expression levels. As discussed at the end of Chapter Three, I cannot be certain that the *EUC1* promoter deletion does not contain activator binding sites for the *ECM3* promoter. However, there is also a positive correlation between *EUC1* and *ECM3* expression in the conditions examined in Chapter Three, even when the *EUC1* promoter has not been mutated. For this reason and others discussed in Chapter Three, we favor a model where *EUC1* transcription positively regulates *ECM3* expression.

5.1.3 *ECM3* expression is positively regulated by the Paf1 complex and methylation of H3 K4

In Chapter Four, I investigated how *EUC1* transcription might be regulating *ECM3* expression. Other genes regulated by intergenic transcription are mediated by local alterations in chromatin. We hypothesized that transcription of *EUC1* might also be regulating *ECM3* expression by altering chromatin across the *ECM3* promoter. I tested this by analyzing *ECM3* expression by northern blot analysis using strains carrying mutations that lead to a variety of defects in chromatin. These analyses revealed that the Paf1 complex and methylation of H3 K4 positively regulate *ECM3* expression. All five members of the Paf1 complex are necessary for this positive regulation. Among histone modifications regulated by the Paf1 complex, the regulatory effects on *ECM3* expression are specific to methylation of H3 K4 and also require

the prerequisite ubiquitylation of H2B K123. Absence of either *EUC1* transcription or Paf1 results in a loss of H3 K4me3 at the *ECM3* locus, indicating a shared role for *EUC1* transcription and the Paf1 complex in promoting methylation of H3 K4 at the *ECM3* locus. Paf1, Rtf1, and Set1 show no obvious correlation in their effects on *EUC1* levels. This may indicate that the *EUC1* RNA product is not responsible for regulation of *ECM3* and that these regulatory effects are instead attributed to the alterations in chromatin, which may take place during transcription of *EUC1*.

The Paf1 complex must have multiple roles in regulation of *ECM3* expression. This is evidenced by genetic data indicating that the Paf1 complex and transcription of *EUC1* have separable roles that have an additive effect in *ECM3* expression. Promoting methylation of H3 K4 is clearly a shared role of *EUC1* transcription and the Paf1 complex, but each of these factors must have additional roles remaining to be identified that contribute to regulation of *ECM3* expression.

5.1.4 Transcription of a long *EUC1* isoform is not sufficient to regulate *ECM3* expression

I observed a larger isoform of *EUC1* in two physiological contexts as well as in the absence of *PAF1*. I investigated potential regulatory roles for this longer *EUC1* isoform in Chapter Four. The long *EUC1* isoform is transcribed in the sense direction relative to the *ECM3* ORF and, like the short *EUC1* isoform, is unstable. As the Paf1 complex is required for proper termination by the Nrd1-Nab3-Sen1 complex (SHELDON *et al.* 2005; TOMSON *et al.* 2011; TOMSON *et al.* 2013), which terminates transcription of CUTs, we hypothesized that the long *EUC1* isoform may arise from transcriptional read-through of an upstream termination sequence. In support of this, the short and long *EUC1* isoforms share transcription start sites.

Furthermore, depletion of Nrd1 from the nucleus produces an *EUC1* isoform of the same size. However, production of the long *EUC1* isoform was not sufficient to alter *ECM3* expression in these conditions. The long *EUC1* isoform may play a spatial or temporal role in regulation of *ECM3* in physiological conditions; however, these data indicate that transcriptional read-through is not sufficient to regulate *ECM3* expression.

5.1.5 Working model for positive regulation of *ECM3* expression by transcription of *EUC1*

Integrating these findings, we propose a model of *ECM3* regulation by intergenic transcription depicted in Figure 27. In this model, *EUC1* transcription positively regulates *ECM3* expression. This is likely to be mediated, at least in part, by the co-transcriptional methylation of H3 K4 across the *ECM3* promoter. We hypothesize that this methylation mark serves as a signal to recruit or activate downstream activators of *ECM3* expression. My preliminary results indicate that this downstream activator could be the SAGA and NuA3 HAT complexes, although these results are complicated by the fact that SAGA and NuA3 also positively regulate *EUC1* expression. This is quite interesting as it presents potential for a feed-forward regulatory mechanism that could have an important role for maintaining or amplifying the amount of *EUC1* and *ECM3* expression.

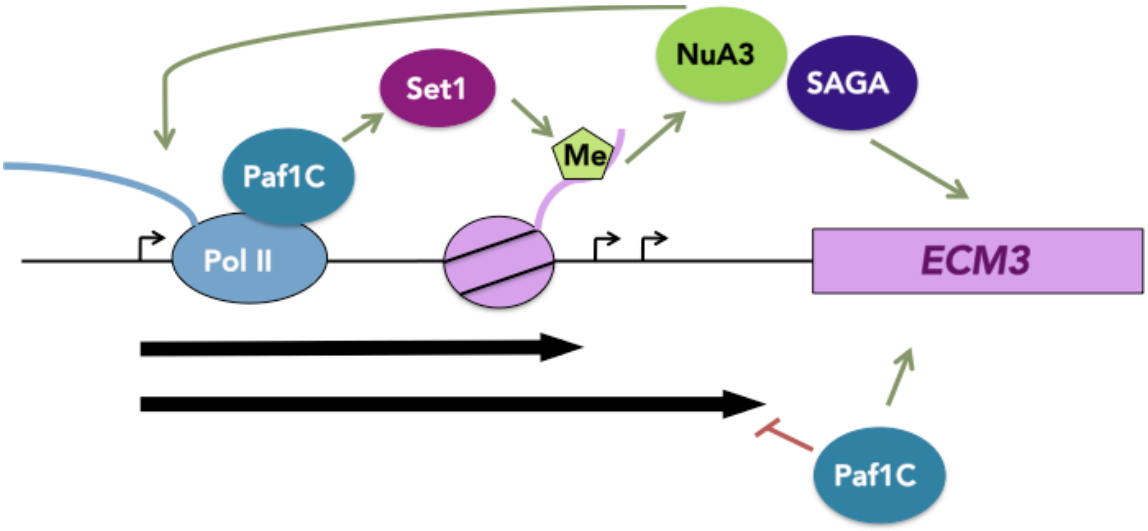

Figure 27. Model of positive regulation of *ECM3* expression by intergenic transcription.

Transcription of *EUC1* results in methylation of H3 K4 across the *ECM3* promoter, which may serve as a signal to downstream activators of *ECM3* expression, putatively the SAGA and NuA3 HAT complexes. These HAT complexes also positively regulate *EUC1* expression. The Paf1 complex plays multiple roles in the regulation of *ECM3* expression. The Paf1 complex promotes H3 K4 methylation, regulates termination of *EUC1*, and has other positive regulatory roles in *ECM3* expression, which remain to be identified.

5.2 FUTURE DIRECTIONS

In addition to beginning to answer some very interesting questions, the exploratory nature of this research has presented many more questions to investigate in the future. This work has identified several new model genes for regulation by intergenic transcription that appear to utilize previously uncharacterized mechanisms to achieve this regulation. These mechanisms would be interesting to explore. In particular, *FET4* provides a model for regulation by intergenic transcription that may be dependent on stability of that intergenic transcript and *KNH1* provides a model for exploring how intergenic transcription could influence promoter directionality and repress cryptic transcripts. As this work has primarily investigated the regulation of *ECM3*, many more directed questions remain to be answered to further characterize this regulatory mechanism. Several of these questions are outlined here.

5.2.1 Determine if positive regulation of *ECM3* expression by methylation of H3 K4 is dependent on *EUC1* transcription

The results presented here are consistent with a model where methylation of H3 K4, possibly during transcription of *EUC1*, leads to downstream activation of *ECM3* expression. Although disruption of *EUC1* transcription lowers H3 K4me3 occupancy at the *ECM3* promoter, these results do not necessarily indicate that the regulatory effect of H3 K4 methylation is dependent on *EUC1* transcription. Observing whether deletion of *SET1* and deletion of the *EUC1* promoter have an additive effect on *ECM3* expression levels could test this idea. If *ECM3* levels were decreased to a similar level in the context of double deletion mutations as the single mutations alone, this would suggest that the regulation of *ECM3* by

methylation of H3 K4 is dependent on *EUC1* transcription. If these mutations have an additive effect on *ECM3* expression, this would suggest that methylation of H3 K4 has regulatory roles in *ECM3* expression that are independent of *EUC1* transcription. Based on these data, I hypothesize that the regulatory effect of H3 K4 methylation are dependent on *EUC1* transcription.

5.2.2 Analyze the occupancy of potential HAT complexes as H3 K4me readers at the *ECM3* locus

My preliminary data suggest roles for the SAGA and NuA3 HAT complexes in positive regulation of *ECM3* expression. SAGA and NuA3 are both capable of recognizing and binding to methylated H3 K4, and these regulators may be recruited to the *ECM3* locus by methylation of H3 K4 across the *ECM3* promoter. However, the interpretation of these results are complicated by the fact that SAGA and NuA3 are also positive regulators of *EUC1* expression, making the effect difficult to separate from simply up-regulating *EUC1* expression which would indirectly lead to positive regulation of *ECM3* expression. On possible way to test this would be to analyze SAGA or NuA3 occupancy at the *ECM3* locus by ChIP. The most specific way to test this would be in the absence of *EUC1* transcription. I hypothesize that the proper recruitment of SAGA and NuA3 to the *ECM3* promoter depends on *EUC1* transcription. The current data alone are not in agreement with this simple model as mutation of SAGA or NuA3 has a much stronger effect on *ECM3* expression than loss of *EUC1* transcription. However, these HAT complexes are multi-subunit proteins that likely associate with chromatin in multiple ways. Thus it is plausible that SAGA and NuA3 occupancy may not be completely lost in the absence of *EUC1* transcription and this may explain why mutation of SAGA or NuA3 have a

much stronger effect on *ECM3* expression levels than loss of *SET1*. This difference could also arise from the fact that *EUC1* expression is not dependent on *SET1*, but is dependent on SAGA and NuA3. It is possible that the function of Set1 is only necessary for *ECM3* expression, while SAGA and NuA3 are necessary for both *EUC1* and *ECM3* expression and loss of either SAGA or NuA3 disrupts a positive feedback mechanism that would have a more dramatic effect on *ECM3* expression.

5.2.3 Determine if *EUC1* transcription is necessary for regulation of *ECM3* expression in a physiological condition

An important question regarding *ECM3* regulation by *EUC1* transcription is whether this is the biologically relevant mechanism that cells normally employ to regulate *ECM3* expression. My preliminary data in Chapter Three suggests that this is possible but these results are simply correlative. The conditions discussed in Chapter Three have only subtle effects on *ECM3* expression. In order to better characterize this effect, I would first need to identify conditions where *ECM3* expression is dramatically altered. It has been reported in the literature that *ECM3* expression is rapidly induced in response to DNA damaging agents (DARDALHON *et al.* 2007). I hypothesize that in these conditions, *EUC1* expression is also induced. To test whether induction of *ECM3* is dependent on *EUC1* transcription, I would analyze *ECM3* expression in the absence of *EUC1* transcription when exposed to DNA damaging agents. Based on my current results, I would expect that *ECM3* would not be induced or not be induced as strongly in the absence of *EUC1* transcription.

5.3 FINAL REMARKS

The results of this research have identified several interesting model genes for previously uncharacterized mechanisms of regulation by intergenic transcription. In particular, this work has supported a positive regulatory role for intergenic transcription in regulation of *ECM3* expression. This regulatory mechanism involves a positive role for methylation of H3 K4, which had previously been reported to have negative regulatory roles at other loci including *GAL1-10*, *DCI1* and *DUR3* (Figure 28). This highlights the idea that regulation by intergenic transcription is likely to be highly diverse across the genome. This also highlights that our current understanding of the regulatory mechanisms of intergenic transcription is very limited, as only a small percentage of loci have been analyzed for these effects. Collectively, these results have expanded our knowledge of how intergenic transcription can regulate gene expression and identified novel candidates for further exploration of the regulatory roles of intergenic transcription.

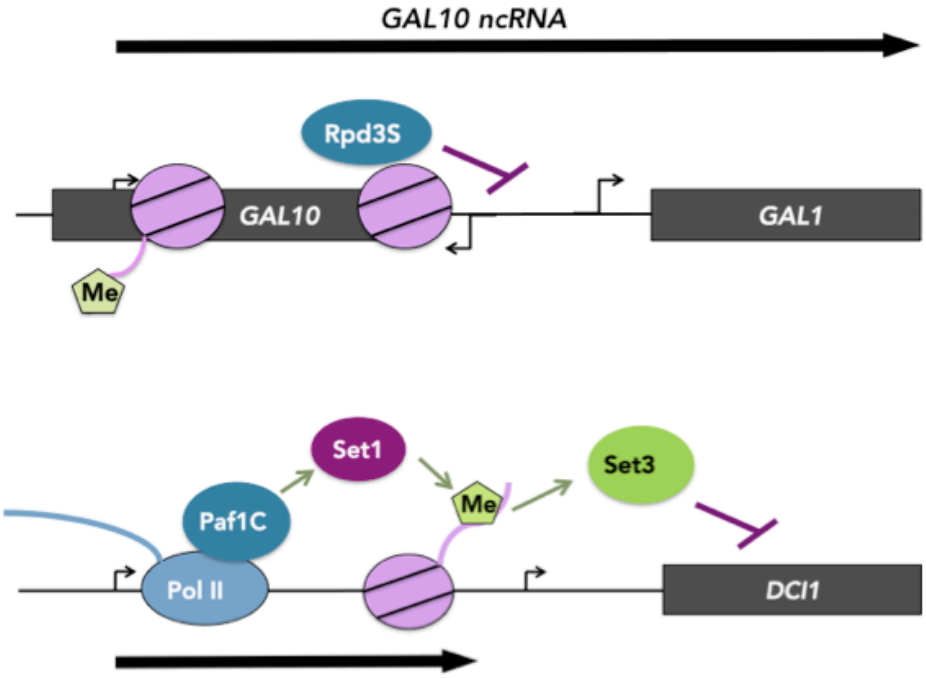

Figure 28. Model of repression by intergenic transcription and methylation of H3 K4 at *GAL1-10*, *DCI1*, and *DUR3*.

At the *GAL1-10* locus, non-coding transcription across the promoters of *GAL1* and *GAL10* results in methylation of H3 K4 that leads to subsequent histone deacetylation by the Rpd3S complex and repression of *GAL1* and *GAL10* expression (HOUSELEY *et al.* 2008; PINSKAYA *et al.* 2009). At the *DCI1* and *DUR3* (not pictured) loci, non-coding transcription across the promoters of these genes results in methylation of H3 K4 that leads to subsequent histone deacetylation by the Set3 complex and repression of *DCI1* and *DUR3* expression (KIM *et al.* 2012).

APPENDIX A

GENOME-WIDE IDENTIFICATION OF GENES REGULATED BY MAINTENANCE OF TRANSCRIPTION-COUPLED NUCLEOSOME OCCUPANCY

A.1 INTRODUCTION

This appendix summarizes the results of a collaborative effort between Travis Mavrich, Paul Yenerall, Sarah Hainer, Joseph Martens, members of the Lars Steinmetz lab, members of the Frank Pugh lab, and myself.

As discussed in the introduction, the act of transcribing a ncRNA, *SRG1*, over the promoter of the *SER3* gene represses expression of *SER3* (MARTENS *et al.* 2004; MARTENS *et al.* 2005). This repression is mediated by histone chaperones, FACT and Spt6, which place nucleosomes over the *SER3* upstream activating sequences during transcription of *SRG1* (HAINER et al. 2011). In light of the knowledge that non-coding transcription is pervasive in eukaryotes, the goal of this project was to determine how many genes might be regulated by a *SER3*-like mechanism across the *S. cerevisiae* genome. To do this, we took advantage of mutations in key regulatory factors that result in strong loss of nucleosomes over the *SER3* promoter and very strongly de-repress *SER3*, including mutations in the histone chaperone, *spt16* and histone H3. We performed genome-wide expression and nucleosome-positioning

assays to identify how many genes in the *S. cerevisiae* genome might be regulated by transcriptional maintenance of nucleosome occupancy over their promoters. These results identified about 20 genes that might be regulated by a similar mechanism to the *SER3* gene. These findings indicate that regulatory mechanisms mediated by non-coding transcription are likely to be repeated throughout the genome. The small fraction of genes with transcription occurring across their promoters that resemble *SER3* expression by these analyses may indicate that there is not likely to be one predominant mechanism that mediates regulation by intergenic transcription and rather there is likely to be a plethora of regulatory mechanisms mediated by intergenic transcription.

A.2 MATERIALS AND METHODS

A.2.1 *S. cerevisiae* strains and media

S. cerevisiae strains used in this study are listed in Table 8. All other strains used in this study are derived from a *GAL2*$^{+}$ S288C isolate using standard genetic crosses and transformations (WINSTON *et al.* 1995). Cells were grown at 30°C in YPD medium (1% yeast extract, 2% peptone, 2% dextrose) until cultures reached a density of 1 X 10^7 to 2 X 10^7 cells per mL for isolation of either RNA or chromatin for use in RT-PCR, northern blotting, microarray expression analysis, MNase-seq analysis, and chromatin immunoprecipitation (ChIP) analysis.

Strain	Genotype	Reference or Source
YEAR003	*MATα*	This study
YEAR004	*MATα*	This study
YEAR004	*MATα*	This study
YEAR007	*MATα spt16-197*	This study
YEAR008	*MATα spt16-197*	This study
YEAR010	*MATa spt16-197*	This study
YEAR012	*MATa spt6-1004*	This study
YEAR013	*MATα spt6-1004*	This study
YEAR018	*MATα spt6-1004*	This study
YTM159	*MATα his3Δ200 leu2Δ0 lys2-128δ trp1Δ63 ura3Δ0 hht1Δ::HHTS-URA3 (hht2- hhf2)Δ::HHTS-HHFS-URA3*	T. Mavrich
YTM163	*MATa his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 hht1Δ::HHTS-URA3 (hht2- hhf2)Δ::HHTS-HHFS-URA3 can1::MFA1pr-HIS3*	T. Mavrich
YTM173	*MATa his3Δ200 leu2Δ0 lys2-128δ trp1Δ63 ura3-52 hht1Δ::HHTS-K122A-URA3 (hht2- hhf2)Δ::HHTS-K122A-HHFS-URA3*	T. Mavrich
YTM175	*MATα his3Δ200 leu2Δ0 lys2Δ0 trp1Δ63 ura3Δ0 hht1Δ::HHTS-K122A-URA3 (hht2- hhf2)Δ::HHTS-K122A-HHFS-URA3*	T. Mavrich
YTM194	*MATa ura3Δ0 hht1Δ::HHTS-URA3 (hht2- hhf2)Δ::HHTS-HHFS-URA3*	T. Mavrich
YTM197	*MATa ura3Δ0 hht1Δ::HHTS-URA3 (hht2- hhf2)Δ::HHTS-HHFS-URA3*	T. Mavrich
YTM201	*MATa ura3Δ0 hht1Δ::HHTS-K122A-URA3 (hht2- hhf2)Δ::HHTS-K122A-HHFS-URA3*	T. Mavrich
YTM202	*MATa ura3Δ0 hht1Δ::HHTS-K122A-URA3 (hht2- hhf2)Δ::HHTS-K122A-HHFS-URA3*	T. Mavrich
YTM203	*MATa ura3Δ0 hht1Δ::HHTS-K122A-URA3 (hht2- hhf2)Δ::HHTS-K122A-HHFS-URA3*	T. Mavrich
YTM210	*MATa ura3Δ0, hht1Δ::HHTS-URA3, (hht2- hhf2)Δ::HHTS-HHFS-URA3*	T. Mavrich

Table 8. *S. cerevisiae* strains used in Appendix A.

A.2.2 Microarray expression analysis

Total RNA was isolated using the acid phenol extraction method described in Chapter Two (AUSUBEL 1987). To ensure that RNA samples were not contaminated with DNA, isolated RNA was treated with DNase using the TURBO-DNA free kit (Ambion). RNA integrity was assayed after DNase treatment by gel electrophoresis and RT-PCR analysis of *SER3* expression using the RETROscript kit (Ambion). For hybridization to arrays, double-stranded cDNA was synthesized from DNase-treated RNA samples using the Affymetrix cDNA synthesis kit and purified using the Affymetrix sample cleanup module. Purified cDNA was then fragmented and labeled for hybridization to arrays using the Affymetrix WT double stranded DNA terminal labeling kit. Labeled samples were processed and hybridized to Affymetrix double stranded *S. cerevisiae* microarrays with 8 base pair resolution by the Genomics and Proteomics Core Laboratories at the University of Pittsburgh. Data were analyzed using Tiling Analysis Software and normalized to the respective wild-type data.

A.3 RESULTS

A.3.1 Identification of genes likely to be repressed by nucleosome occupancy over their promoters

The goal of this project was to identify how many genes in the *S. cerevisiae* genome might be regulated by transcriptionally maintained nucleosome occupancy across their promoters, as observed at the *SER3* gene. Our previous studies have identified that mutations in

the histone chaperones, FACT and Spt6, and in histone H3 (K122A) result in a strong loss of nucleosome occupancy over the *SER3* promoter and strongly de-repress *SER3* expression. Our approach to this question was to identify genes whose expression is changed in strains carrying mutations in FACT, Spt6, and H3 with the assumption that at least some of these genes would display transcription across their promoters.

Travis Mavrich began this work by performing genome-wide expression analysis comparing RNA isolated from wild-type strains and H3 K122A strains. The H3 K122A strains that we used are derivatives of JDY86 from the Boeke lab, which have copy one of the H3 and H4 genes deleted and copy two of the H3 and H4 genes replaced with a synthetic version of the H3 and H4 genes that have distinct nucleotide sequences but the same amino acid sequences as the wild-type genes (DAI *et al.* 2008). Our previous studies suggested that *SER3* expression is sensitive to the dosage of histones as *SER3* is de-repressed in this strain background with only one copy of the H3 and H4 genes even with wild-type alleles (HAINER AND MARTENS 2011a). We speculated that other genes regulated by similar mechanisms might also be sensitive to histone dosage. To avoid this complication, Travis first created strains where both copies of the histone H3 and H4 genes contained the synthetic alleles, maintaining a normal histone dosage. In his wild-type strains, both copies contain wild-type synthetic alleles of H3 and H4. In his H3 K122A strains, both copies contain an H3 K122A allele and a wild-type H4 allele. Sarah Hainer later created similar double synthetic allele strains using a different integration strategy and plasmids engineered by the Boeke lab. Travis later created prototrophic versions of these double copy synthetic histone strains that carried only a *ura3Δ* mutation as the synthetic histone alleles are marked with a *URA3* cassette in order to avoid any changes in gene expression due to

secondary mutations in the strain, which were used in our collaborative studies with the Steinmetz and Pugh labs.

I performed microarray expression analyses comparing RNA isolated from wild-type and *spt6-1004* strains. Microarray expression analyses of the *spt6-1004* allele had previously been reported; however, as this allele is temperature sensitive, these analyses were performed with RNA isolated from *spt6-1004* strains grown at 37°C (CHEUNG *et al.* 2008). Our previous studies indicate that *SER3* is repressed in the presence of the *spt6-1004* allele even at 30°C (HAINER *et al.* 2011). As exposure to heat shock at 37°C has pleiotropic effects, we reasoned that analyzing gene expression in *spt6-1004* strains at 30°C would reduce the number of non-specific targets that we would identify whose expression was altered due to a change in temperature rather than a loss of nucleosome occupancy resulting from a mutated histone chaperone. I first created *spt6-1004* and *spt16-197* mutant strains that were prototrophic and carried no other mutations aside from either the *spt6-1004* or *spt16-197* allele. This was done to increase the chances that any changes in gene expression we observed were due to the mutation in the histone chaperone gene. We chose to initially analyze expression in *spt6-1004* strains with the hope of later analyzing gene expression changes in the context of an integrated *spt16-E857K* allele, as this allele has a much stronger de-repression of *SER3* than the *spt16-197* allele.

Travis performed microarray expression analyses comparing expression of H3 K122A strains to wild-type strains and I performed microarray expression analyses comparing expression of *spt6-1004* strains to wild-type strains. These analyses were performed using double-stranded *S. cerevisiae* microarrays with 8 base-pair resolution purchased from Affymetrix. We prepared cDNA for hybridization using Affymetrix labeling kits and hybridization and data collection were performed at the University of Pittsburgh core facilities.

Data were normalized to the respective wild-type strain and analyzed using Tiling Analysis Software. Travis analyzed the data from the H3 K122A mutant strains, I analyzed the data from the *spt6-1004* strains, and I compared the data sets to each other to compile the following results.

These analyses identified 69 genes that were up-regulated and 230 genes that were down-regulated more than two-fold in the *spt6-1004* strain compared to wild-type. Fitting with our hypothesis that analyzing RNA from an *spt6-1004* strain grown at 30°C should narrow our list of specific targets, the expression analysis performed on *spt6-1004* RNA isolated from cells grown at 37°C identified over 1000 genes that are up-regulated more than two-fold (CHEUNG *et al.* 2008). There were 50 genes up-regulated and 4 genes down-regulated more than two-fold in the H3 K122A strain compared to wild-type. Although we had no reason to think that there should be more genes that are repressed by transcription-coupled nucleosome occupancy over their promoters than activated, the surprisingly few number of genes that are down-regulated in the H3 K122A strain is consistent with that idea.

We reasoned that any genes whose expression changes in both *spt6-1004* and H3 K122A data sets are likely candidates for regulation by nucleosome occupancy over their promoters. Interestingly, 22 genes were up-regulated more than two-fold in both *spt6-1004* and H3 K122A data sets. These genes are listed in Table 9. Importantly, *SER3* is among this list, serving as an internal control for validation of these data. Eight of these 22 genes, including *SER3*, have annotated SUT expression over their promoters, two have annotated CUT expression over their promoters, and several more genes display un-annotated transcripts expressed over their promoters in genome-wide expression analyses (XU *et al.* 2009). Two of these genes have SUTs transcribed in the antisense direction over the body of the gene (XU *et*

al. 2009). Thus 17 of these 22 genes show evidence of non-coding transcription across or near their promoters, and this may be an underestimation as there may be more neighboring non-coding transcripts that have yet to be characterized.

Gene name	*spt6* / WT (log 2)	H3 K122A / WT (log 2)	Annotated SUT	Annotated CUT	Un-annotated expression
YGP1	4.15	1.67	X		
SER3	3.95	4.31			
FIG1	3.91	3.02			
DDR2	3.22	1.16			X
GAD1	2.90	1.21			X
HXT4	2.83	1.20	X		
AGA1	2.30	1.29			X
HXK1	2.23	1.04			
GPH1	2.05	1.24	X		
YRO2	1.85	1.85		X	
TPO2	1.76	1.00			X
IMD2	1.75	1.58		X	
PRM6	1.69	1.97			
FIG2	1.51	1.67			
PRM1	1.51	1.85	X		
NCA3	1.46	1.12	X		
SPI1	1.43	1.08			X
DDI2	1.17	1.14			
RTC3	1.16	1.27	X		
FRE7	1.11	1.24	X		
VPS73	1.09	1.00	antisense		
PRM2	1.00	1.09	antisense		

Table 9. List of genes up-regulated more than two fold in both *spt6-1004* and H3 K122A mutant strains compared to wild-type by microarray expression analysis.

Microarray expression analysis was performed on cDNA synthesized from mid-log phase cultures of *spt6-1004* (YEAR013) or H3 K122A (YTM173) mutant strains compared to wild-type (YEAR003 and YTM159). Values in this table represent the mutant signal averaged over the length of each gene after normalization to the respective wild-type signal on a log 2

scale (value of 1 indicates twice as much signal in mutant strain compared to wild-type). An X in the annotated SUT column indicates that an annotated SUT is located across or near the promoter of that gene (highlighted in pink). Antisense listed in the annotated SUT column indicates that an annotated SUT is transcribed antisense to that protein-coding gene. An X in the annotated CUT column indicates that an annotated CUT is located across or near the promoter of that gene (highlighted in blue). An X in the un-annotated expression column indicates that some un-annotated RNA expression is detected across or near the promoter of that gene (highlighted in green). Data for nearby non-coding transcripts were compiled from (XU *et al.* 2009). Travis Mavrich performed microarray expression analysis of the H3 K122A mutant.

A.3.2 Strand-specific microarray expression analysis of H3 K122A strains

Encouraged by these results indicating that some of these genes are very likely to be regulated by a similar mechanism to the *SER3* gene, we wanted to analyze expression changes in these mutants using strand-specific microarray analysis. This was done in collaboration with the Steinmetz lab who have created their own custom strand specific, high density tiling arrays for the *S. cerevisiae* genome. Fewer changes in gene expression were observed in our H3 K122A microarray expression analyses than our *spt6-1004* microarray expression analyses. For this reason, we thought the K122A mutant might be more specific for the effects we are interested in observing and we selected this mutant for further analyses. Sarah Hainer prepared RNA from H3 K122A strains for hybridization to microarrays, which was performed and analyzed by members of the Steinmetz lab.

The microarray expression analyses performed by the Steinmetz lab identified 48 protein-coding genes that are up-regulated and 37 protein-coding genes that are down-regulated more than two-fold in H3 K122A strains compared to wild-type. One possible explanation for why the Steinmetz microarray was able to identify more genes that are down-regulated in H3 K122A mutant strains is that the analyses they were performing were strand-specific, whereas the inability to distinguish between strands in our previous analysis could have some signals dampened if changes were occurring on both strands at the same locus. Another possible explanation could be attributed to the fact that these arrays were performed on RNA isolated from different strains. The Steinmetz microarray appears to be more sensitive as the fold change in *SER3* expression was greater in the Steinmetz analysis than in our analysis. The genes up-regulated more than two-fold in the H3 K122A microarray analysis performed by the

Steinmetz lab are listed in Table 10. The genes identified as being up-regulated in the H3 K122A mutant in both microarray expression data sets are highlighted in yellow. Several genes identified in the microarray analysis performed by Travis that do not appear in this list are also up-regulated in the Steinmetz array data, but did not meet the two-fold cutoff.

Interestingly, these results show that regulation by intergenic transcription has the ability to regulate expression of both protein-coding and non-protein-coding genes as many snRNAs appear in this list. There were also 17 CUTs and 12 SUTs that were up-regulated more than two-fold in the H3 K122A mutant.

Gene Name	H3 K122A / WT log2 change	H3 K122A / WT Fold increase
SRG1, SER3	3.169970782	9.000285602
RPL13A	3.122995189	8.711947072
SNR6	2.65665334	6.305686038
SNR8	2.304104218	4.938607171
IES6	2.120583252	4.348697183
YOR053W	1.931809788	3.815335136
SNR85	1.867777613	3.64969932
SNR72	1.864627117	3.64173796
HXT4	1.643334304	3.123869756
SNR81	1.604177783	3.040224336
SNR79	1.582637952	2.995170128
SNR78	1.561620358	2.951851936
SNR64	1.547661271	2.923428428
YDR042C, SNR47	1.538983394	2.90589665
YJL047C-A, SNR60	1.453264304	2.738269235
COS12	1.396697488	2.632981688
SNR7-L, SNR7-S	1.335381649	2.523422273
MF(ALPHA)2	1.322365582	2.500758222
SNR69	1.304725514	2.47036722
TPO2	1.27131712	2.413818364
RPS30A	1.262394876	2.398936344
AGA1	1.257032815	2.390036773
SNR71	1.240652478	2.363053804
FIG2	1.232975855	2.350513318
SNR50	1.231930097	2.348810133
PRM2	1.231148571	2.347538097
SNR75	1.220897634	2.330917001
RPL36A	1.213280797	2.318643145
tD(GUC)O	1.202334574	2.301117375
SNR189	1.184247447	2.27244826
SNR63	1.180927483	2.267224862
SNR5	1.17974066	2.26536051
YHR087W	1.179403737	2.264831526
RPS6B	1.177296675	2.261526146
SNR40	1.153770858	2.224946826
SNR56	1.140313535	2.204289229
SNR52	1.130493544	2.189336244
QCR9	1.123423036	2.178632769
SNR13	1.104266269	2.149895101
STP4	1.102863632	2.147805916
MFA1	1.097202988	2.139395168
SNR43	1.081824618	2.116711454
SNR32	1.077937195	2.111015535
SNR77	1.057967918	2.081996896
SOE1	1.03763572	2.052860679
HPF1	1.035073344	2.049217822

SNR3	1.016813035	2.023444158
SNR67	1.004491326	2.006236002

Table 10. Genes up-regulated more than two-fold in H3 K122A mutant strains compared to wild-type in strand-specific microarray expression analyses.

The data in this table reflect the average fold increase in expression of these genes in H3 K122A mutant strains compared to wild-type as measured in high density tiling microarray expression analyses from three biological replicates. Sarah Hainer prepared RNA samples for these analyses. Members of the Steinmetz lab performed microarray hybridization and data analysis. Genes also up-regulated more than two-fold in the microarray expression analysis performed in our lab by Travis Mavrich are highlighted in yellow.

A.3.3 Mapping of nucleosome positions in H3 K122A mutants

Another goal of this project was to correlate the changes in gene expression in the H3 K122A mutant strains with a change in nucleosome occupancy at the promoters of affected genes. To do this, we collaborated with the Pugh lab to map the positions of nucleosomes in an H3 K122A mutant strain across the *S. cerevisiae* genome. Sarah Hainer and Joseph Martens prepared mononucleosomal DNA isolated from H3 K122A and wild-type strains, which was sequenced and mapped to the yeast genome by the Pugh lab. The data generated from samples prepared by Joseph Martens are discussed here.

The results of these nucleosome mapping experiments indicate that globally, there is no significant change in nucleosome positions by metagenomic analysis aligned to transcription start sites (Figure 29A). This is what we would expect based on our previous studies indicating that the H3 K122A mutation affects nucleosome occupancy at highly transcribed regions, but not lowly transcribed regions (HAINER AND MARTENS 2011a). When looking only at the most highly expressed genes (top 10%), a slight reduction in nucleosome occupancy is observed (Figure 29B).

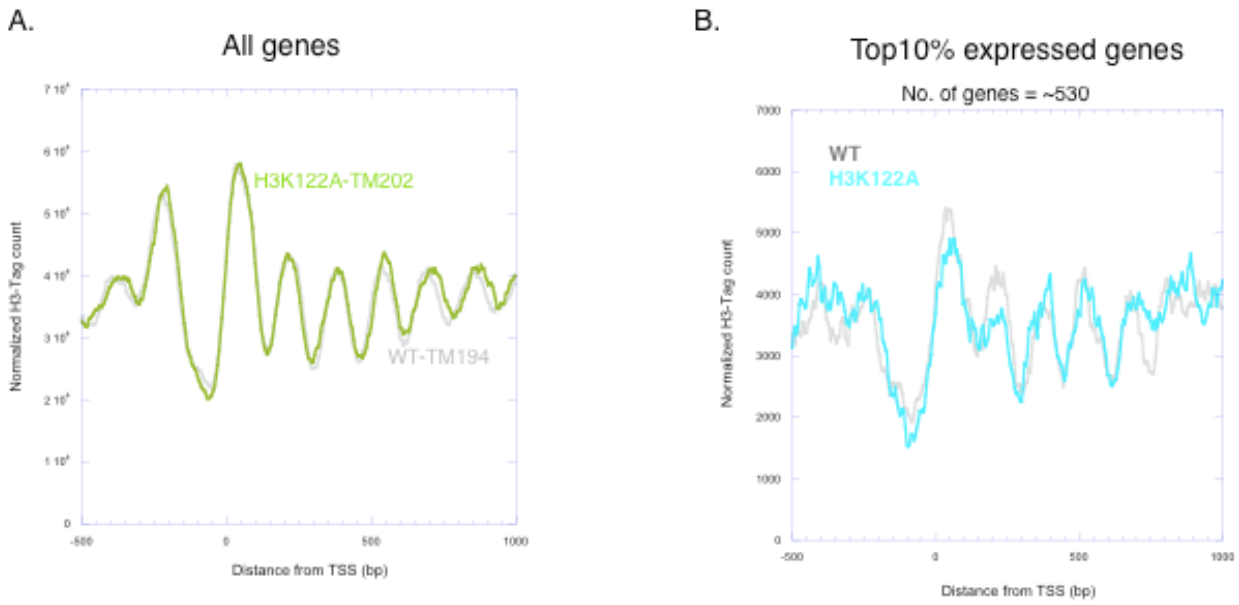

Figure 29. Nucleosome occupancy is decreased in H3 K122A mutant strains compared to wild-type strains specifically at highly transcribed genes.

(A) Metagenomic analysis of H3 occupancy in wild-type (YTM194) and H3 K122A (YTM202) strains of all genes aligned to the transcription start site. (B) Metagenomic analysis of H3 occupancy in wild-type (YTM194) and H3 K122A (YTM202) strains at the top 10% most highly expressed genes aligned to the transcription start site.

A.3.4 Confirmation of gene expression changes in H3 K122A strains

We selected several genes from these genomic analyses to confirm the changes in expression and nucleosome occupancy. Paul Yenerall confirmed the expression of these protein-coding genes by RT-PCR and northern blot analysis. Sarah Hainer confirmed the expression of these protein-coding genes by northern blot analysis and performed chromatin immunoprecipitation of H3, which has yet to be quantified by qPCR. Several of these candidates show evidence of intergenic transcripts across their promoters and northern probes have been designed to examine whether these intergenic transcripts could be playing a role in regulating expression of their neighboring genes.

A.4 DISCUSSION

The data presented in this Appendix have identified a small set of genes that may be regulated by maintenance of transcription-coupled nucleosome occupancy. This was achieved by genome-wide analysis of gene expression and nucleosome position changes in the presence of histone chaperone or histone protein mutations that cause defects in nucleosome occupancy at highly transcribed regions. Interestingly, many of these genes show evidence of transcription occurring across their promoters, which we hypothesize, may be necessary for this regulatory effect. The next step for this project is to integrate the genome-wide expression data with the changes in nucleosome occupancy to determine if we observe a correlation between changes in

nucleosome occupancy and changes in expression of that gene. We anticipate that genes whose expression changes in H3 K122A mutants will also show changes in nucleosome occupancy or perhaps position at that locus. The changes in nucleosome occupancy could have either positive or negative regulatory effects on gene expression. It would be very interesting to take several candidates from this data set and test whether the intergenic transcripts near their promoters are necessary for these regulatory effects as we have observed at the *SER3* locus.

BIBLIOGRAPHY

Adam, M., F. Robert, M. Larochelle and L. Gaudreau, 2001 H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. Mol Cell Biol 21**:** 6270-6279.

Ahn, S. H., M. Kim and S. Buratowski, 2004 Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. Mol Cell 13**:** 67-76.

Arigo, J. T., D. E. Eyler, K. L. Carroll and J. L. Corden, 2006 Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Mol Cell 23**:** 841-851.

Arndt, K. M., and D. Reines, 2015 Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. Annu Rev Biochem.

Ausubel, F. M., 1987 *Current protocols in molecular biology*. Greene Pub. Associates ;

J. Wiley, order fulfillment, Brooklyn, N.Y.

Media, Pa.

Bartel, D. P., 2004 MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116**:** 281-297.

Bartkowiak, B., P. Liu, H. P. Phatnani, N. J. Fuda, J. J. Cooper *et al.*, 2010 CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev 24**:** 2303-2316.

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

Belotserkovskaya, R., S. Oh, V. A. Bondarenko, G. Orphanides, V. M. Studitsky *et al.*, 2003 FACT facilitates transcription-dependent nucleosome alteration. Science 301**:** 1090-1093.

Bevington, S., and J. Boyes, 2013 Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination. EMBO J 32**:** 1381-1392.

Bian, C., C. Xu, J. Ruan, K. K. Lee, T. L. Burke *et al.*, 2011 Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. EMBO J 30**:** 2829-2842.

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

Briggs, S. D., T. Xiao, Z. W. Sun, J. A. Caldwell, J. Shabanowitz *et al.*, 2002 Gene silencing: trans-histone regulatory pathway in chromatin. Nature 418**:** 498.

Buratowski, S., 2009 Progression through the RNA polymerase II CTD cycle. Mol Cell 36**:** 541-546.

Buratowski, S., S. Hahn, L. Guarente and P. A. Sharp, 1989 Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56**:** 549-561.

Callahan, K. P., and J. S. Butler, 2010 TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6. J Biol Chem 285**:** 3540-3547.

Camblong, J., N. Iglesias, C. Fickentscher, G. Dieppois and F. Stutz, 2007 Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell 131**:** 706-717.

Carroll, K. L., R. Ghirlando, J. M. Ames and J. L. Corden, 2007 Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. RNA 13**:** 361-373.

Carrozza, M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson *et al.*, 2005 Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123**:** 581-592.

Castelnuovo, M., S. Rahman, E. Guffanti, V. Infantino, F. Stutz *et al.*, 2013 Bimodal expression of PHO84 is modulated by early termination of antisense transcription. Nat Struct Mol Biol 20**:** 851-858.

Cheung, V., G. Chua, N. N. Batada, C. R. Landry, S. W. Michnick *et al.*, 2008 Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces cerevisiae genome. PLoS Biol 6**:** e277.

Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt and S. Buratowski, 2001 Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes Dev 15**:** 3319-3329.

Chu, Y., R. Simic, M. H. Warner, K. M. Arndt and G. Prelich, 2007 Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. EMBO J 26**:** 4646-4656.

Clapier, C. R., and B. R. Cairns, 2009 The biology of chromatin remodeling complexes. Annu Rev Biochem 78**:** 273-304.

Colin, J., T. Candelli, O. Porrua, J. Boulay, C. Zhu *et al.*, 2014 Roadblock termination by reb1p restricts cryptic and readthrough transcription. Mol Cell 56**:** 667-680.

Consortium, E. P., 2012 An integrated encyclopedia of DNA elements in the human genome. Nature 489**:** 57-74.

Cramer, P., K. J. Armache, S. Baumli, S. Benkert, F. Brueckner *et al.*, 2008 Structure of eukaryotic RNA polymerases. Annu Rev Biophys 37**:** 337-352.

Crisucci, E. M., and K. M. Arndt, 2011 The Roles of the Paf1 Complex and Associated Histone Modifications in Regulating Gene Expression. Genet Res Int 2011.

Cutter, A. R., and J. J. Hayes, 2015 A brief review of nucleosome structure. FEBS Lett 589**:** 2914-2922.

Dai, J., E. M. Hyland, D. S. Yuan, H. Huang, J. S. Bader *et al.*, 2008 Probing nucleosome function: a highly versatile library of synthetic histone H3 and H4 mutants. Cell 134**:** 1066-1078.

Dardalhon, M., W. Lin, A. Nicolas and D. Averbeck, 2007 Specific transcriptional responses induced by 8-methoxypsoralen and UVA in yeast. FEMS Yeast Res 7**:** 866-878.

Davis, C. A., and M. Ares, Jr., 2006 Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 103**:** 3262-3267.

Dover, J., J. Schneider, M. A. Tawiah-Boateng, A. Wood, K. Dean *et al.*, 2002 Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J Biol Chem 277**:** 28368-28371.

Engreitz, J. M., A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman *et al.*, 2013 The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 341**:** 1237973.

Flores, O., H. Lu, M. Killeen, J. Greenblatt, Z. F. Burton *et al.*, 1991 The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc Natl Acad Sci U S A 88**:** 9999-10003.

Formosa, T., 2012 The role of FACT in making and breaking nucleosomes. Biochim Biophys Acta 1819**:** 247-255.

Fox, M. J., H. Gao, W. R. Smith-Kinnaman, Y. Liu and A. L. Mosley, 2015 The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway. PLoS Genet 11**:** e1004999.

Fu, X. D., 2014 Non-coding RNA: a new frontier in regulatory biology. Natl Sci Rev 1**:** 190-204.

Gelfand, B., J. Mead, A. Bruning, N. Apostolopoulos, V. Tadigotla *et al.*, 2011 Regulated antisense transcription controls expression of cell-type-specific genes in yeast. Mol Cell Biol 31**:** 1701-1709.

Gerard, M., L. Fischer, V. Moncollin, J. M. Chipoulet, P. Chambon *et al.*, 1991 Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. J Biol Chem 266**:** 20940-20945.

Ghazal, G., J. Gagnon, P. E. Jacques, J. R. Landry, F. Robert *et al.*, 2009 Yeast RNase III triggers polyadenylation-independent transcription termination. Mol Cell 36**:** 99-109.

Gimeno, C. J., and G. R. Fink, 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.

Ginsburg, D. S., C. K. Govind and A. G. Hinnebusch, 2009 NuA4 lysine acetyltransferase Esa1 is targeted to coding regions and stimulates transcription elongation with Gcn5. Mol Cell Biol 29**:** 6473-6487.

Gudipati, R. K., T. Villa, J. Boulay and D. Libri, 2008 Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nat Struct Mol Biol 15**:** 786-794.

Gupta, R. A., N. Shah, K. C. Wang, J. Kim, H. M. Horlings *et al.*, 2010 Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464**:** 1071-1076.

Gutschner, T., and S. Diederichs, 2012 The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol 9**:** 703-719.

Guttman, M., and J. L. Rinn, 2012 Modular regulatory principles of large non-coding RNAs. Nature 482**:** 339-346.

Hainer, S. J., B. A. Charsar, S. B. Cohen and J. A. Martens, 2012 Identification of Mutant Versions of the Spt16 Histone Chaperone That Are Defective for Transcription-Coupled Nucleosome Occupancy in Saccharomyces cerevisiae. G3 (Bethesda) 2**:** 555-567.

Hainer, S. J., and J. A. Martens, 2011a Identification of histone mutants that are defective for transcription-coupled nucleosome occupancy. Molecular and cellular biology 31**:** 3557-3568.

Hainer, S. J., and J. A. Martens, 2011b Transcription of ncDNA: Many roads lead to local gene regulation. Transcription 2**:** 120-123.

Hainer, S. J., J. A. Pruneski, R. D. Mitchell, R. M. Monteverde and J. A. Martens, 2011 Intergenic transcription causes repression by directing nucleosome assembly. Genes & development 25**:** 29-40.

Hardy, S., P. E. Jacques, N. Gevry, A. Forest, M. E. Fortin *et al.*, 2009 The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. PLoS Genet 5**:** e1000687.

Harrow, J., A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans *et al.*, 2012 GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 22**:** 1760-1774.

Hartzog, G. A., and J. Fu, 2013 The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. Biochim Biophys Acta 1829**:** 105-115.

He, X. J., Z. Y. Ma and Z. W. Liu, 2014 Non-coding RNA transcription and RNA-directed DNA methylation in Arabidopsis. Mol Plant 7**:** 1406-1414.

Hirota, K., T. Miyoshi, K. Kugou, C. S. Hoffman, T. Shibata *et al.*, 2008 Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature 456**:** 130-134.

Hobson, D. J., W. Wei, L. M. Steinmetz and J. Q. Svejstrup, 2012 RNA polymerase II collision interrupts convergent transcription. Mol Cell 48**:** 365-374.

Hocine, S., R. H. Singer and D. Grunwald, 2010 RNA processing and export. Cold Spring Harb Perspect Biol 2**:** a000752.

Hongay, C. F., P. L. Grisafi, T. Galitski and G. R. Fink, 2006 Antisense transcription controls cell fate in Saccharomyces cerevisiae. Cell 127**:** 735-745.

Houseley, J., L. Rubbi, M. Grunstein, D. Tollervey and M. Vogelauer, 2008 A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Mol Cell 32**:** 685-695.

Hung, C. L., L. Y. Wang, Y. L. Yu, H. W. Chen, S. Srivastava *et al.*, 2014 A long noncoding RNA connects c-Myc to tumor metabolism. Proc Natl Acad Sci U S A.

Hwang, W. W., S. Venkatasubrahmanyam, A. G. Ianculescu, A. Tong, C. Boone *et al.*, 2003 A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. Mol Cell 11**:** 261-266.

Jensen, L. T., and V. C. Culotta, 2002 Regulation of Saccharomyces cerevisiae FET4 by oxygen and iron. J Mol Biol 318**:** 251-260.

Joh, R. I., C. M. Palmieri, I. T. Hill and M. Motamedi, 2014 Regulation of histone methylation by noncoding RNAs. Biochim Biophys Acta 1839**:** 1385-1394.

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

Keogh, M. C., S. K. Kurdistani, S. A. Morris, S. H. Ahn, V. Podolny *et al.*, 2005 Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell 123**:** 593-605.

Kim, K. Y., and D. E. Levin, 2011 Mpk1 MAPK association with the Paf1 complex blocks Sen1-mediated premature transcription termination. Cell 144**:** 745-756.

Kim, M., S. H. Ahn, N. J. Krogan, J. F. Greenblatt and S. Buratowski, 2004a Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. EMBO J 23**:** 354-364.

Kim, M., N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedea *et al.*, 2004b The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature 432**:** 517-522.

Kim, T., and S. Buratowski, 2009 Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. Cell 137**:** 259-272.

Kim, T., Z. Xu, S. Clauder-Munster, L. M. Steinmetz and S. Buratowski, 2012 Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. Cell 150**:** 1158-1169.

Kireeva, M. L., W. Walter, V. Tchernajenko, V. Bondarenko, M. Kashlev *et al.*, 2002 Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. Mol Cell 9**:** 541-552.

Kravchenko, J. E., I. B. Rogozin, E. V. Koonin and P. M. Chumakov, 2005 Transcription of mammalian messenger RNAs by a nuclear RNA polymerase of mitochondrial origin. Nature 436**:** 735-739.

Krogan, N. J., J. Dover, A. Wood, J. Schneider, J. Heidt *et al.*, 2003a The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell 11**:** 721-729.

Krogan, N. J., M. Kim, S. H. Ahn, G. Zhong, M. S. Kobor *et al.*, 2002 RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. Mol Cell Biol 22**:** 6979-6992.

Krogan, N. J., M. Kim, A. Tong, A. Golshani, G. Cagney *et al.*, 2003b Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol 23**:** 4207-4218.

Kulaeva, O. I., F. K. Hsieh, H. W. Chang, D. S. Luse and V. M. Studitsky, 2013 Mechanism of transcription through a nucleosome by RNA polymerase II. Biochim Biophys Acta 1829**:** 76-83.

LaCava, J., J. Houseley, C. Saveanu, E. Petfalski, E. Thompson *et al.*, 2005 RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell 121**:** 713-724.

Lardenois, A., Y. Liu, T. Walther, F. Chalmel, B. Evrard *et al.*, 2011 Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. Proc Natl Acad Sci U S A 108**:** 1058-1063.

Lee, J. T., 2009 Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. Genes Dev 23**:** 1831-1842.

Lee, W., D. Tillo, N. Bray, R. H. Morse, R. W. Davis *et al.*, 2007 A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet 39**:** 1235-1244.

LeRoy, G., G. Orphanides, W. S. Lane and D. Reinberg, 1998 Requirement of RSF and FACT for transcription of chromatin templates in vitro. Science 282**:** 1900-1904.

Li, B., M. Carey and J. L. Workman, 2007 The role of chromatin during transcription. Cell 128**:** 707-719.

Li, B., L. Howe, S. Anderson, J. R. Yates, 3rd and J. L. Workman, 2003 The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem 278**:** 8897-8903.

Licatalosi, D. D., G. Geiger, M. Minet, S. Schroeder, K. Cilli *et al.*, 2002 Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. Mol Cell 9**:** 1101-1111.

Liu, C. L., T. Kaplan, M. Kim, S. Buratowski, S. L. Schreiber *et al.*, 2005 Single-nucleosome mapping of histone modifications in S. cerevisiae. PLoS Biol 3**:** e328.

Luger, K., M. L. Dechassa and D. J. Tremethick, 2012 New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat Rev Mol Cell Biol 13**:** 436-447.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, 1997 Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389**:** 251-260.

Luse, D. S., 2014 The RNA polymerase II preinitiation complex. Through what pathway is the complex assembled? Transcription 5**:** e27050.

Lussier, M., A. M. White, J. Sheraton, T. di Paolo, J. Treadwell *et al.*, 1997 Large scale identification of genes involved in cell surface biosynthesis and architecture in Saccharomyces cerevisiae. Genetics 147**:** 435-450.

Martens, J. A., L. Laprade and F. Winston, 2004 Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature 429**:** 571-574.

Martens, J. A., P. Y. Wu and F. Winston, 2005 Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. Genes & development 19**:** 2695-2704.

Martin, D. G., D. E. Grimes, K. Baetz and L. Howe, 2006 Methylation of histone H3 mediates the association of the NuA3 histone acetyltransferase with chromatin. Mol Cell Biol 26**:** 3018-3028.

Marzluff, W. F., and R. J. Duronio, 2002 Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. Curr Opin Cell Biol 14**:** 692-699.

Mason, P. B., and K. Struhl, 2003 The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. Mol Cell Biol 23**:** 8323-8333.

Matsui, T., J. Segall, P. A. Weil and R. G. Roeder, 1980 Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem 255**:** 11992-11996.

Mayer, A., M. Lidschreiber, M. Siebert, K. Leike, J. Soding *et al.*, 2010 Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol 17**:** 1272-1278.

Maze, I., K. M. Noh, A. A. Soshnev and C. D. Allis, 2014 Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nat Rev Genet 15**:** 259-271.

McCullough, L., Z. Connell, C. Petersen and T. Formosa, 2015 The Abundant Histone Chaperones Spt6 and FACT Collaborate to Assemble, Inspect, and Maintain Chromatin Structure in Saccharomyces cerevisiae. Genetics 201**:** 1031-1045.

Mueller, C. L., and J. A. Jaehning, 2002 Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. Mol Cell Biol 22**:** 1971-1980.

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

Naar, A. M., B. D. Lemon and R. Tjian, 2001 Transcriptional coactivator complexes. Annu Rev Biochem 70**:** 475-501.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha *et al.*, 2008 The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320**:** 1344-1349.

Neil, H., C. Malabat, Y. d'Aubenton-Carafa, Z. Xu, L. M. Steinmetz *et al.*, 2009 Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457**:** 1038-1042.

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

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

Nordick, K., M. G. Hoffman, J. L. Betz and J. A. Jaehning, 2008 Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II. Eukaryot Cell 7**:** 1158-1167.

Orphanides, G., G. LeRoy, C. H. Chang, D. S. Luse and D. Reinberg, 1998 FACT, a factor that facilitates transcript elongation through nucleosomes. Cell 92**:** 105-116.

Owen, D. J., P. Ornaghi, J. C. Yang, N. Lowe, P. R. Evans *et al.*, 2000 The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. EMBO J 19**:** 6141-6149.

Pearson, E., and C. Moore, 2014 The evolutionarily conserved Pol II flap loop contributes to proper transcription termination on short yeast genes. Cell Rep 9**:** 821-828.

Pelechano, V., J. Garcia-Martinez and J. E. Perez-Ortin, 2006 A genomic study of the inter-ORF distances in Saccharomyces cerevisiae. Yeast 23**:** 689-699.

Penheiter, K. L., T. M. Washburn, S. E. Porter, M. G. Hoffman and J. A. Jaehning, 2005 A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. Mol Cell 20**:** 213-223.

Pfaffl, M. W., 2001 A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29**:** e45.

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

Pokholok, D. K., N. M. Hannett and R. A. Young, 2002 Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Mol Cell 9**:** 799-809.

Pokholok, D. K., C. T. Harbison, S. Levine, M. Cole, N. M. Hannett *et al.*, 2005 Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122**:** 517-527.

Porrua, O., F. Hobor, J. Boulay, K. Kubicek, Y. D'Aubenton-Carafa *et al.*, 2012 In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. EMBO J 31**:** 3935-3948.

Porrua, O., and D. Libri, 2015 Transcription termination and the control of the transcriptome: why, where and how to stop. Nat Rev Mol Cell Biol 16**:** 190-202.

Pruneski, J. A., S. J. Hainer, K. O. Petrov and J. A. Martens, 2011 The Paf1 complex represses SER3 transcription in Saccharomyces cerevisiae by facilitating intergenic transcription-dependent nucleosome occupancy of the SER3 promoter. Eukaryotic cell 10**:** 1283-1294.

Ramachandran, S., and S. Henikoff, 2015 Replicating Nucleosomes. Sci Adv 1.
Reeder, R. H., and R. G. Roeder, 1972 Ribosomal RNA synthesis in isolated nuclei. J Mol Biol 67**:** 433-441.
Rhee, H. S., and B. F. Pugh, 2011 Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147**:** 1408-1419.
Rhee, H. S., and B. F. Pugh, 2012 Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature 483**:** 295-301.
Rinn, J., and M. Guttman, 2014 RNA Function. RNA and dynamic nuclear organization. Science 345**:** 1240-1241.
Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu *et al.*, 2007 Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129**:** 1311-1323.
Rio, D. C., 2011 *RNA : a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
Robzyk, K., J. Recht and M. A. Osley, 2000 Rad6-dependent ubiquitination of histone H2B in yeast. Science 287**:** 501-504.
Roeder, R. G., and W. J. Rutter, 1969 Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature 224**:** 234-237.
Rondon, A. G., H. E. Mischo, J. Kawauchi and N. J. Proudfoot, 2009 Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. Mol Cell 36**:** 88-98.
Ryan, O., R. S. Shapiro, C. F. Kurat, D. Mayhew, A. Baryshnikova *et al.*, 2012 Global gene deletion analysis exploring yeast filamentous growth. Science 337**:** 1353-1356.
Sadowski, M., B. Dichtl, W. Hubner and W. Keller, 2003 Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. EMBO J 22**:** 2167-2177.
Santaguida, S., and A. Musacchio, 2009 The life and miracles of kinetochores. EMBO J 28**:** 2511-2531.
Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein *et al.*, 2002 Active genes are tri-methylated at K4 of histone H3. Nature 419**:** 407-411.
Sawadogo, M., and R. G. Roeder, 1985 Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. Proc Natl Acad Sci U S A 82**:** 4394-4398.
Schaeffer, L., V. Moncollin, R. Roy, A. Staub, M. Mezzina *et al.*, 1994 The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. EMBO J 13**:** 2388-2392.
Schaeffer, L., R. Roy, S. Humbert, V. Moncollin, W. Vermeulen *et al.*, 1993 DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. Science 260**:** 58-63.
Schaft, D., A. Roguev, K. M. Kotovic, A. Shevchenko, M. Sarov *et al.*, 2003 The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. Nucleic Acids Res 31**:** 2475-2482.
Schaughency, P., J. Merran and J. L. Corden, 2014 Genome-wide mapping of yeast RNA polymerase II termination. PLoS Genet 10**:** e1004632.
Schneider, B. L., W. Seufert, B. Steiner, Q. H. Yang and A. B. Futcher, 1995 Use of polymerase chain reaction epitope tagging for protein tagging in Saccharomyces cerevisiae. Yeast 11**:** 1265-1274.

Schneider, C., and D. Tollervey, 2013 Threading the barrel of the RNA exosome. Trends Biochem Sci 38**:** 485-493.

Schulz, D., B. Schwalb, A. Kiesel, C. Baejen, P. Torkler *et al.*, 2013 Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell 155**:** 1075-1087.

Seila, A. C., L. J. Core, J. T. Lis and P. A. Sharp, 2009 Divergent transcription: a new feature of active promoters. Cell Cycle 8**:** 2557-2564.

Sheldon, K. E., D. M. Mauger and K. M. Arndt, 2005 A Requirement for the Saccharomyces cerevisiae Paf1 complex in snoRNA 3' end formation. Mol Cell 20**:** 225-236.

Shi, X., M. Chang, A. J. Wolf, C. H. Chang, A. A. Frazer-Abel *et al.*, 1997 Cdc73p and Paf1p are found in a novel RNA polymerase II-containing complex distinct from the Srbp-containing holoenzyme. Mol Cell Biol 17**:** 1160-1169.

Shirra, M. K., S. E. Rogers, D. E. Alexander and K. M. Arndt, 2005 The Snf1 protein kinase and Sit4 protein phosphatase have opposing functions in regulating TATA-binding protein association with the Saccharomyces cerevisiae INO1 promoter. Genetics 169**:** 1957-1972.

Shuman, S., 2001 Structure, mechanism, and evolution of the mRNA capping apparatus. Prog Nucleic Acid Res Mol Biol 66**:** 1-40.

Simic, R., D. L. Lindstrom, H. G. Tran, K. L. Roinick, P. J. Costa *et al.*, 2003 Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. EMBO J 22**:** 1846-1856.

Simon, J. A., and R. E. Kingston, 2013 Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. Mol Cell 49**:** 808-824.

Singer, V. L., C. R. Wobbe and K. Struhl, 1990 A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes Dev 4**:** 636-645.

Spizzo, R., M. I. Almeida, A. Colombatti and G. A. Calin, 2012 Long non-coding RNAs and cancer: a new frontier of translational research? Oncogene 31**:** 4577-4587.

Squazzo, S. L., P. J. Costa, D. L. Lindstrom, K. E. Kumer, R. Simic *et al.*, 2002 The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. EMBO J 21**:** 1764-1774.

Srikantan, V., Z. Zou, G. Petrovics, L. Xu, M. Augustus *et al.*, 2000 PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. Proc Natl Acad Sci U S A 97**:** 12216-12221.

Steinmetz, E. J., C. L. Warren, J. N. Kuehner, B. Panbehi, A. Z. Ansari *et al.*, 2006 Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol Cell 24**:** 735-746.

Strahl, B. D., P. A. Grant, S. D. Briggs, Z. W. Sun, J. R. Bone *et al.*, 2002 Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol Cell Biol 22**:** 1298-1306.

Struhl, K., 2007 Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol 14**:** 103-105.

Sudarsanam, P., V. R. Iyer, P. O. Brown and F. Winston, 2000 Whole-genome expression analysis of snf/swi mutants of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 97**:** 3364-3369.

Sun, Z. W., and C. D. Allis, 2002 Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature 418**:** 104-108.

Talbert, P. B., and S. Henikoff, 2010 Histone variants--ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol 11**:** 264-275.

Thiebaut, M., E. Kisseleva-Romanova, M. Rougemaille, J. Boulay and D. Libri, 2006 Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol Cell 23**:** 853-864.

Tomson, B. N., E. M. Crisucci, L. E. Heisler, M. Gebbia, C. Nislow *et al.*, 2013 Effects of the Paf1 complex and histone modifications on snoRNA 3'-end formation reveal broad and locus-specific regulation. Mol Cell Biol 33**:** 170-182.

Tomson, B. N., C. P. Davis, M. H. Warner and K. M. Arndt, 2011 Identification of a role for histone H2B ubiquitylation in noncoding RNA 3'-end formation through mutational analysis of Rtf1 in Saccharomyces cerevisiae. Genetics 188**:** 273-289.

Tsai, M. C., O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang *et al.*, 2010 Long noncoding RNA as modular scaffold of histone modification complexes. Science 329**:** 689-693.

Tudek, A., O. Porrua, T. Kabzinski, M. Lidschreiber, K. Kubicek *et al.*, 2014 Molecular basis for coordinating transcription termination with noncoding RNA degradation. Mol Cell 55**:** 467-481.

van Attikum, H., and S. M. Gasser, 2009 Crosstalk between histone modifications during the DNA damage response. Trends Cell Biol 19**:** 207-217.

van Dijk, E. L., C. L. Chen, Y. d'Aubenton-Carafa, S. Gourvennec, M. Kwapisz *et al.*, 2011 XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature 475**:** 114-117.

van Leeuwen, F., P. R. Gafken and D. E. Gottschling, 2002 Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell 109**:** 745-756.

Vanacova, S., J. Wolf, G. Martin, D. Blank, S. Dettwiler *et al.*, 2005 A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol 3**:** e189.

Vasiljeva, L., and S. Buratowski, 2006 Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. Mol Cell 21**:** 239-248.

Vasiljeva, L., M. Kim, H. Mutschler, S. Buratowski and A. Meinhart, 2008 The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat Struct Mol Biol 15**:** 795-804.

Wade, P. A., W. Werel, R. C. Fentzke, N. E. Thompson, J. F. Leykam *et al.*, 1996 A novel collection of accessory factors associated with yeast RNA polymerase II. Protein Expr Purif 8**:** 85-90.

Warner, M. H., K. L. Roinick and K. M. Arndt, 2007 Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification. Mol Cell Biol 27**:** 6103-6115.

Waters, B. M., and D. J. Eide, 2002 Combinatorial control of yeast FET4 gene expression by iron, zinc, and oxygen. J Biol Chem 277**:** 33749-33757.

Watson, J. D., and F. H. Crick, 1953 Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171**:** 737-738.

Weinmann, R., H. J. Raskas and R. G. Roeder, 1974 Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. Proc Natl Acad Sci U S A 71**:** 3426-3439.

Weinmann, R., and R. G. Roeder, 1974 Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes. Proc Natl Acad Sci U S A 71**:** 1790-1794.
West, S., N. Gromak and N. J. Proudfoot, 2004 Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature 432**:** 522-525.
Wilson, M. D., M. Harreman and J. Q. Svejstrup, 2013 Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. Biochim Biophys Acta 1829**:** 151-157.
Winston, F., C. Dollard and S. L. Ricupero-Hovasse, 1995 Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast 11**:** 53-55.
Wlotzka, W., G. Kudla, S. Granneman and D. Tollervey, 2011 The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J 30**:** 1790-1803.
Wood, A., N. J. Krogan, J. Dover, J. Schneider, J. Heidt *et al.*, 2003a Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol Cell 11**:** 267-274.
Wood, A., J. Schneider, J. Dover, M. Johnston and A. Shilatifard, 2003b The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. J Biol Chem 278**:** 34739-34742.
Wyce, A., T. Xiao, K. A. Whelan, C. Kosman, W. Walter *et al.*, 2007 H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex. Mol Cell 27**:** 275-288.
Wyers, F., M. Rougemaille, G. Badis, J. C. Rousselle, M. E. Dufour *et al.*, 2005 Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 121**:** 725-737.
Wyrick, J. J., F. C. Holstege, E. G. Jennings, H. C. Causton, D. Shore *et al.*, 1999 Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. Nature 402**:** 418-421.
Xiao, T., H. Hall, K. O. Kizer, Y. Shibata, M. C. Hall *et al.*, 2003 Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev 17**:** 654-663.
Xu, Z., W. Wei, J. Gagneur, S. Clauder-Munster, M. Smolik *et al.*, 2011 Antisense expression increases gene expression variability and locus interdependency. Mol Syst Biol 7**:** 468.
Xu, Z., W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Munster *et al.*, 2009 Bidirectional promoters generate pervasive transcription in yeast. Nature 457**:** 1033-1037.
Zentner, G. E., and S. Henikoff, 2013 Regulation of nucleosome dynamics by histone modifications. Nat Struct Mol Biol 20**:** 259-266.
Zhang, H., F. Rigo and H. G. Martinson, 2015 Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Mol Cell 59**:** 437-448.
Zhang, Z., J. Fu and D. S. Gilmour, 2005 CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. Genes Dev 19**:** 1572-1580.
Zhu, B., S. S. Mandal, A. D. Pham, Y. Zheng, H. Erdjument-Bromage *et al.*, 2005 The human PAF complex coordinates transcription with events downstream of RNA synthesis. Genes Dev 19**:** 1668-1673.