UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Regulation and function of antisense transcription in Saccharomyces yeast

Permalink
https://escholarship.org/uc/item/6q55h0n6

Author
Mostovoy, Yulia

Publication Date
2014

Peer reviewed|Thesis/dissertation
Regulation and function of antisense transcription in *Saccharomyces* yeast

by

Yulia Mostovoy

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

and the Designated Emphasis

in

Computational and Genomic Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Rachel B. Brem, Chair
Professor Kathleen Collins
Professor Jasper Rine
Professor Daniel Zilberman

Spring 2014
Regulation and function of antisense transcription in *Saccharomyces* yeast

© 2014

by Yulia Mostovoy
Abstract

Regulation and function of antisense transcription in *Saccharomyces* yeast

by

Yulia Mostovoy

Doctor of Philosophy in Molecular and Cell Biology

Designated Emphasis in Computational and Genomic Biology

University of California, Berkeley

Professor Rachel Brem, Chair

Transcription of RNA antisense to protein-coding genes is widespread in genomes from bacteria to human. For most antisense transcription, the methods by which it is regulated and its potential function remain unknown. We sought to address these questions, using budding yeast as a model system, with two complementary approaches. First, we mapped antisense expression in four *Saccharomyces* species. Antisense transcripts conserved across yeasts were predominantly detected at open reading frames in a tail-to-head orientation with respect to the next gene 3' to the reference gene. For such tandem gene pairs, the region between the genes exhibited distinctive signatures of binding by transcription factors, suggesting that these factors could regulate transcription of both the mRNA from the 3' gene and the upstream antisense transcript. Transcription factor deletion experiments supported this hypothesis, conferring decreased expression of both antisense and downstream sense transcripts at such gene pairs; often in these mutants, as an antisense transcript dropped in expression, its host gene mRNA level increased, a hallmark of antisense-mediated repression. To test this model, we focused on the stress-response gene *YKL151C* and its downstream neighbor *GPM1*, which was strongly expressed in rich media. *Cis*-regulatory mutation experiments showed that *YKL151C* antisense expression was co-regulated with *GPM1* and repressed mRNA levels of its overlapping sense transcript. In a second body of work, we used natural variation rather than engineered mutations to access the impact of variants that modulate antisense expression levels. We mapped naturally occurring polymorphisms between yeast strains that showed linkage to sense and antisense expression. We identified cases where such regulatory polymorphisms mediated their effects through transcription factors that bound at the promoters of linked antisense transcripts; in all such cases, expression of the downstream gene also showed linkage to the polymorphism, and, in several cases, sense transcription overlapping the antisense was linked as well. Our studies thus converged on a model in which antisense transcription at one yeast gene frequently originates from and is co-regulated with the promoter of a neighboring gene, mediated by the action of transcription factors; at a fraction of such loci, antisense
transcription acts to repress expression of its overlapping mRNA, enabling the joint control of adjacent genes specialized to opposing conditions.
# Table of contents

**ACKNOWLEDGEMENTS** ii

**CHAPTER 1: Introduction to antisense transcription** 1
- Classes of functional non-coding RNAs 1
- Global antisense transcription 2
- Biogenesis of antisense transcription 3
- Repression of antisense transcription 4
- Mechanisms of function 5

**CHAPTER 2: Transcription factors regulate antisense expression and function in yeast** 7
- Introduction 7
- Results 9
  - A survey of antisense transcription conserved across yeast species 9
  - Transcription factor binding is associated with conserved antisense expression 10
  - Antisense expression requires transcription factors and shows signatures of regulatory function 10
  - Antisense transcription at YKL151C has a repressive function 11
- Discussion 13
- Materials and Methods 16
- Figures 21
- Tables 42

**CHAPTER 3: eQTL mapping of antisense transcription in *Saccharomyces cerevisiae*** 51
- Introduction 51
- Results 52
  - A survey of natural variation in yeast sense and antisense expression 52
  - eQTL effects on neighboring and overlapping transcribed elements 52
- Discussion 55
- Materials and Methods 57
- Figures 59
- Tables 63

**CHAPTER 4: Divergence of iron metabolism in wild Malaysian yeast** 67
- Introduction 67
- Results 69
- Discussion 70
- Materials and Methods 72
- Figures 73

**REFERENCES** 75
Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Rachel Brem, for her patient support and mentorship over the past six years. Her enthusiasm and guidance have been invaluable. I would also like to thank the members of my thesis committee, Dr. Jasper Rine, Dr. Kathleen Collins, and Dr. Daniel Zilberman, for providing advice and feedback along the way.

I would like to thank the members of the Brem lab for helpful discussions and invigorating coffee breaks. Thanks especially to Oh Kyu Yoon for help with his 3’ end RNA-seq protocol and for sharing his linkage data, and to Alex Thiemicke for working so hard on gathering data for this project. Thanks also to members of the Rine and Eisen labs, who lent me their expertise, equipment, and reagents many times over the years.

Finally, I would like to thank my family for their endless support and love and encouragement, and my amazing fiancé Mike for, among many other things, being so patient as I finished this Ph.D.
CHAPTER 1

Introduction to antisense transcription

The function and regulation of non-protein-coding transcription remains a central question in the field of genomics. While protein-coding sequences account for less than 2% of the human genome (International Human Genome Sequencing Consortium 2004), more than 90% of the genome is transcribed on one or both strands (Pertea 2012), including regions of the genome that overlap coding genes on the opposite strand. The question of whether this pervasive transcription is regulated and functional or opportunistic and nonfunctional has been a matter of ongoing debate (Berretta and Morillon 2009; Layer and Weil 2009; Pertea 2012; van Bakel and Hughes 2009). In recent years, this vast non-coding portion of the genome has increasingly been implicated in functional roles (The ENCODE Project Consortium 2012), acting as part of the web of regulation that allows for high levels of organismal complexity (Mattick 2001). A comprehensive understanding of the genome will require thorough characterization of the regulation and the function of non-coding transcription.

Classes of functional non-coding RNAs

Non-coding RNAs have long been known to constitute critical components of cellular infrastructure, including tRNAs used in translation (Rich and RajBhandary 1976) and rRNAs that comprise essential segments of the ribosome (Noller 1991). Small nuclear RNAs perform processes such as regulating splicing (Zhuang and Weiner 1986), or, in the case of small nucleolar RNAs, guiding modifications of other non-coding RNAs (Bachellerie et al. 2002).

In past decades, our stable of known types of non-coding RNAs has grown to include many classes that function in the regulation of gene expression and/or translation. A prominent example of these are microRNAs (miRNAs), short RNAs of ~22bp that were first discovered in C. elegans and have since been identified in organisms throughout eukaryotes (Bartel 2004). Mature processed miRNAs target transcripts via significant yet imperfect complementarity and inhibit their translation. miRNAs have proven to be critical building blocks of eukaryotic complexity, contributing to key developmental processes, tissue identity, and, when misregulated, to many types of cancers (Bushati and Cohen 2007). Other known classes of small regulatory RNAs include Piwi-interacting RNAs (piRNAs), which repress transposition of mobile genetic elements in germ cells (Aravin et al. 2007), and small interfering RNAs (siRNAs), which are produced from double-stranded RNA and trigger the degradation of RNAs that share their sequence (Ghildiyal and Zamore 2009).

Distinct from this multitude of short RNAs is the class of long non-coding RNAs (lncRNAs) that function through a wide variety of mechanisms (Wang and Chang 2011). Notable examples include HOTAIR, which acts as a scaffold for the assembly of chromatin-modifying proteins that it transports to sites across the genome (Chu et al. 2011).
a suite of IncRNAs on the X chromosome that recruit histone modification enzymes in cis during the process of X chromosome inactivation in mammals (Lee 2010); and TERRA, IncRNAs transcribed from telomeres that appear to regulate telomere length and telomerase activity (Cusanelli and Chartrand 2014).

Global antisense transcription

In addition to non-coding RNAs that are transcribed from intergenic or intronic regions, non-coding transcription also occurs antisense to coding regions. This class of transcription occurs in many diverse organisms (Dornenburg et al. 2010; Georg et al. 2009; He et al. 2008; Katayama et al. 2005; Ma et al. 2006; Ni et al. 2010; Teodorovic et al. 2007; Wang et al. 2005; Xu et al. 2009). Although estimates of antisense abundance tend to vary by species and study methodology (Ni et al. 2010), antisense transcription occurs at a substantial fraction of genes. Antisense transcription was detected with multiple lines of evidence at 66% of mouse genes (Katayama et al. 2005). In human, the ENCODE consortium annotated 4220 distinct antisense loci at 21% of coding genes (Harrow et al. 2012); a previous study detected antisense transcription in over 30% of human genes (He et al. 2008). In budding yeast, tiling arrays (Xu et al. 2009) and RNA-seq (Yassour et al. 2010) detected antisense transcription at 16-20% of genes. While some loci generate more antisense transcription than sense (Ni et al. 2010), antisense transcription in most organisms and across most genes is of low-abundance, comprising less than 1% of poly-adenylated transcriptional output in budding yeast (Yassour et al. 2010).

Complicating attempts at detection and quantification of antisense transcription is the observation that many antisense transcripts are rapidly degraded. Disabling the nuclear exosome in budding yeast resulted in the detection of roughly twice as many antisense transcripts as were detected in wild-type cells (Xu et al. 2009); these exosome-degraded transcripts have been called CUTs, or cryptic unstable transcripts. Another study disabled the cytoplasmic degradation factor Xrn1, and identified up-regulated antisense transcripts at an additional ~16% of genes (van Dijk et al. 2011). While CUTs were not detected by tiling array in wild-type cells, the vast majority of them were detected in wild-type cells by deep RNA-seq (Yoon and Brem 2010), highlighting the need for sensitive assays to detect these low-abundance transcripts.

In some cases, degradation of antisense RNAs is regulated in a condition-specific manner and is linked to functional antisense transcription. For example, the antisense transcript at PHO84 is degraded by the nuclear exosome in young yeast cells but escapes degradation and accumulates in aging cells, causing the down-regulation of PHO84 mRNA transcription (Camblong et al. 2007). More broadly, nuclear exosome function is impaired during meiosis via degradation of exonuclease component Rrp6, coinciding with the increased accumulation of antisense and other non-coding RNA (Lardenois et al. 2011). In fission yeast, the nuclear exosome is known to be responsible for degrading meiotic mRNAs during vegetative growth (McPheeters et al. 2009); an analogous situation in budding yeast may underlie the need for Rrp6
downregulation during meiosis. However, the purpose of meiosis-accumulating antisense RNAs in budding yeast is as yet unclear.

**Biogenesis of antisense transcription**

Critical to our understanding of the biology of antisense transcription is the question of its biogenesis and regulation. A subset of antisense RNAs may be opportunistic byproducts of particular genomic environments, e.g. open chromatin or a high local concentration of initiation factors (Struhl 2007). These antisense transcripts may be produced without being functional, so the search for functional antisense transcription would be greatly aided by identifying antisense transcripts that are independently regulated rather than being byproducts.

In fact, many individual functional antisense RNAs have been shown to derive from independently regulated promoters (Gelfand et al. 2011; Heo and Sung 2011; Hongay et al. 2006; Houseley et al. 2008; Nagano et al. 2008; Nishizawa et al. 2008; Pandey et al. 2008; Stavropoulos et al. 2005; Zhao et al. 2013). One prominent example comes from budding yeast, where antisense transcription regulates IME4, a regulator of entry into meiosis. The antisense promoter region harbors a binding site for the a1/α2 transcription factor, which is only expressed in diploid cells and which represses antisense transcription in these cells. In haploids, the antisense transcript is highly expressed and inhibits expression of IME4. Mutant diploid cells that ectopically express the antisense RNA are unable to initiate meiosis (Hongay et al. 2006).

Conversely, many antisense transcripts show signs of being opportunistic. The majority of antisense transcripts in budding yeast originate near the 3’ end of the sense-strand gene, initiating from a nucleosome-free region shared with either the 3’ end of the sense gene or, more commonly, with the promoter of the adjacent downstream gene; this promoter is referred to as bidirectional because it is associated with divergent transcriptional initiation of the downstream sense and the upstream antisense transcript (Neil et al. 2009; Teodorovic et al. 2007; Xu et al. 2009). Among loci where antisense transcripts originate from such bidirectional promoters, expression of the downstream gene is faintly but significantly correlated with upstream antisense expression (Xu et al. 2009). Similarly, in Giardia lamblia, ~50% of transcribed loci harbor antisense transcription, which comprises over 20% of poly-adenylated transcripts in this organism, and largely stems from bidirectional promoters (Teodorovic et al. 2007).

The biology of bidirectional promoters and the general question of how promoters enforce directionality is not yet fully understood. An important insight into bidirectional transcription was the observation that bidirectional promoters in yeast contain two distinct pre-initiation complexes assembled at opposite ends of a nucleosome-free region (Murray et al. 2012; Rhee and Pugh 2012). Earlier analyses of promoters in yeast revealed some sequence determinants of promoter directionality, such as the position of transcription factor binding sites relative to the TATA box (Smale and Kadonaga 2003). Promoter directionality is additionally enforced through the formation of gene loops (Tan-Wong et al. 2012). Furthermore, metazoan genomes contain a
promoter element called BRE that is located upstream of the TATA box and that binds TFIIB; the orientation of BRE relative to the TATA box guides the orientation of the assembling pre-initiation complex (Littlefield et al. 1999).

Nonetheless, bidirectional promoters are widespread in higher eukaryotes, where most active promoters have been observed to generate short antisense-oriented ncRNAs whose expression correlates with the adjacent gene (Jacquier 2009; Seila et al. 2008). In these organisms, directionality appears to be enforced at the level of termination rather than initiation (Almada et al. 2013). On the basis of the data from these diverse eukaryotes, it has been suggested that eukaryotic promoters are intrinsically bidirectional (Jacquier 2009; Neil et al. 2009).

In many cases, both transcripts arising from a bidirectional promoter are functional and not byproducts of opportunistic transcription. This principle is most clearly illustrated in cases where bidirectional promoters include a pair of coding genes rather than a coding and non-coding transcript. In yeast, many genes are transcribed divergently from a single shared nucleosome-free region (Xu et al. 2009). Even in the relatively dispersed human genome, over 10% of genes initiate divergently within 1 kb of one another, with nearly two-thirds of those initiating within 300 bp (Trinklein et al. 2004). Overall, such gene pairs were significantly positively correlated in expression, although a subset showed anti-correlated expression.

It remains unclear to what extent the two transcripts emerging from bidirectional promoters are individually regulated. The mechanism of co-expression between transcripts may involve the sharing of chromatin state and local concentrations of various transcription factors between the two pre-initiation complexes (Chen et al. 2010). In yeast, bidirectional transcripts are weakly but significantly co-regulated on a global level (Xu et al. 2009). This claim has been disputed on the level of individual loci, where some antisense transcripts are expressed at high levels despite low levels of downstream gene transcription, and vice versa (Murray et al. 2012). One study of an individual locus found that, when the TATA box upstream of the sense mRNA was disrupted and sense transcription decreased, bidirectional antisense transcription increased by a large margin (Neil et al. 2009), suggesting that, as in human (Trinklein et al. 2004), expression at some bidirectional transcript pairs in yeast is anti-correlated.

**Repression of antisense transcription**

If antisense transcription frequently arises opportunistically and has the potential for function, some antisense transcripts would likely be deleterious and would need to be repressed. Accordingly, several mechanisms for repression of antisense transcription have been characterized in budding yeast. One mechanism involves the chromatin remodeling factor Isw2 (Whitehouse et al. 2007). This factor repositions the nucleosome at the 5’ end of transcripts, moving it upstream so as to reduce the size of the promoter-associated nucleosome free region. At the 3’ ends of transcripts, Isw2 moves the last nucleosome downstream, again reducing the size of the termination-associated nucleosome free region. Among genes with the latter type of repositioning, more than
75% were positioned adjacent to a downstream gene promoter, suggesting that Isw2 may be involved in repressing bidirectional antisense transcription at these loci. Indeed, loss of Isw2 resulted in the accumulation of many antisense RNAs (Yadon et al. 2010).

Another mechanism of antisense repression involves the modification of histone residues. The histone methyltransferase Set2 associates with the elongating polymerase complex and trimethylates the histone residue H3K36, which is then recognized by the histone deacetylase Rpd3S. This chain of events leads to deacetylation of histones near the 3' ends of genes, which would otherwise be acetylated during the act of transcription. By closing chromatin, the Set2/Rpd3S pathway helps prevent the initiation of cryptic transcription from within genes (Lickwar et al. 2009). It also represses antisense transcription in both fission and budding yeast (Churchman and Weissman 2011; Nicolas et al. 2007).

**Mechanisms of function**

The characterization of functional antisense transcription has revealed a wide diversity of mechanisms. These can be grouped into several broad classes, one of which involves the formation of RNA duplexes with the sense mRNA. In eukaryotes with RNAi machinery, these duplexes can be processed to generate siRNA that targets the locus for regulation (Jin et al. 2008; Okamura et al. 2008; Watanabe et al. 2008). Antisense transcripts can also regulate alternative splicing by binding to sense mRNAs and occluding access to splicing motifs (Beltran et al. 2008). Similarly, antisense RNA can occlude other kinds of regulatory information in the mRNA, such as degradation signals within the UTR, as in the case with p53 (Mahmoudi et al. 2009).

Another class of functional antisense transcripts acts via direct interaction with proteins. Two well-characterized mammalian examples, Air and Kcnq1ot1, direct imprinting by binding to histone-modifying proteins and recruiting them to the encoding locus, where they deposit repressive chromatin marks (Nagano et al. 2008; Pandey et al. 2008). Interestingly, both of these transcripts function in cis, leading to the suggestion that these long antisense RNAs may be localized to their encoding loci via cotranscriptional tethering (Wang and Chang 2011). In yeast, antisense expression at the PHO84 locus is capable of suppressing PHO84 mRNA expression in trans through a mechanism that remains undetermined but that requires the antisense RNA to contain a sequence homologous to the PHO84 promoter region (Camblong et al. 2009), suggesting that the RNA may recruit regulatory factors to the locus via sequence complementarity. There exists precedent for trans-acting recruitment among other long ncRNAs, such as HOTAIR, which serves as a scaffold for the assembly of repressive complexes and guides them to hundreds of sites across the genome (Chu et al. 2011).

A final class of antisense transcription functions via the act of transcription itself, rather than via the RNA that is generated. Consequently, these cases regulate their overlapping genes in cis. One such example from budding yeast is at the GAL1-10 locus, which encodes genes essential for the metabolism of galactose. Antisense transcription at this locus lays down repressive marks across the sense mRNA.
promoters, and is functional in borderline conditions in which the cell needs to commit to metabolism of either glucose or galactose (Houseley et al. 2008). Another example occurs at the IME4 locus, where antisense transcription interferes with sense mRNA expression in cis via a mechanism that requires the two transcripts to overlap (Hongay et al. 2006). Finally, antisense transcription overlapping the promoter of PHO5 helps to maintain its chromatin in an open state, leading to more rapid induction of the sense mRNA (Uhler et al. 2007). More generally, trans-acting antisense regulation in budding yeast appears to be the exception rather than the rule, with only two known cases (Berretta et al. 2008; Camblong et al. 2009), while cis-acting regulation is more prominent. This trend may be partially attributable to the lack of RNAi in budding yeast, eliminating a potential trans-acting regulatory mechanism from these cells.

In these and other cases, the regulatory functions of individual antisense transcripts have been experimentally dissected, but it is unclear to what extent these findings can be generalized, and the regulation and function of the multitude of other cases remain unknown. To address these questions, we have used budding yeast as a model system to study the conservation, biogenesis, regulation, and functions of antisense transcription. In Chapter 2, we examine conservation of antisense transcription across budding yeast species, and develop and test the hypothesis that conserved antisense transcription tends to be regulated by transcription factors binding at bidirectional promoters. Focusing on one promising locus, we experimentally validate the transcription factor-based regulation of antisense transcription and conclude that it acts as a repressor of its overlapping sense transcript. In Chapter 3, we take an analogous approach, where instead of modulating antisense transcription via deletion of transcription factors, we instead harness the natural variation in antisense transcription already present between two diverged strains of yeast. We use this natural variation to dissect the regulation of antisense transcription and identify additional candidate cases of functional antisense transcription. We find that, in Chapters 2 and 3, these different approaches yield convergent and consistent results to the question of antisense regulation and function. Finally, in Chapter 4, we use natural variation between yeast strains to address a different question, regarding the genetic determinants of expression divergence among genes involved in iron metabolism.
Chapter 2

Transcription factors regulate antisense expression and function in yeast


INTRODUCTION

Transcriptional profiling in many organisms has uncovered widespread expression at loci that do not encode proteins or other known functional elements (Berretta and Morillon 2009). Understanding how such transcripts are generated and what functions, if any, they carry out in cells is a key challenge in modern genomics. Many putatively noncoding RNAs are transcribed antisense to coding genes, overlapping them on the opposite strand (Dornenburg et al. 2010; He et al. 2008; Katayama et al. 2005; Wang et al. 2005; Xu et al. 2009). In landmark cases, examples of antisense RNA transcription have been shown to play functional regulatory roles, often repressing transcription of their respective sense genes (Chen et al. 2012; Li et al. 2012; Magistri and Faghihi 2012; Pelechano and Steinmetz 2013; Zhao et al. 2013). Globally, the proportion of antisense transcription in a given genome that plays a regulatory function is a matter of ongoing debate (Layer and Weil 2009; Struhl 2007; Wei et al. 2011).

Critical to the understanding of the cellular role of antisense RNAs is the question of their biogenesis and regulation. Individual cases of functional antisense RNA have been shown to derive from their own distinct and independently-regulated promoters (Hongay et al. 2006; Houseley et al. 2008; Nishizawa et al. 2008; Pandey et al. 2008; Stavropoulos et al. 2005; Zhao et al. 2013), and genomic analyses of particular transcription factors have detected an association between these proteins’ localization and the expression of antisense and other noncoding RNAs (Cawley et al. 2004; Impey et al. 2004). For the vast majority of antisense RNAs, however, dissecting mechanisms of biogenesis and regulation remains an area of active research. Much of the recent literature has focused on promoters that drive expression of a sense mRNA in one direction and a noncoding RNA in the opposite direction on the other strand (Taft et al. 2009). In organisms with compact genomes, such as budding yeast, these promoters account for the majority of antisense transcripts (Neil et al. 2009; Teodorovic et al. 2007; Xu et al. 2009). Bidirectional promoters in yeast harbor two distinct pre-initiation complexes on opposite strands (Rhee and Pugh 2012), with the complex driving antisense transcription regulated independently of or coordinately with the divergently transcribed gene (Murray et al. 2012; Xu et al. 2009). Why these complexes assemble at some intergenic regions and not others is as yet poorly understood, though a number of determinants of antisense expression have been reported (Churchman and Weissman 2011; Tan-Wong et al. 2012; Whitehouse et al. 2007; Yadon et al. 2010).
To investigate the regulation of antisense transcripts and shed light on their potential for function, we took a comparative genomic approach. We first catalogued antisense transcription in four members of the *Saccharomyces* species complex. With the set of antisense transcripts conserved across yeasts, we set out to investigate the mechanisms and biological rationale for antisense expression in cells. We discovered a genome-scale role for transcription factors in antisense expression, and we used our insights as a jumping-off point for genomic and single-locus analysis of the regulatory logic of antisense RNAs in the compact yeast genome.
RESULTS

A survey of antisense transcription conserved across yeast species

To survey the genomic distribution of transcription in yeasts, we used our recently generated expression profiles of wild-type Saccharomyces cerevisiae, S. paradoxus, S. mikatae, and S. bayanus grown in rich media (Schraiber et al. 2013). Antisense-expressing loci in S. cerevisiae and S. paradoxus in this data set agreed well with results from other genomic techniques (Figure 2.1), including many transcripts whose detection was previously reported to depend on mutation of the nuclear exosome (Xu et al. 2009) (Figure 2.1B). Across our complete set of Saccharomyces transcriptomes, we detected antisense expression in at least one species at over half of all genes (57%), with antisense RNAs generally expressed at low levels (Figure 2.2A). At 220 genes, we detected antisense transcription in all four surveyed species, and a further 417 genes had antisense transcription in three species (Figure 2.2B). Given their high expression (Figure 2.3A), we reasoned that the union of the latter two classes would serve as a window onto the attributes of genes with abundant, conserved antisense transcription, which included loci at which antisense expression has previously been reported to play regulatory roles (Figure 2.4).

Our set of genes hosting conserved antisense RNAs exhibited the reduced sense expression (Figure 2.2C), tandem tail-to-head positioning with respect to downstream ORFs (Figure 2.5A-B and Figure 2.3B), and condition specificity (Figure 2.3C) previously seen in analyses of S. cerevisiae antisense transcription (Lardenois et al. 2011; Xu et al. 2009). Many of these attributes were more salient among loci with conserved antisense RNAs than in the larger set of loci at which antisense RNA was detected using data from S. cerevisiae alone. Following (Xu et al. 2009), we anticipated that tandem gene pairs, which included the majority of our detected cases of conserved antisense expression, would be a valid and useful platform for inference of the principles of antisense regulation and function. At such pairs, within the 3’ ends of upstream tandemly-oriented genes expressing conserved antisense RNAs, we observed the expected chromatin and protein occupancy signatures of transcriptional initiation (Goodman et al. 2013; Murray et al. 2012) (Table 2.1-2). We additionally noted that the intergenic region between tandem genes with conserved antisense transcription contained two closely spaced peaks of occupancy of the histone variant H2A.Z (Albert et al. 2007; Mavrich et al. 2008; Raisner et al. 2005), which was lacking at tandem genes without antisense expression (Figure 2.5C); we also observed this dual-peak occupancy pattern at divergent sense gene pairs sharing nucleosome-free regions (Figure 2.6), implicating it as a hallmark of bidirectional transcription. We conclude that at ~10% of yeast genes, antisense expression is detectable across divergent species, with functional, structural, and chromatin-related properties suggestive of a shared mechanism of expression among these antisense RNAs.
Transcription factor binding is associated with conserved antisense expression

We hypothesized that transcription factors were likely to be key players in the regulation of conserved antisense expression. To test this notion, we evaluated patterns of transcription factor binding sites (MacIsaac et al. 2006) within the intergenic regions between tandemly-oriented genes, comparing gene pairs with and without conserved antisense expression at their respective upstream genes. The results, shown in Figure 2.7, revealed a shift in the position of transcription factor binding sites \( (p = 1e^{-4}, \text{Wilcoxon rank-sum test}) \) at tandem gene pairs with conserved antisense expression. At such gene pairs, transcription factor binding sites lay farther from the start of transcription of the downstream mRNA, and closer to the upstream gene of the pair that harbored the antisense transcript (Figure 2.7). Control analysis ruled out a difference between the lengths of intergenic regions of the two groups of loci as a potential confounding factor \( (p = 0.98, \text{Wilcoxon rank-sum test}) \). We additionally noted a higher density of transcription factor binding sites in the intergenic regions of tandem pairs with conserved antisense expression \( (p = 1e^{-4}, \text{Wilcoxon rank-sum test}) \), reflecting more complex regulation at the latter loci. We conclude that conserved antisense expression is associated with a unique genomic signature of transcription factor binding at tandem gene pairs, strongly suggesting a genome-scale role for such factors in antisense regulation.

Antisense expression requires transcription factors and shows signatures of regulatory function

To pursue on a molecular basis the role of transcription factors in antisense biogenesis, we used our genome-scale data to infer the factors regulating individual tandem gene pairs that expressed antisense RNAs, and we evaluated these predictions in genetic experiments in S. cerevisiae. For this purpose we chose as a testbed Ace2, a cell cycle regulator, and Gcr2, a regulator of genes involved in glycolysis and gluconeogenesis. The binding sites of these two transcription factors, when detected in intergenic regions across the genome, were particularly likely to lie far from the start site of the downstream gene \( (p < 0.02, \text{Wilcoxon rank-sum test}) \). To evaluate the effects of these factors on antisense regulation, we identified tandem gene pairs in which the upstream gene of the pair harbored antisense expression and the downstream mRNA, but not the upstream sense mRNA, was directly regulated by Ace2 or Gcr2 according to binding-based (MacIsaac et al. 2006) and functional evidence (Figure 2.8A, Figure 2.9A, and (Hu et al. 2007)). For five such loci for each transcription factor, we tested the hypothesis that Ace2 or Gcr2 was additionally required for antisense transcription at the upstream gene. The results bore out our prediction, with upstream antisense levels at each predicted Ace2 or Gcr2 target decreasing 1.4- to 21-fold in the corresponding transcription factor deletion strain (Figure 2.8B and Figure 2.9B). The impact of these transcription factor deletions on antisense expression at predicted targets was not a consequence of a genome-scale effect, as antisense levels were largely unchanged at loci with no evidence for regulation by these factors (Figure 2.10). Thus, at a given tested tandem pair, either Ace2 or Gcr2 served as an activator of antisense expression.
at the upstream gene as well as sense expression at the downstream gene, validating our genome-scale inferences of transcription factor regulation at antisense loci.

We next asked whether a given case of antisense transcription controlled by Ace2 or Gcr2 might serve a regulatory function, by modulating the mRNA abundance of the overlapping sense gene. In such a case, we would predict that perturbations to antisense expression, via deletion of its activating transcription factor, would impact mRNA expression from the overlapping sense strand. Half the loci that we assayed conformed to this expectation: in these five instances, a drop in antisense expression was coincident with an increase in the sense transcript on the opposite strand (Figure 2.8C and Figure 2.9C). An additional and distinct method of perturbing antisense transcript abundance involves mutation of the nuclear exosome, which guides the 3’-end formation and degradation, rather than the transcriptional initiation, of many antisense transcripts (Castelnovo et al. 2013; Neil et al. 2009). Among the antisense loci in our Ace2 and Gcr2 target set, we observed three cases in which an exosome mutation conferred increased abundance of an antisense RNA and decreased abundance of the mRNA on the opposite strand (Figure 2.11). We conclude that conserved antisense targets of transcription factors include cases whose behavior is consistent with a role as a repressor of the overlapping sense mRNA, underscoring the potential for these antisense loci to act as mediators of regulation.

Antisense transcription at YKL151C has a repressive function

We next aimed to test directly the regulatory function of transcription factor-mediated antisense expression. For this case study, we chose the tandem gene pair comprising YKL151C, which encodes an NAD(P)H repair enzyme; its overlapping antisense transcript, which we named YKL151C_as; and GPM1, which encodes a glycolytic and gluconeogenic enzyme (Figure 2.12A). In the absence of the transcription factor gene GCR2, expression of GPM1 and YKL151C_as dropped to ~10% of wild-type levels, and YKL151C mRNA expression doubled (Figure 2.8). We thus formulated a model in which regulatory input to the promoter shared by GPM1 and YKL151C_as would, via a repressive function of the latter, also modulate YKL151C sense mRNA. Consistent with this model, when comparing wild-type cells grown with either glucose or glycerol as the sole carbon source, conditions known to perturb GPM1 (Roberts and Hudson 2006), we observed co-regulation of GPM1 mRNA and YKL151C_as expression, and anti-correlation of GPM1 and YKL151C mRNA expression (Figure 2.12B). Likewise, as expression of YKL151C_as increased 2.5-fold in an exosome mutant, sense expression at this gene decreased by 20% (Figure 2.11). However, regulation of sense and antisense expression at this locus was separable: mutating the Rpd3L histone deacetylase complex, which associates with the YKL151C promoter (Venters et al. 2011) and is required for YKL151C expression (Lenstra et al. 2011), affected mRNA but not antisense levels at YKL151C (Figure 2.12C). As such, a causal model in which regulation of sense expression alone determined antisense levels at YKL151C was unlikely, instead suggesting a role for the antisense element as a repressor of the overlapping sense mRNA.
To investigate the relationship between \textit{YKL151C\textsubscript{as}} and its overlapping sense gene more directly, we cloned a version of the \textit{YKL151C} locus bearing a mutation in a binding site for the Gcr2 transcription factor, which lay in the intergenic region beyond the end of the 3’ UTR of \textit{YKL151C} (Figure 2.12A). We cultured cells harboring this mutant construct and, separately, a strain carrying a cloned wild-type \textit{YKL151C} locus, and we measured expression from both strands in each case. The results revealed a 23% reduction in \textit{YKL151C\textsubscript{as}} levels in the Gcr2 binding site mutant, whose contrast with the more dramatic response to deletion of the \textit{GCR2} gene (Figure 2.8B) was suggestive of additional indirect regulatory effects of the latter. Importantly, however, consonant with results from wholesale gene deletion, the focused \textit{cis}-regulatory mutation in the \textit{YKL151C\textsubscript{as}} promoter drove an 80% increase in \textit{YKL151C} sense mRNA expression (Figure 2.12D), establishing the disrupted site as a driver of antisense and mRNA expression at \textit{YKL151C} and further supporting a model of antisense repression of the sense.

As an additional independent test of the repressive function of \textit{YKL151C\textsubscript{as}}, we used a method in which disrupting putative binding sites for the Nrd1-Nab3-Sen1 termination complex at an exosome-sensitive locus confers elevated expression (Arigo \textit{et al.} 2006; Castelnuovo \textit{et al.} 2013). Polyadenylated 3’ transcript ends at the \textit{YKL151C} locus in wild-type cells (Schraiber \textit{et al.} 2013) indicated the presence of two 3’ forms of \textit{YLK151C\textsubscript{as}}, one terminating 340bp into the sense ORF and the other extending through the promoter of the sense gene. The latter yielded a predicted long form for the antisense transcript of ~1.2kb, consistent with a previous report (Creamer \textit{et al.} 2011) (Figure 2.12A). We identified nine putative Nrd1 and Nab3 binding sites within a 241-bp region surrounding the short form polyadenylation site for \textit{YKL151C\textsubscript{as}}, and we generated a version of the \textit{YKL151C} locus in which these sites were disrupted by point mutations that left the coding sequence of the \textit{YKL151C} ORF unchanged (Figure 2.13). This construct expressed the long form of \textit{YKL151C\textsubscript{as}} at 51-fold higher levels relative to wild-type (Figure 2.12E), confirming the importance of the mutated sites in regulation of the antisense transcript, and echoing the dramatic increase seen in endogenous \textit{YKL151C\textsubscript{as}} in a strain bearing a mutation in \textit{NRD1} (Creamer \textit{et al.} 2011). Concomitantly, we also observed 7-fold lower abundance of \textit{YKL151C} sense mRNA in the Nrd1-Nab3 site mutant (Figure 2.12E). These results parallel the evidence for a role for \textit{YKL151C\textsubscript{as}} as a driver of sense gene expression that we observed when mutating the \textit{YKL151C\textsubscript{as}} promoter (Figure 2.12D). Taken together, the responses to \textit{cis-} and \textit{trans-}acting perturbations at this locus supported a model of joint regulation of \textit{YKL151C\textsubscript{as}} and \textit{GPM1} at their shared promoter, which indirectly, via the repressive function of the former, also regulates \textit{YKL151C} mRNA.
DISCUSSION

Understanding the prevalence of antisense RNA expression in genomes is a primary challenge in modern genomics. Against the backdrop of elegant molecular studies of individual antisense transcripts (Chen et al. 2012; Li et al. 2012; Magistri and Faghihi 2012; Pelechano and Steinmetz 2013; Zhao et al. 2013), how and why organisms express most antisense RNAs remains poorly understood. In this work, we have identified a set of antisense transcripts conserved across four Saccharomycetes separated by ~20 million years of evolution. We have used this catalog to dissect the role and regulatory logic of transcription factors in yeast antisense expression, including the discovery of a previously uncharacterized antisense RNA functioning as a repressor of YKL151C sense gene expression.

Our approach to identify antisense RNAs with conserved expression is complementary to analyses of noncoding RNA sequence conservation (Cabili et al. 2011; Lee et al. 2008; Moghe et al. 2013; Pauli et al. 2012; Ponjavic et al. 2007) and provides a direct readout of the maintenance of transcriptional activity at a given locus during evolution. Of the thousands of yeast gene loci expressing antisense RNAs, we observed ~10% at which the presence of antisense transcription was reliably detectable across Saccharomycetes grown in rich medium. Although this value may prove to be an underestimate if low-abundance RNAs have escaped detection in some sequencing libraries, the degree of conservation we report here is consistent with analyses of noncoding RNA expression in several taxa (Guohua Xu et al. 2010; Kutter et al. 2012; Necsulea et al. 2014), including a recent study of two closely related yeasts (Goodman et al. 2013). The emerging picture is one in which only a fraction of noncoding RNAs exhibit evidence for selective constraint between species in a given condition, consistent with the suspicion that antisense levels may be functional in at most a few hundred genes in rich medium in S. cerevisiae (van Dijk et al. 2011; Yadon et al. 2010).

Our analysis of conserved antisense transcripts in yeast focused on tandem gene pairs, in each of which an intergenic region drives mRNA expression of the downstream gene and an RNA antisense to the upstream gene (Wei et al. 2011). Such regions harbored a dual-peak signature of histone H2A.Z occupancy that was absent at single-gene promoters, representing a key departure from blanket analyses of all coding loci (Albert et al. 2007; Mavrich et al. 2008; Raisner et al. 2005) and implicating dual peaks of H2A.Z as markers of promoter bidirectionality (Rando and Chang 2009). In genomic and single-gene analyses, respectively, S. cerevisiae lacking H2A.Z and its remodeling factor Swr1 exhibited few changes in antisense expression (Table 2.3) when cultured in rich medium. Future studies of yeast response to environmental perturbation will likely uncover additional effects of H2A.Z occupancy at antisense loci, as has been observed at sense genes (Adam et al. 2001; Zhang et al. 2005).

Our analyses revealed genome-scale signatures of transcription factor binding at tandem gene pairs expressing antisense RNAs, and in single-gene experiments, we established the requirement for these factors in antisense regulation. A global role for transcription factors as drivers of antisense expression would be in keeping with known
cases of transcription factor control of individual functional long noncoding RNAs (Hongay et al. 2006; Houseley et al. 2008; Nishizawa et al. 2008; Pandey et al. 2008; Stavropoulos et al. 2005; Zhao et al. 2013). In our data, transcription factors that bind to conserved sites far from these downstream target genes were among the most likely to generate upstream antisense RNAs, plausibly via interaction with the more distal of two pre-initiation complexes assembled in the downstream gene promoter's nucleosome-free region (Rhee and Pugh 2012). In principle, any transcription factor targeting a downstream gene and its antisense product could be co-opted during evolution as an indirect regulator of the upstream sense gene.

Indeed, our discovery of \textit{YKL151C} as as a functional antisense element highlights the potential for a regulatory logic involving both up- and downstream genes of tandem pairs. The protein product of \textit{YKL151C}, the upstream gene of a tandem pair, is a dehydratase that repairs hydration damage to the redox reaction cofactors NADH and NADPH (Marbaix et al. 2011) and is induced in response to oxidative stress, as well as in many other stress conditions (Gasch et al. 2000). \textit{GPM1}, the downstream gene of the tandem pair, encodes phosphoglycerate mutase, a glycolysis and gluconeogenesis enzyme (Lam and Marmur 1977). \textit{GPM1} is repressed during utilization of non-fermentable carbon sources, in response to carbon starvation, and during stationary phase (Bradley et al. 2009a; Gasch et al. 2000), all conditions in which cells have increased requirements for the oxidative stress response (Longo et al. 1996) and induce \textit{YKL151C} (Bradley et al. 2009a; Gasch et al. 2000). The emerging picture is one in which yeast use starvation and stress as triggers for three regulatory events at this locus: repression of \textit{GPM1}, conceivably as part of a program to minimize the energetic cost of glycolytic enzyme expression (Fraenkel 1982); downregulation of the repressor \textit{YKL151C} as; and relief of repression of its target, \textit{YKL151C} mRNA, to maximize NAD(P)H repair activity. We hypothesize that antisense-mediated regulation serves as an additional layer of control of \textit{YKL151C} mRNA, over and above the activity of stress-induced transcription factors that govern transcriptional initiation at its 5' end. A benefit for this fine-tuning would be consistent with the significant fitness consequences of modest changes in \textit{YKL151C} gene dosage (Davey et al. 2012; Pir et al. 2012), and the observation that \textit{YKL151C} confers an advantage in some conditions (Davey et al. 2012; Jayakody et al. 2011; Pir et al. 2012; Samanfar et al. 2013) and is deleterious in others (Postma et al. 2009).

Notably, in each of two other well-characterized cases of functional antisense expression in yeast, \textit{PHO5} (Uhler et al. 2007) and \textit{GAL10} (Houseley et al. 2008), the gene hosting the antisense RNA is positioned upstream of another gene in the same biochemical pathway. Furthermore, among the loci at which our transcription factor mutation experiments revealed signatures of repressive function by antisense transcription (Figure 2.8C), in several cases the two genes of a tandem pair had functions in common. For example, at one locus, the upstream gene \textit{CHS7} encodes a post-translational regulator of Chs3 (Trilla et al. 1999), a chitin synthase responsible for producing the chitin ring during cell division (Shaw et al. 1991); the downstream gene, \textit{DSE2}, is a putative glucosidase involved in cytokinesis (Colman-Lerner et al. 2001). At another locus, the upstream gene \textit{PDX1} encodes a subunit of the pyruvate
dehydrogenase complex, which converts pyruvate into acetyl-CoA (Behal et al. 1989), and the downstream gene, *TDH3*, encodes glyceraldehyde-3-phosphate dehydrogenase, a key enzyme in the interconversion between glucose and pyruvate (McAlister and Holland 1985); the latter two gene products are inversely regulated in response to carbon starvation (Bradley et al. 2009a). It is thus tempting to speculate that neighboring gene pairs with related functions, although generally rare in the yeast genome (Hall and Dietrich 2007), may be prime candidates for the model of biologically relevant, joint regulation of sense mRNAs with functional noncoding RNAs that act on neighboring genes (Lapidot and Pilpel 2006).

Our results provide several clues to the mechanism by which *YKL151C_as* acts as a repressor of its overlapping sense gene. We observed that the long form of *YKL151C_as* extends through the *YKL151C* sense promoter (Figure 2.12A), consistent with a model in which antisense transcription interferes directly with polymerases in this region or sets up a repressive chromatin environment that shuts down sense expression (Pelechano and Steinmetz 2013). As would be expected under any such potential cis-acting model, we detected no evidence for trans-acting regulation by an exogenous copy of *YKL151C_as* at the endogenous sense locus (Figure 2.14). Furthermore, our experiments using two distinct cis-regulatory mutations, each of which activates or represses *YKL151C_as* to a different degree, both revealed changes in expression of the overlapping mRNA (Figure 2.12D and E). The latter findings suggest a graded relationship between antisense and sense levels at *YKL151C*, rather than the threshold function observed in analyses of trans-acting RNA regulators in yeast (Castelnuovo et al. 2013) and other systems (Levine et al. 2007). Further work will elucidate the molecular basis of *YKL151C_as* repression, and that of other candidate antisense regulators controlled by transcription factors at tandem gene pairs.

Strategies to dissect the function of antisense RNAs, and the causal relationship between sense and antisense expression, are at a premium in the modern literature. Our work establishes the utility of analysis of transcription factor regulation, beyond focused studies of known antisense biogenesis factors (Gelfand et al. 2011; Nishizawa et al. 2008), to characterize antisense regulation and function. We anticipate that transcription factor control of an antisense transcript may often serve as a signpost for its regulatory role. In systems where such a principle holds, our transcription factor-centered analysis approach will be a powerful tool in the search for functional noncoding RNAs.
MATERIALS AND METHODS

Strains and growth conditions. Yeast strains used in this study are listed in Table 2.4. Strains were grown at 30°C in YPD medium (Ausubel et al. 1995) to log phase (between 0.65-0.75 OD at 600 nm), except where indicated. Uracil dropout medium (Amberg et al. 2005) was used for experiments with strains that harbored plasmids. To measure the effect of glycerol metabolism on the YKL151C locus, cells were grown in YPG medium (1% yeast extract, 2% peptone, 2% glycerol v/v).

RNA processing and RNA-seq. For RNA-seq and qPCR experiments, total RNA was isolated by the hot acid phenol method (Ausubel et al. 1995) and treated with Turbo DNA-free (Ambion) according to the manufacturer's instructions.

To investigate the role of H2A.Z in antisense expression on a genomic scale in Table 2.3, RNA was isolated and processed as above from two replicates each of S. cerevisiae swr1Δ htz1Δ strain JRY9006 and the isogenic wild-type strain JRY9002 (Halley et al. 2010). Samples were sequenced using 50 bp paired-end modules on an Illumina HiSeq 2000 at the Vincent J. Coates Genomic Sequencing Laboratory at the University of California, Berkeley.

Transcript annotation. Sequence data from wild-type S. cerevisiae, S. paradoxus, S. mikatae, and S. bayanus were taken from (Schraiber et al. 2013). Mapping was performed as described in that study, with several modifications. Reads from each species were mapped only to that species’ genome, and read counts were generated for antisense as well as for sense transcripts. For most loci, antisense transcript features were defined as extending from 300 bp 5’ of the open reading frame (ORF) to the 3’ end of the ORF, on the strand opposite to the gene. In the case of convergently oriented genes, each gene is oriented tandemly with respect to the antisense region of its convergent partner, presenting potential ambiguity about the assignment of reads. In such cases of convergently oriented genes, we shifted the boundaries of the antisense feature to exclude the 500 bp downstream of the adjacent ORF.

Given orthology relationships for genes across yeasts from (Scannell et al. 2011), we discarded those with characteristics that would preclude interspecies comparisons of antisense expression. In particular, we eliminated from analysis genes that had 1) antisense features (as defined above) whose lengths were either shorter than 100 bp or less than half the length of the defined sense region, or 2) either sense or antisense features whose lengths differed by more than 10% between species. The final analyzed set retained 3914 genes with orthologs in all species. We considered antisense transcription to be detectable in a given species if the final normalized, averaged expression value was five or more counts. For 84% of antisense features detected in S. cerevisiae and at least one other species, the primary 3’ form in the former was also expressed in the latter (data not shown), indicating that 3’ end positions of antisense features were generally conserved across species.
Processing of 3’ end RNA-seq data from strains JRY9002 and JRY9006 was performed as above.

**Gene group enrichment tests.** In tests of functional enrichment in Figure 2.3C, genes with conserved antisense expression were defined as those with detected antisense expression in at least three species. We tested for enrichment of these genes among biological process Gene Ontology slim terms relative to the genome (downloaded from [http://downloads.yeastgenome.org/curation/literature/go_slim_mapping.tab](http://downloads.yeastgenome.org/curation/literature/go_slim_mapping.tab)) using Fisher’s exact test, excluding terms that contained fewer than five genes from our set of filtered orthologs. We performed multiple-testing correction with the method of Benjamini and Hochberg (Benjamini and Hochberg 1995). We separately used Fisher’s exact test to evaluate enrichment, relative to the genome, of genes with conserved antisense expression among genes with TATA boxes in their promoters (Basehoar et al. 2004) and genes that were components of the yeast environmental stress response (Gasch et al. 2000).

**Histone modification analysis.** Histone modification data (Pokholok et al. 2005) was downloaded from [http://web.wi.mit.edu/young/nucleosome](http://web.wi.mit.edu/young/nucleosome). For analysis of enrichment at the 3’ ends of genes in Table 2.1, histone modification levels were averaged within the last 500 bp of the gene’s transcript boundaries (Xu et al. 2009). Linear regression was then performed for each type of histone modification, with abundance of the modification regressed against sense expression, antisense expression, and antisense conservation (encoded as the number of species in which antisense expression was detected, from 0 to 4).

**Regulatory protein enrichment.** Genome-wide occupancy data for regulatory proteins in *S. cerevisiae* (Venters et al. 2011) were downloaded from [http://atlas.bx.psu.edu/cj/occ/occ_data.html](http://atlas.bx.psu.edu/cj/occ/occ_data.html). For each factor, we compiled the set of genes that exhibited binding at 25°C to the probe set located at the 3’ ends of open reading frame boundaries. Limiting our analysis to genes in a tandem orientation with respect to their downstream neighbor, in Table 2.2 we used Fisher’s exact test to assess the overlap of each set of bound genes with the set of genes with conserved antisense expression, relative to the overlap of the bound genes with the set of all tandemly-oriented genes. We used the method of Benjamini and Hochberg (Benjamini and Hochberg 1995) to perform multiple testing correction and determine the proteins that had a corrected enrichment $p < 0.01$.

**H2A.Z localization analysis.** Genome-wide H2A.Z occupancy levels in *S. cerevisiae* (Albert et al. 2007) were downloaded from [http://h2az.atlas.bx.psu.edu](http://h2az.atlas.bx.psu.edu). To analyze these occupancies with respect to antisense expression, we first collated the set of tandem gene pairs where the upstream gene harbored conserved antisense expression and, separately, the set of tandem gene pairs with no antisense expression. Considering each of these two sets in turn, for Figure 2.5C we aligned all gene pairs of the set at the transcription termination site of the upstream gene from (Xu et al. 2009) to generate a composite. For each composite, H2A.Z occupancy was averaged within 50-bp moving windows across the region from 500 bp upstream to 500 bp downstream of
the alignment position. Each window average was normalized by the number of genes used in the composite. This value was then further normalized by the maximum across all positions in the composite in order to yield percentages of H2A.Z occupancy across the composite locus. To analyze H2A.Z occupancies with respect to divergent (head-to-head) gene pairs in Figure 2.6, we used data from (Xu et al. 2009) to collate the set of divergent gene pairs whose transcription initiated from shared nucleosome-free regions and, separately, the set of divergent gene pairs whose transcription initiated from separate nucleosome-free regions. We aligned the gene pairs of each set at the transcription start sites of the downstream genes to generate a composite, and analyzed and normalized H2A.Z occupancies in moving windows as above.

**Transcription factor binding site analysis.** Inferred transcription factor binding sites (Maclsaac et al. 2006) were downloaded from [http://fraenkel.mit.edu/improved_map](http://fraenkel.mit.edu/improved_map). Sites were included for analysis if they were bound at \( p < 0.001 \) according to (Maclsaac et al. 2006) and were not filtered for conservation. For Figure 2.7, we collated the set of tandem gene pairs harboring conserved antisense expression and aligned all pairs of the set according to the transcription start sites of the downstream genes from (Xu et al. 2009); we tallied transcription factor binding sites across this composite locus, averaging the number of binding sites within a 50-bp moving window. Transcription factor binding site frequencies were then normalized to the number of gene pairs. Separately, we then collated the set of tandem gene pairs with no detectable antisense expression and repeated the alignment and binding-site average analysis.

**Exosome mutant analysis.** To analyze expression in response to exosome mutation in Figure 2.11, transcriptional profiling data were downloaded from (Xu et al. 2009). We tabulated the expression of each sense feature in each sample as the average hybridization across sense-strand probes within the body of the open reading frame. We defined the end of each antisense feature as the median of the 3' end positions of *S. cerevisiae* antisense-strand reads from (Schraiber et al. 2013) at that locus; we then tabulated the expression of each antisense feature in each sample as the average hybridization across antisense-strand probes between the antisense feature end and the 3' end of the sense open reading frame.

**Strand-specific qPCR.** Strand-specific quantitative RT-PCR was performed using a protocol adapted from (Bessaud et al. 2008). cDNA synthesis was performed using a transcript- and strand-specific primer that was tagged with an 18 bp exogenous sequence. This tagged sequence was used as one of the primers in the subsequent quantitative PCR reaction to ensure the specificity of the amplification. Primers used to amplify either strand of *S. cerevisiae YKL151C* targeted the region between positions 165211 and 165288, with the exception of Figure 2.12E, where primers targeted the region between positions 165569 and 165726. As an internal control, the sense strand of *ACT1* was also amplified in every reaction with a second exogenous tag, with the exception of experiments comparing expression between cells grown in YPD and YPG medium in Figure 2.12B; the latter experiments instead used the sense strand of *SCR1* as a control, which did not change between conditions (data not shown). Primers used in this study are listed in Table 2.5.
Reverse transcription reactions were performed as follows: 1 µl of 200 pM transcript-specific tagged primer, 1 µl of 200 pM tagged ACT1 or SCR1 primer, 1 µl of 10 mM dNTP, and 2 µg of RNA in DEPC-treated water were mixed in a 12 µL total volume and incubated at 65°C for 5 minutes and 4°C for 1 minute. To each reaction was added 1 µl of 0.125 µg/µl actinomycin D (Perocchi et al. 2007), 1 µl of 0.1 M DTT, 4 µl of 5x First Strand Synthesis Buffer (Invitrogen), 1 µl of RNaseOUT (Invitrogen), and 1 µl of SuperScript III (Invitrogen). Reactions were incubated at 55°C for 50 minutes and 70°C for 15 minutes, then cooled to 4°C. 1 µl RNaseH (Invitrogen) was added and the reaction was incubated at 37°C for 20 minutes and stored at 4°C. Reactions were purified using the QIAquick PCR Purification Kit (Qiagen) and eluted in 50 µl EB buffer.

qPCR reactions were set up as follows: 3.67 µl purified cDNA, 1 µl of 10 µM tag primer, 1 µl of 10 µM transcript-specific primer, 0.15 µl ROX dye, and 12.5 µl of 2x DyNAmo HS SYBR Green qPCR master mix (Finnzymes) were mixed in 25 µl total volume. Amplification reactions were performed in triplicate on an Mx3000p Stratagene qPCR machine.

The efficiency of every primer pair was evaluated using a standard curve of cDNA generated by qPCR as described above, and only primer pairs with efficiencies inside the range of 85%-115%, and those yielding a single peak in the dissociation curve, were used for analyses. For each qPCR experiment, final target quantities were calculated by normalizing to levels of an internal control reference gene (SCR1 in Figure 2.12B and ACT1 in all other experiments) using the method of (Pfaffl 2001) after averaging Ct’s across the three technical replicates. When replicates for a single experiment were grown on different days, quantities for each set of replicates were normalized by dividing each quantity by the ratio of the mean reference-normalized quantity of the respective set of replicates to the mean reference-normalized quantity of all the replicates for that experiment.

**Cloning and mutagenesis of the YKL151C locus.** We cloned YKL151C and its flanking intergenic regions from S. cerevisiae strain BY4724 (Baker Brachmann et al. 1998) into a plasmid using sequence- and ligation-independent cloning (Li and Elledge 2007). The locus was amplified by PCR (see Table 2.5 for primers) and cloned into the Smal site of plasmid pRS316 (Sikorski and Hieter 1989), yielding plasmid pYM007. This wild-type construct and all mutant versions (see below) harbored 8 TA repeats of a TA tract located 38 nt downstream of the YKL151C open reading frame which, in the reference genome, contained 17 repeats (yeastgenome.org).

To identify and mutate the Gcr2 binding site downstream of YKL151C in Figure 2.12D, we scanned the 3' intergenic region using a position weight matrix for Gcr1 (Gordan et al. 2011), Gcr2's DNA-binding partner, and identified the highest-scoring sequence AAGAGGAAGCTC, located 170 bp from the 3' end of the YKL151C open reading frame. We mutated this sequence in pYM007 to AAGAGTGAGATC (see Table 2.5 for primer sequences) using the Quickchange Site-Directed Mutagenesis kit (Agilent). Mutated and wild-type plasmids were each separately transformed into a yeast strain bearing a

To disrupt Nrd1/Nab3 motifs in the 3’ end of *YKL151C_as* in Figure 2.12E, we synthesized six 60-mer oligos that overlapped each other by 20nt (Elim Biopharmaceuticals, Inc., Hayward, CA) (see Table 2.5 for primer sequences); these oligos encoded mutations in nine Nrd1/Nab3 motif-matching sites (Carroll *et al.* 2004) (Figure 2.13) located on the antisense strand. The oligos were assembled in equimolar ratios using Gibson Assembly Master Mix (New England Biolabs, Inc.) according to the manufacturer’s instructions. The product was PCR amplified using the two outer oligos as primers. Separately, the rest of the gene and its flanking intergenic regions were PCR amplified using primers YKL151C_cloning_fwd and exomut-r1 for the 5’ end and exomut-f2 and YKL151C_cloning_rev for the 3’ end. These three products were purified using the QIAquick PCR Purification Kit (Qiagen) and assembled together with SmaI-digested pRS316 in equimolar ratios in a new Gibson reaction. The resultant plasmid was transformed into a yeast strain bearing a deletion of the endogenous copy of *YKL151C* (Winzeler *et al.* 1999) (Invitrogen).

To test whether *YKL151C_as* functioned in trans in Figure 2.14, we mutated positions 654-663 of the *YKL151C* ORF, corresponding to the 3’ end of the sense qPCR primer binding site, from AGGTCAGTCA to GGGACAAAGT (Figure 2.14A). Substitutions were selected to be synonymous with respect to the gene’s amino acid sequence. We introduced these mutations into pYM007 using Quikchange mutagenesis, yielding plasmid pYM017. To verify that the mutations allowed for reliable differentiation of mutant and original sequences using qPCR (Figure 2.14B), we subjected equal total amounts of wild-type plasmid, mutant plasmid, or a 1:1 mixture of the two to pseudo-reverse transcription by carrying out 10 rounds of linear amplification using *YKL151C* sense RT primer. Samples were then processed for qPCR as described above, using either mutant or original sense qPCR primers. To detect mutant antisense RNA in cells, reverse transcription was performed using the mutant antisense RT primer, followed by qPCR amplification with the original antisense primer. Plasmid pYM017 and, separately, the empty plasmid pRS316 were transformed into yeast strain BY4724.
**FIGURES**

(A) This study  SUTs

960 252 71

(B) This study  CUTs

991 221 197

(C) This study  Goodman et al.

41 455 1829

**Figure 2.1 Antisense transcripts detected in this study overlap with results of previous surveys.** (A-B) Each panel reports a comparison between the antisense transcripts detected in wild-type *S. cerevisiae* in this study and those identified in wild-type (stable unannotated transcripts, SUTs) or exosome mutant (cryptic unstable transcripts, CUTs) *S. cerevisiae* strains by (Xu et al. 2009). In a given panel, numbers outside and inside intersecting regions report the number of genes with antisense transcripts detected only in the indicated analysis or in both analyses, respectively. (C) Numbers outside and inside intersecting regions report genes with antisense transcripts detected in *S. cerevisiae* and *S. paradoxus* in only this study or an expression profiling data set that includes non-polyadenylated transcripts (Goodman et al. 2013), or in both analyses, respectively. For each previously published data set, only genes interrogated in our study were considered for analysis. In all comparisons, $p < 2e-16$, Fisher’s exact test.
Figure 2.2 A survey of antisense expression in *Saccharomyces* yeasts. (A) The x-axis reports the normalized number of RNA-seq reads that mapped to the strand opposite a given open reading frame in *S. cerevisiae*; the y-axis reports the number of genes with the antisense read count value on the x. Genes with zero antisense reads were excluded from analysis. (B) Text in each circle reports the number and proportion of genes with antisense expression detected in at least the number of species indicated, from among *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. bayanus*. (C) The x-axis reports the number of species, among *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. bayanus*, in which a given antisense transcript was detected. The y-axis reports the distribution of mRNA expression of the host genes across the antisense transcripts whose conservation is indicated on the x. For each distribution, the median is reported.
as a thick horizontal line, the box encompasses the upper and lower quartiles, and the thin horizontal bars denote the inter-quartile range (upper quartile – lower quartile) multiplied by 1.5.
Figure 2.3 Conserved antisense transcripts are highly expressed and enriched in functional categories and at tandem genes. (A) For antisense transcripts detected in the number of species indicated on the x-axis, expression levels for each transcript were averaged across detected species. For each distribution, the median is reported as a thick horizontal line, the box encompasses the upper and lower quartiles, and the thin horizontal bars denote the inter-quartile range (upper quartile – lower quartile) multiplied by 1.5. (B) Each bar represents the proportion of genes falling in tandem gene pairs (see Figure 2.5A for a schematic), for a set of genes whose antisense expression was conserved to a given extent across Saccharomyces. The x-axis reports the number of species, among *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. bayanus*, in which a given antisense transcript was detected. The genome-wide proportion of genes falling in tandem pairs is indicated with a dashed line. Asterisks denote a difference from the genome-wide average that is significant at $p < 0.01$ by Fisher’s exact test. (C) Each row reports the nominal $p$-value results of a test for enrichment of conserved antisense expression in a group of genes of common function. The first row reports the only biological process Gene Ontology slim category (*Ashburner et al. 2000*) significantly enriched for genes hosting conserved antisense expression in a Fisher’s exact test at $p < 0.05$, after correction for multiple testing using
the method of Benjamini-Hochberg (Benjamini and Hochberg 1995). The second and third rows report enrichment of genes hosting conserved antisense expression, in a Fisher’s exact test, among TATA-regulated genes (Basehoar et al. 2004) and environmental stress response genes (Gasch et al. 2000).
Figure 2.4 Functional noncoding RNAs are conserved across species. Each cartoon and set of plots reports expression across Saccharomyces of one noncoding RNA with previously characterized regulatory function. Each cartoon at top represents a genomic locus, with black bars representing aligned coordinates for each species, and the host gene of the functional RNA in magenta. Open reading frames shown above the coordinates are encoded on the same strand as the host gene, and those below the coordinates are on the opposite strand. Each plot at bottom reports 3' RNA-seq reads from one species from (Schraiber et al. 2013). The x-axis reports the genome position.
of mapped reads on the scale of the coordinates for the respective species in the top cartoon; the y-axis reports the number of reads mapping to each position. Each grey bar represents one mapped transcript end read, with blue and red bars representing poly-adenyl tails of reads mapping to the same strand as the host gene of interest and reads mapping to the opposite strand, respectively. The region of relevant noncoding RNA expression is boxed in red. (A) Host gene *PHO84* (Camblong *et al.* 2007); antisense expression is conserved in all four species and extends through the neighboring gene *YML122C*. (B) Host gene *GAL10* (Houseley *et al.* 2008); antisense expression is conserved in all four species. (C) Host gene *ZIP2* (Gelfand *et al.* 2011); antisense expression extending part-way through the gene is conserved in all four species. (D) Host gene *PHO5* (Uhler *et al.* 2007); antisense expression is detected in each species except *S. mikatae*. 
Figure 2.5 Dual peaks of occupancy of the histone variant H2A.Z at antisense promoters. (A) Schematic of the architecture of a tandem gene pair with antisense transcription. The blue, cyan, and red arrows denote upstream sense transcription, downstream sense transcription, and antisense transcription, respectively. (B) Each histogram represents the start sites for one class of RNA at tandem gene pairs with conserved antisense expression. The x-axis, shown in (C), reports the start site position of a given RNA (Xu et al. 2009), and the y-axis reports the density of RNAs with a start site position as indicated on the x. Transcript identifiers refer to the architecture in (A). (C) Each trace represents a composite of the occupancy of histone H2A.Z on the DNA between pairs of tandem genes in S. cerevisiae, with each color representing a set of tandem pairs with distinct characteristics of antisense expression. The x-axis reports position on the DNA relative to the upstream mRNA’s transcript end site, for a given tandem gene pair; the y-axis reports H2A.Z occupancy (Albert et al. 2007) averaged within a 50-bp moving window centered on the indicated position on the x, normalized to the maximum observed value. Conserved antisense, tandem gene pairs with antisense expression detected in the upstream gene in any three of S. cerevisiae, S. paradoxus,
S. mikatae, and S. bayanus; no antisense, tandem gene pairs without detectable antisense expression in the upstream gene in any of the four species.
Figure 2.6 Head-to-head sense gene pairs have two peaks of occupancy of H2A.Z in their intergenic regions. (A) Schematic of the architecture of a divergent gene pair. Blue and cyan arrows denote upstream and downstream sense transcription, respectively. (B) Each trace reports a composite of the occupancy of histone H2A.Z on the DNA of regions between pairs of divergent genes in *S. cerevisiae*. Each color represents a set of divergent pairs with distinct characteristics of the intergenic region as delineated by (Xu et al. 2009): for a given pair, either the two genes share a nucleosome-free region (NFR), or each gene’s promoter contains its own NFR. The x-axis reports position on the DNA relative to the downstream mRNA’s transcript start site for a given divergent gene pair; the y-axis reports H2A.Z occupancy (Albert et al. 2007), averaged within a 50-bp moving window centered on the indicated position on the x and normalized to the maximum observed value.
Figure 2.7 Dense transcription factor motifs at tandem gene antisense promoters are distal to the downstream gene. (A) Schematic of the architecture of a tandem gene pair with antisense transcription. Yellow boxes represent gene transcript boundaries. The blue, cyan, and red arrows denote upstream sense transcription, downstream sense transcription, and antisense transcription, respectively. Positions of the end of upstream sense transcription and the start of antisense transcription, with respect to the transcriptional start site of the downstream gene, reflect median values across features detected in *S. cerevisiae* on the scale of the x-axis of (B) (this study and (Xu *et al.* 2009)). (B) Each trace reports the positions of transcription factor binding sites in *S. cerevisiae* in the intergenic regions of tandem gene pairs. Each color represents the results of analyses of a set of tandem pairs with distinct characteristics of antisense expression. The x-axis reports position on the DNA, relative to the downstream mRNA transcript start site of a given tandem gene pair; the y-axis reports the number of inferred transcription factor binding sites across all loci of the indicated set, averaged within a 50-bp moving window centered on the position on the x and normalized by the total number of gene pairs in the set. Conserved antisense, tandem gene pairs with antisense expression detected in the upstream gene among any three of *S. cerevisiae*,

31
S. paradoxus, S. mikatae, and S. bayanus; no antisense, tandem gene pairs without detectable antisense expression in the upstream gene in any of the four species.
Figure 2.8 The transcription factors Gcr2 and Ace2 are required for antisense expression at predicted targets. In the cartoons at left, yellow rectangles represent tandem genes, and green ovals represent either Ace2 or Gcr2 binding in the intergenic region between the genes. Each row reports expression, measured by strand-specific quantitative RT-PCR in *S. cerevisiae*, of the class of transcripts schematized in the cartoon at left, in tandem gene pairs with evidence for Gcr2 (left data panels) or Ace2 (right data panels) binding and function. In a given row, the left-hand data panel reports expression changes between a wild-type (WT) and GCR2 mutant (gcr2△) strain, and the right-hand data panel reports changes between a wild-type and ACE2 mutant (ace2△) strain. In a given data panel, each set of bars reports measurements of the indicated transcript and each bar color reports measurements from the indicated strain. In x-axis labels, each integer identifies one tandem gene pair, e.g. locus 1 represents the tandem pair at which YKL151C is the upstream gene and GPM1 is the downstream gene (see Figure 2.12A). (A) mRNA levels of downstream genes in tandem pairs. (B) Antisense RNA levels of upstream genes in tandem pairs. (C) mRNA levels of upstream genes in tandem pairs. For each transcript, all abundance measurements were normalized against the mean expression measurement in the wild-type strain for presentation. Error bars represent propagated standard error from 4-6 biological replicates each of wild-type and mutant strains. *p < 0.05, Student’s t-test.
Figure 2.9 The transcription factors Gcr2 and Ace2 are required for antisense expression at predicted targets. Data are those of Figure 2.8, reported without normalization to wild-type expression for each transcript. Labels and symbols are as in Figure 2.8 except that raw expression measurements are listed below each set of bars.
Antisense transcripts not targeted by Ace2 and Gcr2 are largely unaffected by their deletion. Each panel reports expression, measured by strand-specific quantitative PCR, of antisense transcripts in tandem gene pairs (see top and Figure 2.5A for a schematic) in one comparison of a pair of S. cerevisiae strains. In a given panel, each set of bars reports expression of one antisense transcript with no evidence for Gcr2 or Ace2 binding or function (Hu et al. 2007; MacIsaac et al. 2006), with each color reporting measurements from one strain background. (A) Comparison of wild-type (WT) to a GCR2 mutant (gcr2Δ). (B) Comparison of wild-type to an ACE2 mutant (ace2Δ). Error bars represent standard error from four biological replicates each of wild-type and mutant strains. * p < 0.05, Student’s t-test.
Figure 2.11 Nuclear exosome mutation perturbs antisense expression and, at some loci, expression of overlapping sense mRNA. Each panel reports expression in *S. cerevisiae* from one strand of the upstream genes of tandem gene pairs from Figure 2.8. In a given set of bars, each shade reports expression at the indicated locus in one strain background, either wild-type (WT) or an strain harboring a deletion of the exosome component *RRP6* (*rrp6Δ*). (A) Sense transcripts. (B) Antisense transcripts. Symbols are as in Figure 2.8 except that expression measurements were taken from (Xu et al. 2009). * p < 0.05, Student’s t-test.
Figure 2.12 YKL151C mRNA is regulated by antisense expression. (A) Yellow rectangles represent open reading frames, and blue, cyan, and red arrows represent upstream sense, downstream sense, and antisense transcripts, respectively. Transcript end positions were taken from (Schraiber et al. 2013) and start sites from (Xu et al. 2009). The location of a Gcr2 binding motif (Gordan et al. 2011) is also indicated. Vertical red lines represent Nab3/Nrd1 motifs near the end of the short form of the antisense. Horizontal black lines labeled P1, P2, and P3 represent qPCR amplicon locations. (B) Each set of bars reports expression from one strand of endogenous YKL151C or GPM1 in wild-type S. cerevisiae by quantitative RT-PCR, using primers targeting the regions labeled P2 and P3, respectively, in (A); each shade reports measurements in one growth medium containing the indicated sugar as the sole carbon source. (C) Each set of bars reports expression from one strand, measured by quantitative RT-PCR using
primers targeting the region labeled P2 in (A), of plasmid-borne YKL151C in S.
cerevisiae harboring a deletion of the endogenous locus; each shade reports results
from a version of the YKL151C locus with either a wild-type antisense promoter (WT) or
one bearing a mutation in the Gcr2 binding motif at the position shown in (A). (E) Each set of bars reports expression from one strand, measured by quantitative RT-PCR using primers targeting the region labeled P1 in (A), of plasmid-borne YKL151C in S.
cerevisiae harboring a deletion of the endogenous locus; each shade reports results
from a version of the YKL151C locus with either a wild-type coding region (WT) or one bearing silent mutations in nine Nab3/Nrd1 motifs at the positions shown in (A). * p < 0.05, Student’s t-test. Error bars represent standard error from at least two biological replicates.
Figure 2.13 Positions of mutated Nrd1/Nab3 sites at the 3’ end of *YKL151C*. At top, the yellow rectangle represents the *S. cerevisiae* *YKL151C* open reading frame, and blue and red arrows represent sense and antisense transcripts, respectively. Transcript end positions were taken from (Schraiber et al. 2013) and start sites from (Xu et al. 2009). Vertical red lines represent the positions of engineered Nrd1/Nab3 mutations. Dotted lines delineate the region whose sequence is shown below. At bottom, for each pair of sequences, the top line depicts the wild-type *S. cerevisiae* sequence and the bottom line shows the engineered mutations, which are demarcated by asterisks. Nucleotides in red represent predicted Nrd1/Nab3 binding sites on the antisense strand (Carroll et al. 2004).
Figure 2.14 *YKL151C* as does not have a detectable trans-acting effect on sense *YKL151C* expression. (A) Schematic showing engineered mutations to distinguish plasmid-encoded *YKL151C* from the endogenous locus. The yellow rectangle represents the *YKL151C* open reading frame, and blue and red arrows represent sense and antisense transcripts, respectively. Black arrows depict the location of primers used for RT-PCR. Below, the sequence of engineered mutations in the forward primer, with the top line representing the original sequence and the bottom line representing the mutated sequence. Mutations are marked with asterisks, and nucleotides within the primer sequence are in blue. (B) Engineered mutations distinguish plasmid-borne from endogenous *YKL151C*. Each bar reports the concentration, measured by quantitative PCR in *S. cerevisiae*, of an amplicon from the DNA of a plasmid bearing the *YKL151C* locus. Each set of bars reports results from amplification of one input DNA sample: original DNA sequence, a plasmid bearing wild-type *YKL151C*; mutant DNA sequence, a plasmid bearing the silent mutation in the *YKL151C* coding region in (A); equal mixture, a 50:50 mixture of the wild-type and mutated plasmid DNA samples. Each shade reports results from one PCR primer set: original primers, primers complementary to the wild-type *YKL151C*; mutant primers, primers complementary to the mutant version. For presentation, each measurement was normalized to the
abundance of the wild-type DNA amplified with original primers. Error bars represent propagated standard error from three different concentrations of input DNA. (C) Each panel reports expression, measured by strand-specific quantitative RT-PCR, of either the endogenous copy of YKL151C or the plasmid-borne copy from the construct in (A). In a given panel, each shade reports expression measured in S. cerevisiae cells bearing one plasmid: either a plasmid harboring the construct in (A) or an empty plasmid. Error bars represent standard error from four biological replicates of each strain. * p < 0.05, Student’s t-test.
### Table 2.1 Relationships between 3' end histone modifications, RNA expression, and antisense conservation.

Each column represents the outcome of a linear regression across genes relating three variables: the dependent variable was the DNA occupancy of the indicated histone modification in the last 500bp of a given gene as measured in *S. cerevisiae* (Pokholok et al. 2005), and the independent variables were attributes of sense and antisense expression at the gene locus as indicated. Antisense conservation, a categorical variable scoring the detection of antisense expression in three or more species. Asterisks or periods indicate the significance of the designated variable in the multiple regression as determined by partial t-test; colors indicate the direction of correlation between the histone modification and the predictor. Blank cells indicate $p > 0.1$; $.$, $p < 0.1$, $*$, $p < 0.05$; **, $p < 0.01$, *** $p < 0.001$. 

<table>
<thead>
<tr>
<th>Predictor</th>
<th>H3K9ac</th>
<th>H3K14ac</th>
<th>H4ac</th>
<th>H3K4me1</th>
<th>H3K4me2</th>
<th>H3K4me3</th>
<th>H3K36me3</th>
<th>H3K79me3</th>
</tr>
</thead>
<tbody>
<tr>
<td>sense expression</td>
<td>***</td>
<td>***</td>
<td>**</td>
<td>*</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>antisense expression</td>
<td>.</td>
<td>.</td>
<td>*</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>antisense conservation</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

Legend:
- **Positive correlation**
- **Negative correlation**
- **No correlation** ($p > 0.1$)
<table>
<thead>
<tr>
<th>Protein</th>
<th>Complex/Description</th>
<th>Enrichment p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bdf1</td>
<td>TFIID, SWR1</td>
<td>1.03E-24</td>
</tr>
<tr>
<td>Rad3</td>
<td>TFIIH</td>
<td>3.31E-13</td>
</tr>
<tr>
<td>Ssl2</td>
<td>TFIIH</td>
<td>5.90E-11</td>
</tr>
<tr>
<td>Spt16</td>
<td>FACT</td>
<td>5.53E-10</td>
</tr>
<tr>
<td>Bdf2</td>
<td>TFIID</td>
<td>6.94E-10</td>
</tr>
<tr>
<td>Tfg1</td>
<td>TFIIF</td>
<td>3.31E-09</td>
</tr>
<tr>
<td>Htz1</td>
<td>H2AZ</td>
<td>7.85E-09</td>
</tr>
<tr>
<td>Ssl1</td>
<td>TFIIH</td>
<td>1.46E-08</td>
</tr>
<tr>
<td>Tfb1</td>
<td>TFIIH</td>
<td>2.47E-08</td>
</tr>
<tr>
<td>Rsc9</td>
<td>RSC</td>
<td>3.07E-08</td>
</tr>
</tbody>
</table>

**Table 2.2 Occupancy of DNA-binding proteins is enriched at loci with conserved antisense expression.** Each row reports enrichment of experimentally measured binding, in *S. cerevisiae*, of the indicated transcriptional regulator at the 3’ ends of open reading frames of the upstream genes of tandem pairs (see Figure 2.5A for a schematic) with conserved antisense expression. *p*-value, result of Fisher’s exact test comparing protein binding at genes in tandem pairs with and without conserved antisense expression. Nominal *p*-values are shown; tests for all factors shown are significant at *p* < 0.01 after Benjamini-Hochberg multiple testing correction (Benjamini and Hochberg 1995). Binding data were taken from (Venters *et al.* 2011).
### RNA-seq

<table>
<thead>
<tr>
<th>Gene</th>
<th>strand</th>
<th>Fold-change (swr1Δ htz1Δ / wild-type)</th>
<th>p-value</th>
<th>p-value adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>YDR334W</td>
<td>sense</td>
<td>0.025092335</td>
<td>2.18E-08</td>
<td>8.49E-05</td>
</tr>
<tr>
<td>YFL012W</td>
<td>antisense</td>
<td>0.022717764</td>
<td>2.21E-08</td>
<td>8.49E-05</td>
</tr>
<tr>
<td>YFL012W</td>
<td>sense</td>
<td>0.008161697</td>
<td>8.50E-08</td>
<td>0.000218</td>
</tr>
<tr>
<td>YDR318W</td>
<td>sense</td>
<td>0.082880972</td>
<td>1.37E-05</td>
<td>0.026378</td>
</tr>
<tr>
<td>YHR189W</td>
<td>sense</td>
<td>50.98836892</td>
<td>4.44E-05</td>
<td>0.068303</td>
</tr>
<tr>
<td>YMR025W</td>
<td>sense</td>
<td>0.104111648</td>
<td>0.000127654</td>
<td>0.163674</td>
</tr>
<tr>
<td>YNL154C</td>
<td>sense</td>
<td>7.834410194</td>
<td>0.000192226</td>
<td>0.199813</td>
</tr>
<tr>
<td>YBR058C</td>
<td>sense</td>
<td>0.041310921</td>
<td>0.000207786</td>
<td>0.199813</td>
</tr>
</tbody>
</table>

### qPCR

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>swr1Δ</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MET13</td>
<td>1.07E-03</td>
<td>7.18E-04</td>
<td>0.01454</td>
</tr>
<tr>
<td>YNR014W</td>
<td>5.29E-04</td>
<td>2.72E-04</td>
<td>0.018793</td>
</tr>
<tr>
<td>MET2</td>
<td>3.71E-04</td>
<td>2.69E-04</td>
<td>0.011679</td>
</tr>
<tr>
<td>RIM4</td>
<td>4.81E-04</td>
<td>2.20E-04</td>
<td>0.000265</td>
</tr>
<tr>
<td>CYB2</td>
<td>3.77E-04</td>
<td>2.38E-04</td>
<td>0.005659</td>
</tr>
</tbody>
</table>

**Table 2.3 Transcriptional effects upon deletion of HTZ1, the gene encoding H2A.Z, and SWR1.** RNA-seq: Each row reports the expression response, measured by RNA-seq, of the indicated transcript to the loss of HTZ1 and SWR1. Transcripts with an adjusted p-value below 0.2 are shown. Fold-change, the average expression across replicates of a swr1Δ htz1Δ strain of *S. cerevisiae* divided by the average expression across replicates of a wild-type *S. cerevisiae* strain; p-value, significance of the difference in expression evaluated by the negative binomial method (Anders and Huber 2010); p-value adjusted, p-values corrected for multiple testing with the method of Benjamini-Hochberg (Benjamini and Hochberg 1995). qPCR: Each row reports the average expression, measured by strand-specific quantitative RT-PCR in *S. cerevisiae* in four biological replicates, of one antisense transcript with a peak of H2A.Z at its 5’ end according to genome-wide localization measurements (Albert et al. 2007). WT, wild-type; swr1Δ, strain bearing a deletion of the SWR1 gene. p-value, significance of the difference in expression evaluated by Student’s t-test.
<table>
<thead>
<tr>
<th>Strain ID</th>
<th>name</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>JRY9002</td>
<td>WT</td>
<td>MATa HTZ1-3FLAG::KanMX GAL1-GFP::HIS3MX</td>
<td>Halley et al. 2010</td>
</tr>
<tr>
<td>JRY9005</td>
<td>swr1Δ</td>
<td>MATalpha swr1Δ::SpHIS5MX HTZ1-3FLAG::KanMX GAL1-GFP::HIS3MX</td>
<td>Halley et al. 2010</td>
</tr>
<tr>
<td>JRY9006</td>
<td>swr1Δ htz1Δ</td>
<td>MATa swr1Δ::SpHIS5MX htz1Δ::KanMX GAL1-GFP::HIS3MX</td>
<td>Halley et al. 2010</td>
</tr>
</tbody>
</table>

**W303 background**

<table>
<thead>
<tr>
<th>Strain ID</th>
<th>name</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BY4716</td>
<td>WT</td>
<td>MATalpha lys2Δ0</td>
<td>Baker Brachmann et al. 1998</td>
</tr>
<tr>
<td>BY4724</td>
<td>WT</td>
<td>MATa lys2Δ0 ura3Δ0</td>
<td>Baker Brachmann et al. 1998</td>
</tr>
<tr>
<td>14088</td>
<td>ace2Δ</td>
<td>MATalpha leu2Δ0 lys2Δ0 ura3Δ0 ace2Δ::KanMX</td>
<td>Invitrogen; Winzeler et al. 1999</td>
</tr>
<tr>
<td>12013</td>
<td>gcr2Δ</td>
<td>MATalpha leu2Δ0 lys2Δ0 ura3Δ0 gcr2Δ::KanMX</td>
<td>Invitrogen; Winzeler et al. 1999</td>
</tr>
<tr>
<td>17234</td>
<td>pho23Δ</td>
<td>MATalpha leu2Δ0 lys2Δ0 ura3Δ0 pho23Δ::KanMX</td>
<td>Invitrogen; Winzeler et al. 1999</td>
</tr>
<tr>
<td>15001</td>
<td>ykl151cΔ</td>
<td>MATalpha leu2Δ0 lys2Δ0 ura3Δ0 ykl151cΔ::KanMX</td>
<td>Invitrogen; Winzeler et al. 1999</td>
</tr>
</tbody>
</table>

**S288C background**

<table>
<thead>
<tr>
<th>Strain ID</th>
<th>name</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>JRY9002</td>
<td>WT</td>
<td>MATa HTZ1-3FLAG::KanMX GAL1-GFP::HIS3MX</td>
<td>Halley et al. 2010</td>
</tr>
<tr>
<td>JRY9005</td>
<td>swr1Δ</td>
<td>MATalpha swr1Δ::SpHIS5MX HTZ1-3FLAG::KanMX GAL1-GFP::HIS3MX</td>
<td>Halley et al. 2010</td>
</tr>
<tr>
<td>JRY9006</td>
<td>swr1Δ htz1Δ</td>
<td>MATa swr1Δ::SpHIS5MX htz1Δ::KanMX GAL1-GFP::HIS3MX</td>
<td>Halley et al. 2010</td>
</tr>
</tbody>
</table>

Table 2.4 Strains used in this study.
<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT1 RT sense</td>
<td>TGACCTCGCTTCTCGCTAGACAAACGCTTGGATGG</td>
</tr>
<tr>
<td>ACT1 sense qPCR</td>
<td>TTGAGAGTGGCCAGAAG</td>
</tr>
<tr>
<td>SCR1 RT sense</td>
<td>TGACCTCGCTTCTCGCTAGCTGCCCAGAAGATCG</td>
</tr>
<tr>
<td>SCR1 sense qPCR</td>
<td>TTGTGGGACCCGTCCTTTTCT</td>
</tr>
<tr>
<td>ACT1 or SCR1 qPCR primer #2 (tag)</td>
<td>TGACCTCGCTTCTGCTA</td>
</tr>
<tr>
<td>qPCR primer #2 (tag) for targets other than ACT1 or SCR1</td>
<td>CAGTGAAGGTCGCCCTA</td>
</tr>
<tr>
<td>KSS1 sense RT</td>
<td>CAGTGAAGGTCGCAATCCAAAGGGCTAAAGAG</td>
</tr>
<tr>
<td>KSS1 sense qPCR</td>
<td>TTGCAGAGGTGAAGGATG</td>
</tr>
<tr>
<td>KSS1 antisense RT</td>
<td>CAGTGAAGGTCGCTACATTGGCAAGAGGTGCT</td>
</tr>
<tr>
<td>KSS1 antisense qPCR</td>
<td>ATCCCAAGGGCTAAAGAG</td>
</tr>
<tr>
<td>CHS7 sense RT</td>
<td>CAGTGAAGGTCGCCCTACAGTGCTGATCCATGCTT</td>
</tr>
<tr>
<td>CHS7 sense qPCR</td>
<td>CATGCAGTGGTGTTCCG</td>
</tr>
<tr>
<td>CHS7 antisense RT</td>
<td>CAGTGAAGGTCGCCCTACATTGGCAAGAGGTGCT</td>
</tr>
<tr>
<td>CHS7 antisense qPCR</td>
<td>GGTCAGGTCGCCATGCT</td>
</tr>
<tr>
<td>YLR287C sense RT</td>
<td>CAGTGAAGGTCGCCCTACATTCACCATGTCTGATCAATCCAT</td>
</tr>
<tr>
<td>YLR287C sense qPCR</td>
<td>ACAATTTGAGCCGAAGAAG</td>
</tr>
<tr>
<td>YLR287C antisense RT</td>
<td>CAGTGAAGGTCGCCCTACATTGGCAAGAGGTGCT</td>
</tr>
<tr>
<td>YLR287C antisense qPCR</td>
<td>TCCCATGTGTCAATCCAT</td>
</tr>
<tr>
<td>CWH41 sense RT</td>
<td>CAGTGAAGGTCGCCCTACATGGTTGAGTTTTTGAGGAT</td>
</tr>
<tr>
<td>CWH41 sense qPCR</td>
<td>ATGAGGTTACGCTCCCG</td>
</tr>
<tr>
<td>CWH41 antisense RT</td>
<td>CAGTGAAGGTCGCCCTACATTGGAGGTGCTACGTCCCG</td>
</tr>
<tr>
<td>CWH41 antisense qPCR</td>
<td>TGAGGTGATTTTGAGG</td>
</tr>
<tr>
<td>YNR068C sense RT</td>
<td>CAGTGAAGGTCGCCCTACATTTCAGGATGTTACGATCCACACC</td>
</tr>
<tr>
<td>YNR068C sense qPCR</td>
<td>GCTACCATAGTGCCAGAAGTTT</td>
</tr>
<tr>
<td>YNR068C antisense RT</td>
<td>CAGTGAAGGTCGCCCTACAGCTACATAGTGCCAGAAGGTTT</td>
</tr>
<tr>
<td>YNR068C antisense qPCR</td>
<td>TTTCCAGCAGATCCACACC</td>
</tr>
<tr>
<td>Gene</td>
<td>RT Sequence</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>YKL151C</td>
<td>CAGTGAAGGTCGCCTACATCCTACCTACCTTTGTGCTG</td>
</tr>
<tr>
<td>YKL151C</td>
<td>CAGTGAAGGTCGCCTACATGGAAGGTCAGTC</td>
</tr>
<tr>
<td>YKL151C</td>
<td>GTGGTGGAAGGACAGGTCAGTC</td>
</tr>
<tr>
<td>YKL151C</td>
<td>TCCTACCTTCTTGTGCTG</td>
</tr>
<tr>
<td>YKL151C</td>
<td>CAGTGAAGGTCGCCTACATCGACG</td>
</tr>
<tr>
<td>YKL151C</td>
<td>TACACGCTGAGGTACGGTCAGTC</td>
</tr>
<tr>
<td>YKL151C</td>
<td>CAGTGAAGGTCGCCTACATCGACGCAG</td>
</tr>
<tr>
<td>YKL151C</td>
<td>TACACGCTGAGGTACGGTCAGTC</td>
</tr>
<tr>
<td>YKL151C</td>
<td>CAGTGAAGGTCGCCTACATCGACGCAG</td>
</tr>
<tr>
<td>PEX2</td>
<td>CAGTGAAGGTCGCCTACATCGACGCAG</td>
</tr>
<tr>
<td>PEX2</td>
<td>TACACGCTGAGGTACGGTCAGTC</td>
</tr>
<tr>
<td>PEX2</td>
<td>CAGTGAAGGTCGCCTACATCGACGCAG</td>
</tr>
<tr>
<td>PEX2</td>
<td>TACACGCTGAGGTACGGTCAGTC</td>
</tr>
<tr>
<td>PEX2</td>
<td>CAGTGAAGGTCGCCTACATCGACGCAG</td>
</tr>
<tr>
<td>PEX2</td>
<td>TACACGCTGAGGTACGGTCAGTC</td>
</tr>
<tr>
<td>PEX2</td>
<td>CAGTGAAGGTCGCCTACATCGACGCAG</td>
</tr>
<tr>
<td>PEX2</td>
<td>TACACGCTGAGGTACGGTCAGTC</td>
</tr>
<tr>
<td>PEX2</td>
<td>CAGTGAAGGTCGCCTACATCGACGCAG</td>
</tr>
<tr>
<td>PEX2</td>
<td>TACACGCTGAGGTACGGTCAGTC</td>
</tr>
<tr>
<td>PEX2</td>
<td>CAGTGAAGGTCGCCTACATCGACGCAG</td>
</tr>
<tr>
<td>PEX2</td>
<td>TACACGCTGAGGTACGGTCAGTC</td>
</tr>
<tr>
<td>YKL044W</td>
<td>CAGTGAAGGTCGCCTACATCGACGCAG</td>
</tr>
<tr>
<td>YKL044W</td>
<td>TACACGCTGAGGTACGGTCAGTC</td>
</tr>
<tr>
<td>YKL044W</td>
<td>CAGTGAAGGTCGCCTACATCGACGCAG</td>
</tr>
<tr>
<td>YKL044W</td>
<td>TACACGCTGAGGTACGGTCAGTC</td>
</tr>
<tr>
<td>Gene</td>
<td>Sense/Antisense</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>YKL044W</td>
<td>sense qPCR</td>
</tr>
<tr>
<td>TDH3</td>
<td>sense RT</td>
</tr>
<tr>
<td>TDH3</td>
<td>sense qPCR</td>
</tr>
<tr>
<td>FBA1</td>
<td>sense RT</td>
</tr>
<tr>
<td>FBA1</td>
<td>sense qPCR</td>
</tr>
<tr>
<td>TPI1</td>
<td>sense RT</td>
</tr>
<tr>
<td>TPI1</td>
<td>sense qPCR</td>
</tr>
<tr>
<td>BUD9</td>
<td>sense RT</td>
</tr>
<tr>
<td>BUD9</td>
<td>sense qPCR</td>
</tr>
<tr>
<td>DSE2</td>
<td>sense RT</td>
</tr>
<tr>
<td>DSE2</td>
<td>sense qPCR</td>
</tr>
<tr>
<td>CTS1</td>
<td>sense RT</td>
</tr>
<tr>
<td>CTS1</td>
<td>sense qPCR</td>
</tr>
<tr>
<td>SCW11</td>
<td>sense RT</td>
</tr>
<tr>
<td>SCW11</td>
<td>sense qPCR</td>
</tr>
<tr>
<td>DSE4</td>
<td>sense RT</td>
</tr>
<tr>
<td>DSE4</td>
<td>sense qPCR</td>
</tr>
<tr>
<td>FKS3</td>
<td>antisense RT</td>
</tr>
<tr>
<td>FKS3</td>
<td>antisense qPCR</td>
</tr>
<tr>
<td>YIH1</td>
<td>antisense RT</td>
</tr>
<tr>
<td>YIH1</td>
<td>antisense qPCR</td>
</tr>
<tr>
<td>STU2</td>
<td>antisense RT</td>
</tr>
<tr>
<td>STU2</td>
<td>antisense qPCR</td>
</tr>
<tr>
<td>REX2</td>
<td>antisense RT</td>
</tr>
<tr>
<td>REX2</td>
<td>antisense qPCR</td>
</tr>
<tr>
<td>CAP2</td>
<td>antisense RT</td>
</tr>
<tr>
<td>CAP2</td>
<td>antisense qPCR</td>
</tr>
<tr>
<td>HMS2</td>
<td>sense RT</td>
</tr>
<tr>
<td>HMS2</td>
<td>sense qPCR</td>
</tr>
<tr>
<td>HMS2</td>
<td>antisense RT</td>
</tr>
<tr>
<td>HMS2</td>
<td>antisense qPCR</td>
</tr>
<tr>
<td>MET13</td>
<td>antisense RT</td>
</tr>
<tr>
<td>MET13</td>
<td>antisense qPCR</td>
</tr>
<tr>
<td>Name</td>
<td>Sequence</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>YKL151C_cloning_fwd</td>
<td>CTAGAACTAGTGGAATCCCCCGGAATGTTACGTTGTTTCG</td>
</tr>
<tr>
<td>YKL151C_cloning_rev</td>
<td>ATATCGAATTCTGCAGCCGGTTGACCTGAAAATGGCAGGACCTTCTCC</td>
</tr>
<tr>
<td>mutate_gcr2_fwd</td>
<td>CTTCTGCTTTTTCTTTAAGAAGAGAGATGGGTATCCATTTCC</td>
</tr>
<tr>
<td>mutate_gcr2_rev</td>
<td>GGAGAAGGGAATGCCATTGGAGATGGGTATCCATTTCC</td>
</tr>
<tr>
<td>mutate_qPCR_fwd</td>
<td>TTTCACTATCGGGTGAAATAGTGGAATTTTTTTCTGATCTTACGACCTTTC</td>
</tr>
<tr>
<td>mutate_qPCR_rev</td>
<td>AATGTTGAGGACATGGAATATGGAATTTTTTTCTGATCTTACGACCTTTC</td>
</tr>
<tr>
<td>exomut-r1</td>
<td>ACTATCGGGTGAAATATTTTTTTCTGATCTTACGACCTTTC</td>
</tr>
<tr>
<td>exomut-o1</td>
<td>AATATTTCTCACCAGTAGTAAGTGAAAGGATATTTCTGAGGACCTTTC</td>
</tr>
<tr>
<td>exomut-o2r</td>
<td>TATAGCTCCTGTAAGATCTCACCAGTAGTAAGTGAAAGGATATTTCTGAGGACCTTTC</td>
</tr>
<tr>
<td>exomut-o3</td>
<td>TATAGCTCCTGTAAGATCTCACCAGTAGTAAGTGAAAGGATATTTCTGAGGACCTTTC</td>
</tr>
</tbody>
</table>
| exomut-o4r     | TCGTTAGAAGATTCTCCCTTT  
|               | TCCTCCTGCTCACAAATTTTA  
|               | AAGTCATACATTGCACGA     |
| exomut-o5     | AAGGGAGAATCTTCTAACGAT  
|               | AAACCCTTGAAAAACTGGGTA  
|               | GACTATGCTATGTTGAGT     |
| exomut-o6r    | GGCGGGAGCATTCTCGTGTG   
|               | ATTGTGCAGCCTGCATAGCAA  
|               | CTCAACATAGCATAGTCTA    |
| exomut-f2     | TGCTATGCAGGCTGCACAATC  
|               | ACACGAGAATGCTCCCGCCT  
|               | AGGATTTAAGGCTAAGGGA    |
|               | Amplify 3’ end of YKL151C locus to clone into Smal site of pRS316; pair with YKL151C_cloning_rev |

**Table 2.5** Primers used in this study.
CHAPTER 3

eQTL mapping of antisense transcription in *Saccharomyces cerevisiae*

INTRODUCTION

Antisense transcription, defined as non-protein-coding transcription that overlaps coding genes on the opposite strand, is ubiquitous across the tree of life (Dornenburg *et al.* 2010; He *et al.* 2008; Katayama *et al.* 2005; Wang *et al.* 2005; Xu *et al.* 2009). Detailed case studies have uncovered strong evidence of functional antisense transcription that regulates its overlapping gene through a wide variety of mechanisms (Chen *et al.* 2012; Li *et al.* 2012; Magistri and Faghihi 2012; Pelechano and Steinmetz 2013; Zhao *et al.* 2013). On a genomic scale, however, the proportion of antisense RNAs in a given genome with regulatory function remains unknown (Layer and Weil 2009; Struhl 2007; Wei *et al.* 2011).

Knowledge of how a given antisense RNA is regulated is often a first critical step in generating molecular tools that allow tests of the transcription’s function. A common approach towards this end is to modulate expression of a given antisense transcript through a characterized regulatory pathway and then test the effect on the overlapping transcript. Such strategies include perturbing histone modification complexes (Castelnuovo *et al.* 2014), other regulatory proteins (Camblong *et al.* 2007; Xu *et al.* 2009; Yadon *et al.* 2010), or transcription factor binding sites that directly regulate expression of the antisense RNA (Hongay *et al.* 2006; Houseley *et al.* 2008; Nishizawa *et al.* 2008; Zhao *et al.* 2013). In yeast and other relatively compact genomes, antisense RNA is frequently transcribed from a bidirectional promoter between two tandem or tail-to-head oriented genes (Xu *et al.* 2009). Mutation of transcription factors that regulate such bidirectional promoters has proven to be a productive approach to identifying functional antisense transcription (Mostovoy *et al.* 2014).

In this study, we sought to harness a complementary method, in which natural genetic variants rather than engineered mutations are used to dissect the regulation, and ultimately the potential for function, of antisense transcription. We used a cross between two genetically distinct yeast strains as a testbed for this approach: we assayed antisense expression in the cross progeny and mapped changes in this quantitative trait to DNA sequence variants inherited from the two parents. Focusing on cases of direct regulation of antisense transcription, we were able to observe anticorrelation between expression of these antisense RNAs and the mRNAs on the overlapping strand, and thus identify candidate cases of functional antisense transcription.
RESULTS

A survey of natural variation in yeast sense and antisense expression

We performed strand-specific RNA-seq on 63 meiotic segregants of a cross between the \textit{S. cerevisiae} lab strain BY4716 and the wild vineyard isolate RM11-1a (Brem \textit{et al.} 2002). For each segregant, we mapped RNA-seq reads to reference genomes with Bowtie (Langmead \textit{et al.} 2009); normalization, with correction for library size as implemented in EDASeq (Risso \textit{et al.} 2011), yielded expression estimates of 5077 sense and antisense features across the panel of strains. We used the expression of each feature in turn across the segregants, and genotype at each of 1245 marker loci (Brem and Kruglyak 2005), as input into a genetic linkage analysis calculation, using a standard Wilcoxon linkage test with experiment-wide multiple testing correction by permutation. Markers that significantly link to transcript expression levels are referred to as expression quantitative trait loci (eQTLs); in total, we detected 1842 linkages between eQTLs and transcripts at a false discovery rate of 10% (Table 3.1). We considered any transcript whose expression showed linkage to a marker within 10kb to represent a candidate case of locally acting, or \textit{cis}-regulatory, expression variation. Such \textit{cis}-eQTLs were among the minority in our data set (Table 3.1), with both expression of both sense and antisense transcripts most often showing linkage in \textit{trans}, \textit{i.e.} to loci at positions other than their own encoding loci. However, antisense transcripts had a higher proportion of linkage to \textit{cis} eQTLs than did sense transcripts, in agreement with a previous report (Chen \textit{et al.} 2012). Also as expected (Brem \textit{et al.} 2002), a given eQTL was often linked to expression of many transcripts, plausibly reflecting its activity as a master regulator controlling a large suite of downstream targets (Figure 3.1). The set of eQTLs showing linkage to expression of sense transcripts largely recapitulated results from non-strand-specific eQTL mapping in this cross (Brem \textit{et al.} 2002), and loci linked to expression of antisense transcripts also mirrored these eQTL hotspots (Figure 3.1). The eQTL that linked to the highest number of transcripts, both sense and antisense, was a marker in the transcription factor gene \textit{HAP1}, long considered a strong candidate for a causal \textit{trans}-acting regulatory variant between BY4716 and RM11-1a on account of the Ty1 insertion in the former (Gaisne \textit{et al.} 1999).

\textbf{eQTL effects on neighboring and overlapping transcribed elements}

At a given gene, expression from one strand can negatively regulate levels of the RNA from the opposite strand. This relationship could involve functional antisense transcription repressing the overlapping sense mRNA or, conversely, sense transcription could repress the level of potentially non-functional or opportunistic antisense transcription (Struhl 2007; Xu \textit{et al.} 2011). In either case, expression of the sense and antisense transcripts would, if affected by DNA sequence polymorphisms segregating in our cross, both show linkage to the same eQTL. We thus reasoned that screening for co-linkage by sense mRNAs and their overlapping antisense RNAs would provide a useful starting point from which to identify candidate cases of function by the latter. Among genes expressing antisense RNA, for 156 we observed that expression of
both the sense and antisense transcripts linked to the same eQTL, a strong enrichment over the expectation by chance (4.7 shared linkages, \( p < 0.001 \); Figure 3.2). For half of these loci, the eQTL allele associated with high expression of a sense mRNA was associated with low levels of the antisense overlapping it (Figure 3.2), potentially reflecting a repressive or interfering activity by one strand on the other. For the other cases of co-linkage of expression of sense and antisense, eQTL effects on the two strands were positively correlated, for which one compelling explanation could be chromatin-modifying activity governing expression at both strands (Figure 3.2).

Many antisense transcripts in yeast derive from bidirectional promoters between tandemly-oriented genes (Xu et al. 2009). We anticipated that eQTLs controlling the trans-regulatory input to the promoter of such a tandem gene pair would often show linkage to expression of antisense at the upstream gene and mRNA of the downstream sense (see Figure 3.2 for schematic). To test this, we surveyed the eQTL linkage patterns among tandemly-oriented genes where the upstream gene harbored antisense transcription. The results revealed 70 tandem gene pairs at which upstream antisense and downstream sense expression linked to the same eQTL (Figure 3.2), a strong enrichment over the chance expectation (2.3 loci with shared linkages, \( p < 0.001 \)). Of these co-regulated adjacent sense and antisense transcripts, for 43 (61%) the eQTL allele associated with high expression of the downstream mRNA was also associated with high levels of the upstream antisense; for the remainder, eQTL effects on downstream sense and upstream antisense were anti-correlated (Figure 3.2). We expected that in some cases, as an eQTL modulated expression of the upstream antisense at a tandem gene pair, the overlapping mRNA—that is, the sense expression of the upstream gene of the pair—would also be affected. This was borne out at 25 loci, at which a single eQTL linked to expression of all three transcripts of the tandem gene pair (Figure 3.2, far right); as in our whole-genome analyses, expression of antisense and the overlapping sense were correlated in some cases and anti-correlated in others. We conclude that trans-acting eQTLs can have coherent regional effects on expression of sense and antisense at tandem gene pairs, highlighting their potential as tools to discover the machinery that coordinates strand-specific regulatory events.

To further dissect the mechanisms by which eQTLs influence antisense expression, we developed an analysis strategy focused on transcription factors as mediators of eQTL activity. We first catalogued the causal genes underlying eQTLs, from the extensive previous literature validating trans-acting effects of individual polymorphisms in this cross (Table 3.2). We next mined our linkage data and previous maps of the yeast regulatory network (see Methods for details) to infer transcription factors that potentially mediate the effect of each causal eQTL gene, which we call components of the eQTL pathway (Table 3.2). Lastly, for each gene whose mRNA and/or antisense expression linked to an eQTL, we tabulated the instances of binding, in regions flanking the gene, by transcription factors in the eQTL pathway (see schematic in Table 3.3), with binding measurements from (MacIsaac et al. 2006). The results revealed that, for a given gene whose sense mRNA expression linked to an eQTL, transcription factors in the pathway often occupied the region upstream of this sense transcript (39% of loci, \( p < 0.001 \); Table 3.3) and almost never downstream (4% of genes, \( p = 0.9 \); Table 3.3), likely
reflecting eQTL effects on the mRNA mediated by transcription factor binding at the gene promoter. By contrast, for a given gene whose antisense expression linked to an eQTL, transcription factors in the pathway occupied regions up- and downstream of the gene at equal frequency (20% of genes for each class of regions, both \( p < 0.001 \); Table 3.3).

Of particular interest in the latter data were the 11 cases in which downstream of an open reading frame, in the putative promoter of an antisense whose expression linked to an eQTL, we observed occupancy of transcription factors in the eQTL pathway (Table 3.3). This suite of genes served as a rich source of testable hypotheses for transcription factor binding sites that could regulate antisense expression directly. For each of these 11 antisense transcripts targeted by one or more transcription factors in an eQTL pathway, the hosting gene was oriented in tandem with respect to the next open reading frame downstream (Figure 3.3A). At nine of these tandem gene pairs, expression of the mRNA of the downstream gene showed linkage to the same eQTL as did the upstream antisense itself (Figure 3.3B, right column). These data provide a first, strongly suggestive line of evidence that at each locus, the implicated transcription factor acts to regulate the mRNA expression of the downstream gene of a tandem pair as well as the antisense expression at the upstream gene of the pair (Mostovoy et al. 2014).

We next sought to investigate signatures of regulatory function among the 11 antisense transcripts we had identified based on linkage of their expression to eQTLs, and on transcription factor binding in their putative promoters. For this purpose, we examined the expression of the mRNA overlapping each antisense, i.e. the upstream gene of each tandem gene pair (see schematic in Figure 3.3A). Expression of this upstream mRNA, in two cases, linked significantly to the eQTL to which expression of the overlapping antisense also showed linkage (Figure 3.3B, left column). For each of two loci, the eQTL allele associated with high expression of the antisense conferred low levels of the upstream mRNA, a signature of potential repressive function by the antisense on its overlapping sense gene (Figure 3.3B, left column). At both of these loci, the same eQTL allele was associated with high expression of the antisense and the downstream mRNA, consistent with a model where the antisense transcription is co-regulated with the downstream gene and represses its overlapping sense transcript. We propose future experiments (detailed in Discussion) to test this model, particularly for the \( KSS1-BUD9 \) locus.
DISCUSSION

Our study dissected cis and trans-acting regulation for sense and antisense transcripts throughout the genomes of two strains of *Saccharomyces cerevisiae*. We identified eQTLs for sense transcripts at 26% of genes and for antisense transcripts at 9.3% of genes. Our analysis was intrinsically limited by the variation present between the two parent strains; studies using different or more divergent parental strains would likely uncover additional eQTL-transcript relationships (Gilad *et al.* 2008). Nevertheless, our approach was a powerful complement to binding-based assays for identifying transcription factor regulation of antisense transcription. The regulatory relationships uncovered here highlight emerging principles of antisense RNA regulation. Half of the antisense-containing loci for which we identified transcription factor regulators were bound by those regulators at the promoter of the sense transcript. At these loci, modulation of sense transcription may have an effect on overlapping antisense transcription (Xu *et al.* 2011); meanwhile, antisense expression at these loci may represent nonfunctional transcriptional noise (Struhl 2007). In contrast, eleven other loci with eQTL-linked antisense transcription were bound in the antisense promoter region by their eQTL-associated transcription factors. Strikingly, all of these loci were oriented tail-to-head with respect to their downstream gene, and 82% of the downstream genes were linked to the same eQTL. This result is consistent with studies implicating bidirectional promoters as the predominant source of antisense transcription in yeast (Xu *et al.* 2009) while challenging a model of independently-regulated antisense transcription at such divergent promoters (Murray *et al.* 2012). Bidirectional promoters typically contain two distinct pre-initiation complexes (Murray *et al.* 2012; Rhee and Pugh 2012); at loci where divergent transcription is co-regulated, one plausible model involves transcription factors binding between the pre-initiation complexes and interacting with both.

Of the two candidate cases of functional antisense transcription that we identified, one involves a gene of unknown function, *YNL146W*. The other, overlapping the gene *KSS1*, would be particularly interesting for further investigation. *KSS1* codes for a mitogen-activated protein kinase (MAPK) that regulates the cellular decision to undertake filamentous growth (Madhani *et al.* 1997), a growth phase involving unidirectional linear budding that diploid budding yeast cells enter when starved for nitrogen (Gimeno *et al.* 1992). The adjacent downstream gene is *BUD9*, which encodes a protein that is critical for selection of the bud site: Bud9p guides bud formation at the proximal bud site (Harkins *et al.* 2001), and deletion of *BUD9* causes increased usage of the distal bud site (Harkins *et al.* 2001). During filamentous growth, the distal bud site is preferentially used, and *BUD9* mRNA levels are reduced by approximately 50% (Taheri *et al.* 2000). *KSS1* and *BUD9* are therefore specialized for opposing conditions, one involving use of both proximal and distal bud sites and the other using only the distal site. It thus would be biologically plausible for antisense transcription co-regulated with *BUD9* to be a repressor of *KSS1* expression; in our data, we find that these three transcripts are all significantly linked to an eQTL at the gene *AMN1*, at which the BY strain carries a loss-of-function mutation (Yvert *et al.* 2003). Amn1 is a negative regulator of Ace2, a transcription factor that regulates genes involved in daughter cell...
separation during mitosis. Ace2 and other cell cycle regulators bind at the promoter of Bud9 (MacIsaac et al. 2006). Interestingly, we previously deleted ACE2 as part of a search for functional antisense transcription, and identified KSS1 and one other locus as having potentially functional antisense transcription (Mostovoy et al. 2014). The expression changes observed at this locus in the ACE2 deletion strain were very similar in magnitude and direction to expression changes seen in segregants with the RM allele of AMN1, consistent with a loss-of-function mutation at this locus being causative for the eQTL we observe in that region.

Future follow-up experiments on the KSS1-BUD9 locus can determine whether antisense transcription overlapping KSS1 acts as a functional repressor of KSS1 mRNA expression. We propose two experiments to test this hypothesis. In both, a plasmid bearing the KSS1-BUD9 locus and flanking intergenic regions will be transformed into a BY strain bearing a deletion of that locus. For the first experiment, this plasmid will be modified such that the Ace2 binding sites in the BUD9 promoter (MacIsaac et al. 2006) are mutated. This plasmid will then be separately transformed into the kss1 bud9 deletion strain and expression will measured from the wild-type and mutated plasmids with quantitative real-time PCR. If our model is correct, the strain harboring the mutated plasmid will, when compared to the wild-type plasmid, have lower levels of expression of BUD9 mRNA and KSS1 antisense transcription, and higher levels of KSS1 mRNA expression. In the second experiment, we will make use of the targeting of KSS1 antisense transcription by the nuclear exosome (Xu et al. 2009), which is recruited by the RNA binding proteins Nrd1 and Nab3 (Arigo et al. 2006). The wild-type plasmid will be modified such that Nrd1/Nab3 recognition motifs (Creamer et al. 2011) near the 3′ end of the KSS1 antisense transcript are mutated. When this plasmid is transformed into the kss1 bud9 deletion strain and its expression is compared to expression from the wild-type plasmid, we would expect to see increased levels of KSS1 antisense transcription and reduced expression of KSS1 mRNA (Mostovoy et al. 2014), while BUD9 expression may or may not be effected.

This study applied natural variation to the search for functional antisense transcription and identified regulatory modulation of antisense transcription on a biologically relevant scale. Future studies of eQTL regulation of antisense transcription can provide a complement to mutation- and binding-based studies, analogous to the combinatorial approaches that have illuminated the regulation of mRNA transcription (Zhu et al. 2008).
MATERIALS AND METHODS

Strains, growth conditions, and RNA-seq. Segregants of the BYxRM cross used in this study are listed in Table 3.4. Segregants and parent strains BY4716 and RM11-1a were grown at 30°C in YPD medium (Ausubel et al. 1995) to log phase (between 0.65-0.75 OD at 600 nm). For one biological replicate of each segregant strain and four replicates of each parent strain, total RNA was isolated by the hot acid phenol method (Ausubel et al. 1995) and treated with Turbo DNA-free (Ambion) according to the manufacturer’s instructions. Samples were sequenced using paired-end modules at the Vincent J. Coates Genomic Sequencing Laboratory at the University of California, Berkeley.

Data processing. Reads were processed and mapped as described (Mostovoy et al. 2014), with the following exceptions. Reads, consisting of paired mates, were separately mapped to both the BY (S288C version R61, (Engel et al. 2013)) and RM (RM11-1a, downloaded from http://www.broadinstitute.org/annotation/genome/saccharomyces_cerevisiae/) genomes using Bowtie (Langmead et al. 2009), yielding matches to both genomes for 96% of mate pairs. Those mate pairs were evaluated to determine whether they mapped to orthologous regions in both genomes, using whole genome alignments obtained via FSA (Bradley et al. 2009b). Mate pair mappings to the two genomes were considered orthologous if, for both mates, the mapping to the BY genome fell within 100bp of the mapping to the RM genome. Mate pairs that did not map to orthologous regions in the two genomes were discarded, while the remainder of mate pairs were combined with reads that mapped to only one of the RM or BY genomes for further processing. We determined sense and antisense expression for genes as described (Mostovoy et al. 2014). Expression levels were normalized across all samples using the upper-quartile between-lane normalization method implemented in EDASeq (Risso et al. 2011).

Linkage analysis. Segregant genotypes at 2956 markers across the genome were those used in (Brem and Kruglyak 2005) and were downloaded from http://blogs.ls.berkeley.edu/bremlab/data/genotypes_forrelease_1_20_05/. We updated the marker positions to correspond to the R61 version of the S. cerevisiae S288C genome by using BLAST (Altschul et al. 1990) to re-align each marker probe sequence to the R61 genome. To remove redundant markers, we merged adjacent markers that had identical ancestry across our sequenced segregants, disregarding sites with missing data, yielding a total of 1245 distinct markers. To identify linkages between markers and transcripts, for every combination of transcript and marker, we used the Wilcoxon rank-sum test to determine whether the transcript’s expression level differed significantly between segregants with BY or RM inheritance at the marker. We determined a p-value cutoff for significance by permuting the names of segregants associated with genotypes, such that genotypes for a given segregant no longer matched the segregant’s expression levels; linkage analysis was then performed as described above. The most significant p-value for each transcript was determined for each of 10 permutations. The p-value cutoff was set to the value that, genome-wide, gave a false discovery rate of 10% in the permuted dataset. This permutation process
was performed separately for sense and antisense transcripts, yielding \( p \)-value cutoffs of \( 8.7 \times 10^{-5} \) and \( 4.1 \times 10^{-5} \), respectively.

The linkages described in Table 3.1 were determined by including only the strongest of significant linkages per chromosome for each transcript, so as to avoid counting multiple markers within a single broad linkage region as multiple linkages; the number of linkages reported in Table 3.1 is therefore likely to be an underestimate of real linkages as it does not allow for multiple distinct linkages per chromosome. For both that table and downstream analysis, \( cis \)-linkages were defined in the following manner: for a given transcript, all markers within 10kb of the corresponding gene’s boundaries were classified as being in \( cis \). Additionally, the two markers flanking the gene on either side were considered to be in \( cis \) regardless of their distance to the gene.

For the histogram in Figure 3.1, the genome was divided into bins of 20kb, with the bin at the end of each chromosome being equal to the lesser of 20kb or the remaining length of the chromosome. Bins were assigned a linkage when, for a given transcript, the strongest significant \( trans \)-acting linkage region on that chromosome fell within the bin boundaries; linkage regions were defined as the span of genome covered by the strongest significant linkage, taking into account merged markers that span the region between the individual markers, as well as adjacent markers with equally strong linkages to the transcript.

We evaluated the statistical enrichment of overlapping transcripts in Figure 3.2 by assigning random antisense transcripts to each sense transcript being evaluated in the real dataset and then determining whether they linked to the same eQTLs. We performed this permutation 1000 times for each category of overlaps, generating a null distribution to which we compared our observed level of overlap, yielding an empirical \( p \)-value.

Regions linked in \( trans \) to many transcripts were compared to a manually curated list of previously identified causative variants underlying QTLs in this cross to yield the eQTLs in Table 3.2. To generate a set of transcription factors that could mediate the effect of eQTL genotype on expression levels, three possibilities were considered: 1) the causative gene could itself encode a transcription factor, 2) the eQTL could be linked to expression of one or more transcription factors, and/or 3) the causative gene could be known to act upstream of one or more transcription factors in a post-transcriptional matter, as determined by searching the literature. The application of these three criteria yielded the list of transcription factors listed in Table 3.2.

Positions of transcription factor binding were taken from (MacIsaac et al. 2006), using the set of binding events detected at the level of \( p<0.05 \) with no requirement of sequence conservation of the motif. Genes with sense or antisense transcripts linking to one of the eQTLs in Table 3.2 were interrogated in the 500bp flanking their coding region for binding of the eQTL-associated transcription factors. To determine the false positive rate and empirical \( p \)-value, the analysis was repeated 1000 times for random genes in place of the linked genes, while using the same set of transcription factors.
Figure 3.1 Histogram of trans-acting eQTLs for sense and antisense transcripts. We separated the S. cerevisiae genome into bins of 20kb (x-axis) and counted the number of transcripts that linked in trans to each bin (y-axis). We did this analysis separately for sense transcripts, reported in blue (top), and antisense transcripts, reported in red (bottom). eQTLs from Table 3.2 are highlighted.
Figure 3.2 eQTL linkages shared among neighboring and overlapping transcribed elements. For each category of neighboring or overlapping transcript, we counted the number of loci where the respective transcripts significantly linked to the same eQTL. We report three categories of loci: first column, genes where both sense and antisense transcription had significant linkage to the same eQTL; second column, tandemly-oriented gene pairs where upstream antisense transcription and downstream mRNA transcription linked to the same eQTL; third column, tandemly-oriented gene pairs where transcription of upstream antisense RNA, upstream mRNA, and downstream mRNA all linked to the same eQTL. For each category, we separated loci into those where a given allele of the linked eQTL was associated with higher expression of each of the linked transcripts (black bars) or with a combination of higher or lower expression across the linked transcripts (grey bars). In cartoon diagrams below the chart, yellow boxes represent genes, blue arrows denote mRNA transcripts that overlap antisense transcription, red arrows denote antisense transcripts, and cyan arrows denote mRNA transcripts at tandemly-oriented genes downstream of antisense-harboring genes.
Figure 3.3 Linkage and expression at loci with eQTL-linked and directly regulated antisense transcripts. (A) Diagram depicting the organization of loci with eQTL-linked antisense transcripts where eQTL-associated transcription factors (TFs) bound in the region corresponding to the antisense promoter. Yellow boxes represent genes, the blue arrow represents mRNA transcription, the red arrow represents antisense transcription that is linked to the eQTL, and the cyan arrow represents mRNA transcription of the adjacent downstream gene. The dashed arrow between the eQTL and TF represents an inference that the TF mediates the transcriptional effects of different alleles of the eQTL. The solid black arrow between the TF and the intergenic region between the genes represents detected binding in that region (MacIsaac et al.)
2006). (B) Each row represents one locus where antisense transcription was linked to an eQTL whose associated TF bound at the antisense promoter region, as depicted in A. First column, eQTL to which the antisense transcript was linked. In the last three columns, the color represents the average expression of the indicated transcript among segregants inheriting the RM allele of the antisense-linked eQTL divided by the average expression among segregants inheriting the BY allele. Asterisks indicate significant linkage between the respective transcript and the eQTL in the first column.
### Table 3.1 Summary of significant linkages.

For sense and antisense transcripts, columns report statistics regarding significant linkages, defined as the most significant linkage per chromosome with a $p$-value below the permutation-based $p$-value cutoff.

<table>
<thead>
<tr>
<th></th>
<th>Transcripts with at least one linkage</th>
<th>Trans linkages</th>
<th>Cis linkages</th>
<th>Total linkages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense</td>
<td>1297</td>
<td>870</td>
<td>498</td>
<td>1368</td>
</tr>
<tr>
<td>Antisense</td>
<td>470</td>
<td>251</td>
<td>223</td>
<td>474</td>
</tr>
<tr>
<td>QTLs</td>
<td>TFs through which they act</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAT</td>
<td>Mcm1, MATα1, MATα2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHO84</td>
<td>Pho2, Pho4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEU2</td>
<td>Leu3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS2</td>
<td>Lys14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>URA3</td>
<td>Ppr1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAP1</td>
<td>Hap1, Rox1, Mot3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA1</td>
<td>Kar4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMN1</td>
<td>Ace2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2 Genes inferred to harbor the causative variant behind eQTLs and the transcription factors which mediate their effects. Probable causative variants were inferred by searching the literature on the RM/BY cross. The MAT, LEU2, LYS2, URA3, and HAP1 causative variants were identified in (Brem et al. 2002), GPA1 and AMN1 in (Yvert et al. 2003), and PHO84 in (Perlstein et al. 2007). Transcription factors identified in the literature as acting downstream of the gene underlying each eQTL were as follows: Mcm1, MATα1, MATα2 (Haber 1998); Pho2, Pho4 (Oshima et al. 1996); Leu3 (Friden and Schimmel 1987); Lys14 (Feller et al. 1994); Ppr1 (Losson and Lacroute 1981); Rox1, Mot3 (Sertil et al. 2003); Kar4 (Kurihara et al. 1996). Expression of ROX1, MOT3, KAR4, and ACE2 mRNA were linked to their respective eQTLs.
<table>
<thead>
<tr>
<th></th>
<th>Sense targets of eQTLs (expectation by chance)</th>
<th>Antisense targets of eQTLs (expectation by chance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF binding upstream of gene</td>
<td>58 (7)</td>
<td>11 (2)</td>
</tr>
<tr>
<td>TF binding downstream of gene</td>
<td>6 (4)</td>
<td>11 (1)</td>
</tr>
</tbody>
</table>

**Table 3.3. Binding by eQTL-associated transcription factors at intergenic regions flanking linked transcripts.** Among genes with sense and/or antisense transcripts linked to eQTLs in Table 3.2, cells report the number of genes where eQTL-associated transcription factors bound either upstream or downstream of the gene. Numbers in red report the average number of random genes at which the transcription factors bound in the specified region among 1000 permutations, where the number of random genes tested was the same as in the corresponding real dataset. For diagrams below each column, yellow boxes represent genes, blue arrows represent mRNA, red arrows represent antisense RNA, dashed lines represent linkage, and solid arrows between eQTLs and TFs indicate an inferred relationship where TFs mediate the transcriptional effect of eQTLs.
<table>
<thead>
<tr>
<th>10 1  c</th>
<th>13 5  c</th>
<th>3 1  d</th>
<th>5 5  d</th>
<th>7 4  c</th>
<th>8 7  b</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 3  c</td>
<td>14 2  c</td>
<td>3 3  d</td>
<td>6 1  d</td>
<td>7 5  d</td>
<td>9 1  d</td>
</tr>
<tr>
<td>10 4  d</td>
<td>1 1  d</td>
<td>3 4  d</td>
<td>6 2  b</td>
<td>7 6  c</td>
<td>9 2  d</td>
</tr>
<tr>
<td>11 1  a</td>
<td>1 3  d</td>
<td>3 5  d</td>
<td>6 3  c</td>
<td>7 7  c</td>
<td>9 3  d</td>
</tr>
<tr>
<td>11 2  d</td>
<td>1 4  c</td>
<td>4 1  c</td>
<td>6 4  d</td>
<td>7 8  d</td>
<td>9 4  d</td>
</tr>
<tr>
<td>11 3  b</td>
<td>1 5  c</td>
<td>4 3  d</td>
<td>6 5  d</td>
<td>8 1  a</td>
<td>9 5  d</td>
</tr>
<tr>
<td>12 1  d</td>
<td>2 2  d</td>
<td>4 4  d</td>
<td>6 6  d</td>
<td>8 2  d</td>
<td>9 6  d</td>
</tr>
<tr>
<td>12 2  b</td>
<td>2 3  d</td>
<td>5 1  d</td>
<td>6 7  d</td>
<td>8 3  a</td>
<td>9 7  d</td>
</tr>
<tr>
<td>13 1  a</td>
<td>2 4  a</td>
<td>5 2  d</td>
<td>7 1  d</td>
<td>8 4  c</td>
<td></td>
</tr>
<tr>
<td>13 3  b</td>
<td>2 5  d</td>
<td>5 3  d</td>
<td>7 2  c</td>
<td>8 5  b</td>
<td></td>
</tr>
<tr>
<td>13 4  a</td>
<td>2 6  d</td>
<td>5 4  d</td>
<td>7 3  d</td>
<td>8 6  c</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4 Segregants of the BYxRM cross used in this study.
CHAPTER 4

Divergence of iron metabolism in wild Malaysian yeast


INTRODUCTION

Heritable genetic differences at the molecular level are widespread between populations and species. Much of the divergence observed in DNA sequence, gene expression, and other molecular attributes may bear little relationship to macroscopic traits, and the outstanding challenge in the field is to detect signatures of biological relevance in catalogs of molecular data. Although linkage or association methods can map the determinants of trait differences within species to loci that affect gene expression (Loehlin and Werren 2012; Rebeiz et al. 2009; Shapiro et al. 2004; Tung et al. 2009; Wittkopp et al. 2009), these methods cannot be applied to comparisons between reproductively isolated populations. In landmark cases, focused candidate gene studies have implicated regulatory changes in trait differences between species (Booth et al. 2010; Hanikenne et al. 2008; Marcellini and Simpson 2006; Werner et al. 2010). The ultimate goal is to predict and test phenotypes ab initio from genome-scale expression profiles of divergent individuals. To date, little precedent has been set for the experimental validation of such a pipeline.

In this work, we set out to identify genetic determinants of phenotypic differences between strains of budding yeast. Nectar of the flowers of Malaysian bertam palm trees hosts a yeast community whose fermentation is consumed by mammalian pollinators (Wiens et al. 2008). S. cerevisiae strains collected from this nectar are reproductively isolated from the rest of the species (Cubillos et al. 2011), so approaches other than standard mapping methods are required to investigate the evolution of this unique yeast population. An initial unbiased transcriptional profiling experiment was performed by another student in the Brem laboratory; for details, refer to (Lee et al. 2013). These data demonstrated that Malaysian yeast cultured in standard laboratory conditions repressed iron scavenging genes and activated genes involved in resistance to iron toxicity (referred to as iron resistance genes) relative to wine/European isolates (Figure 1 in (Lee et al. 2013)), suggesting that Malaysian yeast are particularly sensitive to iron. Consistent with this model, Malaysian strains had a marked growth defect in iron-supplemented media relative to wine/European isolates (Figure 3 in (Lee et al. 2013)); moreover, iron resistance genes in the Malaysian yeast, and not in any other population, showed an elevated rate of protein evolution relative to other genes in the genome (Table 2 in (Lee et al. 2013)).

Candidate regulators of the Malaysian iron response were identified by searching for components of the iron metabolism gene network containing nonsynonymous coding
polymorphisms unique to the Malaysian population. Two of the hits were known to be essential in laboratory yeast for growth in high-iron conditions (Li et al. 2001; Li et al. 2008): the transcription factor Yap5, which activates genes in response to treatment with excess iron, and Ccc1, a transporter that sequesters iron ions in the vacuole (Figure 1D in (Lee et al. 2013)). A third candidate was the transcription factor Aft1 (Figure 1D in (Lee et al. 2013)), a protein that has been characterized for its activity, in low levels of iron, as a regulator of the iron-scavenging genes that were differentially expressed in Malaysian yeast (Lee et al. 2013; Shakoury-Elizeh et al. 2004). All three of these genes harbored nonsynonymous mutations that were private to the Malaysian strains (Figure S1 in (Lee et al. 2013)); in hemizygote strains that contained either a copy of the Malaysian or the wine/European allele of one of these three genes but were otherwise isogenic, the Malaysian alleles of CCC1 and YAP5 compromised growth in high-iron conditions relative to the wine/European alleles (Figure 2 in (Lee et al. 2013)), mirroring the known effects of engineered null mutations at these loci (Li et al. 2001; Li et al. 2008). We set out to test whether these genes were additionally responsible for the expression differences we observed between Malaysian and wine/European iron scavenging and iron resistance genes.
RESULTS

To investigate the role of our candidate genes in divergent regulation of iron response genes in Malaysian yeast, we used reciprocal hemizygote analysis (Steinmetz et al. 2002) for each candidate locus as follows. In the diploid hybrid formed by mating a haploid Malaysian strain to a haploid wine/European, we knocked out each allele in turn of the gene of interest. The resulting hybrid strains in such a pair were isogenic to one another throughout the genome except at the hemizygous locus, where each strain harbored only one of the two alleles from the parent strains. Any phenotypic differences between the strains of a hemizygote pair could thus be ascribed to variation at the manipulated site. We cultured reciprocal hemizygote strains for AFT1, YAP5, and CCC1 in high-iron conditions and in a complete medium control, along with the wild-type parents, and we measured expression of iron starvation and iron resistance genes in each case.

Expression measurements supported a model of losses of function at Malaysian orthologs of CCC1 and YAP5: the Malaysian allele of CCC1 drove upregulation of iron starvation genes and iron resistance genes in standard complete medium and high-iron conditions, respectively, whereas the Malaysian allele of YAP5 reduced expression of iron resistance genes in high-iron conditions (Figure 4.1, 4.2), consistent with the behavior of hypomorphs at these loci (Li et al. 2001; Li et al. 2008; Lin et al. 2011). For AFT1, the Malaysian allele had a significant regulatory effect in reciprocal hemizygotes, driving upregulation of iron starvation genes during growth in complete medium (Figure 4.1, 4.2) in a manner consistent with a gain-of-function mutation (Shakoury-Elizeh et al. 2004). We conclude that the Malaysian genome harbors alleles with both activating and repressing effects on iron metabolism, reflecting a complex genetic model that involves YAP5 and CCC1 as well as AFT1.
DISCUSSION

Against a backdrop of hundreds of comparative transcriptomic studies in many species, the power of expression divergence as a predictor of trait variation has remained largely invalidated. To establish Malaysian yeast as a test bed for the reverse ecology approach, we generated expression profiles from cultures grown in standard conditions, and we used these data as a point of departure for a study of regulatory and growth phenotypes in the presence of excess iron. Although several elegant studies have characterized growth traits in yeast populations (Hodgins-Davis et al. 2012; Warringer et al. 2011; Zorgo et al. 2012), to date the degree of divergence between strains in iron sensitivity has been unknown. Our expression measurements, growth assays, and molecular genetic manipulations confirmed the iron sensitivity of Malaysian yeast and the role of three candidate genes, *AFT1*, *YAP5*, and *CCC1*, in regulatory and growth phenotypes.

What are the evolutionary forces that have driven divergence of iron metabolism in Malaysian yeast? The behavior of the Malaysian *YAP5* and *CCC1* alleles as losses of function, and the sequence-based evidence for non-neutral evolution of the coding regions of iron resistance genes, support a model of specialization by Malaysian strains to a low-iron environment. In one evolutionary scenario, reduced exposure to high-iron conditions in the Malaysian niche would have relaxed the strength of purifying selection on the iron toxicity response, allowing Malaysian yeasts to accumulate mutations in this gene network. Additionally, given the gain-of-function behavior of the Malaysian allele of *AFT1*, it is tempting to speculate that negative regulatory control of this activator of the iron starvation response has been eliminated in Malaysian yeast to sidestep iron-sensing mechanisms and raise expression of iron transporters and scavengers as would be suitable for a constant, low-iron environment. By contrast, the avid iron uptake and functional vacuolar iron storage of wine/European strains would reflect a need for flexible iron homeostasis machinery that can respond more fully to environmental change. Taken together, our findings provide a case in which the rare and often deleterious mutations that litter wild yeast genomes (Zorgo et al. 2012) follow a compelling evolutionary logic. A broader involvement of additional metals is suggested by the growth defect of a Malaysian isolate in high-copper medium (Hodgins-Davis et al. 2012). The emerging picture is one in which the Malaysian yeast population has experienced unique evolutionary pressures on metal metabolism, highlighting the palm flower niche of these microbes as a driver of evolutionary change.

Although we confirmed that our candidate genes are associated with divergent expression of iron metabolism genes, our results leave open the question of the molecular basis of the parental expression patterns. In reciprocal hemizygote experiments, the most striking regulatory effects of our candidate genes were in directions that opposed the patterns of divergence in parental homozygotes: Malaysian alleles at the *YAP5* and *AFT1* transcription factors drove downregulation of iron resistance genes and upregulation of iron starvation genes, respectively, contrasting with the high and low expressions of these gene sets, respectively, in Malaysian parental strains. Given the prevalence of transgressive segregation in the genetics of
gene expression variation within species (Gibson and Weir 2005), we favor a model in which the regulatory program of Malaysian homozygotes is an indirect regulatory response to a suite of defects in iron resistance genes in Malaysian homozygotes, one which obscures the transgressive contribution of YAP5 and AFT1 in the purebred context. However, our experiments do not rule out the possibility that alleles of YAP5 and AFT1 exert different effects in the parental background than they do in the hybrid. Future work will establish the potential for epistasis among our candidate loci and between each gene and the genetic background using the loci we have validated here as a springboard to deepen our mechanistic understanding of the regulatory and growth behaviors of Malaysian isolates.

Our work makes clear that expression profiles can be used as a powerful hypothesis generator for the study of organismal trait variation. This reverse ecology paradigm is likely to be most successful when coherent regulatory change in well-annotated pathways underlies, or responds to, change in a phenotype. With the increased availability of functional genomic resources in many taxa, the expression-based approach will likely prove to be broadly applicable in the genetic dissection of differences between populations and species.
METHODS

The Malaysian (UWOPS03.461.4) and wine/European (BC187) isolates used in this study were obtained from the National Collection of Yeast Cultures.

Quantitative PCR. Two biological replicates of each parental strain, two biological replicates of each of two independent transformants of each AFT1 and YAP5 reciprocal hemizygote, and two sets of CCC1 reciprocal hemizygote strains, each comprising two biological replicates of each of two independent transformants, were grown to mid-log phase in YPD medium at 30°C. For iron-treated samples, FeSO4 was added to a final concentration of 5 mM. After 30 min of incubation at 30°C, total RNA was isolated and treated with DNase I as described for RNA-seq, except that for iron-treated samples, RNA was purified on RNeasy columns (Qiagen). For each sample, cDNA was synthesized using SuperScript III reverse transcriptase (Life Technologies) and diluted to 1 ng/µL. Maxima SYBR green (Thermo Scientific) was used for quantitative PCR, with two to three technical replicates for each gene in each strain, on the Mx3000P system (Agilent). Cycle thresholds (Ct) for each replicate were obtained using the MxPro software (Agilent), and gene expression levels relative to ACT1 were calculated using the 2^{-\Delta\Delta Ct} method (Schmittgen and Livak 2008). As a further normalization step for ease of comparison between experiments, the vector of expression measures of a given gene in a given strain background and replicate set was centered with respect to the median across the set. Statistical significance of differential expression between genotypes for a given regulon was assessed by a paired Wilcoxon test. YAP5 reciprocal hemizygotes did not differ significantly with respect to iron starvation gene expression, and we detected little impact of variation at AFT1, YAP5, or CCC1 on iron resistance genes in synthetic complete medium (data not shown).
Figure 4.1 Regulatory impact of variation in AFT1, CCC1, and YAP5 between Malaysian and wine/European yeast. Each panel reports the regulatory effects of variation at iron metabolism genes in yeast cultured in one growth medium. In a given panel, each row reports comparisons between the effects of Malaysian and wine/European genotypes on expression of an iron starvation (Gasch et al. 2004) or iron resistance (Lin et al. 2011; Pimentel et al. 2012) target gene. Color in each cell represents a ratio of the expression measurements from two strains. A given cell in the first column reports the ratio, for the indicated target gene, of the expression in a homozygous Malaysian strain (UWOPS03.461.4), as a median across replicates (n = 2), to that in a homozygous wine/European strain (BC187). A given cell in each remaining column reports expression as a median across replicates in a Malaysian-wine/European hemizygote (UWOPS03.461.4 × BC187) bearing the Malaysian allele of the indicated variant locus (CCC1, n = 8; AFT1, n = 4; YAP5, n = 4) relative to the median expression in the hemizygote bearing the wine/European allele. (A) Iron starvation genes in synthetic complete medium. (B) Iron resistance genes in synthetic complete medium supplemented with 5 mM FeSO4. Values at the top of each column report the results of a paired Wilcoxon test for the significance of the differences in expression, across genes of the regulon, between the indicated genotypes.
Figure 4.2 Regulatory impact of variation in AFT1, CCC1, and YAP5 between Malaysian and wine/European yeast. Data are as in Figure 4.1 except that distributions of expression measurements across experimental replicates are shown. Each panel reports expression of iron-starvation (Gasch et al. 2004) and iron-resistance (Lin et al. 2011; Pimentel et al. 2012) genes as ratios of the levels measured in two strains derived from a Malaysian (UWOPS03.461.4) and a wine/European strain (BC187). At left in each row, each bar reports the median across replicates ($n = 2$) of the ratio of expression of the indicated gene between a Malaysian homozygote and a wine/European homozygote. In each remaining panel, each bar reports the median across replicates ($CCC1, n = 8; AFT1, n = 4; YAP5, n = 4$) of the ratio of expression of the indicated gene between a Malaysian-wine/European hemizygote bearing the Malaysian allele of the indicated variant locus, and the hemizygote bearing the wine/European allele. Error bars report 95% confidence intervals. (A) Iron-starvation genes in synthetic complete medium. (B) Iron-resistance genes in synthetic complete medium supplemented with 5 mM FeSO$_4$. 


REFERENCES


