AN EXAMINATION OF THE MECHANISM AND CONSERVATION OF ADJACENT
GENE CO-REGULATION IN S. CEREVISIAE AND HIGHER EUKARYOTES

By

Jeffrey R. Arace

Faculty Advisor: Dr. Michael A. McAlear

A Dissertation submitted to the Faculty of Wesleyan University in partial
fulfillment of the requirements for the degree of Master of Arts

Middletown, Connecticut  September 2013
Acknowledgments

Three years ago, I spent my entire summer doing work-study in the MB&B department at Wesleyan. After working in each of the labs in the department, I knew that I wanted to join the McAlear lab. At the end of the summer I finally worked up enough courage and asked Dr. McAlear if I could join his lab. Over the last three years and especially these last few months, I have grown immensely as a student, a scientist, and as a person under his guidance. Without his knowledge, seemingly endless analogies, and constructive criticism none of this work would have been possible. I cannot thank him enough for the opportunity he has given me.

I would also like to thank all the past and present McAlear lab members, especially James Arnone and Sara Kass-Gergi. I am extremely grateful for these two and all the support and advice they have given me along the way.

Finally, I would like to thank my friends and family for their unwavering support throughout the past few years. Without their encouragement, I would not be here today.
# Table of Contents

**Acknowledgements**

**Table of Contents**  
1

**Index of Figures**  
2

**Index of Tables**  
3

**Abstract**  
4

**Introduction**  
5

Section 1: Budding yeast as a model system  
6

Section 2: Transcriptional regulation in eukaryotes  
12

Section 3: The Ribosome Biogenesis Pathway  
26

Section 4: Adjacent Gene Co-regulation  
36

**Materials and Methods**  
43

**Results**  
50

Section 1: Adjacent positioning of co-regulated genes is widely conserved  
50

Section 2: The ability of the *MPP10* promoter to regulate transcription of its adjacent gene is specific to *YJR003C*  
55

Section 3: Co-regulation of the *MPP10-YJR003C* gene pair is not mediated by nucleosome re-positioning  
57

**Discussion**  
65

**References**  
76
Index of Figures

**Figure 1:** Preinitiation (PIC) complex assembly

**Figure 2:** Nucleosome architecture of a canonical yeast gene

**Figure 3:** Chromatin remodeling complexes regulate transcription by exposing or occluding transcription factor binding sites

**Figure 4:** Examples of proteins with domains that recognize specific histone modifications

**Figure 5:** Structure of the *S. cerevisiae* ribosome

**Figure 6:** Nutrient signaling pathways control ribosome biogenesis

**Figure 7:** Gene networks required for ribosome biogenesis

**Figure 8:** The *MPP10*-*YJR003C* locus

**Figure 9:** The transcriptional response of *MPP10* and *YJR003C* is controlled from the promoter of *MPP10*

**Figure 10:** Proper transcriptional response of *YJR003C* depends on its immediate adjacency to *MPP10*

**Figure 11:** Active LEU2 transcription is required for the co-regulation of *MPP10* and *YJR003C*

**Figure 12:** Phylogenetic relationship of the eukaryotic lineages in this study

**Figure 13:** Transcription of *KANr* is not regulated from the promoter of *MPP10*

**Figure 14:** Determination of *GAL* NB protected mononucleosome sized MNase-I digested chromatin
Figure 15: Stress-response associated nucleosome displacement occurs at the HSP104 promoter

Figure 16: Nucleosome mapping at the MPP10 and YJR003C promoter regions

Figure 17: Nucleosome mapping in the snf2Δ mutant

Index of Tables

Table 1: Yeast strains used in this study

Table 2: Primer pairs used in the nucleosome scanning assay

Table 3: Gene pairing is a general feature of many co-regulated gene families in S. cerevisiae

Table 5. Gene ontologies that did not show significant gene pair adjacencies

Table 4: RRB and RP genes exist as immediately adjacent pairs in many eukaryotes
**Abstract**

Coordinated cell growth and development largely depends on the appropriate regulation of ribosome production. In budding yeast, ribosome production is primarily regulated at the level of transcription. This process involves the coordinated expression of three independently co-regulated gene sets that are necessary to produce the 79 ribosomal proteins (RPs), the 4 heavily processed and modified rRNAs, and the rRNA and ribosome biosynthesis (RRB) genes. Interestingly, in a variety of yeast species, a significant fraction of the RRB and RP genes exist throughout the genome as immediately adjacent gene pairs. Further analysis revealed that the set of paired genes in both regulons are much more tightly co-regulated than the unpaired genes are during changing cellular conditions. Here, we extend this initial observation and provide evidence for adjacent gene pairing in a variety of other functionally related gene sets in yeast. Moreover, we discovered adjacent gene pairing in the ribosome production pathways of higher eukaryotes, including humans. Previous genetic analysis of the yeast *MPP10-YJR003C* RRB gene pair has shown that the transcriptional response of both genes is controlled from the *MPP10* promoter. Our results show that the ability of the *MPP10* promoter to control the transcription of its neighboring gene is specific to *YJR003C*. Finally, we find that the co-regulation of this gene pair is not mediated by nucleosome remodeling events at either promoter.
Introduction

Section 1: Budding yeast as model organism

Model organisms are extremely valuable in the field of biological research. The properties of model organisms—rapid growth, genetic manipulability, production of large numbers of offspring, and ease of cultivation—make them ideal for studying the fundamental properties of living cells. With the availability of inbred lines, tried and true procedures, and communities of shared resources, many evolutionarily conserved processes, such as cell growth and division, metabolism, and gene expression, can easily be studied in model organisms. The bacterium *Escherichia coli*, the budding yeast *Saccharomyces cerevisiae*, the nematode worm *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, and the mouse *Mus musculus* are among the oldest and most studied model systems to date. Each of these organisms has unique properties that make them ideal for specific areas of study.

*S. cerevisiae* has proven itself to be an excellent model system. As a single cellular eukaryote, *S. cerevisiae* shares many of the benefits of working with bacteria. Yeast are inexpensive, easy to grow, and have a short life span. They are non-pathogenic and can be handled with relatively few precautions. These benefits, combined with the relative simplicity of the yeast genome, make it perfect for conducting biochemical and genetic experiments.
As a “surprisingly typical” eukaryote, yeast shares the fundamental characteristics of all multicellular organisms such as membrane bound organelles, cytoskeletal organization, signal transduction pathways, second messenger systems, an ATP/NADH based metabolic economy, and chromosome mechanics. Its genetic information is stored in DNA, which is located in the nucleus and organized on chromosomes. DNA is transcribed into RNA by one of three RNA polymerase complexes and the RNAs are processed, capped, and polyadenylated before they are translated by ribosomes into protein (Struhl, 1983). Altogether, the properties of budding yeast make it an ideal system to which classical and molecular genetics and biochemistry techniques can be applied in order to elucidate the relationship between gene structure and protein function (Botstein and Fink, 1988).

**Genetic and biochemical approaches in yeast**

The impact of classical yeast genetic studies over the last 70 years cannot be overemphasized. The existence of a stable haploid and diploid state, a rare characteristic in microorganisms, is extremely beneficial for geneticists. Haploid strains are ideal for isolating and studying recessive mutations, whereas diploid strains allow for complementation tests and the analysis of mutations of essential genes (Sherman, 1991). By observing the phenotypes of the four meiotic products (spores) that result from the sporulation of a diploid yeast cell, geneticists can determine the relative location of genes on each chromosome.
This analysis allowed Lindegren and his colleagues to publish the first genetic map of *S. cerevisiae* in 1949 (Lindegren, 1949).

Another benefit of yeast in the field of classical genetics is the relative ease with which mutants can be isolated. Geneticists can randomly mutagenize a population of yeast cells, isolate mutants with interesting phenotypes, and further characterize them through mating analysis. Conditional mutations that only affect cell viability under a certain set of environmental conditions can be isolated to study the effect of mutations in essential genes (Bjornsti, 2002). For example, temperature-sensitive mutants impair cell viability at high (non-permissive) temperatures but have little or no effect on viability at lower (permissive) temperatures. This property allows geneticists to study mutations that would not be found in a broad mutagenesis screen. Another classical yeast genetics approach to identifying new genes is looking for mutations that suppress or enhance the phenotypes of other mutations. The existence of suppressor or enhancer mutations can be used to identify interactions, both direct and indirect, between two gene products (Struhl, 1983).

The breakthrough of recombinant DNA technology, originally pioneered in *E. coli* in the 1970s, radically transformed yeast genetics from its classical beginnings to a modern molecular field (Botstein and Fink, 1988). Recombinant DNA technology allows scientists to isolate and clone specific regions of the yeast genome. Once isolated, the DNA can be mutated *in vitro* and introduced back into the yeast genome. The versatility of *S. cerevisiae* allows researchers
different ways to introduce cloned DNA into cells. Cloned DNA can be introduced as part of a low copy plasmid that contains an autonomously replicating sequence (ARS) derived from endogenous yeast plasmids. The addition of yeast centromeric sequences to the plasmid allows the cloned DNA to segregate properly during mitosis and to be passed on to daughter cells (Struhl, 1983).

Transforming DNA can also be integrated directly into the yeast genome through homologous recombination. The relatively high rate of homologous recombination and the availability of selectable and counter-selectable markers in yeast allows researchers to insert foreign DNA into the yeast genome by flanking it with cloned yeast sequences. Additionally, homologous recombination can be exploited to replace wild-type alleles with mutant alleles or to delete an allele completely (Rothstein, 1991). This approach can also be used to insert inducible promoters upstream of a given gene or to introduce an epitope/GFP tag within an ORF (Bjornsti, 2002). The ease and precision with which the yeast genome can be genetically manipulated has allowed for the development of a vast array of techniques useful for studying the biochemical function of gene products, the biological consequences of gene mutations, and the relationship between gene structure and protein function (Botstein and Fink, 2011).

In addition to genetics, yeast is an ideal model for biochemical studies. Genes encoding proteins of interest can be cloned into plasmids and overexpressed in yeast. Biochemists can then lyse the cells and purify the
desired protein. The rapid doubling time of *S. cerevisiae*—approximately 90 minutes in rich medium—allows biochemists to grow large cultures within a short period of time, which is crucial when purifying proteins expressed at low levels. In addition to structural studies, biochemists use a variety of techniques such as crosslinking, co-immunoprecipitation, and co-fractionation to study protein-protein interactions in yeast. The development of the two hybrid system allows researchers to identify protein-protein interactions in yeast by simple galactose selection (Fields and Song, 1989). Over the past 20 years this system has evolved and has been used to determine a wide variety of protein-protein interactions (Botstein and Fink, 2011).

**Genomics and yeast**

In addition to being the premier eukaryotic model for genetic and biochemical studies, *S. cerevisiae* has paved the way for entirely new fields such as systems biology and functional genomics. Following the publication of the complete *S. cerevisiae* genome in 1996, researchers have focused on studying the functions and interactions of genes and proteins on a genome wide level (Botstein and Fink, 2011).

The genome of *S. cerevisiae* was the first of any eukaryotic organism to be completely sequenced. Altogether, the haploid *S. cerevisiae* genome is $1.2 \times 10^7$ bp long and contains some 5,885 protein coding genes on 16 chromosomes. With an average gene size of 1.5 kb and relatively few introns, the yeast genome is much more compact (1 ORF/2kb) than higher eukaryotes such as *C. elegans* (1
ORF/6kb) or humans (1 ORF/30kb). In addition to the ~6000 proteins coding genes, the yeast genome contains approximately 140 tandem ribosomal RNA repeats on chromosome XII as well as 40 small nuclear RNA genes and 275 transfer RNA genes scattered throughout the genome (Goffeau et al., 1996).

With the genome completely sequenced, researchers developed novel techniques for measuring gene and protein activity on a genome-wide scale. A new strategy using DNA microarrays was developed to measure genome-wide differential gene expression levels. By hybridizing fluorescently-labeled cDNA to an array of oligonucleotide probes representing each predicted yeast ORF, scientists are able to examine gene expression under a variety of different conditions (Lashkari et al., 1997). DNA microarray analysis combined with chromatin immunoprecipitation (ChIP) can be used to determine the precise location of DNA binding proteins throughout the entire yeast genome (Ren et al., 2000). This technique, known as ChIP-on-chip, is extremely useful for identifying transcriptional regulatory networks in yeast (Lee et al., 2002). In 2002, Giaever and colleagues constructed an almost complete collection of yeast deletion mutants (96% of annotated ORFs). This collection allows researchers to study the effect of loss of gene function for nearly any gene in the yeast genome (Giaever et al., 2002). Similarly, green fluorescent protein (GFP) fusion libraries have been created that allow researchers to visualize the subcellular location of most yeast proteins (Huh et al., 2003).
In 1996, a basic understanding of the biological role was known for 30% of the 6,000 yeast genes. Today, this number has grown to over 85%, higher than any other eukaryote (Botstein and Fink, 2011). Furthermore, 60% of yeast genes have human homologs and 25% of cloned human disease genes have a cloned yeast homolog. In fact, many genes and proteins that cause human disease are studied in yeast. For example, the causative genes for hereditary nonpolyposis colorectal cancer (HNPCC), hMSH2 and hMLH1, were first identified in yeast as DNA mismatch repair proteins (Strand et al., 1993).

Given the observed conservation among amino acid sequence and protein function across yeast and other eukaryotic organisms, such as humans, scientists reasoned that annotated protein function could be transferred from one species to another. In 2000, the Gene Ontology (GO) Consortium was created to organize and systemize information about the biological processes, cellular components, and molecular functions of proteins from all eukaryotes (Ashburner et al., 2000). Unsurprisingly, the majority of catalogued GO annotations are derived from studies conducted in yeast (Botstein and Fink, 2011).

**Section 2: Transcriptional regulation in eukaryotes**

The molecular, genetic, and biochemical advantages of yeast make it an excellent organism for studying gene regulation. In eukaryotes, a significant amount of gene regulation occurs at the transcriptional level. The process of transcription
involves three different families of multi-protein RNA polymerase complexes: RNA Pol I, RNA Pol II, and RNA Pol III. RNA Pol I and III are found in the nucleolus and are responsible for transcribing the 35S pre-rRNA precursor and the 5S pre-rRNA precursor, respectively. RNA Pol III is also responsible for transcribing the tRNA genes. RNA Pol II, on the other hand, transcribes protein-coding genes into mRNA (Grummt, 2003; Smale and Kadonaga, 2003; White, 2011). Due to the high level of conservation of these processes, most of the mechanisms that control transcription in yeast are similar across eukaryotes.

The RNA Pol II Core Promoter

Transcription of a particular protein-coding gene requires RNA Pol II to be correctly positioned at the transcriptional start site. There are many factors, both cis- and trans-, that are involved in the recruitment of RNA Pol II to the correct initiation site. One of these factors is the core promoter, a region of DNA that extends ~35 bp upstream and/or downstream of the transcription start site responsible for directing the accurate initiation of transcription by RNA Pol II. The sequences found in the core promoter are responsible for binding factors that facilitate and allow RNA Pol II to bind. Additionally, the core promoter is responsible for interpreting activating or repressing signals from factors that bind at distal sites (Smale and Kadonaga, 2003).

The core promoter of many eukaryotic genes contains an upstream sequence called a ‘TATA’ box that is found 25-30 bp upstream of the initiation site. Interestingly, in S. cerevisiae, the TATA box can be found up to 120 bp
upstream from the start site (Smale and Kadonaga, 2003). Although initial estimates of yeast promoters containing the TATA sequence were low (20%), recent bioinformatics analysis has revealed that over 99% of yeast gene promoters have a TATA-like sequence that contain only 1 or 2 mismatches to the consensus sequence (Rhee and Pugh, 2012).

The TATA box is recognized by the TATA binding protein (TPB), a component of the transcription factor TFIID. Once TBP is bound, it recruits the basal transcription machinery to the core promoter. The basal transcription machinery is defined as the factors that are minimally essential for transcription to occur in vitro and includes the general transcription factors TFIID, TFIIA, TFIIB, TFIIE, TFIIF, and TFIIH as well as RNA Pol II. When the basal transcription machinery assembles at the core promoter, the preinitiation complex (PIC) is formed (Figure 1) (Smale and Kadonaga, 2003).

In addition to TFIID, the SAGA complex, a large multifunctional coactivator, also has the ability of delivering TBP to gene promoters. Along with the basal transcription machinery and TFIID, the majority of yeast genes are dependent upon the SAGA complex for their proper transcription (Huisinga and Pugh, 2004). In fact, genome-wide examinations have revealed that over 99% of the measurable yeast genome is positively regulated by the overlapping contribution of SAGA and TFIID. The SAGA complex contains five distinct modules: a recruitment module (Tra1), an acetylation module (Gcn5, Ada2, Ada3), a TBP interaction unit (Spt3, Spt8), a deubiquitination module (Ubp8,
Sus1, Sgf11 and Sgf73), and an architecture unit (Spt7, Spt20, Ada1, TAF5, -6, -9 and -12) (Koutelou et al., 2010). Once recruited to the gene promoter, Gcn5 acetylates lysine residues on the N-terminal tails of histone proteins, which allows for additional transcription factors to bind. Additionally, the Spt3 subunit has the ability to recruit TBP, further contributing to the formation of the PIC (Koutelou et al., 2010).

**Figure 1. Preinitiation (PIC) complex assembly.** (A) The TATA sequence is located in the promoter region. (B) The TBP-containing transcription factor TFIID recognizes and binds to the TATA sequence. (C) TFIID binding allows for the subsequent binding of the general transcription factors (TFIIB, TFIIE, TFIIF, and TFIH) and RNA Pol II. (D) The PIC is complete once all of these factors are assembled. (Strachan and Read, 2011)
In addition to the basal transcription machinery, cis- regulatory elements and the trans- acting sequence-specific factors that bind to them also control gene expression in eukaryotes. The cis- regulatory elements consist of regulatory motifs found upstream of the core promoter. In S. cerevisiae, these elements are typically found several hundred base pairs upstream of the translation start site and are highly conserved (Gasch et al., 2000; Harbison et al., 2004). Sequence-specific transcription factors recognize and bind these elements, ultimately assisting in the recruitment of RNA Pol II to the initiation site. Transcription factors can behave as either activators or repressors of transcription. Under different environmental conditions, different combinations of activators and repressors will bind in order to ensure that the gene is transcribed at the proper level (Venters and Pugh, 2009). Genome wide studies of transcription factor binding have revealed complex gene regulatory networks that cells use to co-regulate large sets of genes under certain conditions (Lee et al., 2002).

**Chromatin management and its effects on gene regulation**

Eukaryotic cells are faced with the daunting task of fitting extremely long linear strands of DNA into a relatively small membrane-bound nucleus. For example, a human cell contains 2 meters of DNA that must fit into a nucleus that ranges from 5 to 20 μm (Jansen and Verstrepen, 2011). In order to accomplish this, linear DNA is wrapped around histone proteins. Histone proteins are highly basic and are attracted to negatively charged DNA. This favorable interaction
between DNA and histone proteins allows DNA to be condensed about 10,000 fold (Jansen and Verstrepen, 2011). In eukaryotes, 147 bp of DNA is wrapped 1.65 times around a histone octamer containing two copies of each of the four core histones (H2A, H2B, H3, and H4). Each histone protein is comprised of a globular domain and an unstructured amino terminal ‘tail’ that extends from the surface and is capable of being post-translationally modified. Nucleosomes exist as repeating units joined by short stretches of DNA called linker DNA that is bound by the linker histone H1 (Jansen and Verstrepen, 2011). Before transcription of a eukaryotic gene occurs, the chromatin at the locus is decondensed in order to allow the basal transcription machinery access to the promoter. In general, there are two distinct chromatin states found in eukaryotic genomes: euchromatin and heterochromatin. Euchromatin is characterized as a relatively relaxed environment. This ‘open’ region of the genome is where most actively transcribed genes are found. In contrast, heterochromatin is a much more compact and condensed chromatin structure where most inactive genes are found (Bannister and Kouzarides, 2011). The packaging of DNA into chromatin not only allows the cell to fit the entirety of its DNA inside the nucleus, it also provides the cell with another level of transcriptional regulation.

Initial studies showed that nucleosomes have an overall repressive effect on transcription in vitro and in vivo (Han and Grunstein, 1988; Lorch et al., 1987). It is now understood that nucleosomes affect gene expression by controlling DNA accessibility. In genomic regions where nucleosome occupancy
is high, DNA accessibility is low, making it difficult for transcription factors, the basal transcription machinery and RNA polymerase to bind. In regions where occupancy is low, DNA is accessible to these factors and transcription can occur (Tsankov et al., 2010)

Genome-wide nucleosome mapping studies have shown that nucleosome position in eukaryotes is anything but random. Proper genome function and gene regulation require nucleosomes to be properly positioned. In *S. cerevisiae*, over 90% of gene promoters contain regions of DNA with relatively low nucleosome occupancy. These stretches of DNA, called nucleosome free regions (NFRs), are roughly 150 bp in length and found in the promoter region (Figure 2). Many of the *cis*-regulatory elements, such as transcription factor binding sites and TATA boxes, are found within the NFR and therefore accessible to various DNA binding proteins (Struhl, 1985; Yuan et al., 2005). Flanking the NFR are the strongly positioned -1 and +1 nucleosomes. The -1 nucleosome is defined as the first nucleosome located upstream of the NFR and is responsible for controlling the accessibility of regulatory elements at that position.

Immediately downstream of the NFR is the +1 nucleosome. This nucleosome architecture is generally covered across all eukaryotic species (Jansen and Verstrepen, 2011). As transcription occurs, the +1 nucleosome is evicted but then rapidly returns to its position after RNA Pol II passes by (Jiang and Pugh, 2009). Additionally, genome scale mapping studies have revealed the presence of a 3’ NFR at the end of most genes where RNA Pol II terminates transcription.
However, the functionality of the 3’ NFR remains to be studied (Figure 2) (Jansen and Verstrepen, 2011).

**Figure 2. Nucleosome architecture of a canonical yeast gene.** The ~150 bp 5’ nucleosome free region (5' NFR), which contains the majority of the cis-regulatory elements, is surrounded by the strongly positioned -1 and +1 nucleosomes. The +1 nucleosome is located immediately downstream of the NFR. A 3’ NFR is located downstream of the coding region. Red arrow depicts the transcription start site (TSS) (Jansen and Verstrepen, 2011)

Studies have shown that local DNA sequences play an important role in directing nucleosome positions. In order for linear DNA to be wrapped around the histone octamer, its helical structure must be severely bent. It is believed that GC-rich sequences facilitate nucleosome formation by increasing DNA flexibility. On the other hand, poly (dA:dT) sequences increase the rigidity of DNA and behave as nucleosome excluding elements. As it turns out, NFRs are often enriched for one or more these poly (dA:dT) sequences (Anderson and Widom, 2001; Iyer and Struhl, 1995). AT content has proven to be a strong predictor of nucleosome occupancy; the higher the AT content the lower the nucleosome occupancy. Presumably, these AT rich regions prevent nucleosome
formation because the relatively rigid DNA helix is more resistant to bending (Jansen and Verstrepen, 2011).

**Trans- factors involved in nucleosome positioning**

The difference between genome-scale maps of *in vitro* nucleosome positions (where the effects of *trans-* factors are eliminated and nucleosome position is determined entirely on the intrinsic DNA sequence) and *in vivo* nucleosome positions indicates that nucleosome position is not determined by nucleic acid sequence alone. *Trans-* acting factors such as RNA polymerase, transcription factors, chromatin remodeling complexes, histone variants, and posttranslational histone modifications all contribute to nucleosome positioning (Jansen and Verstrepen, 2011).

Active transcription by RNA polymerase affects nucleosome positions in the promoter, coding, and terminator regions of genes. Nucleosomes must be evicted or repositioned in order for RNA polymerase to move along the DNA strand. Specifically, the -1 nucleosome is evicted to allow RNA polymerase to bind at the initiation site and the +1 nucleosome is evicted as RNA polymerase passes by. In fact, studies have shown that higher levels of transcription correlate with wider NFRs. Similarly, highly transcribed genes show decreased nucleosome occupancy throughout their coding regions (Koerber et al., 2009; Shivaswamy et al., 2008)

Transcription factors also play an active role in shaping chromatin structure. Transcription factors and nucleosomes compete with one another to
gain access to DNA. As a result, regions of DNA that are bound to transcription factors may no longer have the ability to form nucleosomes. Some transcription factor binding sites are actually more depleted of nucleosomes in vivo than in vitro, suggesting that transcription factor binding actually creates or extends the NFR. For example, NFRs that contain binding sites for the transcription factors Abf1, Rap1, and Reb1 show increased nucleosome occupancy when the respective transcription factor is deleted (Field et al., 2008; Han and Grunstein, 1988; Kaplan et al., 2009). It is suggested that the “chromatin reorganizing” activity of these transcription factors is achieved by their ability to recruit chromatin remodeling complexes that effectively displace surrounding nucleosomes (Jiang and Pugh, 2009).

Chromatin remodelers are defined as large, multiprotein complexes that hydrolyze ATP in order to move, evict, or restructure nucleosomes. In vitro, chromatin remodelers control the spacing of nucleosomes on DNA and can move nucleosomes to less favorable positions (Rippe et al., 2007; Yang et al., 2006). There are four families of chromatin remodelers: SWI/SNF (Switch/sucrose-nonfermentable), INO80/SWRI (INOsitol requiring 80/Sick With Rat8 ts 1), ISWI (Imitation SWItch), and CHD (Chromatin organization modifier, Helicase, and DNA-binding domains) (Venters and Pugh, 2009). These families are highly conserved across eukaryotes. All of these complexes have a DNA dependent ATPase domain that serves as a motor to break DNA-histone contacts.
Additionally, they all possess the ability to recognize covalent histone
modifications and transcription factors (Clapier and Cairns, 2009).

Members of the SWI/SNF family of chromatin remodelers positively
regulate transcription by sliding and/or ejecting nucleosomes at many loci. For
example, the RSC chromatin remodeling complex, a member of the SWI/SNF
family, is involved in exposing the TATA box at gene promoters by moving the -1
nucleosome (Figure 3). Deleting the catalytic subunit of RSC causes shortening of
the NFR in RSC occupied promoters (Jansen and Verstrepen, 2011). The
INO80/SWRI family, in addition to playing a role in transcriptional activation, is
also involved in DNA repair and resolving stalled replication forks. Unlike the
SWI/SNF and INO80/SWRI families, the ISWI family is involved is optimizing
nucleosome spacing to facilitate chromatin assembly and is a repressor of
transcription (Venters and Pugh, 2009). Specifically, ISWI remodelers block
transcription initiation sites by moving nucleosomes towards the 5' NFR,
overriding the intrinsic nucleosome excluding sequences in this region (Figure 3)
(Jansen and Verstrepen, 2011). The role of the CHD family of chromatin
remodelers is poorly understood. Chd1 mutants exhibit only minor changes in
transcription, suggesting that perhaps Chd1 acts in parallel pathways with other
chromatin remodelers. Interestingly, Chd1 is a component of the SAGA complex.
However, the extent to which Chd1 coexists and functions with SAGA at specific
genes is not well understood (Venters and Pugh, 2009).
Figure 3. Chromatin remodeling complexes regulate transcription by exposing or occluding transcription factor binding sites. The SWI/SNF remodeling complex moves nucleosomes in order to allow sequence specific factors (horseshoe shaped protein) to bind to the promoter. The ISWI complex essentially reverses this process by increasing nucleosome occupancy at the promoter region. (Varga-Weisz, 2001)

**Posttranslational histone modifications**

Histone proteins are capable of extensive posttranslational modifications. These modifications of their protruding N-terminal tails and their globular domains play essential roles in transcriptional regulation by directly altering histone-DNA interactions, influencing protein-protein interactions, and recruiting regulatory factors such as chromatin remodeling complexes (Bannister and Kouzarides, 2011).
An example of a posttranslational histone modification is acetylation. The opposing forces of histone acetyltransferases (HATs) and histone deacetylases (HDACs) control the acetylation levels on the N-terminal tails of histone proteins. As is the case with most histone-modifying enzymes, both HATs and HDACs are often found as members of large multiprotein complexes such as SAGA and NuRD, respectively, that confer sequence specificity. Using acetyl CoA as a cofactor, HATs transfer acetyl groups to lysine chains. The addition of acetyl groups neutralizes the lysine’s positive charge and effectively reduces the overall positive charge of the histone, which weakens histone interactions with negatively charged DNA. A weaker interaction means a less compact chromatin structure, leaving DNA more accessible to regulatory factor and transcription machinery binding. Not surprisingly, acetylated histones are enriched at transcription enhancer elements and in gene promoters. Additionally, acetylated lysines are bound by bromodomains found in chromatin remodelers such as the SWI/SNF complex (Figure 4) (Bannister and Kouzarides, 2011).

Histones are also subject to a variety of other modifications including methylation, phosphorylation, ubiquitinylation, sumoylation and ADP-ribosylation, all of which regulate transcription by altering histone dynamics and by recruiting additional regulatory proteins (Figure 4). Similar to acetylation, phosphorylation of histones reduces their positive charge, which alters chromatin structure. Methylated histones are recognized by PHD fingers within the ING protein family. ING proteins, in turn, recruit additional chromatin
modifiers such as HATs and HDACs (Champagne and Kutateladze, 2009).

Similarly, in humans, chromodomains within the chromatin remodeler CHD1 recognize and bind tri-methylated H3K4 and alter the local nucleosome positions (Sims et al., 2005).

Figure 4. Examples of proteins with domains that recognize specific histone modifications (Bannister and Kouzarides, 2011)

**Chromatin remodeling in response to stress**

Clearly, chromatin is not static, but rather a dynamic scaffold capable of responding to environmental cues. Transcriptional activation or repression in response to stress depends, in part, on histone modifications. During heat shock, SAGA and the HDAC, Rpd13, are recruited to specific gene promoters to modulate transcription of the target genes. In *S. cerevisiae*, five of the seven main
chromatin remodeling complexes including Swi/Snf, Swr1p, Isw1, Isw2, and Chd1 are involved in the stress response (Tsukiyama et al., 1999; Uffenbeck and Krebs, 2006; Zhang et al., 2005). Genes that are activated in response to environmental stress are typically repressed during normal growth conditions as a result of a well-positioned nucleosome in their promoter (Lee et al., 2007). The transcriptional activation of these genes in response to stress requires chromatin remodelers to evict these nucleosomes. Shivaswamy et al. generated nucleosome maps for yeast cells grown under different physiological conditions and showed that specific nucleosomes were introduced, evicted, or re-positioned under environmental stress. However, on a genome-wide scale, no major changes in nucleosome position were observed. These results suggest that nucleosome displacement in response to stress is unique and that the relationship between nucleosome dynamics and RNA polymerase binding is more complex than initially believed (Shivaswamy et al., 2008).

Section 3: The Ribosome Biogenesis Pathway

In order to ensure survival and proliferation, yeast cells sense their external environment and modify their transcriptional, metabolic, and developmental pathways accordingly. Microarray technology allows researchers to quantitatively assess mRNA levels under different environmental conditions, including progression through the cell cycle and stress response, for nearly all of
the known yeast genes (Gasch et al., 2000). Comparison and analyses of these data sets is useful for identifying clusters of genes that are coordinately regulated under a given condition. The members of a given cluster are often targets of a particular transcriptional control pathway. In many cases, a set of co-regulated genes will share common promoter motifs that serve as targets for transcription factors and other DNA binding proteins (Hughes et al., 2000). Identifying and characterizing sets of co-regulated genes provides insight into the mechanisms by which the cell controls and coordinates gene expression under a variety of conditions.

**Production and assembly of the ribosome**

One of the most highly regulated pathways in yeast is the production of ribosomes (Warner, 1999). The ribosome is comprised of two subunits; the smaller 40S subunit is required for correct tRNA-mRNA recognition while the larger 60S subunit is home to the peptidyl transferase center, which catalyzes the peptidyl transferase reaction (Figure 5). Together, these subunits are capable of translating an mRNA template into the proper polypeptide (Steitz, 2008). Conserved across all forms of life, the ribosome is comprised of some 79 ribosomal proteins (RPs) and four heavily processed and modified rRNAs (Warner, 1999).
As a cell grows and divides it must duplicate its genetic material, increase its cellular size, and double all of its cellular components. All of these processes require the cell to increase its level of protein synthesis. The dividing cell's increased demand for protein synthesis is met by an increase in ribosome production (Lempiäinen and Shore, 2009; Rudra and Warner, 2004). In fact, it is estimated that a mature yeast cell contains roughly 200,000 ribosomes and

**Figure 5. Structure of the *S. cerevisiae* ribosome.** The molecular model for yeast 80S ribosome is shown. rRNA and RPs are shown in yellow and orange for the 40S subunit, respectively. rRNA and RPs are shown in gray and blue for the 60S subunit, respectively. (Armache et al., 2010)
during periods of rapid growth can produce up to 2,000 ribosomes per minute (Warner, 1999).

Ribosome production is a complex pathway that depends on the activity of hundreds of gene products that are necessary to produce the four heavily processed and modified rRNAs and 79 ribosomal proteins (RPs) that are assembled together in the final ribosome. The rDNA genes, which account for 10% of the total yeast genome, exist as single tandem array of ~150 identical repeats on chromosome XII and are responsible for producing the rRNAs. RNA Pol I-initiated transcription of these repeats represents almost 60% of total cellular transcription (Warner, 1999). Each rDNA gene encodes for a large 35S rRNA precursor transcript that is subsequently processed into the mature 25 S, 18 S, and 5.8 S rRNAs as well as a separate 5S rRNA transcript (reviewed in (Kressler et al., 1999). In addition to the rRNAs, a yeast ribosome contains 79 ribosomal proteins (RPs) that are encoded by 138 genes. Roughly 90% of total mRNA splicing events and nearly 50% of total RNA Pol II transcription is devoted to the RPs, despite the fact that they account for only 2% of yeast genes (Warner, 1999). Ribosome biogenesis also depends on the transcription of almost 300 rRNA and ribosome biosynthesis (RRB) genes. These gene products are required for the proper transcription and processing of the rRNAs and their assembly, along with the RPs, into the final ribosome (Wade et al., 2001, 2006).
**Ribosome Synthesis and Nutrition**

Because ribosome biogenesis is such a major consumer of the cell’s recourses, it is absolutely critical that the genes in this pathway are regulated effectively under changing environmental and cellular conditions. In general, when environmental conditions are favorable and nutrients are available, yeast cells grow and divide. Nutrient-sensing pathways are responsible for transducing the presence of nutrients and growth factors into growth. Because increasing overall cellular translation rate is required for entry into the cell cycle, many of these pathways target ribosome biogenesis (Figure 6)(Warner, 1999).

The Ras/PKA pathway senses glucose levels and initiates signaling cascades that ultimately lead to cell division and growth. Increased glucose levels activate a guanine nucleotide-binding protein called Ras. Specifically, addition of glucose to cells increases the levels of GTP-bound Ras. In its GTP bound state, Ras activates adenyl cyclase, an enzyme that converts ATP to cyclic AMP (cAMP), an intracellular second messenger. An increase in the cellular levels of cAMP causes activation of protein kinase A (PKA). By activating a variety of transcription factors and DNA binding proteins, PKA stimulates a transcriptional change in roughly a third of yeast genes. Activation of these genes inevitably leads to cell growth (Lippman and Broach, 2009; Schneper et al., 2004). Constitutively active PKA has been shown to cause the doubling of several RP mRNAs. Conversely, inactive PKA causes reduced RP mRNA levels
and reduced ability to induce RP transcription in response to glucose increase (Klein and Struhl, 1994).

Figure 6. Nutrient signaling pathways control ribosome biogenesis. Nutrient availability controls the ribosomal protein (RP) genes (Pol II transcribed), the rRNA genes (Pol I and Pol II transcribed), and the ribosome biogenesis (RRB) genes (Pol II transcribed). Not all participants for each pathway are shown and interactions are indicated by lines (Zaman et al., 2009)

Another pathway involved in nutrient sensing is the TOR (target of rapamycin) pathway. TOR complex 1 (TORC1), a kinase complex,
phosphorylates target proteins that, in turn, either activate or repress certain transcription factors. When conditions for growth are met, TORC1 functions by activating the transcription of genes involved in protein synthesis and ribosome biogenesis. Additionally, Tor1 has been shown to directly bind to the 35S rDNA promoter in a nutrient-dependent manner and is necessary for 35S rRNA synthesis (Li et al., 2006). Inhibition of the TOR pathway by rapamycin causes transcriptional repression of rRNA and RP genes (Figure 6) (Powers and Walter, 1999; Zaragoza et al., 1998).

Together the TOR and ras/PKA pathways control ribosome biogenesis by activating or repressing transcription factors involved in ribosome production. For example, both pathways inhibit Maf1 (Figure 6). Maf1 is a transcriptional repressor of RNA Pol III, which is responsible for transcribing 5S rRNA, tRNAs, and a handful of small nuclear RNAs. Without Maf1, yeast cells are unable to repress RNA Pol III transcription in response to nutrient deprivation, stress, or rapamycin treatment. The activity of Maf1 is dependent upon its subcellular localization, which is regulated by nutrients. Specifically, phosphorylated Maf1 is retained in the cytoplasm, whereas dephosphorylated Maf1 enters the nucleus and inhibits RNA Pol III transcription. The phosphorylation of Maf1 is regulated by the competing activities of PKA and protein phosphatase 2A (PP2A), which itself is inhibited by TORC1 (Zaman et al., 2009). Similarly, the Sch9p kinase is a major downstream target of both the TORC1 and ras/PKA pathways (Figure 6). Overexpression of SCH9 induces transcription of ribosome biogenesis genes,
whereas selective inhibition of an ATP analog sensitive allele of \textit{SCH9} (sch\textsuperscript{ aos }) inhibits their expression (Zaman et al., 2009). These are just a few scenarios whereby the ras/PKA and TOR pathways intersect to control ribosome biogenesis.

\textbf{Transcriptional Regulation of Ribosome Biogenesis}

Ribosome biogenesis requires the RPs and the mature rRNAs to be produced in roughly equimolar amounts. To ensure this, the ribosome biogenesis genes are tightly co-regulated. In \textit{S. cerevisiae}, ribosome production is controlled primarily at the level of transcription and involves the coordinated expression of three independently co-regulated gene sets; the ribosomal proteins (RP) genes, the rRNA genes, and the rRNA and ribosome biosynthesis (RRB) genes.

The ribosomal protein genes contain binding sites for Rap1p, a transcription factor that can behave as both an activator and as a repressor, and Abf1. Of the 138 RP genes, 127 bind Rap1p and 8 bind Abfp1. Only three RP genes do not bind either factor (Bosio et al., 2011). A typical RP gene promoter contains a pair of Rap1p or Abfp1 binding sites located 200-500 bp upstream of the transcription start site. Rap1 controls transcription by establishing the 5’ NFR and by recruiting additional transcription factors. These factors, including Sfp1p, Fh11p, and Ifh1p, have all been found to be important for the regulated expression of RP genes (Lempiäinen and Shore, 2009)

Production of the four rRNAs begins with the RNA Pol I & III directed transcription of the rDNA genes that encode for the 35S and 5S precursor rRNAs,
respectively (Fromont-Racine et al., 2003). The rDNA genes exist as 150 tandemly repeated copies located on chromosome XII (Planta, 1997). During growth, roughly half of these repeats are actively transcribed while the other half remain inactive. Carbon source downshift causes a decrease in the number of ‘active’ repeats. This action is mediated by the histone deacetylase Rpd3. Carbon downshift, nitrogen starvation, and rapamycin treatment also cause a decrease in the transcriptional activity of Pol I. The key regulatory process affecting the rate of rRNA transcription involves the ability of Pol I to form a stable complex with the general transcription factor Rrn3. Increased glucose levels cause an increase in Rrn3, and TORC1-dependent signaling stabilizes the Rrn3-Pol I complex (Figure 6) (Zaman et al., 2009).

**The ribosome and rRNA biosynthesis regulon**

Ribosome biogenesis also requires a large number of proteins and RNAs that are not components of the final ribosome. Analyses of microarray data from transition through meiosis, mitosis, heat shock and osmotic shock identified this novel set of genes involved in ribosome biogenesis (Wade et al., 2001, 2006). This set of 282 tightly co-regulated genes is known as the ribosome and rRNA biosynthesis (RRB) regulon (Figure 7). It is important to note that this regulon is distinct from the RP regulon and contains genes that encode for proteins and RNAs involved in a variety of roles in the rRNA biosynthesis pathway, including transcription factors for RNA Pol I and III, RNA helicases, and RNA-modifying enzymes, and RNA endo- and exonucleases. These factors ensure that the 2 pre-
rRNAs are correctly transcribed, processed, modified and folded into the mature ribosome along with the structural ribosomal proteins. Unlike the ribosomal proteins and rRNAs, the RRB gene products do not remain with the ribosome for the duration of its life.

**Figure 7. Gene networks required for ribosome biogenesis.** Production of the mature ribosome requires the coordinated expression of the RP, rRNA, and RRB gene networks (Adapted from James Arnone)

The RRB regulatory controls are distinct from the RPs and rRNAs. MEME program analysis revealed that the RRB genes are significantly enriched for PAC (polymerase A(I) and C(II)) and RRPE (ribosomal RNA processing element) promoter sequence motifs that are found between 50 and 200 bp upstream of the translation initiation site, indicating that RRB genes are likely under the control of a common transcriptional regulatory mechanism (Wade et al., 2006). The PAC and RRPE are required for proper regulation of the RRB regulon (Wade
et al., 2001). The PAC and RRPE motifs are found in 47% and 55% of gene promoters, respectively. Furthermore, both motifs occur together in 25% of the RRB genes. The probability that the observed PAC and RRPE motif distribution is due to chance alone is $1.6 \times 10^{-130}$ and $2.1 \times 10^{-75}$, respectively (Wade et al., 2006). A comparison of 14 different fungal genomes revealed that the RRPE and PAC motifs represent the fourth and fifth most highly conserved cis-sequences, respectively (Gasch et al., 2004).

Recent studies have identified the binding factors for the PAC and RRPE motifs. Stb3 (Sin three binding protein) binds to the RRPE motif and Pbfi (Tod6) and Pbf2 (Dot6) binds to the PAC motif (Freckleton et al., 2009; Liko et al., 2007; Zhu et al., 2009). All three of the transcription factors have been found to behave as repressors of the RRB genes (Liko et al., 2010; Lippman and Broach, 2009). Additionally, Stb3, Tod6, and Dot6 interact with the Sin3-Rpd3 complex, a HDAC necessary for the repression of RRB genes during stress response in yeast (Alejandro-Osorio et al., 2009; Kasten and Stillman, 1997; Shevchenko et al., 2008). Sch9, an AGC family kinase and downstream target of the ras/PKA and TORC1 pathways, promotes transcription of RRB genes by directly phosphorylating, and thus inhibiting, Stb3, Dot6, and Tod6 (Huber et al., 2011).

**Section 4: Adjacent gene co-regulation**

Shortly after the RRB regulon was discovered, an interesting observation was made about the genomic locations of the RRB and RP genes. In *S. cerevisiae,*
15% (P-value = 4.1x10^{-4}) of RRB genes and 13% (P-value = 1.1x10^{-4}) of RP gene exist throughout the genome as immediately adjacent pairs (Arnone and McAlear, 2011; Arnone et al., 2012). Interestingly, these gene pairs exist in all possible orientations (i.e. divergent, tandem, convergent), not just in the divergent orientation, as is the case with two genes sharing a bi-directional promoter. This observation has also been extended to the RP and RRB regulons in the widely divergent fungal lineages *C. albicans* and *S. pombe*. For example, 25% (P-value = 2.3x10^{-15}) of the 168 RRB genes and 21% (P-value = 1.4x10^{-14}) of the 118 RP genes are paired in *C. albicans* (Arnone and McAlear, 2011).

The functional significance of this gene pairing was determined by comparing the transcriptional correlation between unpaired genes, genes that are paired but not adjacent, and adjacentely paired genes in response to a heat shock. For both the RRB and RP regulons, it was shown that the whole set of genes that are paired, which includes the paired but not adjacent genes and the adjacent pairs, are significantly more tightly co-regulated as a group than the unpaired genes. Tighter co-regulation of the paired gene set was observed under a variety of changing conditions that cause either the repression or activation of the RRB and RP regulons (Arnone et al., 2012). This finding suggests that one of the ways a cell regulates the expression ribosome biogenesis genes is by distributing them as adjacent pairs throughout the genome.
Identifying the mechanism of adjacent gene co-regulation.

Given the functional significance of adjacent gene pairing, a considerable amount of work has focused on uncovering the mechanism responsible for co-regulating these gene pairs. Previous analysis of the cis- and trans-acting elements that mediate adjacent gene co-regulation have focused on the *MPP10-YJR003C* RRB gene pair (Figure 8). This gene pair is particularly interesting because the genes are convergently oriented and the conserved RRB promoter motifs, PAC and RRPE, are found within the promoter of *MPP10* only (the promoter of *YJR003C* does not contain any identifiable transcription factor binding motifs) (Arnone and McAlear, 2011).

![Figure 8. The MPP10-YJR003C locus.](image)

Located on Chromosome X in *S. cerevisiae*, this gene pair is separated by a 348 bp intergenic region. The thin arrows represent the transcribed region for each gene. R=RRPE motif and P=PAC motif.

These studies showed that mutating the PAC and RRPE motifs of *MPP10* inhibited its ability to properly respond to heat shock along with the rest of the RRB regulon. Surprisingly, these mutations in the *MPP10* promoter also abrogated the transcriptional response of *YJR003C*, despite the fact that *YJR003C* is convergently oriented to *MPP10* and its promoter lies some 4 kbp away (Figure 9). These results indicate that the transcription of *MPP10* and *YJR003C* is regulated from the promoter of *MPP10* (Arnone and McAlear, 2011).
Figure 9. The transcriptional response of \textit{MPP10} and \textit{YJR003C} is controlled from the promoter of \textit{MPP10}. Gene expression levels were monitored for \textit{MPP10} (red), \textit{YJR003C} (yellow), and the unpaired RRB gene \textit{EBP2} (blue) in wild-type (A) and \textit{ΔRRPE ΔPAC} (B) strains following a 30°C to 37°C heat shock. (Arnone and McAlear, 2011)

The regulated expression of \textit{YJR003C} was also abrogated when a 3 kbp \textit{KAN^r URA3 pCORE} reporter cassette was inserted into the intergenic region of \textit{MPP10} and \textit{YJR003C} (Figure 10). The insertion of this cassette, did not however, affect the regulated expression of \textit{MPP10}. These findings reveal that the immediate adjacency must be maintained for the adjacent gene co-regulation of \textit{MPP10} and \textit{YJR003C} to occur. Supporting this assertion is the observation that all RRB (and RP) genes pairs found in the \textit{S. cerevisiae} genome exist as immediately adjacent pairs with no non-RRB (or non-RP) ORFs between them (Arnone et al., 2012).
Figure 10. Proper transcriptional response of \textit{YJR003C} depends on its immediate adjacency to \textit{MPP10}. (A) The relative position of the $\text{KAN}^{\rm r}$-$\text{URA3}$ cassette inserted between \textit{MPP10} and \textit{YJR003C}. The gene expression levels of \textit{MPP10}, \textit{YJR003C}, and \textit{EBP2} were monitored during heat shock (B) and osmotic shock (C). (Arnone and McAlear, 2011)

In a similar fashion, insertion of an actively transcribed \textit{LEU2} gene (i.e. cells grown in SC-Leu media) between \textit{MPP10} and \textit{YJR003C} disrupted the regulated expression of \textit{YJR003C} but not \textit{MPP10} (Figure 10B &D). Inserting the same \textit{LEU2} gene, this time growing the cells in YPD media (i.e. conditions where \textit{LEU2} is repressed), did not disrupt the regulated expression of either \textit{MPP10} or \textit{YJR003C} (Figure 10A & C) (Arnone, 2012). Adjacent gene co-regulation was also not abrogated when an RNA Pol III transcribed tRNA-Thr gene or when a Ty1 nucleosome barrier element was inserted into the intergenic region (Arnone, 2012). Altogether, these results indicate that the co-regulation of \textit{MPP10} and
YJR003C is dependent, at least in part, on the integrity of MPP10’s promoter, the adjacency of the two genes, and the exclusion of an actively transcribed RNA Pol II transcript between them.

Figure 11. Active LEU2 transcription is required for the co-regulation of MPP10 and YJR003C. The gene expression levels of MPP10, YJR003C, and EBP2 were monitored during heat shock under LEU2 repressive conditions (media containing leucine; A and C) and LEU2 activating conditions (media lacking leucine; B and D) (Arnone, 2012)
The present work is focused on determining the extent of adjacent gene co-regulation in *S. cerevisiae* and higher eukaryotes and identifying more of the *cis*- and *trans*- elements involved co-regulating these gene pairs. We found that many metabolically related gene families in *S. cerevisiae*, as well the ribosome production pathways in a variety of complex eukaryotic species, exhibit this phenomenon of gene pairing. We also determined that the *MPP10* promoter does not regulate the expression of *KANr* in a strain where *YJR003C* was replaced with *KANr*. Finally, we found that changes in promoter nucleosome positions do not correlate with the transcriptional coupling of the *MPP10-YJR003C* gene pair.
Materials and Methods

Yeast strains

A complete list of all strains used in this study and their relevant genotypes is listed in Table 1. Strain YMM13 (MATα leu2Δ1 trp1Δ63 ura3-52) was used as a wild-type and is the parent strain for the ΔRRPEΔPAC mutant that was generated previously by Dr. James Arnone. The ΔRRPEΔPAC mutant was generated using the delitto perfetto (Storici and Resnick, 2006) technique, in which the promoter of MPP10 was replaced by homologous recombination with the pCORE reporter cassette. The pCORE cassette was subsequently replaced by a homologous region of the MPP10 promoter containing an Xhol restriction site and an AatII restriction site in the RRPE and PAC motifs, respectively. The single deletion mutant strains were purchased from the Open Biosystems Yeast Deletion Collection.

Table 1. Yeast strains used in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>yMM13</td>
<td>MATα leu2Δ1 trp1Δ63 ura3-52</td>
<td>(Wade et al., 2001)</td>
</tr>
<tr>
<td>yMM514</td>
<td>MATα leu2Δ1 trp1Δ63 ura3-52</td>
<td>(Arnone and McAlear, 2011)</td>
</tr>
<tr>
<td>yMM557</td>
<td>MATα leu2Δ1 met15Δ0 ura3Δ0 yjr003cΔ::KANr</td>
<td>Open Biosystems</td>
</tr>
<tr>
<td>yMM579</td>
<td>MATα leu2Δ1 met15Δ0 ura3Δ0 snf2Δ::KANr</td>
<td>Open Biosystems</td>
</tr>
</tbody>
</table>
Culture conditions for heat shock

Strains were grown in YPD (1% yeast extract, 2% peptone, 2% dextrose) medium to early to mid-log phase (optical density at 600nm: 0.40 to 0.90) prior to the induction of the stress response. A heat shock time course was induced by growing cultures at 30°C and transferring cells to either 37°C or 39°C medium (Gasch et al., 2000). Cultures then continued to grow at 37°C or 39°C and aliquots of cells were taken at each time point.

RNA isolation and expression analysis

15 mL aliquots of yeast were obtained across a heat shock time course and washed at 4°C to remove the medium. RNA was obtained by a hot acid phenol extraction (Ausubel et al., 1994) with the following modifications: samples were extracted twice with phenol and once with chloroform and then ethanol-precipitated. RNA was then resuspended in diethyl pyrocarbonate (DEPC) treated water. A total of ten micrograms of RNA was treated with DNase I according to the manufacturer’s instructions (DNA-free: Ambion) and was verified by PCR using primers directed to the ACT1 coding region. cDNA was generated with oligo(dT) primers using the Retro-script kit according to the manufacturer’s instructions (Ambion).

Linear range conditions were identified by the titration of cDNA template for PCR, followed by native PAGE. Quantitative PCR (qPCR) was then performed across the heat shock time course, and the products were analyzed by native
PAGE stained with Sybr Gold (Invitrogen). Images were obtained on a Typhoon phosphoimager scanner (Molecular Dynamics) and quantified using the manufacturer’s ImageQuant software. Each expression profile represents the normalized average to ACT1.

Mapping of nucleosome positions

Nucleosome positions were mapped using the nucleosome scanning assay as described in (Hainer et al., 2011). First, 100mL aliquots of cells were fixed with 2% formaldehyde for 30 minutes and then the reaction was quenched with 125mM glycine for ten minutes. The cells were then washed once in TBS buffer and spheroplasted with Zymolyase 20T for 40 minutes (until approximately 85% of cells had spheroplasted). Spheroplasts were then washed twice and re-suspended in 2mL of MNase digestion buffer. Next, 360uL aliquots of spheroplasts were digested with limiting concentrations of MNase I (New England Biolabs) for 40 minutes. The digestion reaction was stopped by the addition of TE-SDS buffer and crosslinks were reversed by incubating the samples overnight at 65°C in the presence of Proteinase K (New England Biolabs). DNA was recovered by phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation. RNA was removed from the sample by treating the samples with RNase A for 60 minutes, at which point the DNA was again extracted by phenol/chloroform/isoamyl extraction and ethanol precipitation. The DNA samples were then resuspended in TE buffer. The digestion reactions were then visualized on a 1.0% agarose gel and the sample that resulted in the
generation of mononucleosomal sized fragments was subsequently analyzed by real time PCR. Real time PCR was performed on an Applied Biosystems 7300 instrument utilizing SYBR-Green chemistry (Life Biosystems) and analyzed using the manufacturer’s software. 20uL reactions from each chromatin preparation were run and outliers were removed based on the manufacturer’s criteria prior to analysis. The ratio of a nucleosome-protected to a nucleosome-depleted region within the GAL10 locus was used as a control, and nucleosome positioning data was determined as previously described (Hainer et al., 2011).

The PCR primers used for nucleosome mapping are presented in Table 2. Each nucleosome profile represents the average of 6 RT-PCR reactions from 2 separate chromatin extractions +/- the S.E.

**Table 2. Primer pairs used in the nucleosome scanning assay**

<table>
<thead>
<tr>
<th>Locus</th>
<th>Forward Primer</th>
<th>Anneal Site*</th>
<th>Reverse Primer</th>
<th>Anneal Site*</th>
<th>Midpoint*</th>
</tr>
</thead>
<tbody>
<tr>
<td>YJR003C</td>
<td>GGTACTTTTGAACAGCCGCTTG</td>
<td>-634</td>
<td>GATACGGACAATGGCTCTTG</td>
<td>-475</td>
<td>-554.5</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CTCATCTAGGAAAGGAAATGG</td>
<td>-566</td>
<td>CATAAGTTGAAATCATGAGAAG</td>
<td>-32</td>
<td>-499</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CAGGAAGCATTGGGCCTTGATC</td>
<td>-498</td>
<td>GATGGACGAGCTGCTGTGAG</td>
<td>-388</td>
<td>-443</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CTTCTCTCATGATTCTAACCATTG</td>
<td>-456</td>
<td>GACCTCAATCTAATCTGTCAC</td>
<td>-325</td>
<td>-390.5</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CTACAGCTGAGCTCGTGGATC</td>
<td>-410</td>
<td>GAATCAAAAAAGGCGAGATG</td>
<td>-272</td>
<td>-341</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CATACCTGATCGCTGATC</td>
<td>-346</td>
<td>GACGGAAAAAGGTAAGAATG</td>
<td>-205</td>
<td>-275.5</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CATCCTGCTCTTCTTCTTC</td>
<td>-294</td>
<td>CTAATTGAAATCTGCGGAGT</td>
<td>-182</td>
<td>-238</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CTACCTCTATCTCTCTCTC</td>
<td>-224</td>
<td>GAAACAGCTCCGGTAAATG</td>
<td>-105</td>
<td>-164.5</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CATTACCGAAAGGCTGTTTC</td>
<td>-124</td>
<td>GATGGGGCTTACGCTTAATG</td>
<td>-5</td>
<td>-64.5</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CATCAGGAAGAGTCTGAGT</td>
<td>-64</td>
<td>GAGAGAGAAAGCCTCTCTTATTAG</td>
<td>57</td>
<td>-3.5</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CTAAAGGCTAAGCCGATTC</td>
<td>-26</td>
<td>GCTTTGATTATGCTGTTTAC</td>
<td>92</td>
<td>33</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CATTCTGCGGCTATCGTATC</td>
<td>15</td>
<td>GATTGACGCTTTGAGAATG</td>
<td>129</td>
<td>72</td>
</tr>
<tr>
<td>YJR003C</td>
<td>ATAGAAAAGCATATACAAAAAGC</td>
<td>68</td>
<td>CATTCTGGATAGAGTATG</td>
<td>181</td>
<td>124.5</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CTTCTGATCGATCTGATC</td>
<td>108</td>
<td>CATATTGATAGCCGGGATC</td>
<td>243</td>
<td>175.5</td>
</tr>
<tr>
<td>YJR003C</td>
<td>CATACCTCTATCAGGGAAATG</td>
<td>160</td>
<td>GATGGCGTGTCAAGATC</td>
<td>280</td>
<td>220</td>
</tr>
<tr>
<td>YJR003C</td>
<td>GATGCCGCGATGGCTACAAATATG</td>
<td>219</td>
<td>GCTTTGATTGCGGGAAGTGAAG</td>
<td>343</td>
<td>281</td>
</tr>
<tr>
<td>Locus</td>
<td>Forward Primer</td>
<td>Annel Site*</td>
<td>Reverse Primer</td>
<td>Annel Site*</td>
<td>Midpoint*</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-------------</td>
<td>----------------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>YJR003C</td>
<td>GTGATCTTGACTGAGCCTG</td>
<td>259</td>
<td>CAAGAGCTTTGGACTCTGCTG</td>
<td>383</td>
<td>321</td>
</tr>
<tr>
<td>MPPI0</td>
<td>TATGACTATCATCCATGCTGAAAG</td>
<td>640</td>
<td>GAGGCTTTGCGAGCTCTC</td>
<td>506</td>
<td>573</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GTATTGGACCTGCGTTGATGCTG</td>
<td>584</td>
<td>GCTATCAAAAGAGCACAC</td>
<td>460</td>
<td>522</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GAAAGACTGCGAAAGGCTCCTCC</td>
<td>526</td>
<td>GAAATGCGCTCTTGCAGCTTC</td>
<td>400</td>
<td>463</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GTTGTACATTAGTGTTGGTGTG</td>
<td>321</td>
<td>GAAAAGGGGCTCTATTTATG</td>
<td>313</td>
<td>396</td>
</tr>
<tr>
<td>MPPI0</td>
<td>CATAGCTCTTCTGACTTCTG</td>
<td>-215</td>
<td>GGCTAATCTGATTCGTCAG</td>
<td>-24</td>
<td>-119.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GAAAGACTGCGAAAGGCTCCTCC</td>
<td>-61</td>
<td>GCTTAAACTCTTTTAGAAGTGGC</td>
<td>86</td>
<td>12.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GTGAGAAGACGTGCTTGTACCAG</td>
<td>-30</td>
<td>CTTCAGGTATTGATAACTGAATC</td>
<td>117</td>
<td>43.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GTCAGAAGGCTGAGAGAGACGC</td>
<td>3</td>
<td>CATGAAACTGATACCTTGTGAC</td>
<td>166</td>
<td>84.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GTCATGATCATATCTTCTG</td>
<td>145</td>
<td>GGGTGAGACATCTTTGCTC</td>
<td>269</td>
<td>207</td>
</tr>
<tr>
<td>MPPI0</td>
<td>TTTCTGTAGATATGATGCTTTC</td>
<td>206</td>
<td>GTTCACCTGACCTGTTCAC</td>
<td>340</td>
<td>273</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GAAAGCTGTGATTTCTGATGAAATG</td>
<td>61</td>
<td>GCTTAAACTCTTTTAGAAGTGGC</td>
<td>86</td>
<td>12.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>CACCGAGGCCCCGGAATTCG</td>
<td>-657</td>
<td>GTGCGATCGAAGTCTTCCTTC</td>
<td>-546</td>
<td>-601.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>ATGTGATGCTGATCAAGAGGG</td>
<td>145</td>
<td>GGGTTGGAGTTGGATCTTGTG</td>
<td>501</td>
<td>554.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GCATGAGCACTTCTAGATGAT</td>
<td>3</td>
<td>CATGAAACTGATACCTTGTGAC</td>
<td>166</td>
<td>84.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GAAATGACTGAGCGATCAG</td>
<td>-566</td>
<td>GCTTAAACTCTTTTAGAAGTGGC</td>
<td>-453</td>
<td>-509.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>CAAGAATGAACTTCTGACGGC</td>
<td>-522</td>
<td>GAAATAGTGGAAGATTAGTACTG</td>
<td>-413</td>
<td>-467.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GGGTTAAAACCTTTCTGAGCA</td>
<td>-474</td>
<td>GGATGGAGGTTGCTTATTAAT</td>
<td>-358</td>
<td>-416</td>
</tr>
<tr>
<td>MPPI0</td>
<td>CTACCTATTACCTTATTATTC</td>
<td>-434</td>
<td>TACCTGTCTAGAATAATTCTTG</td>
<td>-279</td>
<td>-356.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>ATACAGTGAATTCTGCTTGTG</td>
<td>-379</td>
<td>CTTCAGACTGAGCGAGAATCGAG</td>
<td>-203</td>
<td>-291</td>
</tr>
<tr>
<td>MPPI0</td>
<td>CCAGAATTTTCTTCTAGAGTTTA</td>
<td>-300</td>
<td>GTGGAGCCGGCTCTGGCTTTAC</td>
<td>-161</td>
<td>-230.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>CAACAGGGGCCTACCTCAAG</td>
<td>-224</td>
<td>GGATTTGAGCAAGTTATGAGAC</td>
<td>-108</td>
<td>-166</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GTAAAAGCGCAGAACGGCGCAAC</td>
<td>-183</td>
<td>TGGCTGATTTGATCTCAAGG</td>
<td>-52</td>
<td>-117.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GGAGTATGATATCGCTTCTG</td>
<td>-132</td>
<td>CGTATATATCTTGTGTAAG</td>
<td>-5</td>
<td>-685</td>
</tr>
<tr>
<td>MPPI0</td>
<td>CCGTTGAATGTAGCAGCA</td>
<td>-71</td>
<td>GCCAAGCTGAAATCTAGTATTGACC</td>
<td>53</td>
<td>-9</td>
</tr>
<tr>
<td>MPPI0</td>
<td>CTACAGTGAGCATAAAATATACG</td>
<td>-29</td>
<td>GTATAGGTTGAATTGGATG</td>
<td>100</td>
<td>35.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GGCGTCTGATGTTTACCTTGTG</td>
<td>29</td>
<td>CTGATCATCCTCTGGCTTTC</td>
<td>142</td>
<td>85.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>CATCCACAATTAACACCTTAC</td>
<td>79</td>
<td>GATGATATGCGAAAGAGGCTC</td>
<td>190</td>
<td>134.5</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GAAAGAGCGGAAAGTGGATCAG</td>
<td>121</td>
<td>GCAGGTTGGCCGTTGAGAATTC</td>
<td>245</td>
<td>183</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GGCGTTACAGCTATGATC</td>
<td>171</td>
<td>CCCCAGGCAACTGCTGAG</td>
<td>279</td>
<td>225</td>
</tr>
<tr>
<td>MPPI0</td>
<td>GAATTCCTCAAGAGCGAACCCTG</td>
<td>224</td>
<td>GCGCTATAAATGAGTCCCTG</td>
<td>337</td>
<td>280.5</td>
</tr>
</tbody>
</table>

Table 2 Continued.

*Coordinates relative to the translation start site of each gene
Identifying sets of functionally related genes

In order to investigate the frequency of adjacent gene pairing within *S. cerevisiae*, we selected 28 sets of functionally related gene for analysis (Arnone et al., 2012). These sets were defined previously by their gene ontology (GO) and downloaded from the *Saccharomyces* Genome Database. The gene sets chosen represent a variety of pathways that are involved in metabolism and environmental response. Additionally, the gene sets represent a wide range of sizes, from 282 genes in the RRB regulon to 8 genes in the purine biosynthesis pathway.

Identifying the RP and RRB genes in higher eukaryotes.

The ribosomal proteins were defined as all genes whose products are considered structural constituents of the ribosome (including those that are cytosolic, chloroplastic, apicoplast, and mitochondrial). The rRNA and ribosome biosynthesis regulon in *S. cerevisiae* was defined as described previously (Arnone et al., 2012), consisting of 282 genes. The homologues were identified for the genes of the RRB regulon using the WU-BLAST algorithm to search for conservation of the protein coding sequences from *S. cerevisiae*. The total number of genes used in the calculations included all verified protein coding genes from *H. sapiens* (Consortium, 2004), *D. melanogaster* (Adams et al., 2000), *C. elegans* (Hillier et al., 2005), *A. thaliana* (Bevan and Walsh, 2005), *T. thermophila* (Eisen et al., 2006), *P. falciparum* (Gardner et al., 2002), *G. lamblia* (Morrison et al., 2007), *N. crassa* (Galagan et al., 2003), *A. nidulans* (Galagan et al.,
2005), and *N. gruberi* (Fritz-Laylin et al., 2010). The genomic distributions of these gene sets were manually curated.

**Calculating the statistical significance of gene adjacency**

The statistical significance of the genomic arrangement we found was calculated by using the binomial probability. The chance probability that there would be *j* adjacent genes within a regulon of *M* genes is:

\[
1 - \sum_{k=0}^{j} \left( \frac{M!}{k! \ (M-k)!} \right) (p^k (1-p)^{M-k})
\]

where:

\[
p = \left( \frac{M}{N} \right) \left( 2 - \frac{M}{N} \right)
\]

and *N* is the total number of genes present within each species. The functional *p*-values were then calculated in Mathematica.
Results

Section 1: Adjacent positioning of co-regulated genes is widely conserved

Given the functional significance of gene pairing in ribosome biosynthesis, we asked whether gene pairing is specific to this pathway, or whether it is a more general feature of co-regulated gene families. To answer this question, we examined the genomic locations of genes in a variety of transcriptionally co-regulated gene families involved in different metabolic and stress pathways. The gene families used for this part of the study were compiled by James Arnone and Adam Robbins-Pianka and pairings were curated by Jeffrey Arace, James Arnone, Adam Robbins-Pianka, and Sara Kass-Gergi (Arnone et al., 2012). In many of the gene sets we investigated, including those of carbohydrate metabolism, nitrogen metabolism, purine base metabolism, DNA damage response, response to arsenic, heat shock response, and response to toxin, we found highly significant values of gene pairing in all possible orientations (divergent, tandem, and convergent) (Table 3). Similar to the paired RRB and RP genes, we found that the paired genes in many of these gene families exhibited tighter transcriptional co-regulation than the unpaired genes (Arnone et al., 2012). These results suggest that adjacent gene pairing is a more general feature of metabolically related genes in yeast and is not just limited to ribosome biogenesis. There were, however, many gene families we investigated, such as alcohol metabolism (GO #0006066), cellular respiration (GO #004533),
metabolism of phosphorus (GO #0006793) and sulfur (GO #0006790), the response to acids, (GO #0001101), osmotic stress (GO #0006970), oxidative stress (GO #0034599), and the unfolded protein response (GO #0006986), that did not show a non-random pattern of gene location (Table 4).

In light of a previous observation that a significant level of the RRB and RP genes are paired in the widely divergent fungal linages *S. cerevisiae, C. albicans, and S. pombe*, it seemed reasonable to suggest that gene pairing in these regulons may occur in other eukaryotes as well (Wade et al., 2006). To test this hypothesis we investigated the chromosomal locations of RRB and RP genes in a variety of eukaryotes, including the fungi *Neurospora crassa* and *Aspergillus nidulans*, the flagellates *Giardia lamblia* and *Naegleria gruberi*, the fruit fly *Drosophila melanogaster*, the nematode worm *Caenorhabditis elegans*, the mustard plant *Arabidopsis thaliana*, the ciliate *Tetrahymena thermophila*, the malarial parasite *Plasmodium falciparum*, and humans (Figure 12).

Table 3. Gene pairing is a general feature of many co-regulated gene families in *S. cerevisiae*. (Arnone et al., 2012)

<table>
<thead>
<tr>
<th>Ontology</th>
<th>G.O. Number</th>
<th>No. Genes</th>
<th>Adj. Genes</th>
<th>P-Value</th>
<th>Divergent</th>
<th>Tandem</th>
<th>Convergent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbohydrate Metabolism</td>
<td>0005975</td>
<td>91</td>
<td>9</td>
<td>5.5x10^{-4}</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>0006807</td>
<td>86</td>
<td>8</td>
<td>9.9x10^{-4}</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Purine Base Metabolism</td>
<td>0006144</td>
<td>8</td>
<td>5</td>
<td>1.2x10^{-14}</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>DNA Damage Response</td>
<td>0006974</td>
<td>175</td>
<td>16</td>
<td>3.2x10^{-2}</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Response to Arsenic</td>
<td>004685</td>
<td>8</td>
<td>3</td>
<td>4.8x10^{-7}</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Heat Shock Response</td>
<td>0009408</td>
<td>18</td>
<td>4</td>
<td>7.3x10^{-8}</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Response to Toxin</td>
<td>0009536</td>
<td>27</td>
<td>7</td>
<td>1.0x10^{-10}</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 4. Ontologies that did not show significant gene pair adjacencies

<table>
<thead>
<tr>
<th>Ontology Class</th>
<th>G.O. Number</th>
<th>Number of Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol Metabolism</td>
<td>0006066</td>
<td>46</td>
</tr>
<tr>
<td>Cellular Metabolism</td>
<td>0044237</td>
<td>15</td>
</tr>
<tr>
<td>Cellular Respiration</td>
<td>0045333</td>
<td>10</td>
</tr>
<tr>
<td>Energy Reserve Metabolism</td>
<td>0006112</td>
<td>28</td>
</tr>
<tr>
<td>Phosphorus Metabolism</td>
<td>0006793</td>
<td>47</td>
</tr>
<tr>
<td>Sulfur Metabolism</td>
<td>0006790</td>
<td>11</td>
</tr>
<tr>
<td>Glycolysis</td>
<td>0006096</td>
<td>31</td>
</tr>
<tr>
<td>Osmotic Stress Response</td>
<td>0006970</td>
<td>28</td>
</tr>
<tr>
<td>Salt Stress Response</td>
<td>0009651</td>
<td>15</td>
</tr>
<tr>
<td>Oxidative Stress Response</td>
<td>0034599</td>
<td>69</td>
</tr>
<tr>
<td>Unfolded Protein Response</td>
<td>0006986</td>
<td>11</td>
</tr>
<tr>
<td>Protein Ubiquitylation</td>
<td>0016567</td>
<td>41</td>
</tr>
<tr>
<td>Macroautophagy Response</td>
<td>0016236</td>
<td>22</td>
</tr>
<tr>
<td>Response to Acid</td>
<td>0001101</td>
<td>9</td>
</tr>
<tr>
<td>Drug Response</td>
<td>0042493</td>
<td>24</td>
</tr>
<tr>
<td>Pheromone Response</td>
<td>0019236</td>
<td>37</td>
</tr>
<tr>
<td>Water Response</td>
<td>0009415</td>
<td>32</td>
</tr>
<tr>
<td>Pyrimidine Nucleotide Biosynthetic Process</td>
<td>0006221</td>
<td>10</td>
</tr>
</tbody>
</table>

For our analysis, RP genes were defined as any gene whose products are structural constituents of the ribosome. The RP genes are well annotated in the organisms we chose for our study, which made it easy to determine their genomic locations and assess any instances of adjacent gene pairing (pairings identified by Jeff Arace and Sara Kass-Gergi). The size of the RP regulon varied
from species to species, ranging from as few as 63 genes in *T. thermophila* to as many as 387 genes in *A. thaliana* (due to many gene duplications). Using genome browsers to look up the genomic location of each gene, we found cases of immediately adjacent RP gene pairing in each species we examined (Table 5). The level of RP gene adjacency varied considerably across eukaryotes with less than 2% of the RP genes existing as pairs in *N. gruberi* to over 13% in *P. falciparum*. The incidence of RP gene pairs in the fungi *N. crassa* (19%) and *A. nidulans* (12%) is comparable to that of *S. cerevisiae* (13%). Significant levels of

![Phylogenetic relationship of the eukaryotic lineages in this study](image)

> Figure 12. Phylogenetic relationship of the eukaryotic lineages in this study
RP gene paring were also observed in the model organisms, C. elegans (7%), D. melanogaster (4%), and A. thaliana (12%). Interestingly, a significant level of RP gene adjacency was even observed in humans (3%, P-Value= $2.8 \times 10^{-3}$).

Unfortunately, the RRB genes are not as well annotated as the RP genes in the species we chose for our study. Before we could examine the genomic distribution of the RRB genes, we first had to identify the rRNA processing and ribosome biosynthesis genes for each species in our study. Putative RRB genes

<table>
<thead>
<tr>
<th>Species</th>
<th>Protein Coding Genes</th>
<th>RP Genes</th>
<th>Adjacent RPs</th>
<th>P-Value</th>
<th>RRB Genes</th>
<th>Adjacent RRBs</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. cerevisiae</td>
<td>5,797</td>
<td>180</td>
<td>24</td>
<td>1.1x10^{-4}</td>
<td>282</td>
<td>44</td>
<td>4.1x10^{-8}</td>
</tr>
<tr>
<td>N. crassa</td>
<td>10,082</td>
<td>115</td>
<td>22</td>
<td>3.9x10^{-15}</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>A. nidulans</td>
<td>9,541</td>
<td>132</td>
<td>16</td>
<td>1.5x10^{-7}</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>22,287</td>
<td>144</td>
<td>6</td>
<td>2.8x10^{-3}</td>
<td>118</td>
<td>0</td>
<td>7.1x10^{-1}</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>13,601</td>
<td>165</td>
<td>6</td>
<td>1.1x10^{-4}</td>
<td>110</td>
<td>0</td>
<td>8.3x10^{-1}</td>
</tr>
<tr>
<td>C. elegans</td>
<td>19,735</td>
<td>86</td>
<td>6</td>
<td>1.1x10^{-5}</td>
<td>106</td>
<td>6</td>
<td>1.6x10^{-4}</td>
</tr>
<tr>
<td>A. thaliana</td>
<td>26,207</td>
<td>387</td>
<td>46</td>
<td>1.0x10^{-14}</td>
<td>111</td>
<td>2</td>
<td>6.9x10^{-2}</td>
</tr>
<tr>
<td>T. thermophila</td>
<td>27,424</td>
<td>63</td>
<td>6</td>
<td>1.9x10^{-3}</td>
<td>92</td>
<td>2</td>
<td>2.4x10^{-2}</td>
</tr>
<tr>
<td>P. falciparum</td>
<td>5,268</td>
<td>150</td>
<td>21</td>
<td>2.4 x 10^{-4}</td>
<td>100</td>
<td>4</td>
<td>6.0x10^{-4}</td>
</tr>
<tr>
<td>G. lamblia</td>
<td>6,470</td>
<td>66</td>
<td>2</td>
<td>1.5x10^{-4}</td>
<td>94</td>
<td>0</td>
<td>3.2x10^{-4}</td>
</tr>
<tr>
<td>N. gruberi</td>
<td>15,727</td>
<td>130</td>
<td>2</td>
<td>3.6x10^{-1}</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 5. RRB and RP genes exist as immediately adjacent pairs in many eukaryotes. (Arnone et al., 2012)

were found by lab members James Arnone and Adam Robbins Pianka who used the BLAST algorithm to identify homologues from 100 S. cerevisiae RRB genes in each species. Using this method, they were able to identify between 92 (T. thermophila) and 118 (H. sapiens) putative RRB genes in each of the organisms
in our study (Table 5). The genome assemblies of *N. crassa*, *A nidulans*, and *N. gruberi* are relatively incomplete and were excluded from this portion of the study. Because this method is limited by a small sampling size (i.e. only 100 RRB genes from *S. cerevisiae*), we expect the levels of RRB gene pairing in these organisms to be a drastic underestimate. Despite the limitations of our approach, we did find RRB gene pairing in *A. thaliana* (2%), *C. elegans* (6%), *T. thermophila* (2%), and *P. falciparum* (4%) (Table 5) (pairings identified by Jeff Arace and Sara Kass-Gergi). The observed pairing of RP and RRB genes in these species suggests that the adjacent pairing of ribosome biogenesis genes is widely conserved across the eukaryotic lineage.

**Section 2: The ability of the MPP10 promoter to regulate transcription of the adjacent gene is specific to YJR003C**

It is well known that the members of the RRB regulon are transcriptionally co-regulated across a variety of environmental conditions, including osmotic shock, heat shock, and glucose replenishment (Wade et al., 2006). Analysis of the convergently oriented *MPP10-YJR003C* RRB gene pair has shown that PAC and RRPE promoter motifs in the *MPP10* promoter are necessary to control the regulated expression of both *MPP10* and *YJR003C*. We wished to investigate whether the PAC and RRPE motifs of *MPP10* are capable of regulating an adjacent gene that is not *YJR003C*. 
To answer this question we first obtained a YJR003C deletion strain in which the coding region of YJR003C is replaced with the KanMX cassette (Open Biosystems). This 1.4 kbp cassette contains the Kanr gene from the E. coli transposon Tn903 that is flanked by the promoter and terminator sequences of the highly expressed TEF gene of Ashbya gossypii (Wach et al., 1994). We then measured the expression levels of MPP10, Kanr, and EBP2, an independent PAC and RRPE containing RRB gene, throughout a heat shock. Additionally, the expression level of the ACT1 gene was measured as an internal control. To monitor the changes in expression levels of each gene throughout a heat shock, cells were grown to log-phase in 30°C YPD and then shifted to 37°C. Aliquots of culture were taken at 5, 10, 20, and 30 minutes after the shift to 37°C. Total RNA was extracted from each aliquot and the levels of mRNA transcripts for each gene were measured by reverse transcription-quantitative PCR (RT-qPCR).

In response to heat shock, both MPP10 and EBP2 exhibited their normal transcriptional response that is characteristic of the RRB regulon (Wade et al., 2001). The expression levels of Kanr, however, remained unchanged during stress (Figure 13). The observation that Kanr is not transcriptionally co-regulated with MPP10 indicates that the PAC and RRPE motifs are not capable of overriding the control of the Kanr promoter. This result suggests that the adjacent gene co-regulation of MPP10 and YJR003C, and perhaps all RRB gene pairs, depends on the individual integrity of the gene sequences themselves.
Figure 13. Transcription of KANr is not regulated from the promoter of MPP10. (A) Schematic showing the replacement of YJR003C with the KanMX cassette. R=RRPE, P=PAC (B) RT-qPCR was used to generate gene expression profiles for MPP10, KanMX, and EBP2 following a 30°C to 37°C heat shock.

Section 3: Co-regulation of the MPP10-YJR003C gene pair is not mediated by nucleosome re-positioning

The discovery of nuclease-hypersensitive sites in the 5′ regions of active D. melanogaster heat shock and globin genes provided the first evidence that changes in chromatin structure accompany gene regulation (Noll and Kornberg, 1977). Lee et al. examined nucleosome occupancy changes during heat shock and determined that there is an inverse relationship between the change in promoter nucleosome occupancy and transcriptional change of the downstream
gene (Lee et al., 2004). It is possible that regulating the expression of two adjacent genes, such as MPP10 and YJR003C, is achieved by altering the underlying chromatin structure through the activity of ATP-dependent chromatin remodeling complexes.

Previously, in an effort to determine the trans-acting elements that mediate adjacent gene co-regulation, chromatin-remodeling mutants were screened for phenotypes such that MPP10 is properly regulated in response to heat shock but YJR003C is not. This screen revealed that the SWI/SNF and Chd1 chromatin remodeling complexes, as well the Spt20 subunit of the SAGA HAT complex, are required for the transcriptional coupling of YJR003C to MPP10 (Arnone, 2012).

Given this finding, we investigated a possible role for dynamic nucleosome positioning in adjacent gene co-regulation by mapping the positions of individual nucleosomes across the MPP10 and YR003C promoter regions in wild-type cells during heat shock. Nucleosome positions were mapped using a nucleosome scanning assay (Hainer et al., 2011). Histones and DNA are first cross-linked with formaldehyde, spheroplasted, and then incubated with increasing amounts of MNaseI, a nuclease that preferentially digests linker DNA between nucleosomes, to generate mononucleosomal sized DNA fragments (~150 bp) (Figure 14A). Protection of nucleosome bound DNA was confirmed by monitoring sequences from well-known nucleosome-occupied (GAL1 NB) and nucleosome-free regions (GAL1 NUB) in the GAL1-10 promoter using real-time
PCR (Hainer et al, 2010)(Figure 14B). Isolated DNA that showed significant MNaseI protection of the GAL1 NB region compared to the GAL NUB region was used for subsequent nucleosome mapping at the MPP10 and YJR003C promoter regions. PCR primers were designed to amplify overlapping DNA sequences at the promoter regions of both genes in both undigested and digested samples before and after heat shock. MNaseI protection was measured using real-time PCR and normalized to the amount of MNaseI-protected GAL1 NB template.
Figure 14. Determination of GAL NB protected mononucleosome sized MNaseI digested chromatin. (A) Gel image of MNaseI digested chromatin. Each lane contains DNA extracted from digestions with 0, 0.01, 0.03, 0.1, 0.3, and 1 ul of MNaseI. Lane 1 contains a 2-log DNA ladder (New England Biolabs). The black arrow denotes the 150 bp position. (B) Diagram of the GAL1-10 promoter region. The blue ovals represent the positions of well-characterized nucleosomes at this locus. The black bars mark the control RT-PCR amplified fragments in the nucleosome un-bound (NUB) and a nucleosome bound (NB) regions that are used to measure the relative amount of MNase protection of the GAL1-10 promoter (Hainer et al., 2011)

We began by mapping nucleosomes at the HSP104 promoter, a region that has been previously shown to undergo nucleosome remodeling during heat shock, to confirm that our assay was capable of detecting changes in nucleosome architecture (Shivaswamy et al., 2008). The HSP104 promoter contains the STRE (Stress Response Element) and HSE (Heat Shock Element) cis-regulatory elements that are responsible for binding the Msn4/Msn2 and Hsf1 transcriptional activators, respectively, during heat shock (Uffenbeck and Krebs, 2006) (Figure 15B). Before heat shock, the peaks of MNase protection determined from our assay are consistent with previously published nucleosome positions at the HSP104 promoter (Mavrich et al., 2008). During heat shock, when Msn4/Msn2 and Hsf1 are bound and HSP104 is transcriptionally active, the region is much more sensitive to MNase digestion—indicating a decrease in nucleosome occupancy (Figure 15). These nucleosome remodeling events we detected are consistent with previous studies and verify that our assay can accurately measure changes in nucleosome architecture.
Mnase protection of the \textit{YJR003C} promoter was measured across a roughly 1 kbp region that begins ~300 bp into the upstream adjacent gene \textit{(SAG1)} and ends some 300 bp into the coding region of \textit{YJR003C} (Figure 16C). The observed pattern of Mnase protection reveals peaks of nucleosome occupancy at the 3’ end of \textit{SAG1} and the 5’ end of \textit{YJR003C} that surround an approximately 200 bp wide nucleosome-depleted region located immediately upstream of the \textit{YJR003C} translation start site (Figure 16A). Mnase protection

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure15.png}
\caption{Stress-response associated nucleosome displacement occurs at the \textit{HSP104} promoter. Nucleosome positions were determined across the \textit{HSP104} promoter by a nucleosome scanning assay in the wild-type background (A) before and after a 15 min heat shock at 39°C. Previously published nucleosome positions at the \textit{HSP104} promoter as determined by ChIP are represented as blue ovals (B) (Mavrich et al., 2008) H = Heat Shock Response Element, S = Stress Response Elements (Uffenbeck and Krebs, 2006).}
\end{figure}
of the \textit{MPP10} promoter was measured across a similar 1 kbp region that begins ~400 bp into the upstream adjacent gene (\textit{AVT1}) and extends ~300 bp into the coding region of \textit{MPP10} (Figure 16C). Nucleosome occupancy is high at the 3’ end of \textit{AVT1}, low throughout the \textit{MPP10} promoter region, and high at the 5’ end of the \textit{MPP10} coding region (Figure 16A). To confirm the accuracy of this assay, nucleosome positions at both promoters were compared to previously published nucleosome positions (Mavrich et al., 2008) (Figure 16C). Our results are consistent with previously published nucleosome positions, and both \textit{MPP10} and \textit{YJR003C} exhibit stereotypical yeast promoter nucleosome architecture characterized by peaks of nucleosome occupancy surrounding a 5’ nucleosome-free region (NFR) (Lee et al., 2007).

When we mapped the nucleosomes at the \textit{MPP10} and \textit{YJR003C} promoters during heat shock, which causes the rapid repression of both genes, we did not observe any changes in nucleosome positioning (Figure 16A). To determine if there are changes to this nucleosome architecture when the transcriptional response of \textit{MPP10} and \textit{YJR003C} is abrogated, nucleosome positions were mapped in the ∆RRPE∆PAC mutant before and after heat shock (Figure 16B). At 30°C, nucleosome positions in the ∆RRPE∆PAC and wild-type backgrounds were nearly identical across the \textit{MPP10} and \textit{YJR003C} promoters. At 37°C, the nucleosome positions were also nearly identical in both strains at both promoters. Our finding indicates that deleting the PAC and RRPE promoter motifs does not disrupt nucleosome positioning at either promoter under
steady-state conditions or during heat shock. These results suggest that nucleosome remodeling events are not associated with adjacent gene co-regulation.

Figure 16. Nucleosome mapping at the MPP10 and YJR003C promoter regions. Nucleosome positions were determined across both promoters by a nucleosome scanning assay in the wild-type background (A) and in the ΔRRPE ΔPAC background (B) before and after a 10 min heat shock at 37°C. Previously published nucleosome positions at the MPP10 and YJR003C promoters as determined by ChIP are represented as blue ovals (C; (Mavrich et al., 2008)).

The transcriptional coupling of YJR003C to MPP10 requires the Snf2 chromatin-remodeling complex (Arnone, 2012) Interestingly, the deletion of this subunit did not cause any major changes in the nucleosome positioning at
the MPP10 and YJR003C promoter regions before or during heat shock (Figure 17).

Figure 17. Nucleosome mapping in the snf2∆ mutant. Nucleosome positions were determined across the MPP10 and YJR003C promoters by a nucleosome scanning assay in the snf2∆ background (A) before and after a 10 min heat shock at 37°C. Previously published nucleosome positions at the MPP10 and YJR003C promoters as determined by ChIP are represented as blue ovals (C; (Mavrich et al., 2008)).
Discussion

Adjacent gene pairing is a common feature of functionally related genes in eukaryotes

Previous studies have shown that a highly significant portion of the genes in the RRB and RP regulons in *S. cerevisiae* exist throughout the genome as immediately adjacent pairs (Wade et al., 2006). It was also observed that the set of paired genes in these two regulons are more tightly co-regulated as the cell responds to a variety of different environmental stresses (Arnone et al., 2012).

In this current work, we wished to see whether adjacent gene pairing is specific to the ribosome production pathway, or whether it is a more general feature of functionally related genes in yeast. Because the RRB and RP regulons are transcriptionally co-regulated, we used the property of a common transcriptional response as a parameter when choosing gene sets to investigate. By identifying groups of gene using previously defined gene ontology (GO) terms, we were able to generate a wide variety of gene families, including families that were responsive to a wide range of environmental and cellular stresses and involved in different metabolic pathways. Subsequent mapping of the genomic distributions of the genes revealed that adjacent gene pairing exists in many of these functionally related gene sets.

Similar to the RP genes, many of these gene sets, including the genes involved in heat shock response, DNA damage, carbohydrate metabolism, and
nitrogen metabolism have been extensively studied (Gasch et al., 2000; Godard et al., 2007; Shalem et al., 2008; Zaman et al., 2009). However, until now, the extent to which the genes in these families exist as adjacent pairs has either not been observed or has been underappreciated. It is important to note that we observed incidences of pairing in both large and small gene sets. For instance, 62% of the 8 purine base metabolism genes are paired as well as 10% of the 91 carbohydrate metabolism genes and 9% of the 175 DNA damage response genes.

When adjacent gene pairing was initially discovered in the RRB and RP regulons, the pairs were found equally in all three possible orientations: tandem, divergent, and convergent (Wade et al., 2006). Interestingly, this observation holds true for the adjacently paired gene sets characterized in this study. Further analysis has also shown that in the adjacently paired regulons, including the heat shock response genes, the carbohydrate metabolism genes, the purine base metabolism genes, and the nitrogen metabolism genes, the paired genes are much more tightly correlated than the unpaired genes.

It was also previously shown that there are highly significant levels of RRB and RP gene pairing in widely divergent fungal lineages. For instance, as many as 24% and 20% of the RRB genes are paired in *C. albicans* and *S. pombe*, respectively (Arnone and McAlear, 2011). Given the vast evolutionary distance between these lineages and the functional significance of gene pairing, it was reasonable to assume that the adjacent positioning of ribosome biogenesis genes was not limited to just these fungal species. Indeed, in this work, we found
significant levels of RP gene pairing in animals, plants, ciliates, alveolata, and flagellates. Consistent with gene pairs identified in yeast, the pairs we identified in these species were equally divergently, tandemly, and convergently oriented. Unfortunately, the RRB regulons are poorly annotated in these species. Despite this, we did observe several instances of RRB gene pairing. As these genome databases become better annotated, we expect to discover increasing levels of RRB gene pairing. We also expect, similar to *S. cerevisiae*, to find adjacent gene positioning in these higher eukaryotic species within regulons not associated with ribosome production.

The observation that cells control the co-regulation of large sets of functionally related genes by placing them immediately beside each other is supported by several other findings. A genome-wide study in yeast revealed that adjacent genes showed correlated expression independent of their orientation (Cohen et al., 2000). Also, in many of the co-regulated gene pairs investigated in this study, only one of the genes in the pair contained any relevant upstream activator sequences (UAS). This is consistent with our lab’s finding that often only one gene in an RRB or RP gene pair contains the highly enriched PAC/RRPE or RAP1 promoter motifs, respectively (Wade et al., 2006). Additionally, a closer look at genome-wide deletion collections in yeast revealed that many of the phenotypes assigned to the deletion of a particular gene are actually caused by transcriptional defects in the adjacent gene (Ben-Shitrit et al., 2012). This discovery, termed the Neighboring Gene Effect (NGE), suggests that adjacent
genes may rely on each other to maintain proper transcriptional control.

Genome scale correlations between gene pairs and transcription have also been found in *D.melongaster, A. thaliana, M. musculus,* and *H. sapiens* (Spellman and Rubin, 2002; Williams and Bowles, 2004; Woo et al., 2010).

**Chromatin architecture and adjacent gene co-regulation**

Proper cell growth and development depends on the cell’s ability to effectively regulate the expression of large sets of genes. Jacob and Monod discovered that prokaryotic organisms achieve this by packaging genes in the same functional pathway into tandem arrays called operons (Pardee et al., 1959). They found that this organization allows for the set of genes to be transcribed as a single polycistronic mRNA. With the exception of a few species, operons are not found in eukaryotes (Hurst et al., 2004; Lercher et al., 2003). This does not mean, however, that the chromosomal arrangement of eukaryotic genes is random. Perhaps the most prominent examples of non-random gene order in eukaryotes are the globin and Hox genes. In these gene families, the positions of the genes on the chromosome correspond to their spatial and temporal expression during development (Chopra, 2011; Noordermeer and de Laat, 2008). In yeast and higher eukaryotes, such as *D. melanogaster, M. musculus,* and *H.sapiens,* it has been shown that a gene becomes transcriptionally silenced when inserted near the telomeres (Baur et al., 2001; Gottschling et al., 1990; Pedram et al., 2006; Wallrath and Elgin, 1995). In yeast, this silencing is
caused by the spread of heterochromatin throughout the gene (Gottschling et al., 1990).

Our finding that functionally related genes often exist as transcriptionally co-regulated adjacent pairs further illustrates the non-randomness of eukaryotic gene order. Since this discovery, our research, like that of Jacob and Monod, has focused on identifying the cis- and trans- regulatory elements that are involved in adjacent gene co-regulation. Previous work on the coordinated expression of divergent gene pairs, such as the GAL1-GAL10 locus, has attributed the existence of a shared bi-directional promoter between the two genes as the mechanism of co-regulation (Johnston and Davis, 1984). However, the mechanisms that control the coordinated expression of convergent and tandem gene pairs remain largely unknown. The orientation of these gene pairs suggests that their coordinated expression involves mechanisms beyond the use of a single bi-directional promoter to recruit RNA polymerase II. A recent study in yeast showed that the transcription of SRG1, a non-coding transcript, alters the nucleosome occupancy at the promoter of the adjacent gene, SER3. By altering the nucleosome occupancy at the SER3 promoter, SRG1 transcription represses SER3 transcription (Hainer et al., 2011). It is possible, in a similar manner, that the underlying chromatin architecture at the convergent MPP10-YJR003C RRB gene pair plays a role in the transcriptional co-regulation of the two genes as they respond to stress. In fact, five of the seven chromatin-remodeling complexes in S. cerevisiae are suggested to play a role in the environmental stress response
(Uffenbeck and Krebs, 2006). Supporting this hypothesis is the observation that the Snf2 and Chd1 chromatin remodeling complexes are required for the transcriptional coupling of the MPP10-YJR003C gene pair (Arnone, 2012). The possibility that adjacent gene co-regulation is sensitive to nucleosome architecture, led us investigate the nucleosome positions at the promoter region of both MPP10 and YJR003C using a nucleosome scanning assay.

In a wild-type strain, both promoters exhibit canonical yeast nucleosome architecture characterized by a relatively depleted nucleosome free region (NFR) that is flanked by regions of high nucleosome occupancy (Lee et al., 2007). Induction of the stress response by heat shock, which causes the coordinated repression of the RRB regulon, did not change the nucleosome positioning at either promoter. This observation is consistent with a recent genome wide study that showed the remodeling of individual nucleosomes (i.e. the nucleosome eviction that occurs at the HSP104 promoter) in response to stress is rare (Shivaswamy et al., 2008). Disrupting the promoter of MPP10, which abrogates the transcriptional response of MPP10 and YJR003C, did not alter the nucleosome positioning of either gene promoter in the unperturbed and perturbed states—suggesting that these motifs do not mediate adjacent gene co-regulation by directing nucleosome positions. Surprisingly, MPP10 and YJR003C nucleosome profiles in the snf2Δ mutant (i.e. transcription of YJR003C is uncoupled from MPP10) were identical to wild-type profiles before and after heat shock. This finding also corroborates with a recent study investigating
nucleosome occupancy before and after carbon source downshift (glucose -
glycerol) in wild-type and snf2Δ backgrounds. They observed that under steady-
state conditions (i.e. glucose) only 20% (900/4352) of gene promoters showed
an increase in nucleosome occupancy when SNF2 was deleted and very few
promoters exhibited occupancy decreases. Similar results were observed after
carbon source downshift with only 10% (432/4532) of promoters showing
increased nucleosome occupancy in the snf2Δ mutant (Tolkunov et al., 2011).
These small groups of genes that showed changes in nucleosome occupancy did
not include MPP10 or YJR003C.

Because nucleosome occupancy patterns were not altered in any of the
background strains we investigated before or during stress, we suggest that the
mechanism of adjacent gene co-regulation is not mediated by nucleosome
remodeling. It is, however, possible that small changes in nucleosome position
are occurring in these strains but the resolution of our assay is not high enough
to detect them. Supporting this possibility is a recent genome wide study
showing that deleting specific chromatin remodeling complexes resulted in
small 10-20 bp shifts of promoter nucleosomes (Yen et al., 2012). It would be
interesting to see if we could detect small changes in nucleosome position using
a more sensitive assay, such as ChIP-exo. Another, more likely, possibility is that
Snf2 and Chd1 have an indirect effect on the co-regulation of MPP10 and
YJR003C that is not related to their chromatin remodeling ability. As is the case
with Snf2, Chd1 can also play a role in regulating gene expression without
changing nucleosome occupancy. Lee et al. showed that genes that are negatively regulated by Chd1 (i.e. up-regulated when CHD1 is deleted) did not show changes in promoter nucleosome occupancy in chd1Δ mutants (Lee et al., 2012). Snf2 and Chd1 have also been shown to interact with the SAGA complex. Specifically, Snf2 contains specificity for SAGA acetylated histones and Chd1 has been shown to be a component of the SAGA complex (Chandy et al., 2006; Pray-Grant et al., 2005). Interestingly, a follow up study in our lab showed that Spt20, an integral subunit of the SAGA complex, is also required for the co-regulation of MPP10 and YJR003C. This is an intriguing result and suggests that perhaps these three complexes act together to mediate adjacent gene co-regulation. The identification of these complexes is also particularly interesting because they are general regulatory factors and not specific to the RRB regulon. As mentioned earlier, we have identified adjacent gene co-regulation in many different regulons and in many different species. It is possible that because of their global effect, these complexes mediate adjacent gene co-regulation is these other regulons/species as well.

Although nucleosome re-positioning does not appear to play a role in adjacent gene co-regulation, there are many other elements of chromatin architecture that may be involved. One such element is post-translational histone modifications. The patterns of histone modifications H3K9ac, H4ac, H3K14ac, H3K4me2/3 and H3K9ac have similarities between adjacent genes. Moreover, there over 2,000 chromatin domains, many of which cover only 2-4
genes, with similar acetylation changes under changing environmental conditions. Gene pairs located in the same domain are more likely to be expressed than random pairs and often share similar functions (Deng et al., 2010). It is also possible that higher order chromatin structures, including DNA looping, are involved in adjacent gene co-regulation. DNA loops bring HMR-E silencing factors into physical contact with the promoter of the adjacent gene in order to impart silencing on that gene (Valenzuela et al., 2008). Chromosome conformation capture (3C) has detected genome wide DNA looping interactions and found that co-expressed genes with similar functions preferentially contact with each other. A potential model for the co-regulation of MPP10 and YJR003C involves the formation of a loop that would place the promoters of both genes in contact with each other. With this structure, it would be possible for the promoter elements of MPP10 to recruit the basal transcriptional machinery to the transcription start sites of both genes. In the future, we plan to use the 3C technique to visualize whether gene looping occurs at this locus and whether the loop is disrupted in the mutants that we have previously shown to disrupt co-regulation.

The ability of the MPP10 promoter to control the expression of its neighboring gene appear to be specific to YJR003C

It has been established that the coordinated transcription of the MPP10-YJR003C gene pair is regulated from the promoter of MPP10. We sought to
determine whether the promoter of \textit{MPP10} had the ability to regulate the transcription of a gene other than \textit{YJR003C}. When we replaced \textit{YJR003C} with a cassette containing \textit{KAN}^{r} fused to a constitutively active promoter, we saw that \textit{MPP10} maintained proper transcriptional control but the transcription of \textit{KAN}^{r} was not regulated from the \textit{MPP10} promoter. One interpretation of these results is that either the integrity of the \textit{YJR003C} coding region or the integrity of its promoter region, or both, is required for adjacent gene co-regulation. It is also possible that the \textit{TEF} promoter controlling \textit{KAN}^{r} is too strong to allow for transcriptional coupling with \textit{MPP10} occur. Perhaps, under the control of a weaker promoter, the \textit{MPP10} promoter could regulate \textit{KAN}^{r}. Additionally, this experiment indicates that \textit{MPP10} does not rely on \textit{YJR003C} to be regulated appropriately. In the future, we plan to characterize strains where only the coding region or the promoter region of \textit{YJR003C} have been replaced in order determine which parts of the \textit{YJR003C} locus are required for co-regulation with \textit{MPP10}. Another future experiment is to move \textit{YJR003C} to another location in the genome and monitor its expression levels. It would be interesting to see whether \textit{YJR003C} falls under the transcriptional control of its new neighboring genes or whether it remains transcriptionally coupled to \textit{MPP10}. 
Conclusion

This work extends an initial observation that a significant fraction of the RRB and RP genes in yeast are present as immediately adjacent pairs throughout the genome. Specifically, adjacent gene pairing was discovered in a variety of yeast regulons and in the ribosome production pathways of diverse eukaryotes. We also provide evidence to suggest that the co-regulation of the *MPP10-YJR003C* gene pair is not mediated by nucleosome remodeling. Finally, we find that that ability of the *MPP10* promoter to drive transcription of *YJR003C* requires integrity of the *YJR003C* locus.
References


Lindegren, C.C. (1949). The yeast cell, its genetics and cytology. (St. Louis: Educational Publishers).


