



# Functionally Related Genes Cluster into Genomic Regions That Coordinate Transcription at a Distance in *Saccharomyces cerevisiae*

Alanna Cera,<sup>a</sup> Maria K. Holganza,<sup>a</sup> Ahmad Abu Hardan,<sup>a</sup> Irvin Gamarra,<sup>a</sup> Reem S. Eldabagh,<sup>a,b</sup> Megan Deschaine,<sup>a</sup> Sarah Elkamhawy,<sup>a</sup> Exequiel M. Sisso,<sup>a</sup> Jonathan J. Foley IV,<sup>b</sup> James T. Arnone<sup>a</sup>

<sup>a</sup>Department of Biology, William Paterson University, Wayne, New Jersey, USA

<sup>b</sup>Department of Chemistry, William Paterson University, Wayne, New Jersey, USA

**ABSTRACT** Balancing gene expression is a fundamental challenge of all cell types. To properly regulate transcription on a genome-wide level, there are myriad mechanisms employed by the cell. One layer to this regulation is through spatial positioning, with particular chromosomal loci exerting an influence on transcription throughout a region. Many coregulated gene families utilize spatial positioning to coordinate transcription, with functionally related genes clustering together which can allow coordinated expression via adjacent gene coregulation. The mechanisms underlying this process have not been elucidated, though there are many coregulated gene families that exhibit this genomic distribution. In the present study, we tested for a role for the enhancer-promoter (EP) hypothesis, which demonstrates that regulatory elements can exert transcriptional effects over a broad distance, in coordinating transcriptional coregulation using budding yeast, *Saccharomyces cerevisiae*. We empirically validated the EP model, finding that the genomic distance a promoter can affect varies by locus, which can profoundly affect levels of transcription, phenotype, and the extent of transcriptional disruption throughout a genomic region. Using the nitrogen metabolism, ribosomal protein, toxin response, and heat shock gene families as our test case, we report functionally clustered genes localize to genomic loci that are more conducive to transcriptional regulation at a distance compared to the unpaired members of the same families. Furthermore, we report that the coregulation of functional clusters is dependent, in part, on chromatin maintenance and remodeling, providing one mechanism underlying adjacent gene coregulation.

**IMPORTANCE** The two-dimensional, physical positioning of genes along a chromosome can impact proper transcriptional regulation throughout a genomic region. The transcription of neighboring genes is correlated in a genome-wide manner, which is a characteristic of eukaryotes. Many coregulated gene families can be found clustered with another member of the same set—which can result in adjacent gene coregulation of the pair. Due to the myriad gene families that exhibit a nonrandom genomic distribution, there are likely multiple mechanisms working in concert to properly regulate transcriptional coordination of functionally clustered genes. In this study, we utilized budding yeast in an attempt to elucidate mechanisms that underlie this coregulation: testing and empirically validating the enhancer-promoter hypothesis in this species and reporting that functionally related genes cluster to genomic regions that are more conducive to transcriptional regulation at a distance. These clusters rely, in part, on chromatin maintenance and remodelers to maintain proper transcriptional coordination. Our work provides insight into the mechanisms underlying adjacent gene coregulation.

**KEYWORDS** coregulation, genomics, *Saccharomyces cerevisiae*, gene expression

**Citation** Cera A, Holganza MK, Hardan AA, Gamarra I, Eldabagh RS, Deschaine M, Elkamhawy S, Sisso EM, Foley JJ, IV, Arnone JT. 2019. Functionally related genes cluster into genomic regions that coordinate transcription at a distance in *Saccharomyces cerevisiae*. mSphere 4:e00063-19. <https://doi.org/10.1128/mSphere.00063-19>.

**Editor** Aaron P. Mitchell, Carnegie Mellon University

**Copyright** © 2019 Cera et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to James T. Arnone, [arnonej@wpunj.edu](mailto:arnonej@wpunj.edu).

**Received** 28 January 2019

**Accepted** 26 February 2019

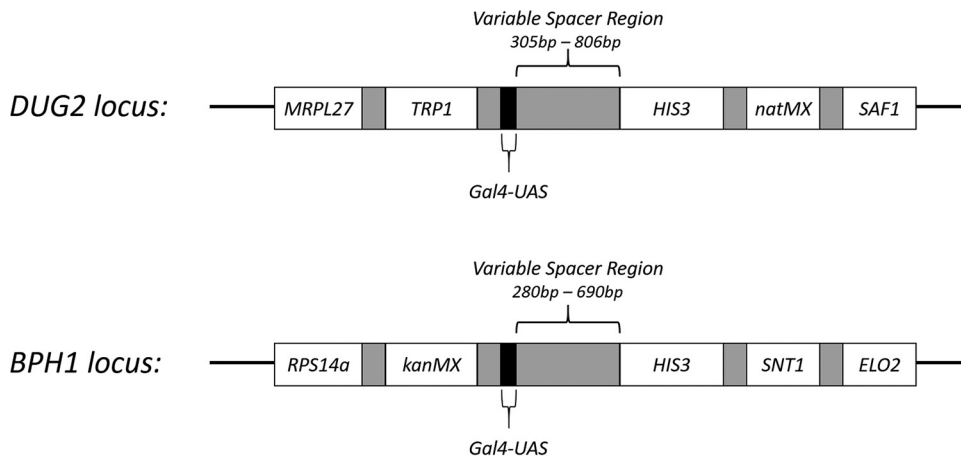
**Published** 13 March 2019

Proper regulation of transcription is essential for cellular survival through the establishment and maintenance of homeostasis within a particular environmental and developmental context. To balance transcription on a genome-wide scale, myriad layers of regulation have evolved and are utilized, including *cis* regulatory DNA sequences, such as promoters and enhancers, that recruit *trans*-acting, DNA-binding transcription factors (1–3). Transcription factors work in either singular or combinatorial fashion to recruit RNA polymerases, ultimately altering the rate of production of RNA to meet cellular demands (4, 5). The accessibility of *cis* regulatory sequences, and the corresponding transcription factors, is altered by the formation of differential chromatin structures resulting from changes in the location, modification, and composition of histones and nucleosome complexes (6–8).

Spatial positioning, the two-dimensional arrangement of genes along the chromosome, can profoundly influence transcription throughout a genomic region. In eukaryotes, genes positioned adjacent to heterochromatin ultimately are converted to heterochromatin and are silenced by either the telomere proximal effect (TPE) or by position effect variegation (PEV) (9–11). Position effects are not limited to gene repression; localized clusters of genes are transcribed and transcriptionally correlated throughout the genome in many eukaryotic organisms, including the *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Arabidopsis thaliana*, and *Danio rerio* genomes, and this is conserved up through humans (*Homo sapiens*) (12–15). This transcriptional correlation is partially mediated by enhancer and promoter promiscuity, termed the enhancer-promoter (EP) theory, whereby the activity of regulatory elements extends over a broad distance to regulate multiple targets scaling in accordance to genome size (16).

One regulatory challenge involves the coordinated expression of functionally related gene families whose composite members are required in stoichiometric levels by the cell. The best characterized examples of this are represented in the coregulated gene families, or regulons, that are required for production of the ribosome. In eukaryotic organisms, each ribosome requires the coordinated production of four highly modified and folded rRNAs, approximately 80 ribosomal proteins (RPs), and more than 200 rRNA and ribosomal biogenesis (RRB) processing and assembly factors (17). Central to translation, and intimately connected to cell growth, the ribosome is tightly regulated in eukaryotes (18). Numerous studies have unraveled many layers coordinating transcription of these components, including the *cis* and *trans* factors specific to the RP and RRB regulons (19, 20). One interesting observation revealed that the RP and RRB genes exhibit a nonrandom, statistically significant distribution throughout the genome; the RP and RRB genes are found clustered with other members of the same regulon (but not with members across the regulons). This genomic distribution is conserved throughout divergent fungal lineages and more complex eukaryotic organisms (21, 22). The members of both the RP and RRB families are predominantly found as pairs. In *S. cerevisiae*, spatial proximity allows for adjacent gene coregulation, whereby the clustered genes are coregulated via shared regulatory elements functioning at a distance (23, 24). Many functionally related gene families exhibit this nonrandom distribution of their members throughout the genome as functional clusters. This characteristic is conserved in broadly divergent eukaryotes and coordinates transcriptional coregulation of these gene families (25).

In this current work, we use the budding yeast, *Saccharomyces cerevisiae*, to empirically test the EP theory as a mechanism underlying adjacent gene coregulation. Our findings reveal that the distance a promoter can activate a gene varies based on both its proximity to a gene and the site of integration. This manifests at the level of transcriptional activation, in changes in growth phenotype, and in the degree of disruption surrounding the integration locus. We report that adjacent gene coregulation is mediated, in part, by the clustering of functionally related genes into genomic regions that are more conducive to transcriptional regulation at a distance compared to their unpaired counterparts. Furthermore, we identify a putative role for chromatin



**FIG 1** Schematic showing the reporter construct at the sites of integration in the strains utilized in this study. The location and spatial arrangement of the reporter constructs at both the *DUG2* locus (top) and the *BPH1* (bottom) locus. The *HIS3* gene was separated from the  $UAS_{Gal}$  by a variable length spacer in each strain. See Table 4 for the complete relevant genotype for every strain used in this study.

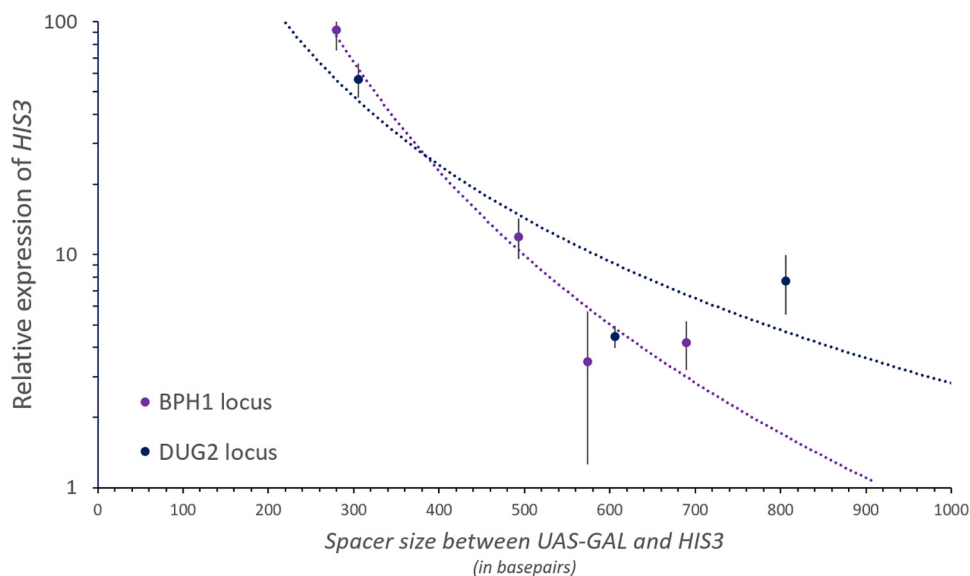
remodeling in the coregulation of functional clusters and identify several chromatin remodeling complexes that are necessary.

## RESULTS

**The level of expression induced by a  $UAS_{Gal}$ -*HIS3* reporter decays at a rate that varies depending on the site of integration.** To test the EP theory, we began by taking advantage of a series of yeast strains that have been previously developed, where an inducible *HIS3* reporter gene was placed under the influence of a galactose-inducible upstream activating sequence ( $UAS_{Gal}$ ) to monitor the effects of transcriptional activation at a distance as a function of genomic location (Fig. 1) (26). The spacing between the *HIS3* gene and the  $UAS_{Gal}$  varied between 280 and 806 bp across a series of yeast strains with the construct integrated into either the *BPH1* locus on chromosome III or into the *DUG2* locus found on chromosome II (26). We began by utilizing quantitative reverse transcription-PCR (qRT-PCR) to measure the relative expression of *HIS3* upon the activation of transcription by the addition of galactose to the media. The level of *HIS3* expression was plotted as a function of spacer size separating the two elements (Fig. 2). When the spacer between *HIS3* and  $UAS_{Gal}$  was smallest (approximately 300 bp), the level of *HIS3* expression was roughly comparable between the two sites of integration. As the size of the spacer increased, a concomitant decrease in the level of *HIS3* expression was observed at each locus. However, the rate of this decrease varied at each site. The rate of decay at the *BPH1* locus was quicker than the rate of decay seen at the *DUG2* locus, and we can extrapolate a complete loss of *HIS3* activation predicted to occur after approximately 920 bp versus 1,500 bp, respectively.

As the difference in the rate of decay observed at the *BPH1* locus and the *DUG2* locus was larger than we initially expected, we characterized the extent of coregulated expression throughout each locus. Using a bioinformatic approach, we calculated the Spearman's correlation coefficient (SCC) throughout the 10-gene window surrounding each site of integration. Gene expression profiles were extracted from the following microarray data sets: cycling cells, a heat shock time course, a DNA damage response time course, an oxidative stress time course, a nitrogen depletion time course, carbon source switching, and nutrient limitation from carbon, nitrogen, phosphorus, or sulfur (27–29). The SCC was calculated for every pairwise gene combination within the window and plotted as a function of genomic distance (Fig. 3).

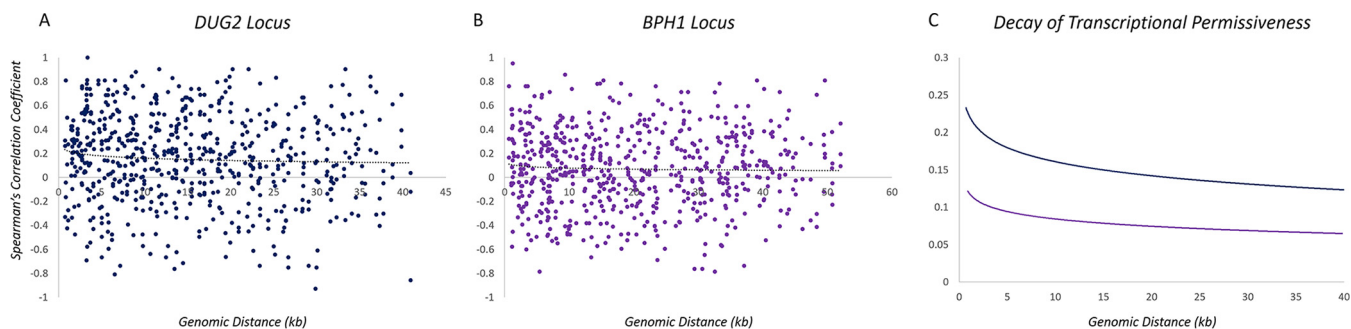
Within both the *BPH1* locus and the *DUG2* locus, both positive correlations and negative correlations (anticorrelations) are observed between specific genes and across different expression conditions (Fig. 3A and B). The transcriptional decay curve was



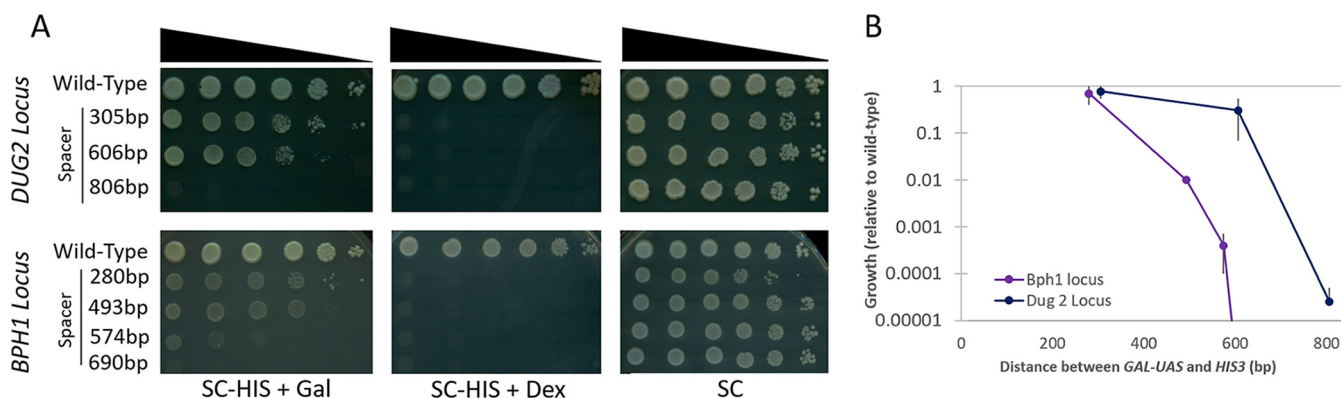
**FIG 2** Relative levels of *HIS3* gene expression and activation at the *BPH1* and *DUG2* loci. Expression of *HIS3* was determined relative to *ACT1* and plotted as a function of the size of the spacer separating the distance of  $UAS_{Gal}$  from the gene. Points on the plot represent the average level of expression, and error bars depict the standard errors of the means. The decay curves are color coded by locus and extrapolated to estimate the x-axis intercepts.

calculated for each locus to determine the extent of coregulated expression throughout each region (Fig. 3A and B). The decay calculations represent the average absolute value of the SCCs. For clarity, the decay curves were overlaid for comparison between each locus as a separate plot (Fig. 3C). Both loci exhibit a positive correlation that decays as a function of distance throughout each window, consistent with previous reports (16). The correlation is higher and decays at a lower rate at the *DUG2* locus ( $y = -0.027 \ln x + 0.2233$ ;  $R^2 = 0.0044$ ) than at the *BPH1* locus ( $y = -0.013 \ln x + 0.1075$ ;  $R^2 = 0.0014$ ), indicating that the *DUG2* locus is more conducive to transcriptional activation across a longer chromosomal distance, validating our hypothesis.

**The differences seen in transcriptional activation at the integration site alter the growth phenotype and the magnitude of disruption to neighboring genes.** In order to observe whether the difference in the activation of transcription at each locus manifests as a difference in the phenotype, we utilized a spotting assay to measure the growth of each strain upon construct activation (Fig. 4A). Tenfold serial dilutions of yeast cells were spotted on agar plates activating the reporter (SC-HIS+Galactose), repressing the reporter (SC-HIS+Dextrose), or on nonselective medium (SC as a positive control). Growth was quantified, compared to the wild type, and plotted as a function of spacer distance (Fig. 4B). Regardless of the site of construct integration, when the



**FIG 3** Extent of transcriptional coregulation across the genomic neighborhood surrounding the site of reporter integration. (A and B) Spearman's correlation coefficient was calculated for every pairwise combination within the 10-gene neighborhood surrounding the *DUG2* locus (A) and the *BPH1* locus (B). (C) The decay curves determined for each locus were color coded and are plotted together for comparison of the permissiveness.



**FIG 4** The distance of transcriptional regulation at a locus alters the growth phenotype. (A) Representative spotting assays comparing the differences in strain growth at the *DUG2* locus (top) and the *BPH1* locus (bottom). Each strain was grown for 72 h to saturation and washed with water, and 10-fold serial dilutions were spotted on the indicated plates. The cultures on the plates were grown for 72 h and imaged. Strain growth was scored as relative to wild-type growth across multiple replicates. (B) Multiple replicates from panel A were averaged and plotted as a function of the size of the spacer insert. The relative growth of each strain is compared to growth of the wild type, and error bars depict standard errors.

spacer size was shortest, the level of growth approached wild-type levels. As the size of the spacer increases, a measurable decrease in growth is observed at both loci, with the level of growth dropping quicker at the *BPH1* locus than at the *DUG2* locus.

The EP theory postulates that transcriptional effects from *cis* regulatory sequences can affect transcription across a genomic region in a nonspecific, promiscuous manner. To test the EP theory, we utilized qRT-PCR to measure transcription of the genes flanking the sites of reporter integration. Gene expression was measured for the genes flanking the sites of integration under conditions activating transcription of the  $UAS_{Gal}$ -*HIS3* construct. The relative expression for the genes flanking the *DUG2* site, *SAF1* and *MRPL27*, and the genes flanking the *BPH1* site, *RPS14A*, *SNT1*, and *ELO2* were measured (Table 1 and Table 2, respectively). At both loci, an increase in transcription is observed for the genes surrounding the integration site upon *HIS3* activation. The magnitude of disruption at the *DUG2* locus was larger than the disruption observed at the *BPH1* locus and varies with insert size; however, this did not appear to be a linear relationship.

It should be noted that the construct integrated at the *DUG2* locus contains two selectable markers used in strain construction, *TRP1* and *NAT<sup>R</sup>*, while the construct integrated at the *BPH1* locus contains the *KAN<sup>R</sup>* marker within the  $UAS_{Gal}$ -*HIS3* reporter (Fig. 1). Though the selectable markers utilized differ at each locus, their expression is constitutive and would not change under the conditions that activate *HIS3* transcription (galactose addition to the media), thus allowing us to monitor the specific effects of *HIS3* activation.

**Functionally related genes cluster together at genomic loci more conducive to coregulation at a distance.** With the observation that the extent and magnitude of transcriptional regulation vary based on both proximity to the promoter and by locus, we tested these factors as a potential mechanism underlying adjacent gene coregulation. We employed a bioinformatic approach to determine the relationship between transcriptional coregulation as a function of distance for four coregulated gene families

**TABLE 1** Transcription levels of the genes flanking reporter integration at the *DUG2* locus on chromosome II

Gene	Transcription level with the following spacer size (bp) <sup>a</sup> :					
	305		606		806	
	RE	SE	RE	SE	RE	SE
<i>SAF1</i>	12.325	1.816	7.192	0.668	15.439	3.568
<i>MRPL27</i>	7.300	4.880	3.831	2.872	1.116	0.275

<sup>a</sup>Transcription levels of genes flanking reporter integration at the *DUG2* locus. RE, relative enrichment (compared to *ACT1*); SE, standard error.



**TABLE 2** Transcription levels of the genes flanking reporter integration at the *BPH1* locus on chromosome III

Gene	Transcription level with the following spacer size (bp) <sup>a</sup> :							
	280		493		574		690	
	RE	SE	RE	SE	RE	SE	RE	SE
<i>RPS14A</i>	1.643	0.738	1.462	0.201	4.140	2.801	0.887	0.446
<i>SNT1</i>	0.999	0.001	9.057	2.120	11.521	2.963	2.637	1.349
<i>ELO2</i>	0.429	0.097	1.893	0.176	4.324	7.948	5.814	3.247

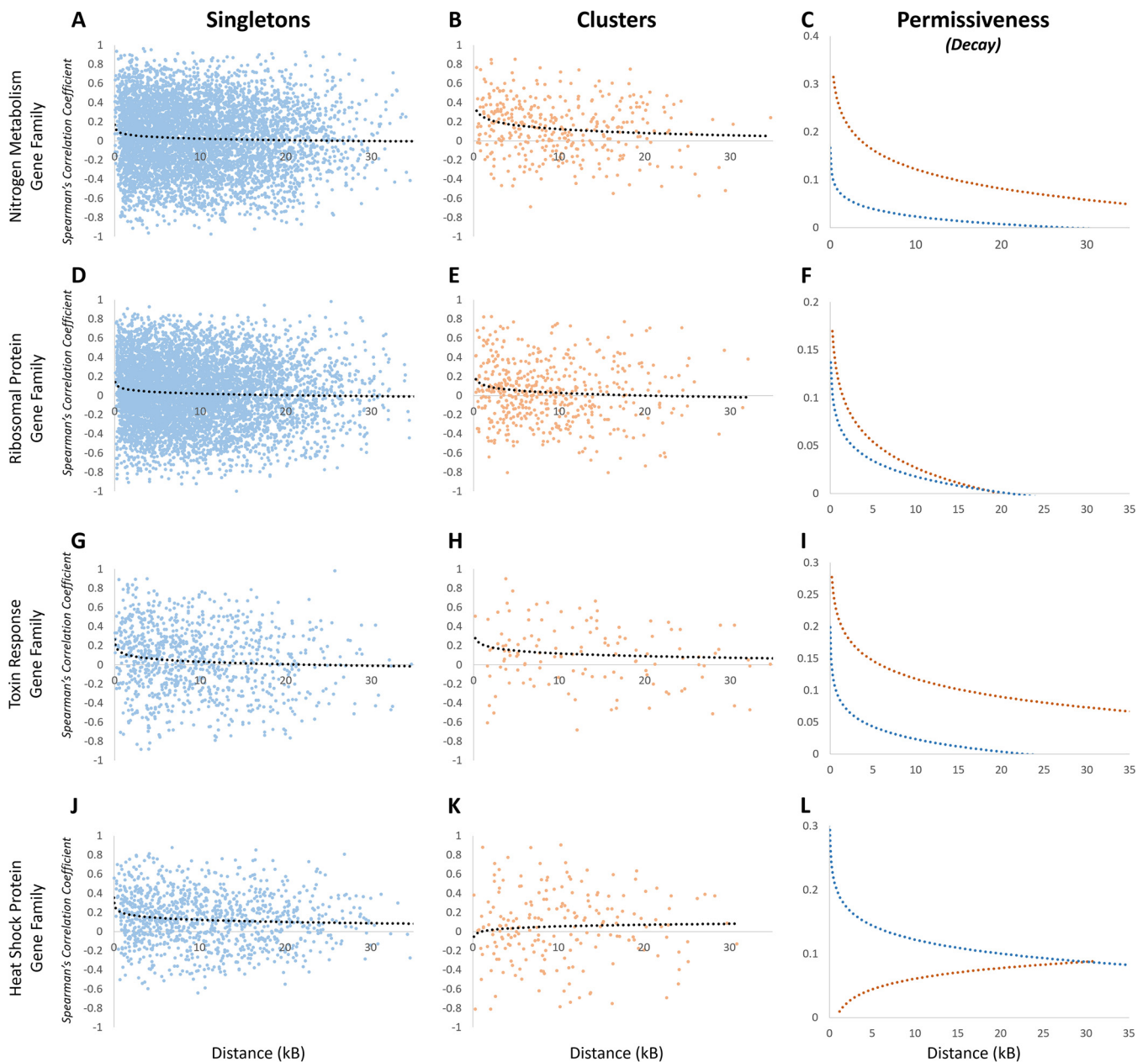
<sup>a</sup>Transcription levels of genes flanking reporter integration at the *BPH1* locus. RE, relative enrichment (compared to *ACT1*); SE, standard error.

exhibiting a nonrandom genomic distribution: the nitrogen metabolism (NM) genes, the ribosomal protein (RP) genes, the toxin response (TR) genes, and the heat shock protein (HS) genes. These families were selected due to their variation in size. These families vary in size from 18 genes (HS), 27 genes (TR), 86 genes (NM), to 129 genes (RP). Additionally, they exhibit reciprocal expression patterns upon induction of the environmental stress response (ESR) in yeast—the NM and RP genes are downregulated, while the TR and HS genes are upregulated (30). When possible, we extracted gene expression profiles for each family member and for the 10 flanking genes (five genes in each direction). Several genes neighbored telomeres, notably members of the TR gene family, and in those cases, our analysis stopped at the last protein-coding gene on the chromosome prior to the telomere. Our analysis focuses on gene expression time courses resulting in the induction of the environmental stress response, including a heat shock time course, DNA damage response time course, oxidative stress time course, a nitrogen depletion time course, and carbon source switching (28).

The Spearman's correlation coefficient was calculated for every pairwise combination of genes at each locus for every family member in each gene set. The SCC was plotted as a function of distance, separating the genes found in functional clusters from the family members found in isolation (the singletons). While significant variability in the SCC measured across the neighborhoods exists, when an exponentially decaying function was applied to the data, the singleton members of all four gene families revealed a positive correlation that decays as a function of distance (Fig. 5A, D, G, and J). For the gene members found as functional clusters throughout the genome, three gene families had a positive SCC: the NM (Fig. 5B), RP (Fig. 5E), and TR (Fig. 5H) genes. The HS genes found as clusters had a negative, anticorrelated SCC that decayed as a function of distance (Fig. 5K).

To directly compare transcriptional activation at a distance for the singleton members to the functionally clustered members of each gene family, the decay curves were extracted and overlaid onto a single plot for clarity. In three gene families, the functionally clustered members are found in genomic locations more conducive to transcriptional activation at a distance than their singleton counterparts—the NM (Fig. 5C), RP (Fig. 5F), and TR (Fig. 5I) genes. Within each of these families, the transcriptional coregulation was greater and extended over a broader genomic distance for the genes found in clusters compared to their singleton counterparts. The HS gene family exhibited the opposite relationship; the unpaired, singleton HS genes localized to more conducive genomic loci than the HS paired genes (Fig. 5L).

**Chromatin remodeling is necessary for adjacent gene coregulation.** Chromatin maintenance provides a desirable model for the coregulation of functionally related, clustered genes. A bioinformatic approach was used to identify putative chromatin remodelers required for coregulation of the clustered genes. This approach took advantage of the large-scale gene deletion study characterizing the differences in gene expression upon deletion of 165 nonessential chromatin remodeling genes (31). Gene expression data were extracted for the members of the NM, RP, TR, and HS gene families, with the extent of gene disruption determined for the clustered members of each family compared to the singleton members. The significance of disruption was



**FIG 5** Comparison of the transcriptional coregulation across genomic regions for the singletons versus the clustered members within functionally related gene families. (A to L) Spearman's correlation coefficient was determined with a 10-gene window for every pairwise combination for the nitrogen metabolism gene family singletons (A) and clusters (B), the ribosomal protein gene family singletons (D) and clusters (E), the toxin response gene family singletons (G) and clusters (H), and the heat shock protein gene family singletons (J) and clusters (K). The decay plots are shown with dotted black lines on each graph, and they are overlaid for comparison (C, F, I, and L).

calculated by a hypergeometric probability density function and focused on remodeling complexes that disrupt the clusters preferentially compared to the rest of the members within each gene family (24). Thus, our analysis looked for chromatin remodeler deletion mutants that resulted in the overrepresentation of clustered genes losing their transcriptional coregulation relative to the rest of the genes within the set. This approach would eliminate the identification of chromatin remodelers that regulate the transcription of an entire gene family and allow for the identification of those remodelers that are specific to the coregulation of clustered family members. Our analysis identified a set of chromatin remodelers that preferentially uncouple transcription of the functionally clustered genes within each family, using a threshold of  $P < 0.05$  as our

**TABLE 3** Genes coding for chromatin remodelers that significantly disrupt the transcription of functional clusters relative to the unpaired members within a gene family

Family and gene	P value
Nitrogen metabolism gene family	
<i>hfi1</i>	0.0003
<i>spt20</i>	0.0011
<i>set3</i>	0.0142
<i>snf12</i>	0.0206
<i>not4</i>	0.0252
<i>bre1</i>	0.0268
<i>set2</i>	0.0288
<i>ies1</i>	0.0303
<i>cac2</i>	0.0303
<i>ssn6</i>	0.0388
<i>sif2</i>	0.0418
<i>ies5</i>	0.0424
<i>hos2</i>	0.0482
<i>sum1</i>	0.0482
Ribosomal protein gene family	
<i>yng1</i>	<0.0001
<i>rtt109</i>	0.0120
<i>med16</i>	0.0221
<i>sas3</i>	0.0257
<i>sdc1</i>	0.0310
<i>ngg1</i>	0.0349
Toxin response gene family	
<i>rsc2</i>	0.0433
Heat shock protein gene family	
<i>hpc2</i>	0.0245

cutoff for transcriptional deviation from wild-type levels of expression (Table 3). The coregulation of the clustered members in the nitrogen metabolism gene family is uncoupled from the rest of the set without the SAGA complex (*hfi1* and *spt20* mutants), the Set3 histone deacetylase complex (*set3* mutant), and the Swi/Snf complex (*snf12* mutant). Likewise, the ribosomal protein gene family clusters require the activity of the NuA3 histone acetyltransferase complex (*yng1* and *sas3* mutants), the Compass complex (*sdc1* mutant), and the SAGA/SLIK complexes (*ngg1* mutant). The toxin response gene family requires the RSC complex (*rsc2* mutant), and the heat shock protein gene family requires the HIR complex (*hpc2* mutant). Altogether, our analysis identifies a role for chromatin remodeling complexes necessary to coordinate the transcription of functionally related gene clusters and underlie adjacent gene coregulation.

## DISCUSSION

### The limit to transcriptional activation at a distance varies by genomic locus.

The work presented in this article expands the growing body of evidence documenting the interconnectedness of gene expression based on spatial positioning. Previous studies have characterized the impact of the position effects seen with the integration of a reporter construct in *Saccharomyces cerevisiae*; these effects result in large differences in expression (up to a 13-fold difference) and can disrupt the expression of the adjacent gene (the neighboring gene effect [NGE]) at a significant number of loci (approximately 7 to 15% of the genome). Transcriptional disruption by the NGE led to erroneous attribution to gene function (32–34). The distance that regulatory elements can function has similarly been characterized, depending on both proximity to a gene (typically 450 bp in yeast) and the activity of the Mediator complex (26, 35). In the present work, we verify and extend these studies, demonstrating that even seemingly subtle differences across genomic loci can profoundly impact the extent of transcriptional activation at a distance.



The understanding of how a promoter regulates transcription on a cellular level has evolved, with recent findings characterizing the EP effect—that *cis* regulatory regions exert their effects over broad distances which can result in alterations in transcriptional regulation throughout a genomic neighborhood (16). The EP effect has been theorized to affect transcription in widely divergent species, from budding yeast up through humans; however, it was empirically tested only in *Caenorhabditis elegans*. Our study builds upon this work and adds to the growing body of evidence that the EP effect may be ubiquitous in eukaryotes. Our work also demonstrates the limiting effects that a genomic locus imposes on this phenomenon, showing that the differences between genomic loci can manifest as phenotypic differences and can result in differences in the transcriptional disruption that vary with genomic neighborhood.

**Adjacent gene coregulation is mediated, in part, by the clustering of genes at loci conducive to transcriptional activation at a distance.** The RP and RRB regulons both exhibit a nonrandom genomic distribution, occurring as functional pairings (21, 36). Functional dissection of one pair of RRB genes, the *MPP10-MRX12* gene pair, revealed a dependency on physical adjacency for their coordinated transcription with the rest of the RRB regulon. The physical separation of *MRX12* from *MPP10* results in the uncoupled expression of *MRX12* (but not *MPP10*) from the rest of the RRB gene family (23). The coregulation of this gene pair depends on two shared promoter motifs found upstream of *MPP10*, which recruits transcription factors to the locus, and results in changes to the chromatin to the region (23). This phenomenon is called adjacent gene coregulation, and follow-up work revealed the dependency of chromatin remodelers for its maintenance at this locus (24). It was subsequently observed that roughly 25% of gene families exhibit a nonrandom positioning throughout the genome of *S. cerevisiae*, predominantly found as functionally related clusters (25).

One challenge to understanding the functional clustering causing adjacent gene coregulation is understanding what mechanisms underlie this process. The gene families found as functional clusters are diverse and varied, belonging to many different ontology classes (25). Each gene family is characterized by its own *cis* and *trans* factors, with many families exhibiting a multitude of additional layers. Here we propose the process that ultimately gives rise to functional clusters may be passive in nature—the result of chance genomic rearrangements over time resulting in two functionally related genes clustering together into a genomic region that is more conducive for coexpression. This would allow for long-range transcriptional control to mediate transcription of the members of a given regulon, effectively streamlining the genome (37). Additional support exists for this interpretation: genes that are physically clustered together are not conserved over evolutionary distance; however, the absolute numbers of clustered genes are comparable (22, 23, 25). For example, *Candida albicans* and *Schizosaccharomyces pombe* each exhibit the same levels as pairing as *S. cerevisiae* for both the RP and RRB genes, although the genes clustered differ across these species (23).

The genic arrangement into clusters can also play a role in regulating mutually exclusive expression at a locus. This is evident in our analysis of the heat shock protein gene family. The expression of one gene interferes with the expression of the neighboring gene; they exhibit a reciprocal expression pattern and are anticorrelated. Such an arrangement has been observed and extensively characterized at the *SER3* locus, which is regulated in a mutually exclusive manner by the adjacent noncoding gene *SRG1* (38, 39). Transcription of *SRG1* inhibits the transcription of *SER3*, with the intergenic repression mediated by nucleosome assembly (40).

**Role of chromatin remodeling in adjacent gene coregulation.** The EP theory demonstrates that *cis* regulatory sequences influence transcription throughout a genomic region (16). Changes to chromatin structure at a locus can modulate transcription throughout a region; we explored this possibility in coordinating the transcription of clustered genes. By taking advantage of the comprehensive screen monitoring gene expression changes after deletion of the nonessential genes comprising

**TABLE 4** Complete list of *Saccharomyces cerevisiae* strains used in this study and relevant genotypes

Strain	Spacer (bp)	Relevant genotype
YJA1508		<i>MAT<math>\alpha</math></i> <i>ura3-52</i> (wild type) (no spacer)
YJA1509	280	<i>MAT<math>\alpha</math></i> <i>his3<math>\Delta</math>200 lys2<math>\Delta</math>128 leu2<math>\Delta</math>0 ura3<math>\Delta</math>0 trp<math>\Delta</math>63 bph1<math>\Delta</math>::kanMX-UASGAL280-HIS3</i>
YJA1511	493	<i>MAT<math>\alpha</math></i> <i>his3<math>\Delta</math>200 lys2<math>\Delta</math>128 leu2<math>\Delta</math>0 ura3<math>\Delta</math>0 trp<math>\Delta</math>63 bph1<math>\Delta</math>::kanMX-UASGAL493-HIS3</i>
YJA1512	574	<i>MAT<math>\alpha</math></i> <i>his3<math>\Delta</math>200 lys2<math>\Delta</math>128 leu2<math>\Delta</math>0 ura3<math>\Delta</math>0 trp<math>\Delta</math>63 bph1<math>\Delta</math>::kanMX-UASGAL574-HIS3</i>
YJA1513	690	<i>MAT<math>\alpha</math></i> <i>his3<math>\Delta</math>200 lys2<math>\Delta</math>128 leu2<math>\Delta</math>0 ura3<math>\Delta</math>0 trp<math>\Delta</math>63 bph1<math>\Delta</math>::kanMX-UASGAL690-HIS3</i>
YJA1515	305	<i>MAT<math>\alpha</math></i> <i>his3<math>\Delta</math>200 leu2<math>\Delta</math>0 ura3<math>\Delta</math>0 trp1<math>\Delta</math>63 dug2<math>\Delta</math>::TRP1-UASGAL305-HIS3-natMX</i>
YJA1516	606	<i>MAT<math>\alpha</math></i> <i>his3<math>\Delta</math>200 leu2<math>\Delta</math>0 ura3<math>\Delta</math>0 trp1<math>\Delta</math>63 dug2<math>\Delta</math>::TRP1-UASGAL606-HIS3-natMX</i>
YJA1517	806	<i>MAT<math>\alpha</math></i> <i>his3<math>\Delta</math>200 leu2<math>\Delta</math>0 ura3<math>\Delta</math>0 trp1<math>\Delta</math>63 dug2<math>\Delta</math>::TRP1-UASGAL806-HIS3-natMX</i>

components of chromatin remodeling complexes, we identified putative regulators of adjacent gene coregulation (31). By focusing our analysis on the remodelers that disproportionately affect functional clusters compared to the singleton members in a gene family, our analysis excluded complexes that are specific regulators of each gene family. Our results implicate several chromatin remodeling complexes that are known regulators of transcription at a distance (Mediator) and coregulation at the *SRG1-SER3* locus (SAGA and the Swi/Snf complexes) (35, 39). We also identify additional regulatory complexes not previously identified in this process, including the NuA3 complex, RSC complex, and HIR complex. Our analysis identified several components from the same remodeling complexes, although there were several examples where only a single component was identified from a complex with multiple components. This observation requires further follow-up analysis and verification and are beyond the scope of this work.

**Conclusion.** The clustering of coregulated genes throughout the genome allows for greater transcriptional coordination by allowing for regulatory mechanisms to reinforce each other throughout a locus. The extension of the EP theory presented here extends the understanding of the wide-spread nature of this phenomenon and helps to complete the picture of how correlated domains of expression arise to streamline transcriptional coregulation. Such a mechanism extends throughout eukaryotic species of diverse lineages and complexity. Our work provides insight into the interconnectivity of spatial positioning in gene expression, with potential applications to the understanding of spatial positioning in humans and disease (41–43).

## MATERIALS AND METHODS

**Yeast strains and growth conditions.** The complete list of *S. cerevisiae* strains utilized in this study and relevant genes can be found in Table 4. For RNA extraction and gene expression analysis, strains of yeast were grown overnight at 30°C in YPAD (1% yeast extract, 2% peptone, 40 mg adenine, 2% dextrose [all per liter]) medium between early and mid-log phase (optical density at 600 nm [OD<sub>600</sub>], 0.40 to 0.90). Cultures of yeast were pelleted, washed once with ddH<sub>2</sub>O, and split prior to resuspension in selective medium (SC-HIS+galactose) and nonselective medium (SC-HIS+dextrose) for 2 or 3 doublings (to mid-log-phase growth) prior to RNA extraction. For the spotting assay, strains of yeast were grown at 30°C in YPAD medium to post-log phase, washed with ddH<sub>2</sub>O, and then serially diluted 10-fold. Three microliters of each dilution was then spotted onto selective plates (SC-HIS+galactose), nonselective plates (SC-HIS+dextrose), and control (SC) plates. Yeast cultures on the plates were grown for 72 h prior to imaging and analysis.

**RNA extraction and gene expression analysis.** Portions (15 ml) of yeast cultures were pelleted at 3,000 (3K) rpm for 3 min at room temperature and washed once with ddH<sub>2</sub>O. RNA was extracted using the ZR Fungal/Bacterial RNA MiniPrep kit (Zymo Research, CA) with the modifications outlined as follows. Pellets of yeast cells were resuspended in 800  $\mu$ l of RNA lysis buffer and transferred to a ZR Bashing Beads lysis tube, and cells were lysed with mechanical shearing by vortex (30-s vortex followed by 60-s incubation on ice for a total of eight cycles). DNase I digestion was performed in a column utilizing the RNase-free DNase I kit (Life Technologies Corporation, CA), and RNA was eluted into 25  $\mu$ l of RNase-free water. The efficacy of DNase I treatment was verified by endpoint PCR, targeting an *ACT1* amplicon.

cDNA was then synthesized using the TaqMan reverse transcription kit (Life Technologies Corporation, CA). DNase-treated RNA (500 to 1,000 ng) was utilized per reaction mixture using oligo(dT) primers, and cDNA was synthesized using a thermocycler according to the manufacturer's instructions. Newly synthesized cDNA was verified by amplification using endpoint PCR targeting the *ACT1* amplicon. Gene expression was then measured by qPCR utilizing PowerUp Sybr Green Master Mix (Life Technologies Corporation, CA) and a final reaction mixture volume of 25  $\mu$ l per reaction. Triplicate RNA reactions were obtained (biological replicates), and at least three technical replicates were performed per sample. Reactions that did not amplify or were nonspecific (e.g., more than a single peak observed on a melt

**TABLE 5** Complete sequences of PCR primers utilized in this study

Primer	Target	Sequence (5' – 3')
prJA0047	<i>HIS3</i> FP	CAGAAGCAGTAGCAGAACAG
prJA0048	<i>HIS3</i> RP	ATGGTCGTCTATGTGTAAGTC
prJA0051	<i>RPS14A</i> FP	AGAGTTACTGGTGGTATGAAG
prJA0052	<i>RPS14A</i> RP	CTGGAGTCTTGGTTCTAGTAC
prJA0053	<i>ELO2</i> FP	CTGGTGAAGGAATGGGTTAC
prJA0054	<i>ELO2</i> RP	AGTTGTTGAACCCACACAGTC
prJA0055	<i>MRPL27</i> FP	TAAAAGAGTGCCGTTGACCAC
prJA0056	<i>MRPL27</i> RP	TGGGGTAACATAGGTTCTGAC
prJA0057	<i>SAF1</i> FP	TCCCAGTGGTTCACAGCATGA
prJA0058	<i>SAF1</i> RP	CCGTGGTATTGACAGTACTC
prJA0063	<i>SNT1</i> FP	AGGGGTGTATTTCCATTAC
prJA0064	<i>SNT1</i> RP	GTGCGTAAAAGGATACTCTG

curve analysis) were discarded from our analysis. Relative expression and fold enrichment were determined using  $2^{\Delta\Delta CT}$  method with *ACT1* as a reference (44). The complete list of PCR primers and their sequences utilized for qPCR analysis can be found in Table 5.

**Determination of transcriptional coregulation at a distance and calculation of the Spearman's correlation coefficient.** The Spearman's correlation coefficient was computed as described previously utilizing the formula:

$$\rho = \frac{\text{cov}(g_1, g_2)}{\sigma_{g_1} \times \sigma_{g_2}}$$

where  $g_1$  and  $g_2$  are the corresponding genes for comparison, cov is their covariance, and  $\sigma$  is their standard deviation (SD) (16). Gene expression profiles were obtained from previously published microarray data sets and analyzed. Data were extracted for the following conditions for analysis: cells cycling through the cell cycle; chemostat growth limiting for carbon, nitrogen, phosphorus, or sulfur; heat shock induction by shifting cultures growing at 30°C to 37°C; oxidative stress induced by the addition of 0.3mM hydrogen peroxide, DNA damage induced by the addition of 0.03% MMS, nitrogen starvation induced by growth in SD media, and carbon source shifted from glucose to glycerol. Prior to the analysis, duplicate data points were averaged together to generate a single time series for each gene and then normalized to *TDH3* expression to allow comparisons across different microarray data sets (27–29). The SCC was calculated for every pairwise combination of genes within a 10-gene window surrounding each locus or until the end of the chromosome was reached. Data were plotted and fit to a decay curve in Excel for analysis.

**Identification of chromatin remodelers that disrupt adjacent gene coregulation.** The data set from the chromatin interaction deletion study was analyzed to identify mutants that preferentially disrupted transcription of the functionally clustered genes relative to that of the unpaired members of the same set (threshold of  $P < 0.05$ ) in the NM, RP, TR, and HS gene sets. This threshold of transcriptional disruption represents the deviation in transcription measured in a mutant strain compared to an isogenic, wild-type strain of yeast. The significance was subsequently determined utilizing a hypergeometric probability density function:

$$P = 1 - \sum_k \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

where  $P$  is the probability,  $K$  is the total number of genes disrupted,  $k$  is the number of genes in the subset disrupted,  $n$  is the number of genes in the subset, and  $N$  is the total number of genes with measured  $P$  values in the microarray data set as previously described (24, 31).

## ACKNOWLEDGMENTS

We offer our heartfelt thanks to Fred Winston (Harvard Medical School, MA) for providing yeast strains utilized in this study and for his words of support, to Adam Robbins-Pianka (Boulder BioConsulting, Inc., CO) for his assistance interpreting the hypergeometric probability density function, and to Kasey-Dee Arnone for her critical reading of the manuscript and helpful discussions throughout the work.

We acknowledge support for this research has come from the Center for Research in the College of Science and Health and for ART from William Paterson University. Support and funding for undergraduate research came from the Louis Stokes Alliance for Minority Participation (LS-AMP) and from the Center for Research in the College of Science and Health at William Paterson University.

## REFERENCES

- Lelli KM, Slattery M, Mann RS. 2012. Disentangling the many layers of eukaryotic transcriptional regulation. *Annu Rev Genet* 46:43–68. <https://doi.org/10.1146/annurev-genet-110711-155437>.
- Shlyueva D, Stampfel G, Stark A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* 15:272–286. <https://doi.org/10.1038/nrg3682>.
- Kim TK, Shiekhattar R. 2015. Architectural and functional commonalities between enhancers and promoters. *Cell* 162:948–959. <https://doi.org/10.1016/j.cell.2015.08.008>.
- Reiter F, Wienerroither S, Stark A. 2017. Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev* 43:73–81. <https://doi.org/10.1016/j.gde.2016.12.007>.
- Todeschini AL, Georges A, Veitia RA. 2014. Transcription factors: specific DNA binding and specific gene regulation. *Trends Genet* 30:211–219. <https://doi.org/10.1016/j.tig.2014.04.002>.
- Johnson DG, Dent SY. 2013. Chromatin: receiver and quarterback for cellular signals. *Cell* 152:685–689. <https://doi.org/10.1016/j.cell.2013.01.017>.
- Weber CM, Henikoff S. 2014. Histone variants: dynamic punctuation in transcription. *Genes Dev* 28:672–682. <https://doi.org/10.1101/gad.238873.114>.
- Venkatesh S, Workman JL. 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol* 16:178–189. <https://doi.org/10.1038/nrm3941>.
- Gottschling DE, Aparicio OM, Billington BL, Zakian VA. 1990. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* 63:751–762. [https://doi.org/10.1016/0092-8674\(90\)90141-Z](https://doi.org/10.1016/0092-8674(90)90141-Z).
- Elgin SC, Reuter G. 2013. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb Perspect Biol* 5:a017780. <https://doi.org/10.1101/cshperspect.a017780>.
- Tennen RI, Bua DJ, Wright WE, Chua KF. 2011. SIRT6 is required for maintenance of telomere position effect in human cells. *Nat Commun* 2:433. <https://doi.org/10.1038/ncomms1443>.
- Cohen BA, Mitra RD, Hughes JD, Church GM. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet* 26:183–186. <https://doi.org/10.1038/79896>.
- Spellman PT, Rubin GM. 2002. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J Biol* 1:5. <https://doi.org/10.1186/1475-4924-1-5>.
- Williams EJ, Bowles DJ. 2004. Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res* 14:1060–1067. <https://doi.org/10.1101/gr.2131104>.
- Tsai HK, Huang PY, Kao CY, Wang D. 2009. Coexpression of neighboring genes in the zebrafish (*Danio rerio*) genome. *Int J Mol Sci* 10:3658–3670. <https://doi.org/10.3390/ijms10083658>.
- Quintero-Cadena P, Sternberg PW. 2016. Enhancer sharing promotes neighborhoods of transcriptional regulation across eukaryotes. *G3 (Bethesda)* 6:4167–4174. <https://doi.org/10.1534/g3.116.036228>.
- Fromont-Racine M, Senger B, Saveanu C, Fasiolo F. 2003. Ribosome assembly in eukaryotes. *Gene* 313:17–42. [https://doi.org/10.1016/S0378-1119\(03\)00629-2](https://doi.org/10.1016/S0378-1119(03)00629-2).
- Lempiainen H, Shore D. 2009. Growth control and ribosome biogenesis. *Curr Opin Cell Biol* 21:855–863. <https://doi.org/10.1016/j.ceb.2009.09.002>.
- Bosio MC, Negri R, Dieci G. 2011. Promoter architectures in the yeast ribosomal expression program. *Transcription* 2:71–77. <https://doi.org/10.4161/trns.2.2.14486>.
- Bosio MC, Fermi B, Dieci G. 2017. Transcriptional control of yeast ribosome biogenesis: a multifaceted role for general regulatory factors. *Transcription* 8:254–260. <https://doi.org/10.1080/21541264.2017.1317378>.
- Wade CH, Umbarger MA, McAlear MA. 2006. The budding yeast rRNA and ribosome biosynthesis (RRB) regulon contains over 200 genes. *Yeast* 23:293–306. <https://doi.org/10.1002/yea.1353>.
- Arnone JT, Robbins-Pianka A, Arace JR, Kass-Gergi S, McAlear MA. 2012. The adjacent positioning of co-regulated gene pairs is widely conserved across eukaryotes. *BMC Genomics* 13:546. <https://doi.org/10.1186/1471-2164-13-546>.
- Arnone JT, McAlear MA. 2011. Adjacent gene pairing plays a role in the coordinated expression of ribosome biogenesis genes MPP10 and YJR003C in *Saccharomyces cerevisiae*. *Eukaryot Cell* 10:43–53. <https://doi.org/10.1128/EC.00257-10>.
- Arnone JT, Arace JR, Soorneedi AR, Citino TT, Kamitaki TL, McAlear MA. 2014. Dissecting the cis and trans elements that regulate adjacent-gene coregulation in *Saccharomyces cerevisiae*. *Eukaryot Cell* 13:738–748. <https://doi.org/10.1128/EC.00317-13>.
- Eldabagh RS, Mejia NG, Barrett RL, Monzo CR, So MK, Foley Jt, Arnone JT. 2018. Systematic identification, characterization, and conservation of adjacent-gene coregulation in the budding yeast *Saccharomyces cerevisiae*. *mSphere* 3:e00220-18. <https://doi.org/10.1128/mSphere.00220-18>.
- Dobi KC, Winston F. 2007. Analysis of transcriptional activation at a distance in *Saccharomyces cerevisiae*. *Mol Cell Biol* 27:5575–5586. <https://doi.org/10.1128/MCB.00459-07>.
- Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 9:3273–3297. <https://doi.org/10.1091/mbc.9.12.3273>.
- Tirosh I, Reikhav S, Sigal N, Assia Y, Barkai N. 2010. Chromatin regulators as capacitors of interspecies variations in gene expression. *Mol Syst Biol* 6:435. <https://doi.org/10.1038/msb.2010.84>.
- Knijnenburg TA, de Winde JH, Daran JM, Daran-Lapujade P, Pronk JT, Reinders MJ, Wessels LF. 2007. Exploiting combinatorial cultivation conditions to infer transcriptional regulation. *BMC Genomics* 8:25. <https://doi.org/10.1186/1471-2164-8-25>.
- Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* 11:4241–4257. <https://doi.org/10.1091/mbc.11.12.4241>.
- Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NA, Margaritis T, van de Pasch LA, van Heesch SA, Brok MO, Groot Koerkamp MJ, Ko CW, van Leenen D, Sameith K, van Hooff SR, Lijnzaad P, Kemmeren P, Hentrich T, Kobor MS, Buratowski S, Holstege FC. 2011. The specificity and topology of chromatin interaction pathways in yeast. *Mol Cell* 42:536–549. <https://doi.org/10.1016/j.molcel.2011.03.026>.
- Wu XL, Li BZ, Zhang WZ, Song K, Qi H, Dai JB, Yuan YJ. 2017. Genome-wide landscape of position effects on heterogeneous gene expression in *Saccharomyces cerevisiae*. *Biotechnol Biofuels* 10:189. <https://doi.org/10.1186/s13068-017-0872-3>.
- Ben-Shitrit T, Yosef N, Shemesh K, Sharan R, Ruppin E, Kupiec M. 2012. Systematic identification of gene annotation errors in the widely used yeast mutation collections. *Nat Methods* 9:373–378. <https://doi.org/10.1038/nmeth.1890>.
- Atias N, Kupiec M, Sharan R. 2016. Systematic identification and correction of annotation errors in the genetic interaction map of *Saccharomyces cerevisiae*. *Nucleic Acids Res* 44:e50. <https://doi.org/10.1093/nar/gkv1284>.
- Reavey CT, Hickman MJ, Dobi KC, Botstein D, Winston F. 2015. Analysis of polygenic mutants suggests a role for Mediator in regulating transcriptional activation distance in *Saccharomyces cerevisiae*. *Genetics* 201:599–612. <https://doi.org/10.1534/genetics.115.181164>.
- Wade C, Shea KA, Jensen RV, McAlear MA. 2001. EBP2 is a member of the yeast RRB regulon, a transcriptionally coregulated set of genes that are required for ribosome and rRNA biosynthesis. *Mol Cell Biol* 21:8638–8650. <https://doi.org/10.1128/MCB.21.24.8638-8650.2001>.
- Arnone JT. 2018. Ribosome biogenesis: streamlining the genome for the efficient production of this biological nanomolecular machine. *Nanomed Nanotechnol J* 2:118.
- Martens JA, Laprade L, Winston F. 2004. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* 429:571–574. <https://doi.org/10.1038/nature02538>.
- Martens JA, Wu PY, Winston F. 2005. Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. *Genes Dev* 19:2695–2704. <https://doi.org/10.1101/gad.1367605>.
- Hainer SJ, Pruneski JA, Mitchell RD, Monteverde RM, Martens JA. 2011. Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev* 25:29–40. <https://doi.org/10.1101/gad.1975011>.
- Ebisuya M, Yamamoto T, Nakajima M, Nishida E. 2008. Ripples from neighbouring transcription. *Nat Cell Biol* 10:1106–1113. <https://doi.org/10.1038/ncb1771>.

42. Mendillo ML, Santagata S, Koeva M, Bell GW, Hu R, Tamimi RM, Fraenkel E, Ince TA, Whitesell L, Lindquist S. 2012. HSF1 drives a transcriptional program distinct from heat shock to support highly malignant human cancers. *Cell* 150:549–562. <https://doi.org/10.1016/j.cell.2012.06.031>.
43. Zhang CQ, Williams H, Prince TL, Ho ES. 2017. Overexpressed HSF1 cancer signature genes cluster in human chromosome 8q. *Hum Genomics* 11:35. <https://doi.org/10.1186/s40246-017-0131-5>.
44. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  method. *Methods* 25:402–408. <https://doi.org/10.1006/meth.2001.1262>.