MBE Advance Access published November 21, 2012

Pervasive Antisense Transcription Is Evolutionarily Conserved in Budding Yeast

Aaron J. Goodman,[†] Evan R. Daugharthy,[†‡] and Junhyong Kim*

Department of Biology, School of Arts and Sciences, University of Pennsylvania

[‡]Present address: Department of Systems Biology, Harvard Medical School, Boston, MA 02115

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: junhyong@sas.upenn.edu.

Associate editor: John H. McDonald

Abstract

Antisense transcription, or transcription on the opposite strand of the same genomic locus as another transcript, has been observed in many organisms, including yeast. Several antisense transcripts are known to be conserved across various species of yeast, and a few antisense transcripts are associated with functional regulation of the sense transcript. We detect antisense transcription from approximately 90% of protein-coding genes, and antisense transcription is generally associated with histone modifications indicative of a transcriptionally active state. The pattern of genome-wide antisense transcription in two species of budding yeast, *Saccharomyces cerevisiae* and *S. paradoxus*, is widely evolutionarily conserved: Antisense transcripts exhibit conserved expression levels and localization with respect to gene annotations. Regions of genes exhibiting conserved antisense transcription also show less sequence divergence than regions of genes without antisense transcription. These findings provide further support that widespread antisense transcription is functional in yeast, and expand the catalog of putative functional antisense transcripts to include nonpolyadenylated transcripts. Because antisense transcripts are less divergent in sequence than expected, they likely contain sequences important to their function.

Key words: comparative transcriptomics, antisense transcription, yeast, *Saccharomyces cerevisiae*, *S. paradoxus*.

Introduction

Antisense transcription, or transcription on the opposite strand of the same genomic locus as another transcript, generates an endogenous class of RNA molecules called *cis*-natural antisense transcripts (*cis*-NATs) or simply antisense transcripts. [The hallmark feature of NATs is their complementary sequence to other endogenously transcribed "sense" RNA molecules (Vanhée-Brossollet and Vaquero 1998); in this case, the antisense transcript contains sequence complementary to the RNA transcribed from the same genomic locus.] Antisense transcription has been observed in many organisms, including bacteria, fungi, protozoa, plants, invertebrates, and mammals (Beiter et al. 2009). Despite the prevalence of antisense transcription, no general pathways have yet been identified for *cis*-NAT function.

Antisense transcription has recently been identified in many strains of hemi-ascomycete yeast, including *Schizosaccharomyces pombe* (Dutrow et al. 2008; Wilhelm et al. 2008; Ni et al. 2010), *Saccharomyces cerevisiae* (Neil et al. 2009; Xu et al. 2009), *S. paradoxus*, *S. mikatae*, *S. bayanus*, *S. castellii*, and *Kluyveromyces lactis* (Yassour et al. 2010). More than 1,500 genes in *S. cerevisiae* have been shown to exhibit transcription from both strands (David et al. 2006). It is thought that some antisense transcription represents transcriptional noise. Noisy transcription can originate from cryptic or weak promoters, or an "open" chromatin milieu favorable to transcription, such as the nucleosome-free region associated with promoters (the source of divergent transcription in yeast generating antisense cryptic unstable transcripts, CUTs) (Johnson et al. 2005; Struhl 2007; Neil et al. 2009; Xu et al. 2009). Because promoter directionality can be influenced by histone modifications, the distinction between regulated and unregulated or "noisy" transcription remains unclear (Churchman and Weissman 2011). The fact that several antisense transcripts have demonstrated functions suggests that at least some portion of antisense transcription events is regulated and functionally significant.

Typically, the functions of *cis*-NATs have been studied in the context of their corresponding sense transcripts (Barrell et al. 1976). As a result, only a small number of antisense transcripts have been implicated in the regulation of well-characterized genes in yeast, for example, *IME4*, *GAL10*, and *PHO84*. For meiosis regulatory gene *IME4*, transcription of a 450 bp region of the antisense strand is necessary to suppress *IME4* transcription; however, the underlying regulatory mechanism is unknown (Hongay et al. 2006; Gelfand et al. 2011). Antisense transcription at the *GAL10* locus is correlated with trimethylation of histone-3 lysine-36 (H3K36me3) via the methyltransferase Set2, as well as histone de-acetylation via the Eaf3-recruited Rpd3S complex, putatively causing repression of *GAL1* and *GAL10* sense transcription (Houseley et al. 2008). Similarly, *PHO84* expression has been shown to be repressed by Hda1-mediated histone de-acetylation in the regions encoding antisense transcripts (Camblong et al. 2007).

There are a number of hypothesized mechanisms for *cis*-NAT-mediated regulation of sense gene expression. The antisense transcription *itself* may regulate sense transcription by

directing chromatin remodeling (Bolland et al. 2004), or by interfering with transcription initiation or elongation on the sense strand (broadly referred to as "transcriptional interference") (Prescott and Proudfoot 2002; Shearwin et al. 2005). Alternatively, the antisense transcript may interact with sense RNA or DNA, functioning as a targeting mechanism or platform for the recruitment of regulatory machinery (Yang and Kuroda 2007; Faghihi and Wahlestedt 2009). Importantly, *S. cerevisiae* lacks essential components of the RNA interference (RNAi) pathway found in most other eukaryotes, so if antisense transcription events have functional consequences in yeast, they must utilize RNAi-independent regulatory mechanisms (Aravind et al. 2000; Drinnenberg et al. 2009).

Experimental observations suggest that antisense transcription is regulated and functionally relevant to gene expression or other cellular processes. If antisense transcription has functional relevance, we expect conservation of the genome-wide pattern of antisense transcription within and between species of yeast. Recently, more than 1,000 polyadenylated (poly(A)$^+$) antisense transcripts have been identified in *S. cerevisiae*, of which six were shown to be conserved in at least one additional species of yeast (Yassour et al. 2010). Of these conserved antisense transcripts, five were additionally shown to exhibit a conserved regulatory effect (anticorrelation in expression levels of the sense and antisense transcripts between mid-log and early stationary growth phases). In addition, 648 poly(A)$^+$ antisense transcripts have been identified in *Schizosaccharomyces pombe*, of which 51% are conserved across two or more other genomes of fission yeasts (Rhind et al. 2011).

These findings are restricted to the evolution of poly(A)$^+$ RNA only (due to poly(A) selection prior to RNA sequencing). The whole transcriptome also contains nuclear, nonpolyadenylated (poly(A)$^-$), and short RNA transcripts. New categories of transcripts such as cryptic unstable transcripts (CUTs) reflect the complexity of RNA metabolism (Wyers et al. 2005). More than 95% of RNA polymerized by polymerase II remains nuclear (Jackson et al. 2000), and exclusively nuclear RNAs have 5-fold greater sequence complexity compared with exclusively cytosolic transcripts (Cheng et al. 2005). Estimates of the size of the poly(A)$^-$ transcriptome vary widely, but poly(A)$^-$ RNAs may account for as much as 70% of primary transcripts (Milcarek et al. 1974; Salditt-Georgieff and Darnell 1982; Cheng et al. 2005; Kapranov et al. 2007; Wu et al. 2008; Yang et al. 2011). Antisense transcripts in particular have been shown to generally lack polyadenylation in both mice and *S. pombe*, a fission yeast (Kiyosawa et al. 2005; Dutrow et al. 2008). Although *S. pombe* is very distantly related to budding yeast, this finding suggests that whole transcriptome approaches may reveal previously uncharacterized antisense transcription in *S. cerevisiae*. In addition, several functional categories of small RNAs (<200 nucleotides) have been characterized (Ghildiyal and Zamore 2009).

By sequencing to nearly 10 times greater depth than previous studies and refraining from size and poly(A) selection, we show that genome-wide antisense transcription is more pervasive in yeast than previously demonstrated and expand the catalog of antisense transcripts to include short and poly(A)$^-$ transcripts. Several histone modifications are also enriched or depleted in the presence of antisense transcription, reflecting increased transcriptional activity and indicating the regulation of antisense transcript biogenesis. We find that antisense transcription—both the expression level and localization within the gene—is evolutionarily conserved within and between two species of budding yeast, *S. cerevisiae* and *S. paradoxus*. Regions of genes exhibiting conserved antisense transcription show less sequence divergence than regions of genes without conserved antisense transcription. These results suggest that antisense transcription is biologically functional.

Materials and Methods

Sample Selection and RNA Collection

To characterize and study the evolution of antisense transcription, we sequenced strand-specific total RNA fragment libraries for four strains of yeast, two natural strains from each of two species, *S. cerevisiae*, and *S. paradoxus*. Strains were selected for diversity in natural environments and haplotype (table 1) (Kuehne and Sniegowski 2005). Yeast was grown in YPD media at 30°C and harvested for RNA extraction during the exponential growth phase (as determined by optical density). RNA was extracted using the TRIzol protocol (Invitrogen Part 15596026), and ribosomal RNA was depleted using the RiboMinus Eukaryote Kit for RNA-seq (Invitrogen Part 100004590).

RNA Sequencing

Eight sequencing libraries were prepared: two biological replicates for each of the four strains of yeast. The SOLiD RNA-seq library construction protocol was used (Life Technologies Part 4452437), with the following modification: The size-selection step was extended to include sequencing templates containing RNA fragments as small as 50 nucleotides. The SOLiD library preparation protocol retains strand specificity by ligating directional adapters to RNA molecules prior to reverse transcription. This prevents spurious second-strand synthesis during reverse transcription from creating artifactual antisense transcripts (Levin et al. 2010). Having intrinsic strand information is preferable to determine a transcript's strand of origin computationally using poly(A) signals or splice-site orientation, because these signals are nonexistent for many yeast genes and for poly(A)$^-$ transcripts. Sequencing was done on the SOLiD 3/3 Plus platform. Four of the libraries were also sequenced in technical replicates for the purpose of methods validation. Sequencing data are available

Table 1. Yeast Strains.

	YPS-2079	YPS-2073	YPS-3434	YPS-3395
Species	*S. cerevisiae*	*S. cerevisiae*	*S. paradoxus*	*S. paradoxus*
Haplotype	B	C	Unknown	E

NOTE.—Four strains of yeast, two from each of the species *Saccharomyces cerevisiae* and *S. paradoxus*, were selected for diversity in natural environments and haplotype.

in the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra) experiment SRP015884.

Alignment of Reads to Reference Genomes

Sequencing reads were mapped to the appropriate species using the BioScope analysis suite (Life Technologies). Reads from *S. cerevisiae* strains YPS-2073 and YPS-2079 were mapped to a strain-specific reference genome, minimizing the effect of sequence divergence on read mapping (Simola D, unpublished data). Both *S. paradoxus* read sets were mapped to the *S. paradoxus* genome reported in Liti et al. (2009). We aligned the *S. paradoxus* and *S. cerevisiae* genomes using LastZ (Harris 2007), and used homology to establish annotations [because Kellis et al. (2003) limited *S. paradoxus* annotations to sequences available in all three species *S. paradoxus*, *S. mikatae*, and *S. bayanus*, and annotations from Vishnoi et al. (2011) were not available at the time of analysis]. Genes were considered homologous if the reading frame was preserved, and if they had a sequence similarity greater than 70%. Therefore, all analysis was done with reference to *S. cerevisiae* annotations (Cherry et al. 1997) [untranslated regions' (UTR) annotations were based on the results of Nagalakshmi et al. (2008)]. Unless otherwise specified, the datasets from biological replicates were pooled, increasing the net sequencing depth of the experiment for each strain.

Calculation of Coverage

For RNA-sequencing, average sequencing depth for a given gene is used as a proxy for gene expression. Because we determined *S. paradoxus* annotations using homology, the number of reads that mapped to each base was calculated (the "pileup"), and all pileups were translated into *S. cerevisiae* coordinates. We then calculated the BPKM (bases per kilobase of gene model per million mapped bases) for each annotated gene, as per Yang et al. (2011) (BPKM for each gene in supplementary tables S1, S2 (normalized) and S3 (unnormalized), Supplementary Material online). Because the *S. cerevisiae* strain YPS-2079 was sequenced to the greatest depth, all specific, noncomparative numbers are in reference to this strain. In our data, biological replicates exhibit a strong genome-wide correlation in both sense and antisense transcription expression levels between biological replicates (table 2). The correlation between biological replicates indicates that our libraries have sufficient sequencing depth to reproducibly detect even lowly expressed antisense transcripts.

Selection of Genes for Analysis

To test our hypotheses about the regulation and function of antisense transcription, we considered three subsets of genes for analysis:

- Verified genes without introns, including UTRs when annotated ($n = 4{,}582$).
- Homologous genes without introns, including UTR when annotated and conserved ($n = 4{,}439$).

Table 2. Correlation of Expression Levels across Biological Replicates.

	YPS-2079	YPS-2073	YPS-3434	YPS-3395
Sense	**0.98**	**0.87**	**0.99**	**0.95**
Antisense	**0.93**	**0.83**	**0.92**	**0.87**

NOTE.—The genome-wide Pearson correlation coefficients for sense and antisense transcript expression levels (BPKM) across biological replicates for each of the experimental strains. All *P*-values $<10^{-20}$.

- Verified, homologous, nonoverlapping genes without introns including UTRs when conserved ($n = 523$) (referred to as the "strict" subset of genes).

For the last, most conservative, subset we defined "nonoverlapping" in such a way as to exclude from analysis antisense transcripts generated by overlapping open reading frames and UTRs. Antisense transcripts generated by these processes may be regulatory, but we created a conservative subset of genes to limit analysis to independently initiated (putatively regulated) antisense transcription. First, all genes with overlapping annotations were removed. Gene pairs that are convergently or divergently transcribed may have overlapping 3′- and 5′-UTR isoforms, respectively, that would be interpreted as antisense transcription, so we required such gene pairs to be separated by at least two standard deviations above the mean UTR length: 618 nucleotides for convergent genes, and 324 nucleotides for divergent genes. We considered gene pairs oriented head-to-tail nonoverlapping if they were separated by at least 471 nucleotides—two standard deviations above the mean 3′-UTR length plus two standard deviations above the mean 5′-UTR length. Antisense transcription in the strict subset of genes may still arise from 3′ or downstream promoter-associated nucleosome-free regions (Xu et al. 2009).

Estimation of Evolutionary Metrics

Conserved expression levels were determined by comparing the variance in expression levels within species and between species. Under no selection, variance in expression levels between lineages are assumed to increase at a constant rate over time. Selective pressure will reduce the rate of divergence, and the significance of the difference in variance was calculated using the *F*-test. We estimated the relative time of species divergence using a maximum likelihood estimate under a molecular clock assumption with the DNAMLK PHYLIP program (Felsenstein 1989), and used previous estimates of *S. paradoxus* (Kuehne et al. 2007), and *S. cerevisiae* (Simola et al. 2010) divergence times. (Uncorrected *P*-values for sense and antisense expression level divergence of genes with homologs in supplementary table S4, Supplementary Material online.)

We estimated the relative rate of nucleotide substitutions in loci exhibiting antisense transcription with the Jukes–Cantor model of nucleotide substitution and gamma distributed rate variation implemented by a custom program. This three-parameter model (two for the gamma prior, one for the rate differential) allows us to estimate the relative rate of evolution of subgenic regions. We defined regions of antisense

transcription as conserved if at least one sequencing read aligned to the homologous region in all four strains. Regions without antisense transcription were defined by a lack of detectable antisense transcription in any of the four strains.

Comparison of Genome-Wide Antisense Transcription and Histone Modification Patterns

We compared the presence of antisense transcription with a genome-wide map of histone modifications using a chi-squared test on the contingency table generated from CHiP-chip data (Pokholok et al. 2005). The 44,290 60-nucleotide tags interrogated in the original study were filtered for location so that only tags in ORFs or overlapping transcript annotations were used, then presence or absence of antisense transcription was determined for each of these tags. Presence or absence of each modification was calculated using the relative occupancy data. Significance was determined using Pearson's chi-squared test with Yates' continuity correction.

Results

Strand-Specific Sequencing Reveals Genome-Wide Antisense Transcription

From whole-transcriptome, strand-specific next generation RNA sequencing, we obtained 50–88 million 50 base-pair sequencing reads per sequencing library, allowing us to detect even low-abundance transcripts. Sequencing reads were mapped to the reference genome of the appropriate species, with 60–65% of total reads able to be aligned. Figure 1 shows the fraction of sequencing reads that aligned to different classes of annotated genomic features. Less than 1% of sequencing reads mapped to unannotated regions. These likely represent the detection of alternative UTR isoforms and unannotated intergenic transcription, including previously undetected poly-$(A)^-$ transcripts.

Although only a small percent of sequencing reads map to unannotated regions of the genome, we detect transcriptional complexity that significantly exceeds current annotations. Seventy-five percent of the *S. cerevisiae* genome sequence is annotated as transcribed from at least one strand of DNA, and 4.3% is annotated as transcribed from both strands (regions expected to exhibit antisense transcription). We detect 84% of the genome as being transcribed from at least one strand of DNA, and 20% as being transcribed from both strands. Therefore, approximately 16% of the genome exhibits unannotated antisense transcription. This result further supports previous observations of genome-wide transcription in *S. cerevisiae* (David et al. 2006).

As expected, sense transcripts were more numerous and abundant than antisense transcripts. Table 3 contains counts of verified, annotated, protein-coding genes with detectable sense and antisense transcription, as well as the median sense and antisense coverage. Figure 2 shows the distribution of sense and antisense coverage. In a previous genome-wide study of poly$(A)^+$ antisense transcripts, 18% of *S. cerevisiae* genes were reported as exhibiting antisense transcription (Yassour et al. 2010). In our data, nearly 90% of all verified,

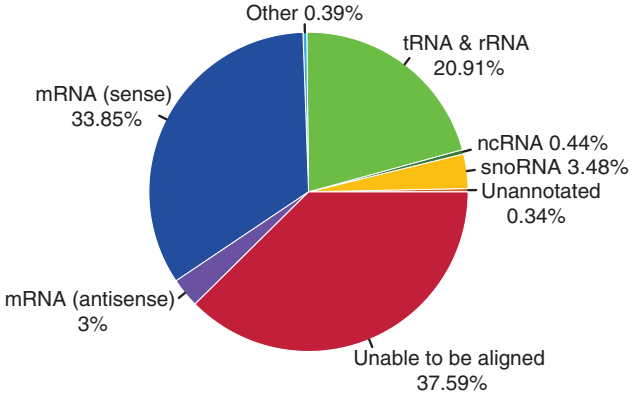

Fig. 1. Percent of sequencing reads mapping to classes of genome annotations. The percents of sequencing reads that align to different classes of annotated genomic features, including messenger RNAs (mRNA), ribosomal and transfer RNAs (rRNA and tRNA), small nucleolar RNAs (snoRNA), noncoding RNAs (ncRNA), other classes of RNAs, and unannotated regions of the genome. Sequencing reads aligned to mRNA in sense and antisense orientations are represented separately. The less than 1% of sequencing reads mapping to unannotated regions of the genome likely represents the detection of both alternative UTR isoforms and previously undetected poly$(A)^-$ transcripts.

annotated genes exhibit some detectable antisense transcription, and over 60% exhibit antisense transcription at a level of 1 BPKM or higher. Because antisense transcripts are generally expressed at a lower level than sense transcripts, this difference may reflect the depth of our sequencing experiment (Yassour et al. 2010 sequenced less than 10 million reads, approximately 10% the number of reads as our deepest sequencing experiment). This difference also likely reflects the inclusion of small and poly$(A)^-$ RNAs in our whole transcriptome study, which suggests that the majority of *cis*-NATs may be poly$(A)^-$ in *S. cerevisiae* and *S. paradoxus*.

Antisense transcription generally covers only a portion of the gene model, in contrast to sense transcription, which was detected evenly over the length of protein coding genes. The aggregate pattern of antisense transcription with respect to the canonical gene model shows that most antisense transcription occurs from the 3′-ends of genes (fig. 3). Figure 4 shows the distribution of the proportion of the gene model with detectable sense and antisense transcription. The majority of sense transcripts cover more than 90% of the annotation. By comparison, most genes have antisense transcription from less than 10% of the annotation. These results indicate that the majority of genome-wide antisense transcription generates RNAs relatively smaller than the sense transcripts and complementary to the 3′ ends. The patterns of antisense transcription we observed are in agreement with those observed by Yassour et al. (2010) (although they are not directly comparable, as that study sequenced S288C, and we sequenced wild strains of yeast, the percent of gene annotation with detectable transcription in that study is correlated with our data, $r^2 = 0.86$ for sense transcription and $r^2 = 0.62$ for antisense transcription; see supplementary table S5, Supplementary Material online).

Table 3. Numbers of Genes with Detectable Transcription.

	>1 Read	>1 BKPM	Median Coverage
Sense	4,515 (93.1%)	4,485 (92.5%)	226
Antisense	4,321 (89.1%)	2,944 (60.7%)	2.0

NOTE.—Of 4,848 verified, annotated, protein-coding genes, the numbers of genes with at least one read in the sense and antisense orientation, and the numbers of genes with sense and antisense read density greater than 1 BPKM are shown, as well as the median read density in BPKM.

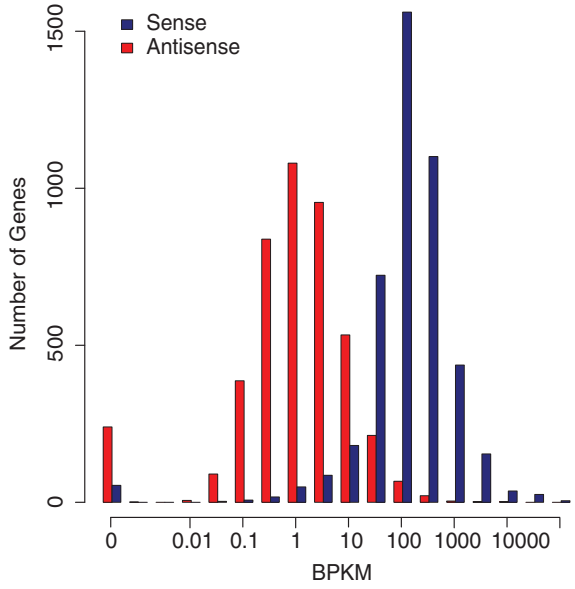

FIG. 2. Distribution of sense and antisense transcription levels. The histogram of the numbers of genes with sense and antisense transcription at logarithmic levels of coverage (BPKM). More than 93% of genes have detectable transcription in the sense orientation, whereas ~90% of genes have detectable transcription in the antisense orientation. The distribution of sense transcription levels is right-shifted when compared with the distribution of antisense transcription levels, indicating that sense transcripts are generally more abundant than antisense transcripts.

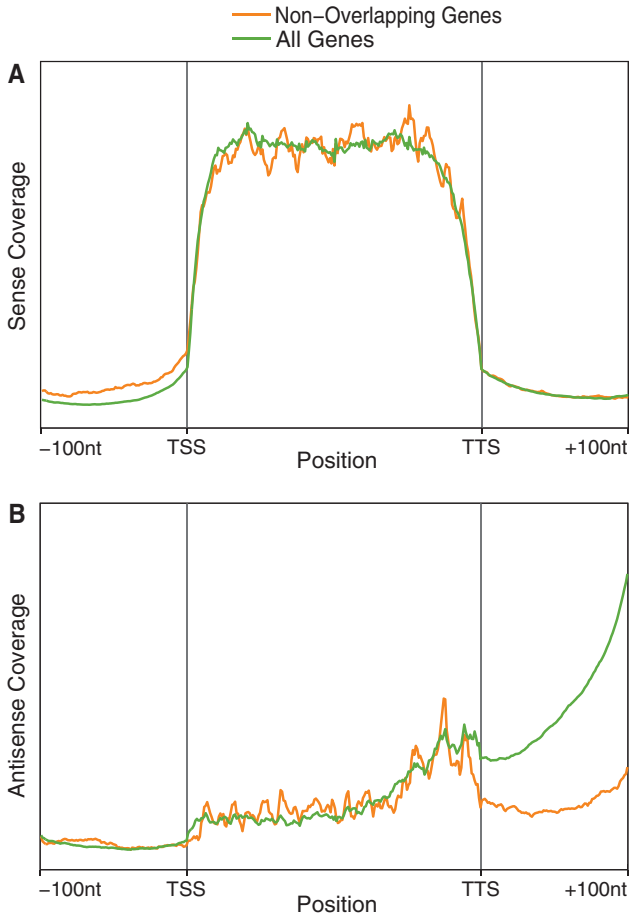

FIG. 3. Normalized density of sense and antisense coverage over a canonical gene model. Normalized density of sense (*A*) and antisense (*B*) coverage is plotted for all genes, as well as for the strict subset of nonoverlapping genes (see Materials and Methods). Mean sense and antisense coverages are normalized across all genes prior to density calculation. Density between the annotated transcription start site (TSS) and TTS is normalized over a canonical gene model that includes the open reading frame and UTRs. Coverage of 100 nucleotides upstream (–) and downstream (+) of the TSS and TTS, respectively, is represented directly (e.g., coverage at –1 nt represents the normalized density of coverage at the first nucleotide upstream of the TSS for all genes). Sense coverage is distributed evenly over the length of gene annotations, with coverage falling off near the TSS and TTS (an artifact of RNA fragmentation [Wang et al. 2009]). Antisense coverage was detected primarily at the 3′-end of gene annotations.

Although the localization of antisense transcription exhibits a strong aggregate pattern with respect to the sense transcript annotation, the levels of antisense transcription appear to be mostly independent of transcriptional activity on the sense strand of the same locus and of transcription from nearby genes. We find sense and antisense transcription levels are slightly negatively correlated ($r = -0.15$, $P < 10^{-20}$, fig. 5), consistent with previous observations (Chen et al. 2005; Werner et al. 2007; Xu et al. 2011). This strongly suggests that antisense transcription is not merely an artifact of sense transcription. Our data also suggest that most antisense transcription does not result from transcriptional "run-through" from long UTRs or nucleosome-free regions, as the levels of antisense transcription are only weakly correlated with the expression levels of nearby genes. The highest correlation occurs with genes in a convergent (tail-to-tail) orientation ($r = 0.12$, $P < 10^{-4}$), and may be partly explained by misannotated or variable 3′-UTR isoforms (Yoon and Brem 2010). To ensure that our genome-wide evolutionary comparative analysis is not affected by run-through transcription, we validated our findings on a strict subset of nonoverlapping genes (referred to as the "strict" subset, see Materials and Methods).

Because of the size of the dataset and gene-level heterogeneity in antisense transcription patterns, we selected three genomic regions for which to display genome browser tracks as representative examples of transcriptional phenomenon (fig. 6). Figure 6*A* shows the strand-specificity of the RNA sequencing method, as most reads align to the genome in the expected orientation, and within annotated regions of the genome. Note that even in this example, a few sequencing reads are aligned in the antisense orientation, which is indicative of genome-wide, low-level antisense transcription. Figure 6*B* shows an example of antisense transcription that

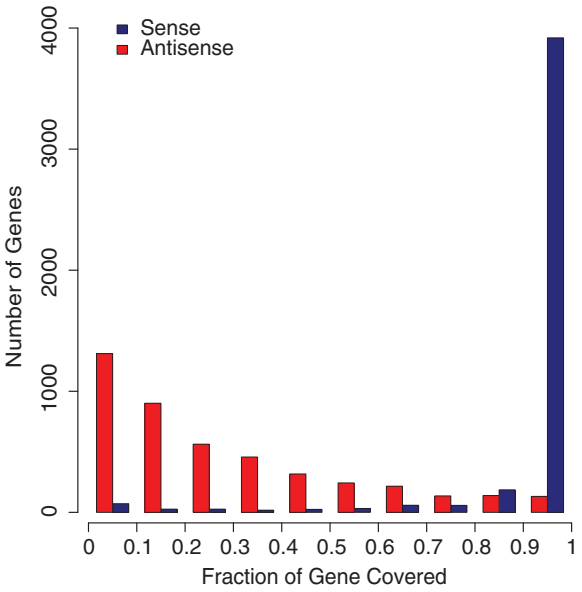

Fig. 4. Distribution of fraction of gene model covered by sense and antisense transcription. The histogram of the proportion of gene annotations with detectable sense and antisense transcription. As expected, the majority of sense transcription covers more than 90% of the annotation. For antisense transcription, the most common class is genes with less than 10% of the annotation covered. The frequency of observation decreases as the percent of the annotation expressing antisense transcription increases (except for genes with antisense transcription from 80% to 90% of the annotation, which are more frequent than genes with antisense transcription from 70% to 80% of the annotation).

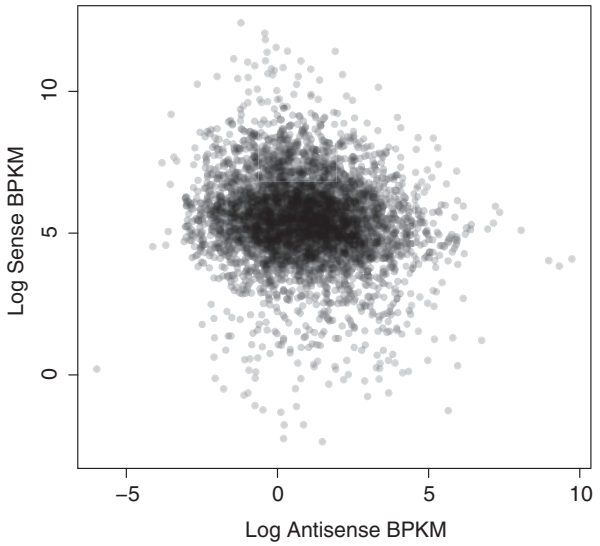

Fig. 5. Comparison of sense and antisense transcription levels. The logarithmic levels of sense and antisense coverage (BPKM) are shown for all verified, annotated, protein-coding genes with nonzero coverage in both orientations. Sense and antisense transcription levels are slightly negatively correlated, $r = -0.15$, $P < 10^{-20}$. In general, the scatter appears to be nearly random.

appears most abundant at the 3′-end, but covers at least half of the gene. In contrast, figure 6C shows an example of highly localized antisense transcription. These figures reflect the diversity of antisense transcription patterns we observe in the data.

The Pattern of Histone Modifications in Regions with Antisense Transcription

Histone modifications result from transcription, as methyl- and aceytl-transferases are recruited during transcription initiation and elongation. They also regulate transcription by causing chromatin state transitions and binding or blocking other transcription regulators. By comparing regions of genes with antisense transcription to regions without antisense transcription, we determined the relative enrichment or depletion of several histone modifications in regions of antisense transcription (see Materials and Methods). In general, the patterns associated with antisense transcription are consistent with active transcription (Pokholok et al. 2005); specifically, enrichment of histone 3 lysine 4 methylation (H3K4me1, $P < 10^{-7}$; H3K4me2, $P < 10^{-59}$; and H3K4me3, $P < 10^{-14}$) and acetylation of histone 3 lysines 9 and 14 (H3K9ac, $P < 10^{-128}$ and H3K14ac, $P < 10^{-44}$) and histone 4 lysines 5, 8, 12, and 16 (H4K5acK8acK12acK16ac, $P < 10^{-88}$). Because these enrichments are against a background of genic regions without antisense transcription, where sense transcription can be presumed to occur, the enrichment may reflect the additional transcriptional activity of antisense transcription.

Interestingly, several histone modifications associated with regions of active transcription were significantly depleted in regions of antisense transcription. Histone 3 lysine 36 trimethylation (H3K36me3), a mark generally associated with regions of active transcription (Pokholok et al. 2005), showed a 10% depletion ($P < 10^{-41}$). H3K36me3 recruits the histone H3/4 de-acetylase complex Rpd3S, which stabilizes the chromatin and thereby reduces the accessibility of the DNA to transcription (Carrozza et al. 2005; Joshi and Struhl 2005). Rpd3S-mediated histone de-acetylation is thought to occur to remove transcription elongation-associated acetylation to suppress spurious intragenic transcription (Kaplan et al. 2003; Carrozza et al. 2005). In addition, histone 3 lysine 79 trimethylation (H3K79me3) is depleted in regions of antisense transcription ($P < 10^{-291}$). This mark is weakly correlated to gene expression, but is strongly anti-correlated to the relative (transcription rate-independent, replication-independent) exchange rate of histone H3 (Gat-Viks and Vingron 2009). Depletion of H3K79m3 may reflect an increased rate of H3 exchange, reducing the level or persistence of H3 acetylation.

Antisense Transcription Is Widely Conserved Between *S. cerevisiae* and *S. paradoxus*

In our data, expression levels of sense and antisense transcripts, respectively, were closely correlated across strains and species (table 4). We tested the hypothesis that the expression levels of individual genes were conserved between the two species using an *F*-test of within and between species variance (scaled by an estimate of relative divergence time of the two species, see Materials and Methods). Levels of sense transcription were found to be conserved across species for 91.0% (strict subset 91.5%) of genes, consistent with previous estimates (Simola et al. 2010). Levels of antisense transcription

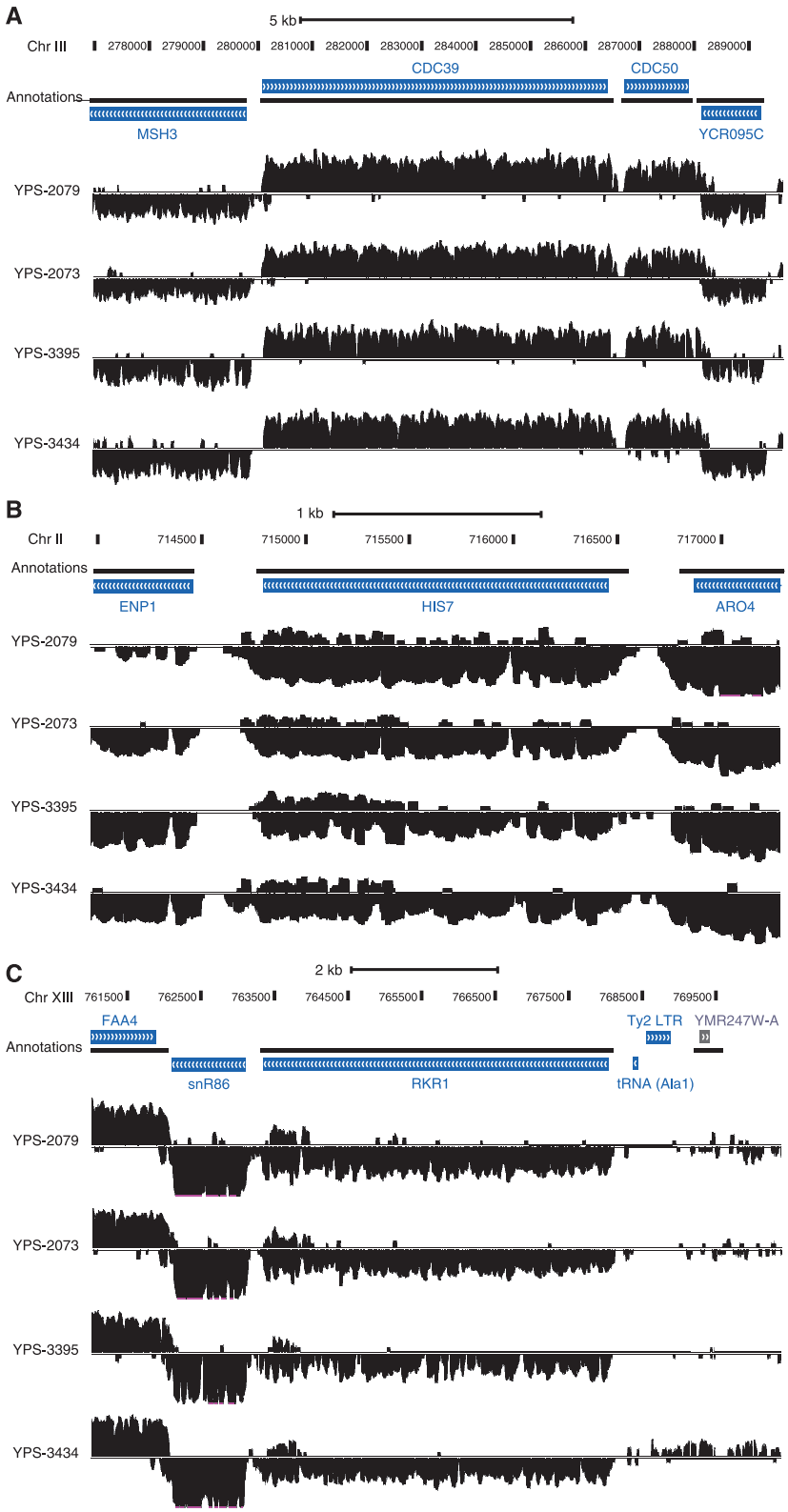

Fig. 6. Stand-specific RNA sequencing detects transcription antisense to protein-coding genes. For each strain of yeast, sequencing reads aligned to the Watson and Crick strands are located above and below, respectively (as are gene annotations). Coverage is shown on a log scale with a cutoff at approximately 3×10^3 sequencing reads. The blue boxes in the "Annotations" section represent open reading frames while the black lines represent mRNA annotations including UTRs. (*A*) In this figure, most sequencing reads map in the expected orientation, but some extremely low-level antisense transcription is observed. Most of the antisense transcription at CDC39 is characterized by the detection of small clusters of clonally-amplified reads, although longer regions of antisense transcription are detected at the 5′-end. Note that CDC50 and YCR095C both have antisense transcription that appears to originate from overlapping 3′-UTRs. (*B*) Antisense transcription is observed at the *HIS7* locus. In YPS-2079, most of the gene annotation is covered by antisense transcription, whereas in the other three strains antisense transcription is generally localized to the 3′ half of the gene annotation. (*C*) Antisense transcription at *RKR1* is highly localized at the 3′-end of the transcript annotation. The gray box represents an unverified ORF.

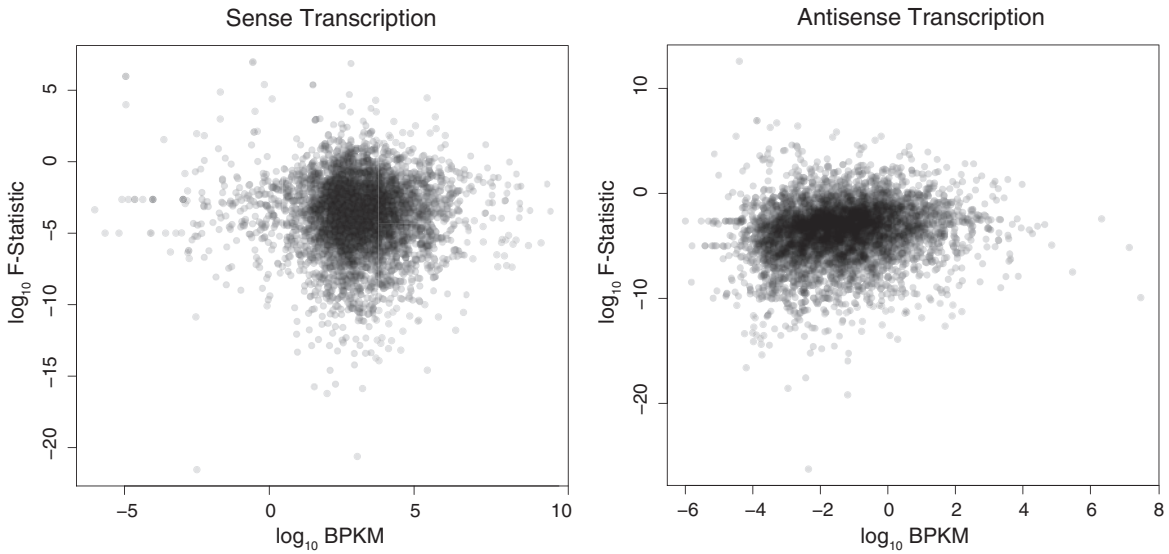

Fig. 7. Conservation of transcription is independent of levels of transcription. We saw only weak correlation between the conservation of level of transcription (F-statistic of within and between species variance in transcript abundance) and the level of transcription itself. This suggests that transcription is conserved for both highly and lowly abundant transcripts. Correlation for sense transcription $R^2 = -0.03$, $P < 0.05$ and antisense $R^2 = 0.13$, $P < 10^{-16}$.

Table 4. Correlation of Expression Levels across Strains and Species.

	YPS-2079	YPS-2073	YPS-3434	YPS-3395
YPS-2079	**1**	**0.94**	**0.87**	**0.86**
YPS-2073	**0.88 (0.83)**	**1**	**0.88**	**0.89**
YPS-3434	**0.69 (0.63)**	**0.70 (0.62)**	**1**	**0.92**
YPS-3395	**0.69 (0.63)**	**0.71 (0.65)**	**0.77 (0.69)**	**1**

Note.—The genome-wide Pearson correlation coefficients for transcript expression levels (BPKM) within and between *Saccharomyces cerevisiae* and *S. paradoxus*. Sense correlations are in the upper right, antisense correlations are in the lower left. Correlations for the strict subset of nonoverlapping genes are presented in parentheses. All P values $< 10^{-645}$ (strict subset $< 10^{-89}$).

were conserved for 86.1% (strict subset 85.5%) of genes. Moreover, conservation of the level of antisense transcription is only weakly correlated with the level of antisense transcription itself ($R^2 = 0.13$, $P < 10^{-16}$; $R^2 = -0.03$, $P < 0.05$ for sense transcription; fig. 7). This suggests that the level of antisense transcription is conserved for both highly and lowly abundant transcripts. Thus, similar to sense transcription levels, antisense transcription in these two species is conserved and potentially under stabilizing selection, indicating functional importance for either the rate of antisense transcription or abundance of the antisense transcript.

As observed earlier, for most genes we detect only a small portion of the gene structure transcribed in the antisense orientation—in other words, antisense transcription is localized within the gene. Of all genic regions that exhibit antisense transcription (more than one nucleotide of contiguous antisense coverage) in at least one strain, 8% overlap with regions of antisense transcription in all four strains. Moreover, 22% of genes exhibit at least one homologous region of antisense transcription in all strains. In these genes, we observe a wide range of conservation (fig. 8). This indicates that not only are antisense transcripts expressed at conserved levels, but also antisense transcription often occurs in homologous regions of genes. For regions with antisense transcription in only two of the four strains, 57% respect the species boundary, that is, are present in both strains of either *S. cerevisiae* or *S. paradoxus*, but absent in the other species. This binomial proportion is significantly higher than the expected 33% if colocalization was random ($P < 10^{-235}$). Furthermore, genes with detectable conserved localization of antisense transcription have median sense and antisense transcription levels of 164 and 11.39 BPKM, respectively—a substantially higher ratio of antisense to sense transcription levels than for the set of all genes (table 3). This suggests that our ability to detect conserved localization of antisense transcription is limited to some extent by the level of antisense transcription, and that conserved localization may be more widespread than we can detect.

Several gene ontology categories were significantly enriched for conserved antisense transcription (Benjamini-Hochberg correction, $P < 0.05$) (Huang et al. 2009a,b). Interestingly, the enriched gene ontology categories are different for sets of genes with antisense transcription evolutionarily conserved either by position or level. The 973 genes with antisense transcription from at least one homologous nucleotide in all four strains are significantly enriched for transferase/methyltransferase activity, chromosomal parts (which includes the genome and all associated proteins), DNA repair/DNA damage, and the enzyme-related terms: hydrolase, metal-binding, and NADP. The 969 genes with significantly conserved, nonzero antisense transcription levels (F-statistic with $P < 0.05$ and nonzero antisense BPKM in YPS-2079) are significantly enriched for phosphoproteins and the terms: nucleus, transport, protein transport, transit peptide, organelle lumen, transcription, ligase, mitochondrion, endoplasmic reticulum, NAD, transcription

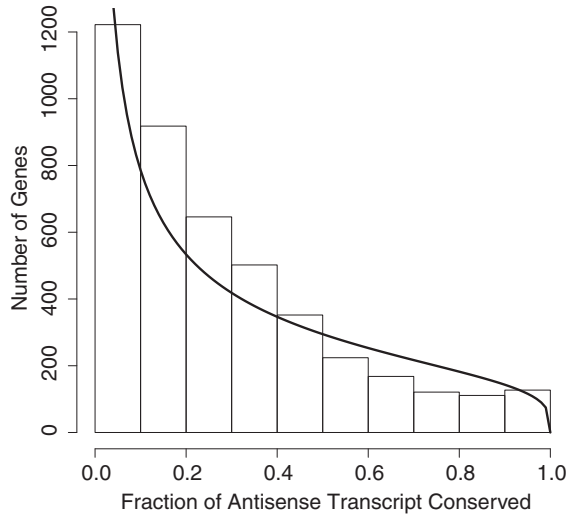

Fig. 8. Fraction of antisense transcript conserved. We observe a wide range of conservation of antisense transcripts. Most genes have a portion of the antisense transcript conserved between *Saccharomyces cerevisiae* and *S. paradoxus*. The frequency of observation decreases as the percent of conserved antisense transcription increases, except for genes with 90–100% conservation, which are more frequent than genes with 80–90% of their antisense transcription conserved.

regulation, chaperone, and coiled-coil. It has been shown that genes with antisense transcription exhibit expression variability in response to genetic and environmental variability (Xu et al. 2011). Our findings further support the observation that housekeeping genes (such as those involved in core metabolic and cellular processes) may not benefit from additional regulation from antisense transcription, and are thus not enriched for conserved antisense transcription.

Antisense Transcripts Are Subject to Sequence-Evolution Constraints

If the localization of antisense transcription is conserved because these regions contain functional motifs, we expect to observe decreased divergence between *S. cerevisiae* and *S. paradoxus* in antisense transcript sequences. To test this hypothesis, we compared the rates of evolution of regions of a gene with and without antisense transcription. On average, regions of antisense transcription conserved across all four strains evolve at 87% the rate of regions of the same genes without antisense transcription (coding sequences only, $P < 10^{-23}$; for the strict subset 82% the rate, $P < 10^{-13}$). In the whole transcript, including UTRs, regions with antisense transcription evolve at 88% the rate of regions of the same genes without antisense transcription ($P < 10^{-49}$; for the strict subset 83% the rate, $P < 10^{-7}$). This result suggests that the sequence content of antisense transcripts is relevant to their function, and therefore is subject to negative selection.

We hypothesize that the sequence content of antisense transcripts may be evolutionarily conserved because they contain sequence or structural motifs, or because they contain regions of complementarity to *trans*-NATs (RNAs transcribed from a separate genomic locus). However, we did not detect any significant sequence motifs in regions of conserved antisense transcription (Hiller et al. 2006). When regions of conserved antisense transcription were compared to the genome using BLAST, we identified one significant hit, but the result was between paralogous genes (Altschul et al. 1997). The lack of significant targeting sequence patterns suggest that antisense transcripts with conserved localization do not target other endogenous RNAs, but may still be functional either through structural motifs or transcriptional interference.

Discussion

The excess of noncanonical transcription detected by increasingly sensitive and high-throughput technologies represents a tremendous reservoir of hidden information in biological systems. Using massively parallel RNA sequencing, we detect antisense transcription from over 4,000 protein-coding genes. Rather than investigating antisense transcription on a gene-by-gene basis, we use an evolutionary comparative genomics approach to hypothesize that genome-wide antisense transcription is regulated, and serves a functional role subject to stabilizing selection. Specifically, antisense transcription is evolutionarily conserved in expression levels and localization relative to the gene; and, regions of genes with conserved antisense transcription exhibit less sequence divergence than regions without antisense transcription. Because the majority of genomic antisense transcription does not result from unannotated gene overlaps or run-through transcription from nearby genes (and we have controlled for these factors by using a strict subset of nonoverlapping genes to test evolutionary hypotheses), we hypothesize that the evolutionary conservation we observe is associated with regulated antisense transcription events.

We detect pervasive, genome-wide antisense transcription: 20% of the genome exhibited transcription from both strands, of which more than 15% cannot be explained by current annotations. Approximately 90% of all verified, protein-coding genes had some detectable level of antisense transcription (91.9% of the strict subset). Although antisense transcripts are extremely varied in abundance, length, and location relative to the sense gene, the majority of antisense transcription occurs at low levels (relative to sense transcription) on the 3′-end of genes and likely generates short, poly(A)$^-$ transcripts. Moreover, our data reveal considerable unannotated low-level transcriptional activity in yeast. Although less than 1% of sequencing reads map to unannotated regions, these reads cover more than 9% of the genome. These results will be useful for improving annotations of *S. cerevisiae* and *S. paradoxus* genomes.

Little is known about the regulation and biogenesis pathways of antisense transcripts. It has been suggested that antisense transcription results from noisy regulation of gene expression; the level and localization of transcriptional noise may be a quality intrinsic to genes as a result of their sequence content (Struhl 2007). Our data suggest that the levels of antisense transcription are independent of the levels of sense and nearby transcription. The aggregate localization

pattern of antisense transcription is greater upstream of the transcription termination site (TTS) than immediately following the TTS, and this pattern becomes more pronounced when potentially overlapping genes are excluded (fig. 3). At the gene level, however, patterns of antisense transcription are inhomogeneous, suggesting that antisense transcription results from potentially diverse mechanisms, including intragenic initiation. This is consistent with a previous finding that long, poly(A)$^+$ antisense transcripts originate primarily from 3′ nucleosome-free regions (~45%) or have biogenesis of unknown origin (~35%). (Yassour et al. 2010).

If antisense transcription results from the chromatin configurations and presence of transcriptional machinery associated with sense transcription (Struhl 2007; Beiter et al. 2009), we expect sense and antisense transcript expression levels to be positively correlated. However, we observe a weak negative correlation between sense and antisense transcript expression levels. This suggests that sense and antisense transcription are independent in regulation but perhaps weakly coupled for polymerase availability. That is, sense and antisense transcription may merely compete for transcriptional machinery, access to regulatory sequences, or clearance for unperturbed transcription elongation (Prescott and Proudfoot 2002; Shearwin et al. 2005). In the most extreme case, the processes or products of sense and antisense transcription may negatively regulate each other (Lapidot and Pilpel 2006), but we do not see evidence for such negative regulation in our data. In certain cases, antisense transcripts may be constitutively transcribed but only circumstantially functional: In the case of *PHO84*, antisense transcription does not suppress sense transcription until the loss of Rrp6 stabilizes the antisense transcript (Camblong et al. 2007).

We find that antisense transcription is correlated with histone modifications indicative of an "open" chromatin milieu. The combined underrepresentation of H3K36me3 and enrichment of H3 and H4 acetylation suggests that some antisense transcription may occur when a lack of Rpd3S-mediated histone de-acetylation allows for transcription from 3′ nucleosome-free regions and cryptic promoters. However, these findings are still consistent with a functional role for antisense transcription. For example, Churchman and Weissman (2011) show that knocking out an essential Rpd3S component increases the level, but does not alter the localization of antisense transcription. This finding, along with our evidence for evolutionary conservation, suggests that antisense transcription is a regulated aspect of RNA metabolism.

The majority of genes exhibit conservation of antisense transcription levels exceeding what can be explained by transcriptional noise alone. Conservation of the antisense transcription levels may result from stabilizing selection on either the abundance of the transcript or the rate of antisense transcription (assuming antisense transcripts are degraded at a constant rate). The set of genes with significantly conserved, nonzero antisense transcription levels is enriched for various gene ontology terms related to intracellular transport, cellular compartments, and protein attributes such as ligase and chaperone. For nearly a quarter of genes, we can detect antisense transcription from homologous regions of genes in all four strains, suggesting the localization of antisense transcription with respect to sense transcription is evolutionarily conserved. However, our ability to detect conserved localization appears to be limited in part by the level of antisense transcription and the sensitivity of RNA sequencing. Our study utilized fragment sequencing libraries, so we did not attempt to identify antisense transcription units—a task that is better suited to paired-end sequencing approaches unavailable at the time of sequencing (Yassour et al. 2010). Therefore, conserved localization may be far more widespread than we can detect. Conserved localization of antisense transcription is expected if antisense transcription events originate from regulatory sequences such as promoters, or if they have a functional role mediated by interactions with proteins (because proteins must recognize specific RNA sequences and structures).

In addition, conserved localization of antisense transcription corresponds to suppressed sequence-level divergence within the conserved region of antisense transcription. This result is highly significant, although the difference in the rate of evolution is smaller than what is expected for overlapping protein-coding genes (Miyata and Yasunaga 1978). This likely reflects the fact that antisense transcripts, being noncoding, are subject to different evolutionary constraints than coding genes. The lack of significant motifs in regions of conserved antisense transcription may reflect the fact that we are unable to detect entire antisense transcription units. Alternatively, functional sequence motifs present in antisense transcripts may be either multitudinous or otherwise not amenable to computational detection, or antisense transcripts may contain functional structural motifs.

Transcription is a critical process in signal transduction—once the transcript is produced its signal, or biological effect, may be rapidly amplified through translation. Antisense transcription may regulate transcription initiation, elongation, or termination on the sense strand (Prescott and Proudfoot 2002; Shearwin et al. 2005). Alternatively, antisense transcription may "activate" genomic DNA sequences by attracting regulatory machinery, such as methyltransferases, to conserved functional motifs (Gelfand et al. 2011). In these cases, sense transcription may be dynamically regulated by the interactions of polymerases or modification and spatial restructuring of DNA, providing additional regulatory layers between canonical regulation of transcription initiation and posttranscriptional regulation of the mRNA. Alternatively, the conserved levels and localization of antisense transcription may reflect regulatory models where the antisense transcript functions as a targeting mechanism or platform for the assembly of regulatory machinery. In these models, reduced sequence divergence in regions of conserved antisense transcription reflects functional motifs or regulatory sequences important to *cis*-NAT biogenesis or processing. These hypotheses are bolstered by the fact that genes with detectable conserved localization of antisense transcription are enriched for gene ontology terms related to DNA modification such as transferase, DNA repair, and chromosomal parts (which includes all genome-associated proteins).

In the context of these results, it seems likely that pervasive genome-wide antisense transcription encompasses many categories of transcripts generated by diverse transcription processes, both regulated and unregulated. Within each of these categories, there may exist subsets of nonfunctional antisense transcripts as well as antisense transcription events that mediate regulation through distinct mechanisms, from transcriptional interference to directing epigenetic modifications. Moreover, the effect size of these regulatory events may be small, as is the case for other regulators. For example, microRNAs can potentially down-regulate hundreds of mRNAs, sometimes subtly (Lim et al. 2005), and may act primarily as buffers or sensors of transcriptional network states (Herranz and Cohen 2010; Su et al. 2011). Additional experiments are necessary to ascertain the extent to, and means by, which antisense transcription adds to the functional complexity of the yeast genome.

Conclusions

We have identified and characterized genome-wide antisense transcription in two species of budding yeast, *S. cerevisiae* and *S. paradoxus*, expanding the library of putative functional antisense transcription events to include those that generate small and nonpoly(A) transcripts. Over the 5–10 million year divergence between *S. cerevisiae* and *S. paradoxus* (Kellis et al. 2003), antisense transcription is widely evolutionarily conserved. We find antisense transcripts exhibit conserved antisense expression levels and conserved localization with respect to gene annotations. Regions where antisense transcription is conserved exhibit reduced sequence divergence, although we did not identify any significant sequence motifs. These results further support the existence of widespread functional antisense transcription with diverse biogenesis pathways and regulatory mechanisms.

Supplementary Material

Supplementary tables S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

E.D. and J.K. conceived of the study. E.D., A.G., and J.K. designed the study. E.D. performed the sequencing and drafted the manuscript. A.G. performed the analysis, and helped to draft the manuscript. J.K. oversaw and coordinated the research and edited the manuscript. All the authors read and approved the final manuscript. The authors acknowledge P. Sniegowski, B. Gregory, and D. Simola for assistance, as well as PGFI and NAPCore for the use of the sequencing facilities. This work was supported, in part, by the Penn Genome Frontiers Institute (PGFI) under a grant with the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analyses, interpretations, or conclusions.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.

Aravind L, Watanabe H, Lipman DJ, Koonin EV. 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc Natl Acad Sci U S A.* 97:11319–11324.

Barrell BG, Air GM, Hutchison CA. 1976. Overlapping genes in bacteriophage phiX174. *Nature* 264:34–41.

Beiter T, Reich E, Williams RW, Simon P. 2009. Antisense transcription: a critical look in both directions. *Cell Mol Life Sci.* 66:94–112.

Bolland DJ, Wood AL, Johnston CM, Bunting SF, Morgan G, Chakalova L, Fraser PJ, Corcoran AE. 2004. Antisense intergenic transcription in V(D)J recombination. *Nat Immunol.* 5:630–637.

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

Carrozza MJ, Li B, Florens L, et al. (11 co-authors). 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123: 581–592.

Chen J, Sun M, Hurst LD, Carmichael GG, Rowley JD. 2005. Genome-wide analysis of coordinate expression and evolution of human *cis*-encoded sense-antisense transcripts. *Trends Genet.* 21:326–329.

Cheng J, Kapranov P, Drenkow J, et al. (19 co-authors). 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308:1149–1154.

Cherry JM, Ball C, Weng S, et al. (11 co-authors). 1997. Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* 387:67–73.

Churchman LS, Weissman JS. 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469: 368–373.

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis R, Steinmetz LM. 2006. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A.* 103: 5320–5325.

Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP. 2009. RNAi in budding yeast. *Science* 326:544–550.

Dutrow N, Nix DA, Holt D, Milash B, Dalley B, Westbroek E, Parnell TJ, Cairns BR. 2008. Dynamic transcriptome of *Schizosaccharomyces pombe* shown by RNA-DNA hybrid mapping. *Nat Genet.* 40: 977–986.

Faghihi MA, Wahlestedt C. 2009. Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol.* 10:637–643.

Felsenstein J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* 5:164–166.

Gat-Viks I, Vingron M. 2009. Evidence for gene-specific rather than transcription rate-dependent histone H3 exchange in yeast coding regions. *PLoS Comput Biol.* 5:e1000282.

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

Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet.* 10:94–108.

Harris R. 2007. Improved pairwise alignment of genomic DNA [PhD thesis]. Philadelphia (PA): Computer Science and Engineering Department, Pennsylvania State University.

Herranz H, Cohen SM. 2010. MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Gene Dev.* 24: 1339–1344.

Hiller M, Pudimat R, Busch A, Backofen R. 2006. Using RNA secondary structures to guide sequence motif finding towards single-stranded regions. *Nucleic Acids Res.* 34:e117.

Hongay CF, Grisafi PL, Galitski T, Fink GR. 2006. Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* 127: 735–745.

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.

Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37:1–13.

Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4:44–57.

Jackson DA, Pombo A, Iborra F. 2000. The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells. *FASEB J.* 14:242–254.

Johnson JM, Edwards S, Shoemaker D, Schadt EE. 2005. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.* 21:93–102.

Joshi AA, Struhl K. 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to pol II elongation. *Mol Cell.* 20:971–978.

Kaplan CD, Laprade L, Winston F. 2003. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301: 1096–1099.

Kapranov P, Willingham AT, Gingeras TR. 2007. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet.* 8:413–423.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423:241–254.

Kiyosawa H, Mise N, Iwase S, Hayashizaki Y, Abe K. 2005. Disclosing hidden transcripts: mouse natural sense–antisense transcripts tend to be poly (A) negative and nuclear localized. *Genome Res.* 15: 463–473.

Kuehne HA, Murphy HA, Francis CA, Sniegowski PD. 2007. Allopatric divergence, secondary contact, and genetic isolation in wild yeast populations. *Curr Biol.* 17:407–411.

Kuehne HA, Sniegowski P. 2005. The genetic structure and biogeography of natural Saccharomyces populations [PhD thesis]. Philadelphia (PA): Biology Department, University of Pennsylvania.

Lapidot M, Pilpel Y. 2006. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep.* 7:1216–1222.

Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods.* 7: 709–715.

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433:769–773.

Liti G, Carter DM, Moses AM, et al. (26 co-authors). 2009. Population genomics of domestic and wild yeasts. *Nature* 458:337–341.

Milcarek C, Price R, Penman S. 1974. The metabolism of a poly (A) minus mRNA fraction in HeLa cells. *Cell* 3:1–10.

Miyata T, Yasunaga T. 1978. Evolution of overlapping genes. *Nature* 272: 532–535.

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

Neil H, Malabat C, d'Aubenton Carafa Y, Xu Z, Steinmetz LM, Jacquier A. 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457:1038–1042.

Ni T, Tu K, Wang Z, Song S, Wu H, Xie B, Scott KC, Grewal SI, Gao Y, Zhu J. 2010. The prevalence and regulation of antisense transcripts in *Schizosaccharomyces pombe*. *PLoS One* 5:1564–1566.

Pokholok DK, Harbison CT, Levine S, et al. (14 co-authors). 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122:517–527.

Prescott EM, Proudfoot NJ. 2002. Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A.* 99: 8796–8801.

Rhind N, Chen Z, Yassour M, et al. (31 co-authors). 2011. Comparative functional genomics of the fission yeasts. *Science* 332:930–936.

Salditt-Georgieff M, Darnell JE Jr. 1982. Further evidence that the majority of primary nuclear RNA transcripts in mammalian cells do not contribute to mRNA. *Mol Cell Biol.* 2:701–707.

Shearwin KE, Callen BP, Egan JB. 2005. Transcriptional interference—a crash course. *Trends Genet.* 21:339–345.

Simola DF, Francis C, Sniegowski PD, Kim J. 2010. Heterochronic evolution reveals modular timing changes in budding yeast transcriptomes. *Genome Biol.* 11:R105.

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

Su WL, Kleinhanz RR, Schadt EE. 2011. Characterizing the role of miRNAs within gene regulatory networks using integrative genomics techniques. *Mol Syst Biol.* 7:490.

Vanhée-Brossollet C, Vaquero C. 1998. Do natural antisense transcripts make sense in eukaryotes? *Gene* 211:1–9.

Vishnoi A, Sethupathy P, Simola D, Plotkin JB, Hannenhalli S. 2011. Genome-wide survey of natural selection on functional, structural, and network properties of polymorphic sites in *Saccharomyces paradoxus*. *Mol Biol Evol.* 28:2615–2627.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 10:57–63.

Werner A, Schmutzler G, Carlile M, Miles CG, Peters H. 2007. Expression profiling of antisense transcripts on DNA arrays. *Physiol Genomics.* 28:294–300.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453:1239–1243.

Wu Q, Kim YC, Lu J, Xuan Z, Chen J, Zheng Y, Zhou T, Zhang MQ, Wu CI. 2008. Poly A-transcripts expressed in HeLa cells. *PLoS One* 3: e2803.

Wyers F, Rougemaille M, Badis G, et al. (12 co-authors). 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly (A) polymerase. *Cell* 121:725–737.

Xu Z, Wei W, Gagneur J, Clauder-Münster S, Smolik M, Huber W, Steinmetz LM. 2011. Antisense expression increases gene expression variability and locus interdependency. *Mol Syst Biol.* 7:468.

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. 2009. Bidirectional

promoters generate pervasive transcription in yeast. *Nature* 457: 1033–1037.

Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. 2011. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol.* 12:R16.

Yang PK, Kuroda MI. 2007. Noncoding RNAs and intranuclear positioning in monoallelic gene expression. *Cell* 128:777–786.

Yassour M, Pfiffner J, Levin JZ, Adiconis X, Gnirke A, Nusbaum C, Thompson DA, Friedman N, Regev A. 2010. Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biol.* 11:R87.

Yoon OK, Brem RB. 2010. Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA* 16:1256–1267.