Genomics 102 (2013) 484–490

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Role of antisense RNAs in evolution of yeast regulatory complexity

Chih-Hsu Lin [a], Zing Tsung-Yeh Tsai [b,c], Daryi Wang [a,*]

[a] Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan
[b] Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica, Taipei 115, Taiwan
[c] Institute of Information Science, Academia Sinica, Taipei 115, Taiwan

ARTICLE INFO

Article history:
Received 29 July 2013
Accepted 26 October 2013
Available online 5 November 2013

Keywords:
Antisense RNA
Molecular evolution
Gene expression regulation
Transcriptome

ABSTRACT

Antisense RNAs (asRNAs) are known to regulate gene expression. However, a genome-wide mechanism of asRNA regulation is unclear, and there is no good explanation why partial asRNAs are not functional. To explore its regulatory role, we investigated asRNAs using an evolutionary approach, as genome-wide experimental data are limited. We found that the percentage of genes coupling with asRNAs in *Saccharomyces cerevisiae* is negatively associated with regulatory complexity and evolutionary age. Nevertheless, asRNAs evolve more slowly when their sense genes are under more complex regulation. Older genes coupling with asRNAs are more likely to demonstrate inverse expression, reflecting the role of these asRNAs as repressors. Our analyses provide novel evidence, suggesting a minor contribution of asRNAs in developing regulatory complexity. Although our results support the leaky hypothesis for asRNA transcription, our evidence also suggests that partial asRNAs may have evolved as repressors. Our study deepens the understanding of asRNA regulatory evolution.

1. Introduction

Antisense RNAs (asRNAs; also known as *cis*-natural antisense transcripts) are transcribed from the opposite DNA strand of the protein-coding sequences and overlap in part with the sense RNAs [1,2]. Considering the low expression level of asRNAs, one hypothesis suggests that asRNAs may result from the leakage of RNA transcription machinery and represent transcriptional noise [3,4]. However, increasing number of evidence has shown that asRNAs participate in eukaryotic gene regulation [1–3]. Particularly, in *Saccharomyces cerevisiae*, the prevailing functional presumption is of a repressive role [5–10], while other functional roles such as activation [11] have been noted. Some asRNAs were found to repress sense RNAs and lead to inverse expression between asRNAs and their corresponding sense RNAs in respect to growth phases, stress conditions, or environmental changes [5–10]. For instance, *IME4*, a gene that mediates MAT and nutritional control of meiosis [12], can be repressed in haploid cells by its asRNA, *RME2*, which inhibits meiosis; low expression of *RME2* in diploid cells results in derepression of *IME4* and thus initiates meiosis [5,9]. However, a genome-wide mechanism of how asRNAs regulate gene expression is not available, nor is there an explanation of why partial asRNAs are not functional. Investigations into the role of asRNAs in regulatory evolution may provide clues to resolve these questions.

Gene expression is regulated by multilayered and interconnected mechanisms, such as transcription factor binding events, histone modification, RNA interference (RNAi), and post-translational modification (PTM). The combinations of these mechanisms contribute to different levels of regulatory complexity for each gene of an organism. Recent studies have shown that regulatory complexity increases through evolutionary time [13–16]. Compared to lineage-specific genes, orthologous genes develop more transcription factor binding sites (TFBSs) as well as more alternative isoforms; older genes also tend to have more physical protein–protein interactions (PPIs), and are more likely to be affected by nonsense-mediated decay and RNA editing [13–16]. In addition, genes of high regulatory complexity (estimated by the number of TFBSs and the degree of physical PPIs) have been shown to be preferentially targeted by miRNAs [17,18]. Therefore, the investigation of the association of asRNA and regulatory complexity may deepen our understanding of the role of asRNAs in gene regulation.

Because regulatory complexity cannot be directly measured, employing surrogate indicators is necessary. Physical PPIs are the foundation of biological signaling, and provide a mechanistic basis for most biological processes in an organism [19]. Therefore, the degree of PPIs has been used to demonstrate evolutionary mechanisms and regulatory complexity of organisms [18–20]. Another useful indicator is the percentage of intrinsically disordered regions (IDRs) in protein sequences. IDRs are flexible protein segments that do not fold completely [21]. This flexibility of protein structure can

Abbreviations: RNAi, RNA interference; PTM, post-translational modification; TFBS, transcription factor binding site; PPI, protein–protein interactions; IDR, intrinsically disordered region; SAL/SRL, the ratio of sense–antisense overlapping length to sense transcript length; non-coding RNA, ncRNA; ORF-T, ORF transcript; BPKM, bases per kilobase of gene model per million mapped bases.

* Corresponding author.
E-mail addresses: jack2030@gmail.com (C.-H. Lin), zing@iis.sinica.edu.tw (Z.T.-Y. Tsai), dywang@gate.sinica.edu.tw (D. Wang).

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.10.008

affect characteristics of regulatory molecules, e.g. decoupling of specificity and affinity, binding diversity, binding commonality, rates of association and dissociation, and precise control and simple regulation of binding thermodynamics [21]. Thus, the percentage of IDR is correlated well with regulatory complexity [22]. Recently, it has been shown that a protein can be modified by more than one form of PTM, and particular PTMs located on specific residues are crucial for the recognition of modular protein domains and the regulation of cellular signaling [23]. It is hence reasonable to assume that a gene that has developed multiple forms of PTM may also have high regulatory complexity. Although no experimental support is available currently, it is interesting to conduct a parallel comparison with this feature included in the evaluation of regulatory complexity. Accordingly, to estimate the regulatory complexity of a gene, we employed 1) the degree of physical PPIs and 2) the percentage of IDRs in its protein sequence as surrogates for regulatory complexity, and we further compared them with the number of PTM forms.

In this study, we explored the contribution of yeast asRNAs in developing regulatory complexity. Data about asRNAs of *S. cerevisiae* is generated by strand-specific paired-end sequencing in three conditions (mid-log phase, early stationary phase, and heat shock), followed by a series of bioinformatics analyses. We first analyzed the association between asRNA distribution and regulatory complexity, reflecting the contribution of asRNAs in developing complex regulation. To unravel the role of asRNAs in regulatory evolution, the asRNA substitution rates and asRNA expression patterns were also investigated. To further explore the fate of asRNAs in evolution, we characterized asRNAs of genes of five different evolutionary ages. Finally, based on the hypothesis that the asRNA mechanism may be RNA–RNA interaction [1], we discussed the mechanisms of asRNA action by analyzing the sense–antisense overlapping region, which is suggested to be the interacting region of sense and asRNAs [1]. Our observations elucidate the role of asRNAs in regulatory evolution and broaden our understanding of the regulatory mechanisms of asRNAs.

2. Results

2.1. Negative association between asRNA distribution and regulatory complexity

Previous studies have indicated that miRNAs preferentially target genes with high regulatory complexity [17,18]. In order to study whether asRNAs have a similar tendency, we investigated the association between asRNA distribution and regulatory complexity. In *S. cerevisiae*, asRNAs have been found to respond to different growth phases, stress conditions, or environmental changes [5–10]. Therefore, we first collected yeast cells in mid-log phase, early stationary phase, and after heat shock treatment, and annotated asRNAs by the transcriptome sequencing of *S. cerevisiae* (see Materials and methods section). Next, for each gene, we estimated the degree of physical PPIs, the percentage of IDRs, and further used the number of PTM forms as a counterpart. Interestingly, as shown in Fig. 1A, the percentage of genes coupling with asRNAs is negatively associated with the degree of physical PPIs of sense genes (Chi-square test of independence, $P = 1.9 \times 10^{-18}$). Similar negative associations of asRNA distribution with the percentage of IDRs or the number of PTM forms were also observed (Fig. 1B, one-sided two-sample proportion test, $P = 2.7 \times 10^{-6}$; Fig. 1C, Chi-square test of independence, $P = 1.2 \times 10^{-6}$). Overall, our results indicate that the negative association between asRNA distribution and the regulatory complexity of sense genes is robust. Hence, the results suggest that asRNAs may frequently appear in genes with simpler regulation, yet tend to be eliminated as genes develop complex regulation.

2.2. Slower evolution of asRNAs of genes under complex regulation

In order to understand the functional significance of asRNAs involved in various regulatory mechanisms, we investigated the rate of evolutionary changes in asRNAs by measuring the number of substitutions compared to orthologous regions in *Saccharomyces paradoxus*. The results show that *S. cerevisiae* asRNAs of genes with more physical PPIs evolve more slowly than those of genes with less PPIs (Fig. 2A, Kruskal–Wallis test, $P = 1.3 \times 10^{-4}$). Similar relationships were discovered between evolutionary rate and the frequency of IDRs or PTMs (Fig. 2B, one-sided Wilcoxon test, $P = 1.3 \times 10^{-3}$; Fig. 2C, Kruskal–Wallis test, $P = 0.031$). As a result, asRNAs of genes involved in more complex regulation tend to evolve more slowly than the others. In addition, all asRNAs evolve more slowly than the non-antisense non-coding RNAs (non-AS ncRNAs) (Fig. 2A, Kruskal–Wallis test, $P = 8.4 \times 10^{-7}$; Fig. 2B, Kruskal–Wallis test, $P = 9.8 \times 10^{-6}$; Fig. 2C, Kruskal–Wallis test, $P = 1.2 \times 10^{-3}$) and the four-fold degenerate sites, which were used as neutral references [24,25] (Fig. 2A, Kruskal–Wallis test, $P < 2.2 \times 10^{-16}$; Fig. 2B, Kruskal–Wallis test, $P < 2.2 \times 10^{-16}$; Fig. 2C, Kruskal–Wallis test, $P < 2.2 \times 10^{-16}$). Our results suggest that asRNAs from genes with more complex regulation are subject to stronger purifying selection.

2.3. Number of asRNAs reducing in yeast genome during evolution

Previous studies have shown that miRNAs tend to accumulate through time, consequently leading to complex regulation [13]. We therefore asked whether asRNAs show a similar tendency. In order to categorize genes by evolutionary time, we followed the method from Liu et al. [26]. We classified yeast genes into five age groups based on the presence of orthologs in divergent phylogenetic groups represented in the database OrthoMCL-DB [27] (see Materials and methods section). The oldest age group, age five, includes 664 of the most ancient genes, which are conserved in all cellular organisms (the common ancestors

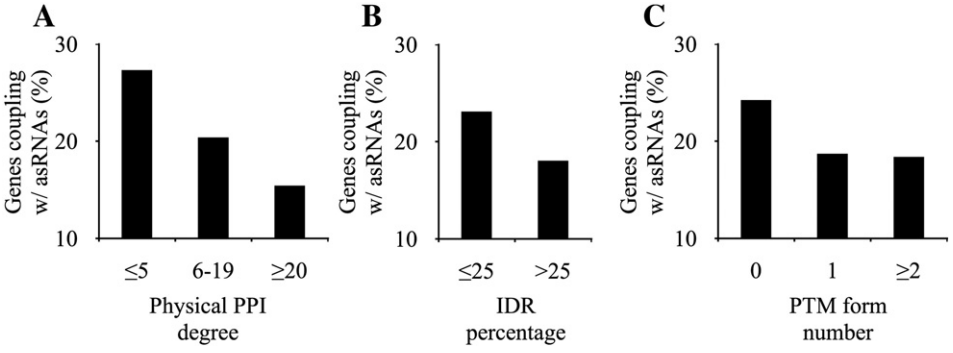

Fig. 1. Association between asRNA distribution and regulatory complexity. Percentages of genes with asRNAs were grouped by different (A) degrees of physical protein–protein interactions (PPIs), (B) percentages of intrinsically disordered regions (IDRs), and (C) numbers of post-translational modification (PTM) forms, including acetylation, ubiquitination, or phosphorylation.

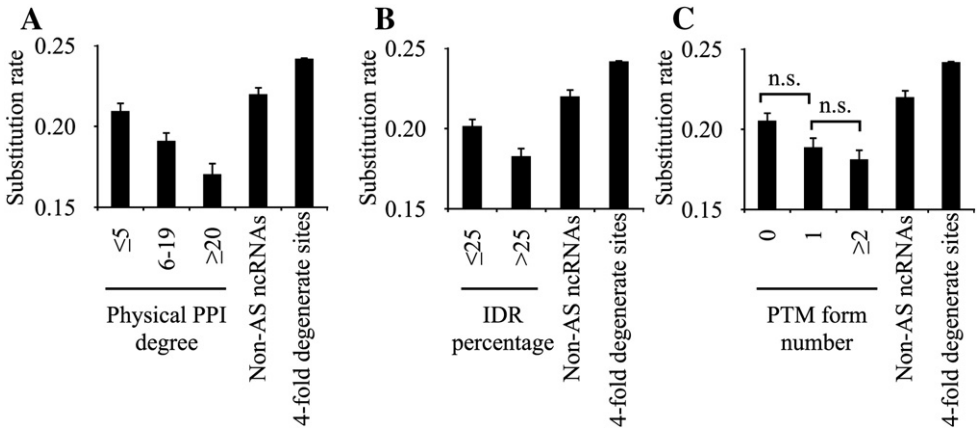

Fig. 2. Substitution rates in asRNAs of genes under different levels of regulatory complexity. Substitution rates in asRNAs were grouped by corresponding sense genes of different (A) degrees of physical PPIs, (B) percentages of IDRs, and (C) numbers of PTM forms. Significance is determined by least significant difference test ($P < 0.05$). Any two bars in each panel are significant except for the bars annotated by "n.s." (not significant). Non-AS ncRNAs, non-coding RNAs that do not overlap with any ORF transcript; 4-fold degenerate sites, four-fold degenerate sites of all ORFs used as a control [24,25].

of Eukaryota, Bacteria and Archaea); age four includes 1610 genes conserved in Eukaryota only; age three includes 691 genes conserved only in the Fungi/Metazoa group (Opisthokonta); age two includes 2224 genes conserved in kingdom Fungi; and age one includes 605 of the youngest genes, conserved only in the genus *Saccharomyces*. Assuming there is a constant ratio for the asRNA gain and loss, we analyzed the asRNAs number among the gene groups of different ages to observe the evolutionary tendency of asRNAs. The results show that genes in the older groups are less likely to have associated asRNAs (Fig. 3; Chi-square test of independence, $P = 3.0 \times 10^{-9}$). Since regulatory complexity accumulates over time [13–16], the negative association between asRNAs and evolutionary age is consistent with the observation in Fig. 1, and suggests that yeast asRNAs tend to be lost during evolution.

2.4. Genes under complex regulation tend to exhibit inverse expression with corresponding asRNAs

Functional sequences tend to be conserved during evolution [28]. Together with the observed asRNA reduction in yeast genome and purifying selection on partial asRNAs, we hypothesized that partial asRNAs may have become functional and may contribute to the complex regulation of sense genes. One of the regulation hypotheses suggests that asRNAs repress sense RNAs, and lead to inverse expression between sense and asRNA pairs under different growth conditions or in different cell types [5–10]. To estimate the inverse expression of sense and asRNA pairs, we calculated the correlation coefficient of their abundance across different conditions (see Materials and methods section). Interestingly, genes with more physical PPI have higher percentage of sense–antisense pairs that show inverse expression (Fig. 4A, Chi-square test of independence, $P = 4.4 \times 10^{-7}$). Moreover, the expression of sense RNAs from genes with more IDRs or more PTM forms also tends to be inversely correlated with the expression of their corresponding asRNAs (Fig. 4B, one-sided two-sample proportion test, $P = 0.028$; Fig. 4C, Chi-square test of independence, $P = 0.025$), regardless of different correlation thresholds (see Supplementary Fig. S1). Taken as a whole, sense–antisense pairs from genes with higher regulatory complexity are more likely to show inverse expression. Thus, the results suggest that a certain proportion of asRNAs might have evolved as repressors when the corresponding sense genes developed complex regulation.

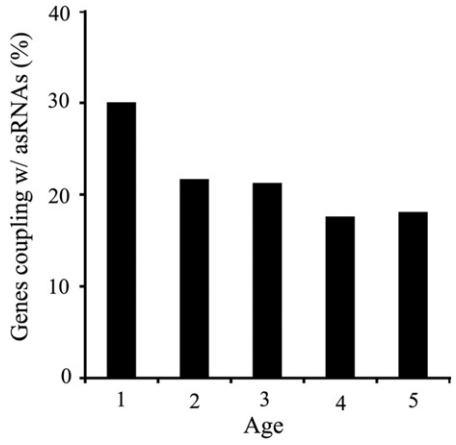

Fig. 3. Percentages of genes coupling with asRNAs at different evolutionary ages. Genes of age one group are the youngest and present only in genus *Saccharomyces*, genes of age two group are present in kingdom Fungi, genes of age three group are present in Fungi/Metazoa group (Opisthokonta), genes of age four group are present in Eukaryota, genes of age five group are most ancient and present in all cellular organisms (the common ancestors of Eukaryota, Bacteria and Archaea).

2.5. Old genes tend to exhibit inverse expression with corresponding asRNAs and less SAL/SRL

As the regulatory machinery is shaped through time [13–16], we explored the potential functions of asRNAs by comparing features of genes of different evolutionary ages. We first analyzed the pattern of inverse expression of sense and asRNAs, which may indicate a repressive regulatory function of asRNAs [5–10]. Subsequently, we calculated the ratio of sense–antisense overlapping length to sense transcript length (SAL/SRL), which we expect will reflect the proportions of sense transcripts interacting with asRNAs. The results show that the older group of genes has the higher percentage of sense and asRNA pairs that exhibit inverse expression (Fig. 5A, Chi-square test of independence, $P = 0.01$), regardless of different correlation thresholds (see Supplementary Fig. S2). In addition, genes in the youngest age group have the highest SAL/SRL (Fig. 5B, Kruskal–Wallis test, $P = 5.2 \times 10^{-5}$), which is inconsistent with the lowest percentage of sense–antisense pairs with inverse expression in this group. Thus, it is likely that RNA–RNA interaction does not solely explain the repression mechanism. Overall, consistent with the observation in Fig. 4, Fig. 5A reveals the significance of the asRNA role as repressors in yeast regulatory evolution. On the contrary, although genes in the

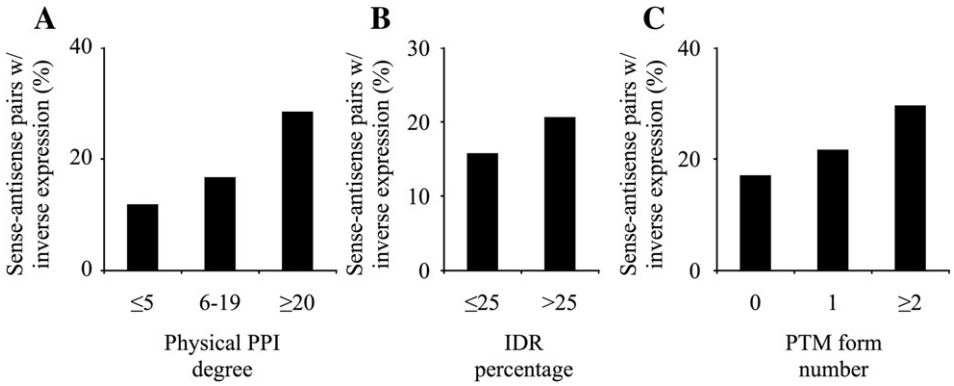

Fig. 4. The tendency of inverse expression between sense and asRNA pairs in genes of different regulatory complexity. Percentages of genes that show inverse expression with corresponding asRNAs were grouped by different (A) degrees of physical PPIs, (B) percentages of IDRs, and (C) numbers of PTM forms.

youngest age group contain the highest percentage of genes coupling with asRNAs (Fig. 3), their percentage of genes with inverse expression is the lowest (Fig. 5A). This phenomenon might be associated with the observed reduction of asRNAs in yeast genome.

3. Discussion

Previous work has elucidated that genes targeted by miRNAs tend to possess high regulatory complexity [17,18]. It is unclear whether genes coupling with asRNAs also exhibit similar tendencies. In this study, we aimed to investigate the asRNA targeting preference in terms of regulatory complexity, as estimated by the degree of physical PPIs, the percentage of IDRs, and the number of PTM forms. Although about one-fifth of the ORF genes in *S. cerevisiae* are coupling with asRNAs in our investigation, we found that the ratio of genes coupling with asRNAs shows a negative association with gene regulatory complexity. The negative association implies that, overall, asRNAs may play a minor role in developing complex regulation. This observation supports previous hypothesis that some asRNAs belong to transcriptional noise [3,4]. This explanation is also supported by our observation that only 175 of 988 pairs of sense and asRNAs exhibit inverse expression [5–10]. Nonetheless, our analyses also show that asRNAs coupling with genes under high regulatory complexity tend to evolve more slowly, and higher percentages of them exhibit inverse expression. Accordingly, we hypothesize that some retained asRNAs may find a niche and tend to function as repressors, which contribute to the complex regulation of sense genes. In contrast, most other asRNAs do not develop a role in gene regulation and may be eliminated in evolution.

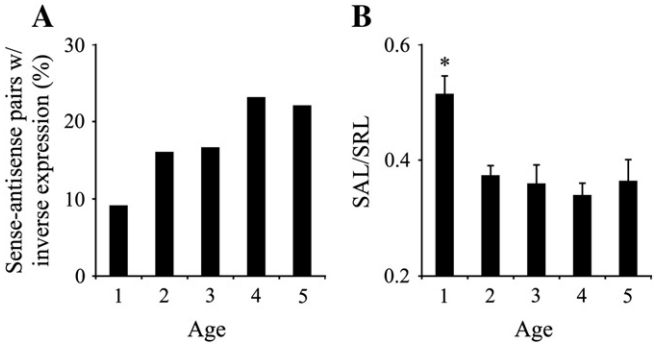

Fig. 5. Analyses showing partial asRNAs function as repressors in yeast evolution. (A) Percentages of genes that exhibit inverse expression with corresponding asRNAs were grouped by different evolutionary ages. (B) The ratios of sense–antisense overlapping length to sense transcript length (SAL/SRL) of genes were grouped by different evolutionary ages. The bar with an asterisk (*) represents significant differences with any other bar, as determined by least significant difference test ($P < 0.05$). Error bars indicate standard error of the mean.

To further support our hypothesis, we investigated the tendency of genes of different evolutionary ages to couple with asRNAs. Indeed, we found that yeast asRNAs become less frequent through evolutionary time. This finding is also consistent with previous studies, which reveal the rapid turnover of asRNAs after divergence of *S. cerevisiae* and *S. paradoxus* [8,29]. In fact, it has been found that half of the six sense–antisense pairs examined do not show conserved expression patterns from mid-log to early stationary phase between *S. cerevisiae* and *S. paradoxus* [8]. Furthermore, we analyzed the data released by Goodman et al. [29] (see Materials and methods section), and found that about one third of asRNAs are not conserved across all four strains in *S. cerevisiae* and *S. paradoxus*, also suggesting the rapid turnover of asRNAs. Consistently, this finding may also explain the higher percentage of genes coupling with asRNA in age one genes (the youngest genes), while only a small portion of them behaves as repressors. As a result, the elimination of asRNAs through evolution provides consistent evidence suggesting that partial asRNAs might be a by-product of transcription (leaky hypothesis) [3,4], at least in the early, function-developing, stage.

In conclusion, our results provide the first evidence showing that asRNAs may not be retained in the genome during evolution. Antisense RNAs may change rapidly during evolution, thus leading to lineage-specific gene regulation, as in the case of intergenic long ncRNA, suggested by Kutter et al. [30]. To further explore asRNA regulatory mechanisms, we compared the sense–antisense expression and SAL/SRL between old and young genes. Our results show that old genes coupling with asRNAs exhibit higher ratios of inverse expression. The finding suggests that asRNAs may tend to play a role as a repressor in regulatory evolution. However, we observed that old genes have unexpectedly low SAL/SRL as compared to young genes. As a relatively large sense–antisense overlap may be required to support the RNA–RNA interaction hypotheses for the mechanism of asRNA regulation [1], this low SAL/SRL finding implies that asRNAs may regulate gene expression through other mechanisms such as transcription collision or the recruitment of histone-modifying enzymes [1]. Although asRNAs coupling with age one genes seem to have higher SAL/SRL, we found that there is no significant difference in the lengths of asRNAs in the gene groups of different ages (Kruskal–Wallis test, $P = 0.13$). Nonetheless, some limitations may cause bias in our conclusions. First, the inverse expression between sense and asRNAs may only represent a subset of functional asRNAs, since asRNAs can act through other mechanisms. Second, the sample size for asRNAs expression is limited, which may lead to experimental bias. Future work should enlarge the sample size by including more different conditions, time points,

and/or cell types. Additional investigations and analyses are necessary to explore the asRNA functions in regulation.

4. Materials and methods

4.1. Yeast strains, growth conditions and treatments

To elucidate the dynamics of asRNAs in yeast, the laboratory strain BY4741 was cultured and sampled in three different growth conditions: mid-log phase, early stationary phase, and after heat-shock treatment. Yeast was grown in rich medium (1.5% yeast extract, 1% peptone, 2% dextrose, 2 g/L SC amino acid mix, 100 mg/L adenine, 100 mg/L tryptophan, 100 mg/L uracil) at 30 °C to an OD_{600} of 0.802 for mid-log phase. Glucose levels were monitored hourly by glucose (HK) assay kit (Sigma-Aldrich, MO, USA), and early stationary phase time points were taken 2 h after the glucose levels reached zero (OD_{600} ~ 2.22). For each condition, 12 ml was harvested and quenched by adding 20 ml pre-chilled liquid methanol to generate a final concentration of 60%. Media was later removed by centrifugation, cells were washed in RNase-free water and stored overnight at −80 °C. Similarly, 12 ml of mid-log culture was collected for heat shock treatment as follows: culture medium was removed by centrifugation. Cells were resuspended in 10 ml, 42 °C pre-warmed medium and put in a 42 °C water bath for 15 min. Heat shocked sample was quenched by adding 20 ml liquid pre-chilled methanol, which was later removed by centrifugation, after which cells were washed in RNase-free water and stored overnight at −80 °C.

4.2. Construction of strand-specific cDNA library

To strengthen the ability to determine the polarity of RNA transcripts, a strand-specific cDNA library was constructed by the dUTP second-strand method [8,31], which was identified as the leading protocol for paired-end sequencing [32]. Briefly, incorporation of dUTP during second-strand cDNA synthesis and subsequent destruction of the uridine-containing strand in the sequencing library were responsible for identifying the orientation of transcripts.

The cDNA library was constructed by TruSeq RNA Sample Preparation Kits v2 based on the guide (Illumina). Five μg of total RNA was diluted with nuclease-free water to 50 μl and heated in a preheated heat block at 65 °C for 5 min to disrupt the secondary structures. PolyA-containing mRNA was purified using 50 μl RNA purification beads for each sample. The purified RNA fragments were reverse-transcribed into first-strand cDNA by SuperScript II Reverse Transcriptase (Invitrogen) and random primers (Illumina). The cDNA/RNA hybrid was precipitated with ethanol and ammonium acetate. The second-strand cDNA synthesis was performed in a 80 μl reaction volume using NEBNext® mrna second strand synthesis module (NEB) and dNTP/dUTP mix (Fermentas, dTTP → dUTP) at 16 °C for 2.5 h. The DNA fragments were purified by QIAquick PCR purification kit (Qiagen) and treated with T4 DNA polymerase, *E. coli* DNA polymerase I Klenow fragment and T4 polynucleotide kinase in a 30 μl reaction using the buffer and the enzyme in the kit (Illumina) at 37 °C for 30 min. The blunt phosphorylated DNA fragments were purified by the MinElute PCR purification kit (Qiagen) and treated with Klenow fragment of DNA polymerase in the presence of dATP to add an adenine overhang to the 3′ end of each strand in a 30 μl reaction using the kit (Illumina) at 37 °C for 30 min. Adapters were ligated to DNA fragments in a 37.5 μl reaction using the kit (Illumina). The ligation products were purified by the MinElute PCR purification kit (Qiagen) and ligation products in the size range from 350 to 450 bp were eluted from the gel. These products were precipitated and resuspended in 21 μl TrueSeq resuspension buffer using QIAquick Gel extraction Kit (Qiagen) and treated with USER™ enzyme mixture (Uracil-Specific Excision Reagent, NEB) to degrade the dUTP in the second strand. The adapter-ligated DNA fragments were amplified by the kit (Illumina). The amplified products were purified by Agencourt AMPure XP (Beckman) and collected in 30 μl resuspension buffer (Qiagen) as the cDNA library for sequencing.

4.3. Identification of ORF transcripts and asRNAs

RNA paired-end sequencing was performed by Illumina Hiseq2000 with the standard protocol. For each sample, sequencing yielded around 11.6 million 101-nucleotide paired-end reads (range from 10.95 million to 12.12 million). To map these 101-nucleotide paired reads, we used Bowtie [33] and TopHat [34] following the published protocol [35] with known genome sequences and annotations from Saccharomyces Genome Database (SGD, http://www.yeastgenome.org) for *S. cerevisiae*. The sequencing reads mapped to unique positions on the genome were further selected from total mapped reads for further analysis (~92% of total mapped reads).

Further, to determine the complete transcriptional landscape, we designed a method to detect all the transcripts. Read depth was used to define the transcript unit as described below, and then the transcript boundaries were manually adjusted. The position with the highest read depth was designated as the starting point to extend each segment. Second, we extended the segment bidirectionally until the read depth mean of the segment was four times higher than its flanking position; the threshold was empirically determined to separate close transcripts. Third, the boundary of this segment was further adjusted to the first and last internal read. Fourth, the regions already identified as transcripts were removed from following cycles. We repeated this process until the highest read depth was less than five. Subsequently, the obtained transcripts overlapping with verified and uncharacterized ORFs in the same orientation were defined as ORF transcripts (ORF-Ts); on the other hand, transcripts distal to verified and uncharacterized ORF were assigned as ncRNAs. Eventually, when an ORF-T and a ncRNA in different orientations overlapped by at least 1 bp, they were coupled as a sense–antisense pair and the ncRNA was defined as an asRNA. In total, we found 1215 genes coupling with 995 asRNAs, including 883 unannotated ncRNAs, 74 dubious ORFs, 22 small nucleolar RNAs, 11 transfer RNAs, and 5 annotated ncRNAs.

4.4. Analyses of PPIs, IDRs and PTMs

We downloaded the physical PPI dataset of *S. cerevisiae* from the database BioGrid (version 3.1.92) [36] to calculate the degree of physical PPIs. In order to analyze the physical PPI degrees appropriately, we divided all genes into three groups of similar sizes based on the 33rd percentile (PPIs = 5), and the 67th percentile (PPIs = 19). As a result, we classified PPIs ≤ 5 as low level (1871 genes), between 6 and 19 as medium level (1991 genes), ≥20 as high level (1932 genes).

We used the percentages of IDRs of each gene as calculated by Kim et al. [37]. The IDRs were predicted by the software DISOPRED [38]. The average of IDR percentages of all genes in the genome is 25%. Therefore, genes with IDR percentages equal to and greater than 25% were categorized into the disordered group, which was suggested to have higher regulatory complexity than the group of other genes with IDR percentages less than 25%.

To identify genes with PTMs, we identified a gene as having protein N-alpha-terminal acetylation, ubiquitination, or phosphorylation if the gene was in the compiled dataset from previous studies [39–43]. Overall, of 5794 verified and uncharacterized ORF genes, there were 1604 genes with N-alpha-terminal acetylation, 918 genes with ubiquitination, and 2504 genes with phosphorylation. The majority, 2460 genes, had no form of PTM, while 1952 genes had only one form of PTM, 1072 genes had two forms of PTM, and 310 genes had all three forms of PTM. We grouped genes by their levels of PTM forms: none (0), single (1) and multiple (≥2) forms of PTM. We proposed that the regulatory complexity of these three gene groups should differ from each other due to following reasons. First, a gene with a PTM form is assumed to be under more complex regulation than a gene without any PTM form.

Second, a gene with multiple PTM forms may be additionally regulated by the combination of different forms of PTM [44] and lead to higher regulatory complexity than genes with only one or no PTM form.

4.5. Estimation for the rate of evolutionary changes in asRNAs

The rate of evolutionary changes in asRNAs was estimated by determining the nucleotide substitution rate without considering insertions or deletions. We aligned the asRNA sequences of *S. cerevisiae* and the orthologous sequences of *S. paradoxus* by the software MUSCLE (version 3.8.31) [45]. We excluded regions overlapping with ORFs and eliminated the cases in which aligned regions were shorter than 50 bp to avoid potential biases of poor alignment. After filtering, 478 asRNAs and 441 non-antisense ncRNAs were obtained for analyses. Moreover, the substitution rate of the four-fold degenerate sites was used as a neutral reference [24,25].

4.6. Gene classification by evolutionary age

In order to investigate the propensity of genes to persist in the genome during evolution, methods of Liu et al. [26] were applied to assign the evolutionary age of genes. We utilized the algorithm GeneTRACE [46] which considers gene loss and horizontal gene transfer events to identify the evolutionary age of each gene. For the input information, the phylogenetic tree from NCBI [47], and the orthologous profile of each gene from the database OrthoMCL-DB (version 5) [27] were used. Based on the results of GeneTRACE, we determined the presence or absence of each orthologous gene at each node of the phylogenetic tree. The genes that were present in all cellular organisms (the common ancestors of Eukaryota, Bacteria and Archaea) were assigned as age five (664 of the most ancient genes), only in Eukaryota assigned as age four (1610 genes), only in the Fungi/Metazoa group (Opisthokonta) as age three (691 genes), only in kingdom Fungi as age two (2224 genes), and only in genus *Saccharomyces* as age one (605 of the youngest genes).

4.7. Analyses of sense–antisense inverse expression and interacting regions

Because the inverse expression between sense and asRNAs was suggested as an indicator for the functionality of asRNAs [5–10], we evaluated the inverse expression by calculating Pearson correlation coefficient of sense and asRNAs expression at mid-log phase, early stationary phase, and after heat shock treatment. Expression data with missing values for any one condition was removed from the analysis. After initial screening, 988 pairs of sense and asRNAs met our criteria and were used in this study. We used BPKM (bases per kilobase of gene model per million mapped bases) as the relative expression abundance [48], and calculated Pearson correlations of BPKMs between sense and antisense expression (threshold for Pearson correlation: less than -0.6). In addition, in order to discuss the regulatory mechanisms of asRNAs, we evaluated the RNA–RNA interaction by calculating SAL/SRL.

4.8. Evaluation of asRNA conservation

To investigate the conservation of asRNAs, we compared the genes with asRNAs across four strains of *S. cerevisiae* and *S. paradoxus*. We downloaded the data of asRNA abundance from Goodman et al. [29]. To increase the confidence of asRNA expression, we only classified genes with asRNAs when the BPKM of asRNAs was more than one. We calculated the ratio of the number of genes with asRNAs conserved across four strains to the average number of genes with asRNAs in each strain.

Authors' contributions

CHL and DW designed the analyses. CHL and ZTYT collected the data. CHL performed the analyses. CHL, ZTYT, and DW wrote the paper. DW was the principal investigator and conceived the analyses. All authors read and approved the final manuscript.

Acknowledgments

We thank Jen-Hao Cheng and Robert Liu for comments on the draft. This work was supported by the National Science Council of Taiwan and Biodiversity Research Center, Academia Sinica.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2013.10.008.

References

[1] M.A. Faghihi, C. Wahlestedt, Regulatory roles of natural antisense transcripts, Nat. Rev. Mol. Cell Biol. 10 (2009) 637–643.
[2] M. Nishizawa, T. Okumura, Y. Ikeya, T. Kimura, Regulation of inducible gene expression by natural antisense transcripts, Front. Biosci. 17 (2012) 938–958.
[3] M. Lapidot, Y. Pilpel, Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms, EMBO Rep. 7 (2006) 1216–1222.
[4] N. Dutrow, D.A. Nix, D. Holt, B. Milash, B. Dalley, E. Westbroek, et al., Dynamic transcriptome of *Schizosaccharomyces pombe* shown by RNA–DNA hybrid mapping, Nat. Genet. 40 (2008) 977–986.
[5] C.F. Hongay, P.L. Grisafi, T. Galitski, G.R. Fink, Antisense transcription controls cell fate in *Saccharomyces cerevisiae*, Cell 127 (2006) 735–745.
[6] J. Camblong, N. Iglesias, C. Fickentscher, G. Dieppois, F. Stutz, Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*, Cell 131 (2007) 706–717.
[7] Z. Xu, W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Münster, J. Camblong, et al., Bidirectional promoters generate pervasive transcription in yeast, Nature 457 (2009) 1033–1037.
[8] M. Yassour, J. Pfiffner, J.Z. Levin, X. Adiconis, A. Gnirke, C. Nusbaum, et al., Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species, Genome Biol. 11 (2010) R87.
[9] B. Gelfand, J. Mead, A. Bruning, N. Apostolopoulos, V. Tadigotla, V. Nagaraj, et al., Regulated antisense transcription controls expression of cell-type-specific genes in yeast, Mol. Cell. Biol. 31 (2011) 1701–1709.
[10] Z. Xu, W. Wei, J. Gagneur, S. Clauder-Münster, M. Smolik, W. Huber, et al., Antisense expression increases gene expression variability and locus interdependency, Mol. Syst. Biol. 7 (2011) 468.
[11] J.P. Uhler, C. Hertel, J.Q. Svejstrup, A role for noncoding transcription in activation of the yeast PHO5 gene, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 8011–8016.
[12] Z. Bodi, J.D. Button, D. Grierson, R.G. Fray, Yeast targets for mRNA methylation, Nucleic Acids Res. 38 (2010) 5327–5335.
[13] M. Warnefors, A. Eyre-Walker, The accumulation of gene regulation through time, Genome Biol. Evol. 3 (2011) 667–673.
[14] J. Capra, K. Pollard, M. Singh, Novel genes exhibit distinct patterns of function acquisition and network integration, Genome Biol. 11 (2010) R127.
[15] F. Theis, N. Latif, P. Wong, D. Frishman, Complex principal component and correlation structure of 16 yeast genomic variables, Mol. Biol. Evol. 28 (2011) 2501–2512.
[16] T. Rito, C. Deane, G. Reinert, The importance of age and high degree, in protein–protein interaction networks, J. Comput. Biol. 19 (2012) 785–795.
[17] Q. Cui, Z. Yu, Y. Pan, E.O. Purisima, E. Wang, MicroRNAs preferentially target the genes with high transcriptional regulation complexity, Biochem. Biophys. Res. Commun. 352 (2007) 733–738.
[18] H. Liang, W.-H. Li, MicroRNA regulation of human protein–protein interaction network, RNA 13 (2007) 1402–1408.
[19] J.-D.J. Han, N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, et al., Evidence for dynamically organized modularity in the yeast protein–protein interaction network, Nature 430 (2004) 88–93.
[20] S. Aris-Brosou, Determinants of adaptive evolution at the molecular level: the extended complexity hypothesis, Mol. Biol. Evol. 22 (2005) 200–209.
[21] V.N. Uversky, C.J. Oldfield, A.K. Dunker, Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling, J. Mol. Recognit. 18 (2005) 343–384.
[22] A. Stein, R.A. Pache, P. Bernadó, M. Pons, P. Aloy, Dynamic interactions of proteins in complex networks: a more structured view, FEBS J. 276 (2009) 5390–5405.
[23] Y.L. Deribe, T. Pawson, I. Dikic, Post-translational modifications in signal integration, Nat. Struct. Mol. Biol. 17 (2010) 666–672.
[24] Y.-D. Li, H. Liang, Z. Gu, Z. Lin, W. Guan, L. Zhou, et al., Detecting positive selection in the budding yeast genome, J. Evol. Biol. 22 (2009) 2430–2437.
[25] Z.T.-Y. Tsai, H.-K. Tsai, J.-H. Cheng, C.-H. Lin, Y.-F. Tsai, D. Wang, Evolution of cis-regulatory elements in yeast de novo and duplicated new genes, BMC Genomics 13 (2012) 717.

[26] Z. Liu, Q. Liu, H. Sun, L. Hou, H. Guo, Y. Zhu, et al., Evidence for the additions of clustered interacting nodes during the evolution of protein interaction networks from network motifs, BMC Evol. Biol. 11 (2011) 133.
[27] F. Chen, OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups, Nucleic Acids Res. 34 (2006) D363–D368.
[28] W.H. Li, Molecular Evolution, Sinauer Associates, Incorporated, 1997.
[29] A.J. Goodman, E.R. Daugharthy, J. Kim, Pervasive antisense transcription is evolutionarily conserved in budding yeast, Mol. Biol. Evol. 30 (2013) 409–421.
[30] C. Kutter, S. Watt, K. Stefflova, M. Wilson, A. Goncalves, C. Ponting, et al., Rapid turnover of long noncoding RNAs and the evolution of gene expression, PLoS Genet. 8 (2012) e1002841.
[31] D. Parkhomchuk, T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, et al., Transcriptome analysis by strand-specific sequencing of complementary DNA, Nucleic Acids Res. 37 (2009) e123.
[32] J. Levin, M. Yassour, X. Adiconis, C. Nusbaum, D. Thompson, N. Friedman, et al., Comprehensive comparative analysis of strand-specific RNA sequencing methods, Nat. Methods 7 (2010) 709–715.
[33] B. Langmead, C. Trapnell, M. Pop, S. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (2009) R25.
[34] C. Trapnell, L. Pachter, S. Salzberg, TopHat: discovering splice junctions with RNA-Seq, Bioinformatics 25 (2009) 1105–1111.
[35] A. Roberts, L. Goff, G. Pertea, D. Kim, D. Kelley, H. Pimentel, et al., Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, Nat. Protoc. 7 (2012) 562–578.
[36] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, BioGRID: a general repository for interaction datasets, Nucleic Acids Res. 34 (2006) D535–D539.
[37] P.M. Kim, A. Sboner, Y. Xia, M. Gerstein, The role of disorder in interaction networks: a structural analysis, Mol. Syst. Biol. 4 (2008) 179.
[38] J.J. Ward, L.J. McGuffin, K. Bryson, B.F. Buxton, D.T. Jones, The DISOPRED server for the prediction of protein disorder, Bioinformatics 20 (2004) 2138–2139.
[39] K. Helsens, P. Van Damme, S. Degroeve, L. Martens, T. Arnesen, J. Vandekerckhove, et al., Bioinformatics analysis of a *Saccharomyces cerevisiae* N-terminal proteome provides evidence of alternative translation initiation and post-translational N-terminal acetylation, J. Proteome Res. 10 (2011) 3578–3589.
[40] W. Lee, M. Lee, J. Jung, K. Kim, D. Kim, SCUD: *Saccharomyces cerevisiae* ubiquitination database, BMC Genomics 9 (2008) 440.
[41] J. Ptacek, G. Devgan, G. Michaud, H. Zhu, X. Zhu, J. Fasolo, et al., Global analysis of protein phosphorylation in yeast, Nature 438 (2005) 679–684.
[42] X. Li, S. Gerber, A. Rudner, S. Beausoleil, W. Haas, J. Villén, et al., Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces *cerevisiae*, J. Proteome Res. 6 (2007) 1190–1197.
[43] C. Stark, T. Su, A. Breitkreutz, P. Lourenco, M. Dahabieh, B. Breitkreutz, et al., PhosphoGRID: a database of experimentally verified in vivo protein phosphorylation sites from the budding yeast *Saccharomyces cerevisiae*, Database 2010 (2010) bap026.
[44] B.A. Benayoun, R.A. Veitia, A post-translational modification code for transcription factors: sorting through a sea of signals, Trends Cell Biol. 19 (2009) 189–197.
[45] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[46] V. Kunin, C. Ouzounis, GeneTRACE-reconstruction of gene content of ancestral species, Bioinformatics 19 (2003) 1412–1416.
[47] E.W. Sayers, T. Barrett, D.A. Benson, E. Bolton, S.H. Bryant, K. Canese, et al., Database resources of the National Center for Biotechnology Information, Nucleic Acids Res. 38 (2010) D5–D16.
[48] L. Yang, M.O. Duff, B.R. Graveley, G.G. Carmichael, L.-L. Chen, Genomewide characterization of non-polyadenylated RNAs, Genome Biol. 12 (2011) R16.