The Role of Transcription Factors at Antisense-Expressing Gene Pairs in Yeast

Yulia Mostovoy[1,2], Alexander Thiemicke[1,3,4], Tiffany Y. Hsu[1,5], and Rachel B. Brem[1,6,*]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, California

[2]Present address: Cardiovascular Research Institute, University of California, San Francisco, CA

[3]Program in Molecular Medicine, Friedrich-Schiller-Universität, Jena, Germany

[4]Present address: Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN

[5]Present address: Graduate Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA

[6]Present address: Buck Institute for Research on Aging, Novato, CA

*Corresponding author: E-mail: rbrem@buckinstitute.org.

Accepted: April 27, 2016

Abstract

Genes encoded close to one another on the chromosome are often coexpressed, by a mechanism and regulatory logic that remain poorly understood. We surveyed the yeast genome for tandem gene pairs oriented tail-to-head at which expression antisense to the upstream gene was conserved across species. The intergenic region at most such tandem pairs is a bidirectional promoter, shared by the downstream gene mRNA and the upstream antisense transcript. Genomic analyses of these intergenic loci revealed distinctive patterns of transcription factor regulation. Mutation of a given transcription factor verified its role as a regulator in *trans* of tandem gene pair loci, including the proximally initiating upstream antisense transcript and downstream mRNA and the distally initiating upstream mRNA. To investigate *cis*-regulatory activity at such a locus, we focused on the stress-induced NAD(P)H dehydratase *YKL151C* and its downstream neighbor, the metabolic enzyme *GPM1*. Previous work has implicated the region between these genes in regulation of *GPM1* expression; our mutation experiments established its function in rich medium as a repressor in *cis* of the distally initiating *YKL151C* sense RNA, and an activator of the proximally initiating *YKL151C* antisense RNA. Wild-type expression of all three transcripts required the transcription factor Gcr2. Thus, at this locus, the intergenic region serves as a focal point of regulatory input, driving antisense expression and mediating the coordinated regulation of *YKL151C* and *GPM1*. Together, our findings implicate transcription factors in the joint control of neighboring genes specialized to opposing conditions and the antisense transcripts expressed between them.

Key words: gene regulation, non-coding RNA, antisense transcription.

Introduction

Transcriptional surveys, in organisms from microbes to humans, have revealed that genes encoded in close proximity tend to be expressed together (Michalak 2008; De Wit and van Steensel 2009; Osbourn and Field 2009). The functional relevance of these gene clusters and their coregulation is the subject of ongoing debate (Lercher et al. 2003; Hurst et al. 2004; Al-Shahrour et al. 2010; Meadows et al. 2010; Weber and Hurst 2011; Diaz-Castillo et al. 2012; Irimia et al. 2012; Xu et al. 2012). Many neighboring coregulated genes fall in regions unaffected by heterochromatic silencing (Talbert and Henikoff 2006), X-chromosome inactivation (Payer and Lee 2008), or imprinting (Ferguson-Smith 2011). Additional potential mechanisms of the coordinated expression of neighboring genes include DNA torsional stress resulting from protein binding, the regional effects of chromatin modifiers, gene looping, transcriptional readthrough and interference, and the activities of non-coding RNAs (ncRNAs) (Raj et al. 2006; De Wit and van Steensel 2009; Deng et al. 2010; He et al. 2012; Grzechnik et al. 2014; Meyer and Beslon 2014; Nguyen et al. 2014). A handful of landmark studies have identified the specific molecular players involved in these and other processes as they manifest at neighboring genes (Ebisuya et al. 2008; Rubin and Green 2013; Arnone et al. 2014; Liao and Chang 2014); for

most loci, however, the machinery mediating coregulation of linked genes is as yet unknown.

Among the most detailed case studies have been those characterizing functional ncRNAs that help spread regulatory signals across multiple genes. In mammals, notable examples of such ncRNA-mediated signals have been characterized at imprinted loci (Nagano et al. 2008; Pandey et al. 2008), where gene silencing spreads across hundreds of kilobases, and during X-chromosome inactivation (Lee 2009). Other examples involve the ncRNA-mediated regulation of gene clusters (Li et al. 2012; Halley et al. 2014). Of the latter, one subclass comprises cases in which, at a tandem pair of genes encoded tail-to-head on the same strand, an antisense transcript overlapping the upstream gene of the pair originates from a bidirectional promoter that also regulates the downstream gene (Wei et al. 2011; Pelechano and Steinmetz 2013).

Such loci have received particular focus in compact genomes such as that of budding yeast (Xu et al. 2009), though examples of this arrangement have also been characterized in mammalian genomes (Pandorf et al. 2006, 2012; Rinaldi et al. 2008). At individual tandem loci, condition-specific transcription factors have been shown to affect transcription of both genes of the pair, mediated in part by the function of the antisense (Ni et al. 2010; Xu et al. 2011; Pandorf et al. 2012; Nguyen et al. 2014). Across the budding yeast genome, tandem gene pairs harboring antisense transcripts often behave as coregulated clusters, with the sense expression of the upstream and downstream mRNAs exhibiting anticorrelation across conditions (Xu et al. 2009, 2011). The latter pattern is consistent with a widespread role for antisense elements as repressors of the upstream genes of tandem pairs in the sense direction (Pelechano and Steinmetz 2013). Although this model remains controversial (Struhl 2007; Layer and Weil 2009; Wei et al. 2011; Wu and Sharp 2013), antisense-expressing loci are prime candidates in the search for the logic of neighboring gene coregulation and its mechanisms.

In budding yeast, a given bidirectional promoter harbors two distinct preinitiation complexes on opposite strands, driving sense and antisense expression, respectively (Rhee and Pugh 2012), either independently or coordinately (Xu et al. 2009; Murray et al. 2012). Why two sets of complexes assemble at some intergenic regions and not others is as yet not fully understood, though a number of determinants of antisense expression have been reported (Whitehouse et al. 2007; Yadon et al. 2010; Churchman and Weissman 2011; Tan-Wong et al. 2012; Alcid and Tsukiyama 2014). We set out to investigate the molecular players that act at yeast bidirectional promoters, and to assess their roles in mRNA and antisense expression in tandem gene pairs. We discovered a genome-scale role for transcription factors at these loci, and we used our results as a jumping-off point for genomic and single-locus analysis of the regulatory logic operating at tandem pairs.

Materials and Methods

Strains and Growth Conditions

Yeast strains used in this study are listed in supplementary table S4, Supplementary Material online. Strain BY4716 (Baker Brachmann et al. 1998) was used as the wild-type strain of *Saccharomyces cerevisiae* unless otherwise indicated. Strains were grown at 30 °C in yeast peptone dextrose (YPD) medium (Ausubel et al. 1995) to log phase (between 0.65 and 0.75 optical density 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, and 2% glycerol v/v).

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 pairs of convergently transcribed genes, read-through sense transcription from one gene into its neighbor would appear indistinguishable from antisense transcription of the latter gene. To filter out such ambiguous reads in convergently transcribed gene pairs, 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 filtered for those with conserved antisense annotation as follows. We eliminated from analysis genes that had 1) antisense features (as defined above) whose lengths were either shorter than 100 bp or less than one-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 3,914 genes with orthologs in all species. We considered antisense transcription to be detectable in a given species if the normalized expression value, averaged across replicates, was five or more. To verify conservation of the boundaries of a given antisense feature which was conserved between *S. cerevisiae* and at least one other species as defined above, we identified the 3′ end position of its most abundant 3′ form in *S. cerevisiae*, and inspected RNA-seq coverage in the other species in which that feature was expressed. In 84% of these cases, we detected 3′ ends at the corresponding genomic position in each species (from among *S. mikatae, S. paradoxus*, and *S. bayanus*) in which the antisense feature was expressed (data not

shown), attesting to the conservation of the 3′ boundaries of these transcripts.

Gene Group Enrichment Tests

In tests of functional enrichment in supplementary figure S2*B*, Supplementary Material online, 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 *S. cerevisiae* biological process Gene Ontology slim terms (Ashburner et al. 2000) relative to the genome 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 (1995). We separately used Fisher's exact test to evaluate enrichment, relative to the genome, of genes with conserved antisense expression among *S. cerevisiae* genes with TATA boxes in their promoters (Basehoar et al. 2004) and genes that were components of the *S. cerevisiae* environmental stress response (Gasch et al. 2000).

Histone Modification Analysis

To evaluate enrichment of histone modifications at the 3′ ends of genes in supplementary table S2, Supplementary Material online, we focused on *S. cerevisiae* due to the relative paucity of data available for other species. We downloaded histone modification data for *S. cerevisiae* from (Pokholok et al. 2005; http://web.wi.mit.edu/young/nucleosome, accessed January 2012), and averaged levels of a given histone modification across the last 500 bp of each 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 (the latter encoded as the number of species in which antisense expression was detected, from 0 to 4).

Regulatory Protein Enrichment

To evaluate enrichment of regulatory protein binding at antisense loci, we used measurements of binding from *S. cerevisiae* as limited data were available for other species. For supplementary table S3, Supplementary Material online, we downloaded genome-wide occupancy data for regulatory proteins in *S. cerevisiae* from (Venters et al. 2011; http://atlas.bx.psu.edu/cj/occ/occ_data.html, accessed November 2011). For each factor, we compiled the set of genes that exhibited binding at 25 °C to the probe sets mapping at the 3′ ends of ORF boundaries. Limiting our analysis to genes in a tandem orientation with respect to their downstream neighbor, 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 (1995) to perform multiple testing correction and determine the proteins that had a corrected enrichment $P < 0.01$.

Transcription Factor Binding Site Analysis

Transcription factor binding sites for *S. cerevisiae* were inferred by combining position weight matrices (Spivak and Stormo 2012) with binding data (MacIsaac et al. 2006) as follows. For each transcription factor with available data in (Spivak and Stormo 2012), the genome was scanned with the corresponding position weight matrix for positions that scored above the recommended cutoff. We filtered these matches to retain only those at the regions upstream of genes for which binding of a transcription factor was detected experimentally at $P < 0.005$ according to (MacIsaac et al. 2006). To integrate these inferred binding sites with antisense loci, we first collated the set of tandem gene pairs harboring conserved antisense expression, which we split into two groups based on whether or not the antisense transcript and the downstream gene initiated from the same nucleosome-free region (NFR). We considered a locus to belong to the latter category when the antisense transcription start site (Xu et al. 2009) was positioned within 300 bp upstream of the downstream gene transcription start site. We generated a composite locus for each set by aligning the gene pairs of the set such that the transcription start sites of the downstream genes (Xu et al. 2009) were at coordinate 0; we then tallied putative transcription factor binding sites, determined as described above, 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 in each set. Separately, we also collated the set of tandem gene pairs with no detectable antisense expression and repeated the alignment and binding-site average analysis.

Coregulation between Transcription Factors and Antisense Targets

To evaluate the coexpression between transcription factors and their inferred antisense transcripts, we first collated all cases in which direct binding of a factor, with significance $P < 0.001$ in the experiments of (Harbison et al. 2004), was detected at a tandem gene pair with conserved antisense expression. A total of 285 relationships between 55 transcription factors and 133 antisense-expressing loci met this criterion. For each factor-locus pair, we calculated the Pearson correlation R between expression levels of the transcription factor and those of the antisense, across measurements from yeast grown in media containing glucose, ethanol, or galactose as the sole carbon source (Xu et al. 2009), finding a median R of −0.18. We then randomly selected 285 random matchings of the 55 factors and 133

antisense-expressing loci; for each we calculated the correlation between expression of the factor mRNA and the antisense for each null factor-locus pair as above. Carrying out this resampling procedure 100,000 times, we took as an estimate of significance the fraction of resampled data sets in which we observed the mean R to be greater than or equal to the R from the true set. Separately, we tabulated the number n_{true} of factor-locus pairs exhibiting correlated expression at $R > 0.2$, which we found to be 111 (39% of the 285 total). For each resampled data set, we then tabulated the number $n_{resample}$ of null factor-locus pairs exhibiting correlated expression at this threshold. The empirical P value was then taken as the fraction of resampled data sets in which we observed $n_{resample} \geq n_{true}$.

Nucleosome Occupancy Analysis

Short read data corresponding to nucleosome occupancy positions in *S. cerevisiae* were downloaded from (Tsankov et al. 2010) and aligned to the version of the *S. cerevisiae* genome used in (Scannell et al. 2011) with Bowtie (Langmead et al. 2009). Following (Tsankov et al. 2010), nucleosome occupancy was considered to include the 100 bp downstream of the left-most aligned base for each read. For supplementary figure S4, Supplementary Material online, we collated the set of tandem gene pairs harboring conserved antisense expression and aligned the pairs according to the transcription start sites of the downstream genes (Xu et al. 2009). We tallied nucleosome occupancy across this composite locus and normalized by the number of genes in the set. Separately, we repeated this analysis for the set of tandem gene pairs with no detectable antisense expression.

Exosome Mutant Analysis

To analyze expression in response to exosome mutation in supplementary figure S7, Supplementary Material online, transcriptional profiling data in *S. cerevisiae* were downloaded from (Xu et al. 2009). 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 ORF. Sense expression was calculated from probes within the body of the ORF.

Strand-Specific qRT-PCR

Total RNA was isolated from yeast cells by the hot acid phenol method (Ausubel et al. 1995) and treated with Turbo DNA-free (Ambion) according to the manufacturer's instructions. Strand-specific quantitative reverse transcription PCR (qRT-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, or 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 4*B*; 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 supplementary table S5, Supplementary Material online.

RT 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 min and 4 °C for 1 min. To each reaction was added 1 μl of 0.125 μg/μl actinomycin D (Perocchi et al. 2007), 1 μl of 0.1 M dithiothreitol, 4 μl of 5× 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 min and 70 °C for 15 min, then cooled to 4 °C. One microliter RNaseH (Invitrogen) was added and the reaction was incubated at 37 °C for 20 min and stored at 4 °C. Reactions were purified using the QIAquick PCR Purification Kit (Qiagen) and eluted in 50 μl EB buffer.

Quantitative PCR (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 2× 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 (SCR or ACT1) 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 (not including the ORF of the adjacent downstream gene *GPM1)*

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 supplementary table S5, Supplementary Material online for primers) and cloned into the SmaI 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* ORF which, in the reference genome, contained 17 repeats (Engel et al. 2014).

To identify and mutate the Gcr2 binding site downstream of *YKL151C*, we used a highly informative position weight matrix for Gcr1 (Gordan et al. 2011), which is Gcr2's DNA binding partner. With this matrix, we identified the highest-scoring sequence AAGAGGAAGCTC, located 170 bp from the 3′ end of the *YKL151C* ORF. We mutated this sequence in pYM007 to AAGAGTGAGATC (see supplementary table S5, Supplementary Material online for primer sequences) using the Quikchange Site-Directed Mutagenesis kit (Agilent). Mutated and wild-type plasmids were each separately transformed into a yeast strain bearing a deletion of the endogenous *YKL151C* gene (Winzeler et al. 1999) (Invitrogen) using standard protocols (Amberg et al. 2005).

To disrupt Nrd1/Nab3 motifs in the 3′ end of *YKL151C_as*, we synthesized six 60-mer oligos that overlapped each other by 20 nt (Elim Biopharmaceuticals, Inc., Hayward, CA) (see supplementary table S5, Supplementary Material online for primer sequences); these oligos encoded mutations in nine Nrd1/Nab3 motif-matching sites (Carroll et al. 2004) (supplementary fig. S8, Supplementary Material online) 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 supplementary figure S9, Supplementary Material online, 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 (supplementary fig. S9*A*, Supplementary Material online). 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 (supplementary fig. S9*B*, Supplementary Material online), we subjected equal total amounts of wild-type plasmid, mutant plasmid, or a 1:1 mixture of the two, to pseudo-RT 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, RT 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 (Baker Brachmann et al. 1998).

Results

A Survey of Antisense Transcription Conserved across Yeast Species

We sought to use antisense-expressing loci as a window onto the regulatory mechanisms that act jointly on neighboring genes in yeast. We reasoned that focusing on evolutionarily conserved antisense expression would bring to the fore the most robust and, potentially, biologically relevant regulatory events. To this end, we surveyed antisense expression in transcriptional 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 previously applied to these species (supplementary fig. S1, Supplementary Material online), including many transcripts whose detection was previously reported to depend on mutation of the nuclear exosome (Xu et al. 2009) (supplementary fig. S1*B*, Supplementary Material online). Across our complete set of *Saccharomyces* transcriptomes, we detected antisense expression in at least one species at over one-half of all genes (57%), with antisense RNAs generally expressed at low levels (fig. 1*A*). At 220 genes, we detected antisense transcription in all four surveyed species; beyond these, an additional 417 genes had antisense transcription in three of the four species (fig. 1*B* and supplementary table S1, Supplementary Material online). For subsequent analyses of conserved antisense transcription, we focused on the union of the latter two classes. We expected that this cohort would serve as a window onto the attributes of genes with relatively abundant conserved antisense transcription (supplementary fig. S2*A*, Supplementary Material online). Our set of conserved antisense transcripts included loci at which antisense expression has previously been reported to play regulatory roles (supplementary fig. S3, Supplementary Material online).

The set of genes hosting conserved antisense RNAs exhibited condition specificity (supplementary fig. S2*B*, Supplementary Material online), reduced sense expression (fig. 1*C*), and "tandem" tail-to-head positioning with respect

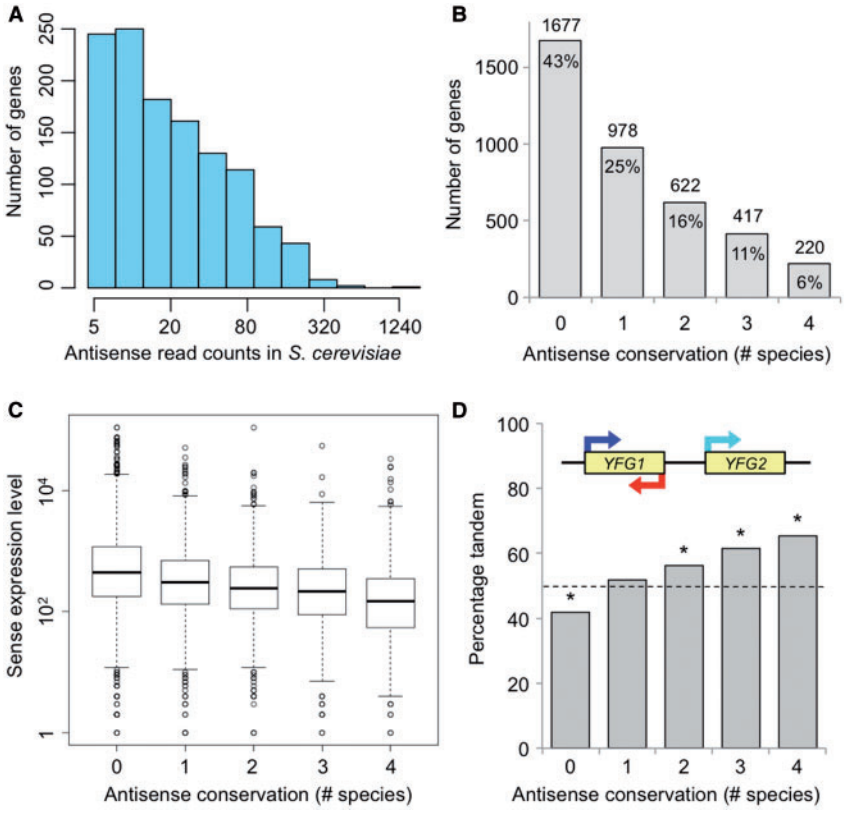

Fig. 1.—A survey of antisense expression in *Saccharomyces* yeasts. (A) The *x*-axis reports the normalized number of RNA-seq reads that mapped antisense to a given gene in *Saccharomyces cerevisiae*; the *y*-axis reports the number of genes with the antisense read count value on the *x*. Genes with zero antisense reads were excluded from analysis. (B) The *x*-axis reports the number of species, among *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. bayanus*, in which antisense transcription for a given gene was detected. The *y*-axis reports the number of genes in each category. Labels above and within each bar indicate the number and percentage of genes in each category, respectively. (C) The *x*-axis reports the same metric as in (B). The *y*-axis reports the distribution of normalized mRNA read counts 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 interquartile range (upper quartile—lower quartile) multiplied by 1.5. The relationship between conservation and mean expression is significant by linear regression at $P = 4.8 e - 05$. (D) The top cartoon depicts a schematic of a tandem gene pair. Yellow rectangles represent ORFs of Your Favorite Gene 1 and 2, and the dark blue, cyan, and red arrows represent upstream sense, downstream sense, and antisense transcription, respectively. Below, each bar represents the proportion of genes falling in tandem gene pairs, for a set of genes whose antisense expression was conserved to a given extent across Saccharomycetes. The *x*-axis reports the same metric as in (B) and (C). 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.

to downstream ORFs (fig. 1*D*), in each case echoing patterns observed in analyses of antisense transcription in *S. cerevisiae* alone (Xu et al. 2009; Lardenois et al. 2011). Many of these attributes were more salient among loci with conserved antisense RNAs than in the larger set of loci ascertained using antisense expression in *S. cerevisiae* alone (fig. 1*C* and *D* and supplementary fig. S2*A*, Supplementary Material online). 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 an informative platform for inference of the principles of regulation of neighboring genes harboring antisense transcription. At tandem pairs where the respective upstream genes were hosts of conserved antisense transcription, the 3′ ends of the upstream genes were enriched for signatures of transcriptional initiation, with respect to histone modification and occupancy of TFII protein components as measured in *S. cerevisiae* (supplementary tables S2 and S3, Supplementary Material online). This pattern is consistent with a role for binding and activity of these factors in the maintenance of active transcription of regions with conserved antisense transcription, mirroring previous analyses of antisense in *S. cerevisiae* alone (Murray et al. 2012; Goodman et al. 2013). Likewise, using nucleosome occupancy data measured in *S. cerevisiae*, at tandem pairs we detected an increase in occupancy at the two nucleosomes upstream of the NFR hosting the promoter of the downstream gene, as expected if this region acted as a bidirectional promoter

for a subset of conserved antisense transcripts (supplementary fig. S4, Supplementary Material online). We conclude that at ~10% of yeast genes, antisense expression is detectable across divergent species, with functional, structural, and chromatin-related properties as measured in *S. cerevisiae* suggestive of a shared mechanism of regulation 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 tandem gene pairs and the antisense transcripts they express. To test this notion, we evaluated transcription factor binding sites (MacIsaac et al. 2006; Spivak and Stormo 2012) at the intergenic regions between tandemly oriented genes, comparing gene pairs with no antisense expression to those with conserved antisense expression at their respective upstream genes. In this analysis, we distinguished between cases in which an antisense transcript initated from the same nucleosome-free region (NFR) as the downstream gene, and cases in which an antisense transcript initiated elsewhere in the intergenic region. The results, shown in figure 2, revealed that the intergenic regions of tandem pairs with conserved antisense expression had a higher density of transcription factor binding sites, reflecting more complex regulation at these loci. Gene pairs where the antisense transcript initiated from a shared NFR had a higher density of binding sites in the 200 bp upstream of the transcription start site of the downstream gene when compared with loci with no antisense transcription (fig. 2; $P < 2e{-}16$, paired Wilcoxon rank-sum test). Loci with antisense transcripts initiating elsewhere had a higher density of binding sites in the region 600 to 200 bp upstream of the transcription start site of the downstream gene ($P < 2e{-}16$, paired Wilcoxon rank-

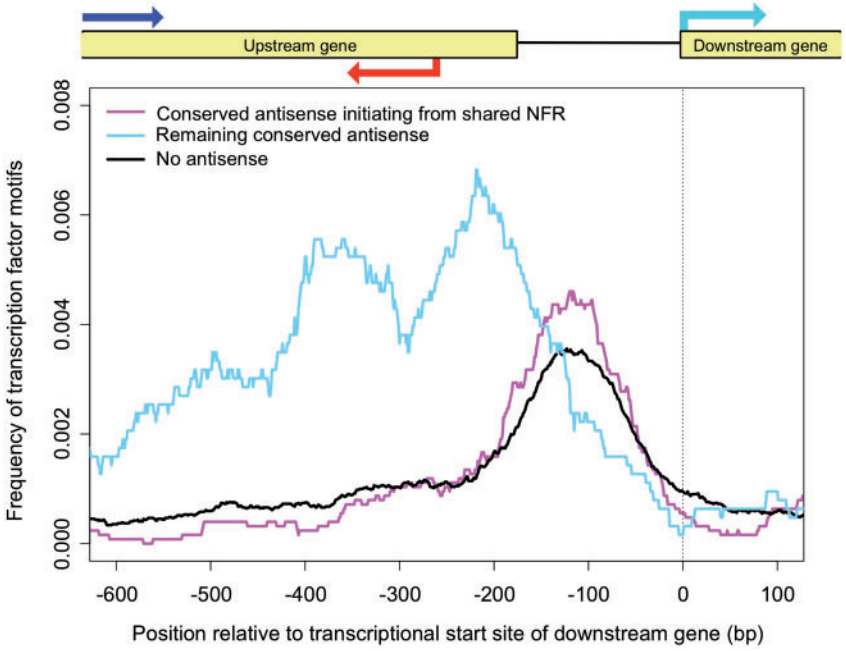

Fig. 2.—Tandem gene antisense promoters have dense transcription factor motifs. Top, schematic of the architecture of a tandem gene pair with antisense transcription. Yellow boxes represent gene transcript boundaries. Blue, cyan, and red arrows denote upstream sense transcription, downstream sense transcription, 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 *Saccharomyces cerevisiae* on the scale of the *x*-axis below (this study and (Xu et al. 2009)). Bottom, 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 initiating from shared NFR: tandem gene pairs with antisense expression detected in the upstream gene among any three of *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. bayanus*, where the antisense transcript initiates within 300 bp upstream of the transcription start site of the downstream gene (Xu et al. 2009). Remaining conserved antisense: tandem gene pairs with conserved antisense transcription at which the antisense transcript initiates outside of the 300 bp upstream of the transcription start site of the downstream gene. No antisense, tandem gene pairs without detectable antisense expression in the upstream gene in any of the four species.

sum test). We conclude that conserved antisense expression is associated with a characteristic genomic signature of transcription factor binding sites at tandem gene pairs.

As an independent strategy to explore the relationship between transcription factors and antisense transcripts, we considered the potential for a parallel with mRNA targets of transcription factors, as the former often respond in kind to up- or downregulation of the latter in response to changes in growth conditions (Margolin et al. 2006; Faith et al. 2007; Marbach et al. 2012). We hypothesized that transcription factor expression would likewise be a predictor of the expression of antisense transcripts at whose loci it binds directly. To test this notion, we tabulated the correlation, across yeast grown in each of a panel of media (Xu et al. 2009), between expression levels of the mRNA of a given transcription factor and expression of antisense transcripts at the loci to which it was observed to bind. These correlations were markedly higher than those of null sets of transcription factors matched with randomly chosen antisense transcripts (resampling $P = 0.00125$), consonant with our prediction. Interestingly, the enrichment was most marked at a fairly modest degree of correlation (resampling $P = 0.012$ for enrichment of transcription factors correlated with antisense expression at Pearson $R > 0.2$; see Materials and Methods), consistent with a model in which transcription factors can serve as partial contributors, alongside other components of the transcriptional and RNA processing machinery, to the regulation of antisense transcripts. Together with our discovery of transcription factor binding site positions at the latter (fig. 2), these findings provide compelling genome-scale support for a role of transcription factors in expression at antisense loci.

Transcription Factors Regulate Tandem Gene Pairs and Their Associated Antisense Transcripts

To pursue the role of transcription factors at antisense-expressing tandem gene pairs on a molecular level, we used our genome-scale data to infer the factors that acted at individual gene pair loci, 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. To evaluate the predicted regulatory impact of these factors at tandem pairs, we identified cases in which the upstream gene of a pair harbored antisense expression, and the downstream mRNA, but not the upstream mRNA, was directly regulated by Ace2 or Gcr2 according to binding-based (MacIsaac et al. 2006) and functional evidence (fig. 3*A*, supplementary fig. S5*A*, Supplementary Material online, and (Hu et al. 2007)). Among loci that fit the latter criteria, we earmarked for validation five with high antisense expression, to maximize the potential for detection in single-gene experiments. At each locus, we tested the hypothesis that Ace2 or Gcr2 was required for antisense transcription at the upstream gene (fig. 3*B*, left-hand cartoon). The results bore out our prediction, with upstream antisense levels at each of the 10 predicted Ace2 or Gcr2 targets decreasing 1.4- to 21-fold in the corresponding transcription factor deletion strain (fig. 3*B* and supplementary fig. S5*B*, Supplementary Material online). The impact of Ace2 or Gcr2 deletion on antisense expression at predicted targets was not a consequence of genome-scale effects, as antisense levels were largely unchanged at loci with no evidence for regulation by these factors (supplementary fig. S6, Supplementary Material online). 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.

We next asked whether Ace2 or Gcr2 served as a regulator at the regional level of a given tandem gene pair. Our prediction was that, although the upstream gene of the pair had no evidence for direct binding by the factor (per (MacIsaac et al. 2006)) and initiated hundreds of base pairs away from the factor's inferred binding site (fig. 3, left-hand cartoons), it could nonetheless respond transcriptionally to mutations in the factor. One-half the loci that we assayed conformed to this expectation: in these five instances, concomitant with a drop in expression of the downstream mRNA and upstream antisense, mutating the factor also triggered induction of the upstream mRNA (fig. 3*C* and supplementary fig. S5*C*, Supplementary Material online). Interestingly, in a strain in which the nuclear exosome, which guides the 3′-end formation and degradation of many antisense transcripts (Neil et al. 2009; Castelnuovo et al. 2013), was disabled by the deletion of exonuclease component *RRP6*, three tandem pairs exhibited both higher antisense RNA levels and lower expression of the upstream mRNA (supplementary fig. S7, Supplementary Material online). These data make clear that Ace2 and Gcr2 function as regulators of pairs of neighboring genes and the antisense transcripts they host, with antisense processing playing a key role in regulation at a subset of these loci.

A Bidirectional Promoter Regulates *YKL151C*, *GPM1*, and the Antisense Transcription between Them

We next aimed to directly test the role of a bidirectional promoter in the regulation of the genes of a tandem pair and their antisense transcription. For this case study, we chose *YKL151C*, which encodes an NAD(P)H repair enzyme; its overlapping antisense transcript, which we named *YKL151C_as*; and the adjacent downstream gene *GPM1*, which encodes a glycolytic and gluconeogenic enzyme (fig. 4*A*). The latter two transcripts both originate from a bidirectional promoter, whereas *YKL151C* initiates 1.1 kb away from the antisense transcription start site (Xu et al. 2009). The locus was notable in that the effects of a transcription factor mutation were echoed by those of an exosome mutant: deleting *GCR2*

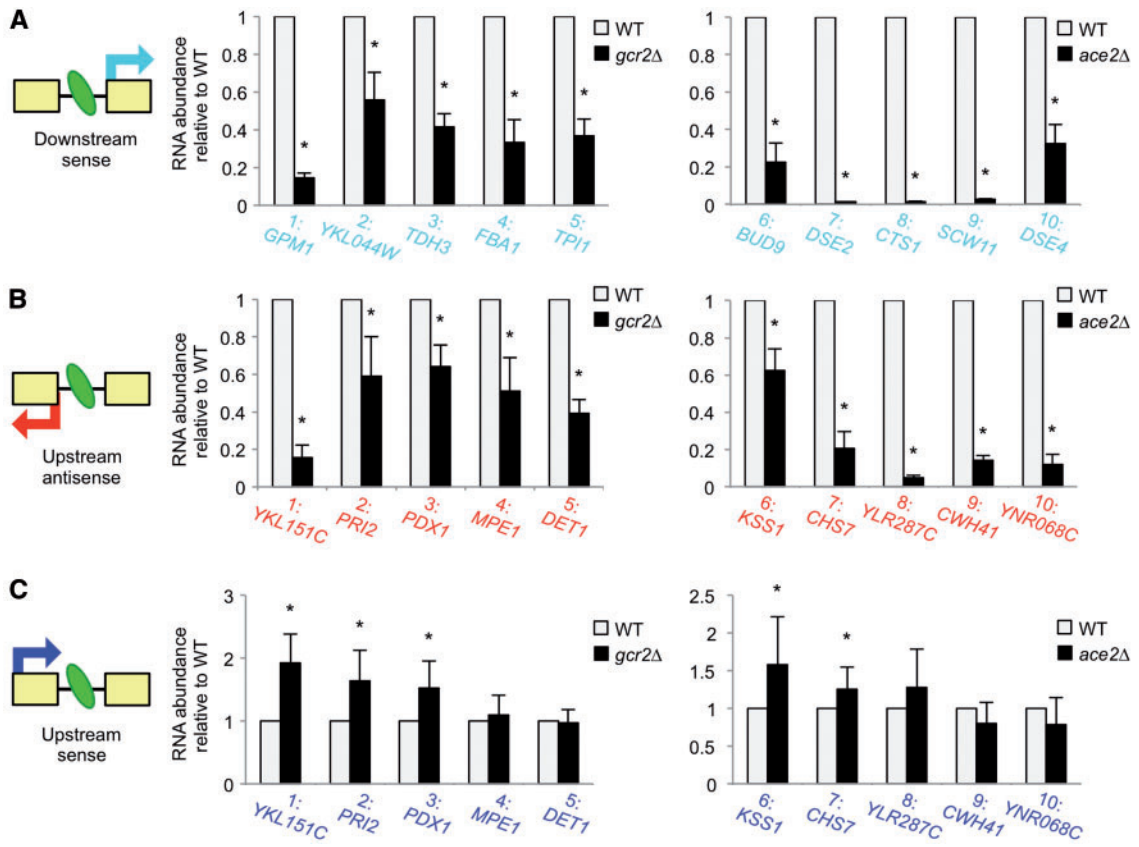

Fig. 3.—The transcription factors Gcr2 and Ace2 are required for antisense expression at predicted targets. In the cartoons at left, yellow rectangles represent tandem genes, and green ovals represent either Ace2 or Gcr2 binding in the intergenic region between the genes. Each row reports expression, measured by strand-specific qRT-PCR in *Saccharomyces 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 (*gcr2Δ*) strain, and the right-hand data panel reports changes between a WT and *ACE2* mutant (*ace2Δ*) 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; for example, locus 1 represents the tandem pair at which *YKL151C* is the upstream gene and *GPM1* is the downstream gene (fig. 3*A*). (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 WT strain; raw data are reported in supplementary figure S5, Supplementary Material online. Error bars represent propagated standard error from 4 to 6 biological replicates each of WT and mutant strains. $*P < 0.05$, Student's *t*-test.

conferred a drop in *GPM1* and *YKL151C_as* expression, whereas *YKL151C* was dramatically upregulated (fig. 3), and mutating the exosome conferred an increase in *YKL151C_as* RNA levels and a drop in sense expression (supplementary fig. S7, Supplementary Material online). We thus formulated a model in which regulatory events at the intergenic region, which is known to control *GPM1* expression (Rodicio et al. 1993), would also directly impact both *YKL151C_as* and the *YKL151C* sense RNA. Consistent with the latter notion, in wild-type cells grown with either glucose or glycerol as the sole carbon source, an environmental change known to perturb *GPM1* (Roberts and Hudson 2006), *GPM1* mRNA and *YKL151C_as* levels were correlated with one another and anticorrelated with those of *YKL151C* (fig. 4*B*). However, regulation of sense and antisense expression at *YKL151C* was separable, because 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* (fig. 4*C*). These data suggested that regulatory activity impacting *YKL151C* sense was not causal for expression effects on its antisense. Rather, we hypothesized that regulatory input at the antisense promoter—which also drives *GPM1* expression—could modulate the distally initiating *YKL151C* mRNA.

To test this notion, we first cloned a version of *YKL151C* and its flanking intergenic regions that bore a mutation in a predicted binding site for the Gcr2 transcription factor, which lay in the intergenic region beyond the end of the 3′ untranslated region of *YKL151C* (fig. 4*A*). We cultured cells harboring

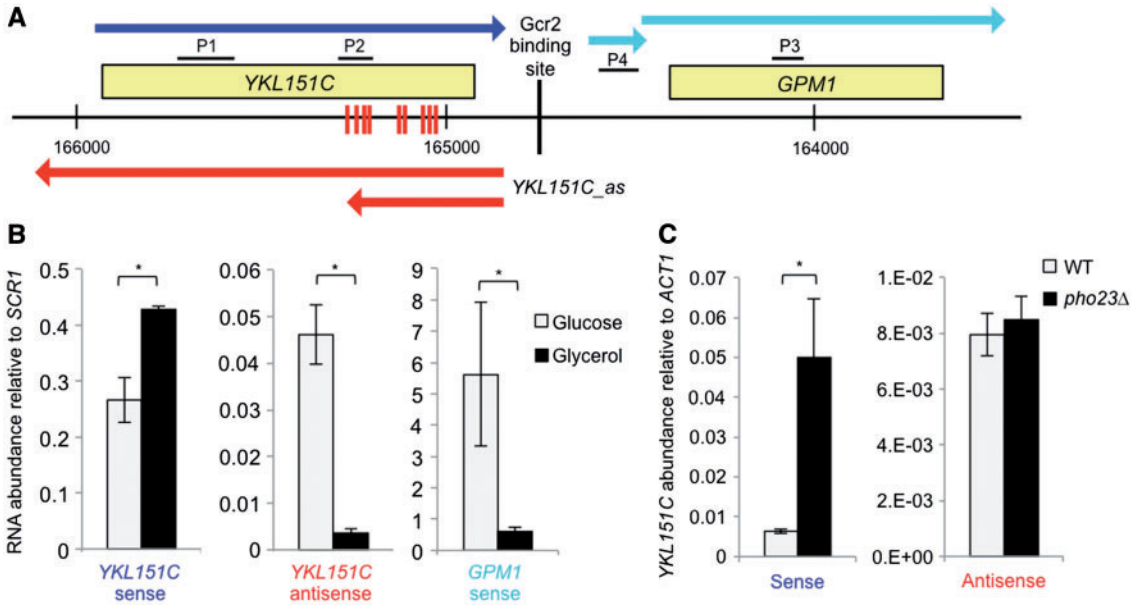

Fig. 4.—*YKL151C* sense and antisense expression is anticorrelated in some conditions but separable in others. (A) To-scale diagram representing features at the *YKL151C-GPM1* locus. Yellow rectangles represent ORFs, and blue, cyan, and red arrows represent upstream sense, downstream sense, and antisense transcripts, respectively. The short cyan arrow upstream of *GPM1* represents non-coding transcript *CUT706*. 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 (MacIsaac et al. 2006) 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–P4 represent qPCR amplicon locations. (B) Each set of bars reports expression from one strand of endogenous *YKL151C* or *GPM1* in WT *Saccharomyces cerevisiae* by qRT-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 of endogenous *YKL151C* by qRT-PCR, using primers targeting the region labeled P2 in (A); each shade reports measurements from one genetic background, either in WT *S. cerevisiae* or a strain lacking the RPD3L subunit gene *PHO23* (*pho23Δ*). $*P < 0.05$, Student's *t*-test. Error bars represent standard error from at least two biological replicates.

this mutant construct and, separately, a strain carrying a cloned wild-type *YKL151C* locus, and we measured expression from both strands in each case. The Gcr2 binding site mutant exhibited a 23% reduction in *YKL151C_as* levels (fig. 5*A*); this dovetailed with the drop in *YKL151C_as* expression seen in the *GCR2* gene knockout (fig. 3*B*), though the latter was more dramatic, as expected if it was the product of both direct and indirect regulatory effects. Likewise, our focused *cis*-regulatory mutation in the *YKL151C_as* promoter evoked an 80% increase in *YKL151C* sense mRNA expression (fig. 5*A*), mirroring results from the *trans*-acting effects of *GCR2* gene deletion (fig. 3*B*). These results establish the *YKL151C_as* promoter as a regulator of both antisense and mRNA expression at *YKL151C*.

As an independent test of the impact of *cis*-regulatory events at the *YKL151C_as* promoter on overlapping sense expression, we used a method in which disrupting putative binding sites for the Nrd1-Nab3-Sen1 termination complex at an exosome-sensitive locus drives elevated RNA abundance of the associated transcript (Arigo et al. 2006; Castelnuovo et al. 2013). Polyadenylated 3′ transcript ends at the *YKL151C* locus in wild-type cells (Schraiber et al. 2013) indicated the presence of two 3′ forms of *YKL151C_as*, one terminating 340 bp 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 ~1.2 kb, consistent with a previous report (Creamer et al. 2011) (fig. 4*A*). We identified nine putative Nrd1 and Nab3 binding sites within a 241-bp region surrounding the short form polyadenylation site for *YKL151C_as*, and we generated a version of the *YKL151C* locus in which these sites were disrupted by point mutations that left the coding sequence of the *YKL151C* ORF unchanged (supplementary fig. S8, Supplementary Material online). This construct expressed the long form of *YKL151C_as* at levels 51-fold higher than wild-type (fig. 5*B*), confirming the importance of the mutated sites in regulation of the antisense transcript, and echoing the dramatic increase seen in endogenous *YKL151C_as* in a strain bearing a mutation in *NRD1* (Creamer et al. 2011). Concomitantly, we also observed 7-fold lower abundance of *YKL151C* sense mRNA in the Nrd1-Nab3 site mutant (fig. 5*B*). Taken together, our results indicate that *cis*-regulatory perturbations to *YKL151C_as* have a robust regulatory effect on *YKL151C*. In light of the known role for the bidirectional promoter in the expression control of *GPM1* (Rodicio et al. 1993), we conclude that this single stretch of DNA functions as a determinant of expression of

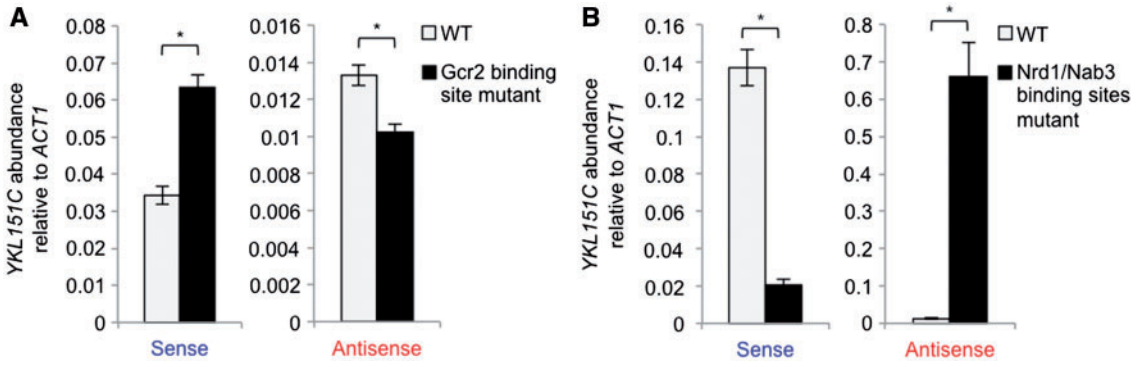

Fig. 5.—Perturbations to *YKL151C* antisense transcription inversely affect overlapping mRNA levels. In a given panel, each pair of bars reports expression from one strand of *YKL151C* in *Saccharomyces cerevisiae*, measured by qRT-PCR of amplicons denoted in figure 3*A*. (A) Each shade reports results from a version of the *YKL151C* locus with either a WT antisense promoter or one bearing a mutation in the Gcr2 binding motif at the position shown in figure 3*A*. The *y*-axis reports expression of the amplicon denoted as P2 in figure 3*A*. (B) Each shade reports results from a version of the *YKL151C* locus with either a WT coding region or one bearing silent mutations in nine Nab3/Nrd1 motifs at the positions shown in figure 3*A*. The *y*-axis reports expression of the amplicon denoted as P1 in figure 3*A*. $*P < 0.05$, Student's *t*-test. Error bars represent standard error from at least two biological replicates.

proximally- and distally initiating transcripts throughout the locus.

Discussion

Understanding the mechanisms and biological relevance of the coregulation of neighboring genes is a key challenge in modern genomics. Studies of tandem gene pairs have reported correlated expression between the antisense and downstream sense transcripts, both of which originate from a bidirectional promoter, and anticorrelation with expression of the distally initiating, upstream sense gene (Xu et al. 2009, 2011; Ni et al. 2010; Chen et al. 2012; Pandorf et al. 2012; Nguyen et al. 2014). In a handful of single-gene studies, the regulatory impact of the bidirectional promoter has been dissected, with an implication of direct regulatory function by the antisense transcript itself (Xu et al. 2011). Yet at many loci, the *cis*-acting source of regional control of gene expression remains unknown, as do the molecular players that mediate any such effect. In this work, we have identified a set of antisense-expressing neighboring gene pairs conserved across four Saccharomycetes separated by ~20 Myr of evolution. We have used our catalog to dissect the role and regulatory logic of transcription factors at these loci, and using the *YKL151C-GPM1* gene pair as a model system, we have established the *cis*-regulatory activity of an intergenic region on the transcripts of the pair.

In our study of antisense-expressing tandem gene pairs, we found that transcription factors with signatures of binding closest to the upstream gene were among the most likely to generate antisense RNAs. Thus, occupancy of a transcription factor at an open promoter may not be sufficient for antisense expression: a specific binding architecture may well be required, potentially to maximize interaction of the factor with the more distal of two preinitiation complexes assembled in the downstream gene promoter's NFR (Rhee and Pugh 2012). As our inferred targeting relationships between a transcription factor and an antisense element bore out in validation experiments, we expect that this strategy will serve as a rich source of candidate regulators for antisense transcripts beyond those studied here. Likewise, in several cases, the transcription factors we inferred to be antisense regulators at tandem gene pairs proved to have effects at the regional level—their deletion perturbed the upstream, distally initiating mRNA of a tandem gene pair as well as the antisense and the downstream mRNA. Our results leave open the mechanism of each such response, which could in some cases proceed from indirect effects of a change in cell state upon transcription factor deletion, and in others from the direct binding of factors at intergenic regions of tandem gene pairs. For the latter factors, the antisense transcripts they stimulate could regulate the upstream genes in *cis* (e.g., by transcriptional interference or laying down chromatin marks) or in *trans* (e.g., by regulating mRNA half-life, localization, or translation) (Faghihi and Wahlestedt 2009; Pelechano and Steinmetz 2013). Alternatively, a factor binding in the intergenic region of a tandem pair could contact the upstream gene's promoter via DNA torsion or looping (Grzechnik et al. 2014; Meyer and Beslon 2014), with the antisense as a non-functional side product. Each model is likely to be at play at a subset of tandem pairs, and in many cases, transcription factor interactions and antisense expression would be a signpost of the coordinated regulation of the up- and downstream genes of the pair.

Our study also reveals the potential for a regulatory logic at antisense-expressing tandem gene pairs. For example, the protein product of *YKL151C*, the upstream gene of a tandem pair controlled by the transcription factor Gcr2, 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). *GPM1*, the downstream gene of the tandem pair, encodes phosphoglycerate mutase, a glycolysis, and gluconeogenesis enzyme (Lam and Marmur 1977). We have demonstrated that the region between the two genes, already known to function as a potent regulator of *GPM1* (Rodicio et al. 1993), also controls *YKL151C*, though the latter initiates hundreds of base pairs away. *GPM1* is repressed during metabolism of non-fermentable carbon sources, in response to carbon starvation, and during stationary phase (Gasch et al. 2000; Bradley et al. 2009), conceivably as part of a program to minimize the energetic cost of glycolytic enzyme expression (Fraenkel 1982). These conditions drive the induction of *YKL151C* (Gasch et al. 2000; Bradley et al. 2009), consistent with an increased requirement for its NAD(P)H repair activity and for the oxidative stress response in general (Longo et al. 1996). Our data suggest that these two regulatory responses to carbon source changes can be coupled through the simultaneous activity of the intergenic region on both genes. This coordinated regulation of two neighboring genes of related function echoes the pattern seen at the yeast *GAL10* locus (Houseley et al. 2008), which is jointly regulated with a neighboring gene of the same pathway via functional antisense expression. It is tempting to speculate that antisense transcription at the *YKL151C-GPM1* locus likewise plays a regulatory role. We detected no evidence for function in *trans* by *YKL151C_as* (supplementary fig. S9, Supplementary Material online). A putative function in *cis* by this transcript could involve modulation of marks by the histone deacetylase complexes Hda1/2/3 and Rpd3, which are required for anticorrelation of overlapping sense and antisense expression at *YKL151C* (Castelnuovo et al. 2014). Alternatively, however, the *YKL151C-GPM1* intergenic region could modulate *YKL151C* by other biophysical or biochemical means. In either case, the intergenic region would likely enable an additional layer of control of *YKL151C*, beyond the regulatory events at its 5′ end that include binding of Rpd3L (Venters et al. 2011) and stress and carbon source-responsive transcription factors such as MSN2/4, HAP1/2/3/4, and ADR1 (MacIsaac et al. 2006).

Importantly, at other tandem pairs that exhibited a regional response to transcription factor deletion (fig. 3C), the two genes of the pair also have functions in common, providing further support for a model of biologically relevant joint regulation. At one locus controlled by the cell cycle regulator Ace2, the upstream gene *CHS7* encodes a posttranslational regulator of Chs3 (Trilla et al. 1999), a chitin synthase responsible for producing the chitin ring during cell division (Shaw et al. 1991), and the downstream gene, *DSE2*, is a putative glucosidase involved in cytokinesis (Colman-Lerner et al. 2001). At another locus, controlled by the metabolic regulator Gcr2, the upstream gene *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. 2009). Each of these loci is thus a compelling candidate case of biologically relevant neighboring gene joint regulation, with the potential involvement of a functional antisense regulator. Together, our results attest to the identification of antisense transcription, and inference of transcription factor control, as a strategy to winnow down the hundreds of coregulated neighboring gene pairs in yeast (Cohen et al. 2000) to the instances that are most likely to yield biological insight.

Strategies to dissect the mechanisms of joint regulation at neighboring genes are at a premium in the modern literature. Our work establishes the utility of analysis of transcription factors, beyond focused studies of known determinants of antisense biogenesis (Nishizawa et al. 2008; Gelfand et al. 2011), to characterize antisense-expressing loci. We anticipate that transcription factor control of a given locus and its antisense expression may often serve as a signpost for the biological relevance of coregulation between the genes at the locus.

Supplementary Material

Supplementary tables S1–S5 and figures S1–S9 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Kevin Doxzen for technical contributions, O.K. Yoon and J.I. Roop for discussions, and L. Lombardi and J. Rine for helpful comments on the article. This work was supported by the National Institutes of Health (R01 GM087432 to R.B.B.).

Literature Cited

Alcid EA, Tsukiyama T. 2014. ATP-dependent chromatin remodeling shapes the long noncoding RNA landscape. Genes Dev. 28:2348–2360.

Al-Shahrour F, et al. 2010. Selection upon genome architecture: conservation of functional neighborhoods with changing genes. PLoS Comput Biol. 6:e1000953.

Amberg DC, Burke D, Strathern JN. 2005. Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual. Cold Spring Harbor (NY): CSHL Press.

Arigo JT, Carroll KL, Ames JM, Corden JL. 2006. Regulation of yeast *NRD1* expression by premature transcription termination. Mol Cell. 21:641–651.

Arnone JT, et al. 2014. Dissecting the *cis* and *trans* elements that regulate adjacent-gene coregulation in *Saccharomyces cerevisiae*. Eukaryot Cell. 13:738–748.

Ashburner M, et al. 2000. Gene Ontology: tool for the unification of biology. Nat Genet. 25:25–29.

Ausubel FM, et al. 1995. Short protocols in molecular biology. New York: John Wiley & Sons.
Baker Brachmann C, et al. 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast 14:115–132.
Basehoar AD, Zanton SJ, Pugh BF. 2004. Identification and distinct regulation of yeast TATA box-containing genes. Cell 116:699–709.
Behal RH, Browning KS, Hall TB, Reed LJ. 1989. Cloning and nucleotide sequence of the gene for protein X from *Saccharomyces cerevisiae*. Proc Natl Acad Sci. 86:8732–8736.
Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 57:289–300.
Bessaud M, et al. 2008. Development of a Taqman RT-PCR assay for the detection and quantification of negatively stranded RNA of human enteroviruses: evidence for false-priming and improvement by tagged RT-PCR. J Virol Methods. 153:182–189.
Bradley PH, Brauer MJ, Rabinowitz JD, Troyanskaya OG. 2009. Coordinated concentration changes of transcripts and metabolites in *Saccharomyces cerevisiae*. PLoS Comp Biol. 5:e1000270.
Carroll KL, Pradhan DA, Granek JA, Clarke ND, Corden JL. 2004. Identification of *cis* elements directing termination of yeast nonpolyadenylated snoRNA transcripts. Mol Cell Biol. 24:6241–6252.
Castelnuovo M, et al. 2013. Bimodal expression of *PHO84* is modulated by early termination of antisense transcription. Nature Struct Biol. 20:851–858.
Castelnuovo M, et al. 2014. Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. Nucleic Acids Res. 42:4348–4362.
Chen H, Rosebrock AP, Khan SR, Futcher B, Leatherwood JK. 2012. Repression of meiotic genes by antisense transcription and by Fkh2 transcription factor in *Schizosaccharomyces pombe*. PLoS One 7:e29917.
Churchman LS, Weissman JS. 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469:368–373.
Cohen BA, Mitra RD, Hughes JD, Church GM. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet. 26:183–186.
Colman-Lerner A, Chin TE, Brent R. 2001. Yeast Cbk1 and Mob2 activate daughter-specific genetic programs to induce asymmetric cell fates. Cell 107:739–750.
Creamer TJ, et al. 2011. Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet. 7:e1002329.
De Wit E, van Steensel B. 2009. Chromatin domains in higher eukaryotes: insights from genome-wide mapping studies. Chromosoma 118:25–36.
Deng Y, et al. 2010. Genome-wide analysis of the effect of histone modifications on the coexpression of neighboring genes in *Saccharomyces cerevisiae*. BMC Genomics 11:550.
Diaz-Castillo C, Xia XQ, Ranz JM. 2012. Evaluation of the role of functional constraints on the integrity of an ultraconserved region in the genus *Drosophila*. PLoS Genet. 8:e1002475.
Ebisuya M, Yamamoto T, Nakajima M, Nishida E. 2008. Ripples from neighbouring transcription. Nat Cell Biol. 10:1106–1113.
Engel SR, et al. 2014. The reference genome sequence of *Saccharomyces cerevisiae*: then and now. G3 4:389–398.
Faghihi MA, Wahlestedt C. 2009. Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol. 10:637–643.
Faith JJ, et al. 2007. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. PLoS Biol. 5:e8.
Ferguson-Smith AC. 2011. Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet. 12:565–575.
Fraenkel DG. 1982. Carbohydrate metabolism. In: Strathern JN, Jones EW, Broach JR, editors. The molecular biology of the yeast Saccharomyces: metabolism and gene expression. Vol. 11. Cold Spring Harbor (NY): CSHL Press. p. 1–37.
Gasch AP, et al. 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell. 11:4241–4257.
Gelfand B, et al. 2011. Regulated antisense transcription controls expression of cell-type-specific genes in yeast. Mol Cell Biol. 31:1701–1709.
Goodman AJ, Daugharthy ER, Kim J. 2013. Pervasive antisense transcription is evolutionarily conserved in budding yeast. Mol Biol Evol. 30:409–421.
Gordan R, et al. 2011. Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights. Genome Biol. 12: R125.
Grzechnik P, Tan-Wong SM, Proudfoot NJ. 2014. Terminate and make a loop: regulation of transcriptional directionality. Trends Biochem Sci. 39:319–327.
Halley P, et al. 2014. Regulation of the apolipoprotein gene cluster by a long noncoding RNA. Cell Rep. 6:222–230.
Harbison CT, et al. 2004. Transcriptional regulatory code of a eukaryotic genome. Nature 431:99–104.
He C, et al. 2012. Young intragenic miRNAs are less coexpressed with host genes than old ones: implications of miRNA-host gene coevolution. Nucleic Acids Res. 40:4002–4012.
Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M. 2008. A ncRNA modulates histone modification and mRNA induction in the yeast *GAL* gene cluster. Mol Cell. 32:685–695.
Hu Z, Killion PJ, Iyer VR. 2007. Genetic reconstruction of a functional transcriptional regulatory network. Nat Genet. 39:683–687.
Hurst LD, Pál C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. Nature Rev Genet. 5:299–310.
Irimia M, et al. 2012. Extensive conservation of ancient microsynteny across metazoans due to *cis*-regulatory constraints. Genome Res. 22:2356–2367.
Lam K, Marmur J. 1977. Isolation and characterization of *Saccharomyces cerevisiae* glycolytic pathway mutants. J Bacteriol. 130:746–749.
Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.
Lardenois A, et al. 2011. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. Proc Natl Acad Sci. 108:1058–1063.
Layer JH, Weil PA. 2009. Ubiquitous antisense transcription in eukaryotes: novel regulatory mechanism or byproduct of opportunistic RNA polymerase? F1000 Biol Rep. 1:33.
Lee JT. 2009. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. Genes Dev. 23:1831–1842.
Lenstra TL, et al. 2011. The specificity and topology of chromatin interaction pathways in yeast. Mol Cell. 42:536–549.
Lercher MJ, Blumenthal T, Hurst LD. 2003. Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. Genome Res. 13:238–243.
Li MZ, Elledge SJ. 2007. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods. 4:251–256.
Li Q, et al. 2012. AS1DHRS4, a head-to-head natural antisense transcript, silences the DHRS4 gene cluster in *cis* and *trans*. Proc Natl Acad Sci. 109:14110–14115.
Liao BY, Chang A. 2014. Accumulation of CTCF-binding sites drives expression divergence between tandemly duplicated genes in humans. BMC Genomics 15(Suppl 1):S8.
Longo VD, Gralla EB, Valentine JS. 1996. Superoxide dismutase activity is essential for stationary phase survival in *Saccharomyces cerevisiae*. J Biol Chem. 271:12275–12280.
MacIsaac K, et al. 2006. An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. BMC Bioinformatics 7:113.

Marbach D, et al. 2012. Predictive regulatory models in *Drosophila melanogaster* by integrative inference of transcriptional networks. Genome Res. 22:1334–1349.
Marbaix AY, et al. 2011. Extremely conserved ATP- or ADP-dependent enzymatic system for nicotinamide nucleotide repair. J Biol Chem. 286:41246–41252.
Margolin AA, et al. 2006. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7(Suppl 1):S7.
McAlister L, Holland M. 1985. Differential expression of the three yeast glyceraldehyde-3-phosphate dehydrogenase genes. J Biol Chem. 260:15019–15027.
Meadows LA, Chan YS, Roote J, Russell S. 2010. Neighbourhood continuity is not required for correct testis gene expression in *Drosophila*. PLoS Biol. 8:e1000552.
Meyer S, Beslon G. 2014. Torsion-mediated interaction between adjacent genes. PLoS Comput Biol. 10:e1003785.
Michalak P. 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics 91:243–248.
Murray SC, et al. 2012. A pre-initiation complex at the 3′-end of genes drives antisense transcription independent of divergent sense transcription. Nucleic Acids Res. 40:2432–2444.
Nagano T, et al. 2008. The air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 322:1717–1720.
Neil H, et al. 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457:1038–1042.
Nguyen T, et al. 2014. Transcription mediated insulation and interference direct gene cluster expression switches. eLife 3:e03635.
Ni T, et al. 2010. The prevalence and regulation of antisense transcripts in *Schizosaccharomyces pombe*. PLoS One 5:e15271.
Nishizawa MF, et al. 2008. Nutrient-regulated antisense and intragenic RNAs modulate a signal transduction pathway in yeast. PLoS Biol. 6:e326.
Osbourn AE, Field B. 2009. Operons. Cell Mol Life Sci. 66:3755–3775.
Pandey RR, et al. 2008. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell. 32:232–246.
Pandorf CE, et al. 2006. Dynamics of myosin heavy chain gene regulation in slow skeletal muscle: role of natural antisense RNA. J Biol Chem. 281:38330–38342.
Pandorf CE, et al. 2012. Regulation of an antisense RNA with the transition of neonatal to IIb myosin heavy chain during postnatal development and hypothyroidism in rat skeletal muscle. Am J Physiol Regul Integr Comp Physiol. 302:R854–R867.
Payer B, Lee JT. 2008. X chromosome dosage compensation: how mammals keep the balance. Annu Rev Genet. 42:733–772.
Pelechano V, Steinmetz LM. 2013. Gene regulation by antisense transcription. Nat Rev Genet. 14:880–893.
Perocchi F, Xu Z, Clauder-Münster S, Steinmetz LM. 2007. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. Nucleic Acids Res. 35:e128.
Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res. 29: e45-e45.
Pokholok DK, et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122:517–527.
Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 4:e309.
Rhee HS, Pugh BF. 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature 483:295–301.
Rinaldi C, et al. 2008. Intergenic bidirectional promoter and cooperative regulation of the IIx and IIb MHC genes in fast skeletal muscle. Am J Physiol Regul Integr Comp Physiol. 295:R208–R218.
Roberts GG, Hudson AP. 2006. Transcriptome profiling of *Saccharomyces cerevisiae* during a transition from fermentative to glycerol-based respiratory growth reveals extensive metabolic and structural remodeling. Mol Genet Genomics. 276:170–186.
Rodicio R, Heinisch JJ, Hollenberg CP. 1993. Transcriptional control of yeast phosphoglycerate mutase-encoding gene. Gene 125:125–133.
Rubin AF, Green P. 2013. Expression-based segmentation of the *Drosophila* genome. BMC Genomics 14:812.
Scannell DR, et al. 2011. The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. G3 1:11–25.
Schraiber JG, Mostovoy Y, Hsu TY, Brem RB. 2013. Inferring evolutionary histories of pathway regulation from transcriptional profiling data. PLoS Comp Biol. 9: e1003255.
Shaw JA, et al. 1991. The function of chitin synthases 2 and 3 in the *Saccharomyces cerevisiae* cell cycle. J Cell Biol. 114:111–123.
Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics 122:19–27.
Spivak AT, Stormo GD. 2012. ScerTF: a comprehensive database of benchmarked position weight matrices for *Saccharomyces* species. Nucleic Acids Res. 40:D162–D168.
Struhl K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol. 14:103–105.
Talbert PB, Henikoff S. 2006. Spreading of silent chromatin: inaction at a distance. Nature Rev Genet. 7:793–803.
Tan-Wong SM, et al. 2012. Gene loops enhance transcriptional directionality. Science 338:671–675.
Trilla JA, Durán A, Roncero C. 1999. Chs7p, a new protein involved in the control of protein export from the endoplasmic reticulum that is specifically engaged in the regulation of chitin synthesis in *Saccharomyces cerevisiae*. J Cell Biol. 145:1153–1163.
Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. 2010. The role of nucleosome positioning in the evolution of gene regulation. PLoS Biol. 8:e1000414.
Venters BJ, et al. 2011. A comprehensive genomic binding map of gene and chromatin regulatory proteins in *Saccharomyces*. Mol Cell. 41:480–492.
Weber CC, Hurst LD. 2011. Support for multiple classes of local expression clusters in *Drosophila melanogaster*, but no evidence for gene order conservation. Genome Biol. 12:R23.
Wei W, Pelechano V, Järvelin AI, Steinmetz LM. 2011. Functional consequences of bidirectional promoters. Trends Genet. 27:267–276.
Whitehouse I, Rando OJ, Delrow J, Tsukiyama T. 2007. Chromatin remodelling at promoters suppresses antisense transcription. Nature 450:1031–1035.
Winzeler EA, et al. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science 285:901–906.
Wu X, Sharp PA. 2013. Divergent transcription: a driving force for new gene origination? Cell 155:990–996.
Xu C, Chen J, Shen B. 2012. The preservation of bidirectional promoter architecture in eukaryotes: what is the driving force? BMC Syst Biol. 6(Suppl 1):S21.
Xu Z, et al. 2009. Bidirectional promoters generate pervasive transcription in yeast. Nature 457:1033–1037.
Xu Z, et al. 2011. Antisense expression increases gene expression variability and locus interdependency. Mol Syst Biol. 7:468.
Yadon AN, et al. 2010. Chromatin remodeling around nucleosome-free regions leads to repression of noncoding RNA transcription. Mol Cell Biol. 30:5110–5122.

Associate editor: Mar Alba