SCIENCE CHINA
Life Sciences

Tsinghua Special Issue
• RESEARCH PAPER •

January 2014 Vol.57 No.1: 22–35
doi: 10.1007/s11427-013-4583-0

Global signatures of protein binding on structured RNAs in *Saccharomyces cerevisiae*

YANG YuCheng[1], UMETSU Jumpei[1,2] & LU Zhi John[1*]

[1]*MOE Key Lab of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing 100084, China;*
[2]*Department of Biological Information, Tokyo Institute of Technology, Tokyo 152-8850, Japan*

Received October 16, 2013; accepted November 21, 2013; published online December 23, 2013

Protein binding is essential to the transport, decay and regulation of almost all RNA molecules. However, the structural preference of protein binding on RNAs and their cellular functions and dynamics upon changing environmental conditions are poorly understood. Here, we integrated various high-throughput data and introduced a computational framework to describe the global interactions between RNA binding proteins (RBPs) and structured RNAs in yeast at single-nucleotide resolution. We found that on average, in terms of percent total lengths, ~15% of mRNA untranslated regions (UTRs), ~37% of canonical non-coding RNAs (ncRNAs) and ~11% of long ncRNAs (lncRNAs) are bound by proteins. The RBP binding sites, in general, tend to occur at single-stranded loops, with evolutionarily conserved signatures, and often facilitate a specific RNA structure conformation *in vivo*. We found that four nucleotide modifications of tRNA are significantly associated with RBP binding. We also identified various structural motifs bound by RBPs in the UTRs of mRNAs, associated with localization, degradation and stress responses. Moreover, we identified >200 novel lncRNAs bound by RBPs, and about half of them contain conserved secondary structures. We present the first ensemble pattern of RBP binding sites in the structured non-coding regions of a eukaryotic genome, emphasizing their structural context and cellular functions.

RNA binding protein, non-coding RNA, UTR, RNA structure, *Saccharomyces cerevisiae*

Citation: Yang YC, Umetsu JP, Liu ZJ. Global signatures of protein binding on structured RNAs in *Saccharomyces cerevisiae*. Sci China Life Sci, 2014, 57: 22−35, doi: 10.1007/s11427-013-4583-0

RNA binding proteins (RBPs) play important roles in regulating almost all post-transcriptional stages of gene expression, such as splicing, stability, localization and translation [1−4]. Many mRNAs are regulated by one or more RBPs as co-regulators. In *Saccharomyces cerevisiae*, there are approximately 600 annotated and predicted RBPs, but relatively few of them have been systematically studied to determine their regulatory targets [5]. In addition to mRNAs, numerous non-coding RNAs (ncRNAs) are targets of RBPs [6,7]. Previous studies revealed that RNAs were regulated by many RBPs post-transcriptionally [8]. It is widely accepted that altering the expression of RBPs will change cellular physiology profoundly. Studies using animal models have revealed that RBPs are involved in many human diseases, such as neurologic disorders and cancers [9].

Although the mechanisms that confer the specificity of RBP-RNA interaction are poorly understood, it is clear that this specificity is determined by both the primary sequence and secondary structure of the target RNA [10,11]. The secondary structures recognized by some RBPs are known. For example, the SAM domain of the yeast post-transcriptional regulator Vts1p recognizes the shape of the SRE of the RNA ligand [12].

Currently, the most direct and powerful approach to profiling RBP-RNA interactions is UV crosslinking and immunoprecipitation (CLIP) of RNA-protein complexes,

*Corresponding author (email: zhilu@tsinghua.edu.cn)

combined with next-generation sequencing (CLIP-seq, also called HITS-CLIP) [13]. Recently, CLIP-seq has been modified to detect RBP binding sites at single-nucleotide resolution by PAR-CLIP (photoactivatable-ribonucleosideenhanced CLIP) [14] and iCLIP (individual-nucleotide resolution CLIP) [15]. In the standard PAR-CLIP procedure, the photoactive nucleotides 4-thiouridine (4sU) or 6-thioguanosine (6sG) are incorporated into the transcripts of growing cells to increase the UV-crosslinking efficiency between protein and RNA [14]. Robust and sensitive computational approaches are critical for RBP binding site identification from CLIP data. Several recently published methods address the problem of site identification from CLIP data [16–18]. Previous studies involving CLIP data mainly focused on identifying mRNA binding sites for specific RBPs, but one recent study developed gPAR-CLIP (global PAR-CLIP) to capture binding sites for all types of RBPs in budding yeast [19]. However, it mainly focused on RBP binding signatures on mRNAs, while the functional and structural elements of RBP binding sites on structured ncRNAs and UTRs are not fully studied.

Here, we present a computational framework for identifying and characterizing transcriptome-wide RBP binding sites in structured non-coding regions, and detecting novel ncRNAs bound by proteins. By adding the gPAR-CLIP data of total RNA (Table S1 in Supporting Information), we emphasize non-coding regions, including canonical ncRNAs, UTRs and novel ncRNAs. More specifically, we highlight the RNA structural context bound by RBPs, which was less analyzed in the previous paper [19]. In addition, our study is featured by integrating more high-throughput data and manually curated data about the structures of non-coding RNAs in a eukaryotic genome.

By summarizing all proteins' binding effects, our transcriptome-wide analysis reveals that RBP binding sites are enriched in regions of evolutionarily conserved structure and sequences with unpaired nucleotides. RBP binding regulates secondary structure and tends to result in an open conformation at local binding sites. We investigated the RBP binding pattern on each individual type of canonical ncRNA. We also found that RBP binding sites on the eukaryotic expansion segments of rRNA derive from some unknown proteins and may be important in regulating translation. We identified specific nucleotide modifications associated with RBP binding to tRNA. In addition, we predicted structural motifs involved in regulating mRNA stability and localization in 3′ and 5′ UTR sequences, respectively. We also examined the structural dynamics of RBP binding sites in UTRs upon glucose and nitrogen deprivation. Novel ncRNAs with RBP binding sites, especially evolutionarily conserved ones, were also identified and characterized. Overall, our analyses provide a global profile of the binding signatures of all kinds of RBPs on structured ncRNAs and UTRs in yeast. These findings suggest that RBP binding regulates the structural conformation of ncRNAs at the co-transcriptional and post-transcriptional levels, and may carry out important biological functions.

1 Materials and methods

1.1 Processing of gPAR-CLIP-seq data

Global PAR-CLIP (gPAR-CLIP) features a global map of transcriptome-wide RBP binding sites *in vivo*, rather than those of specific RBPs as in the normal PAR-CLIP procedure. In total, 10 libraries were used in our study, and were grouped into different types according to experimental conditions or assays: (i) light-poly(A)-RNA libraries; (ii) light-poly(A)-RNA libraries under glucose starvation; (iii) light-poly(A)-RNA libraries under nitrogen starvation; (iv) light-total-RNA libraries; and (v) heavy-total-RNA libraries. 'Light' means the RNPs are lighter than the 40S ribosome based on sedimentation in a sucrose gradient, while 'heavy' means they are heavier than the 40S ribosome (Table S1 in Supporting Information). gPAR-CLIP libraries in (i)–(iii) have been published previously (in GSE43747), and undergone poly(A) selection described in [19]. gPAR-CLIP libraries (iv)–(v) are firstly publicly available in this study (in GSE48888). These libraries did not experience poly(A) selection, and thus contain total RNAs in yeast.

The raw reads were stripped of the 3′ adaptor sequence using FASTX-Toolkit. Reads that did not contain the adaptor sequence or contained an ambiguous nucleotide were discarded. The remaining gPAR-CLIP reads were then sorted into libraries according to their 6-nt barcodes using FASTX. Finally, the barcode sequences in the reads were clipped, and reads that were poly(A/T) or shorter than 18 nucleotides in length were removed. Reads from duplicated experiments were combined for subsequent data analysis. Overall, 96.8% of gPAR-CLIP reads were left and used in the following analyses.

1.2 Transcriptome-wide identification of RBP binding sites in the *S. cerevisiae* genome

The *S. cerevisiae* strain S288C genome (Version R64-1-1) was downloaded from the *Saccharomyces* Genome Database (SGD, http://www.yeastgenome.org). The processed reads from the previous step were mapped to the genome using Bowtie [20]. We allowed at most two mismatches for one read because of the U-to-C conversions induced in PAR-CLIP. Reads from the processed libraries, other than heavy-total-RNA libraries, were mapped to the genome, allowing at most 10 matching locations (parameters: -v 2 -m 10 -best -strata). Because of the multiple rDNA repeats in the yeast genome, reads from the heavy-total-RNA libraries were aligned to the rDNA consensus region (chrXII, from 451575 to 460711), allowing just one matching location (parameters: -v 2 -m 1 -best -strata).

RBP binding sites were then defined from mapped reads

in the yeast genome using PARalyzer, which is designed specifically for PAR-CLIP data [18]. PARalyzer relies on the U-to-C transitions introduced by the PAR-CLIP technology and the read density at the cross-link sites. It can generate high-resolution RNA-protein interaction sites with a high signal-to-noise ratio. First, PARalyzer grouped the overlapped reads for further analysis. A group had to contain at least 10 reads with conversions at two or more locations. For all positions, if there were at least five reads at the position, kernel density estimates were calculated according to read counts both with and without T-to-C conversions. Clusters were defined as regions whose conversion (U to C) density was greater than the non-conversion density. A nucleotide with a signal value (U to C) higher than a background value (U to U) was defined to be bound by RBPs. For each nucleotide in the clusters, a binding affinity score was calculated as follows:

$$\text{binding affinity score}=\frac{\text{signal value}}{\text{signal value+background value}}.$$

A nucleotide is considered to be bound by RBPs if its score is larger than 0.5. The maximum score in a cluster was defined to be the RBP binding affinity of the cluster in our further analyses. Finally, we used BEDtools [21] to overlap the RBP binding sites with different genomic elements, which were annotated by SGD (March 2012) and a study of UTRs [22].

1.3 Mapping RBP binding sites on secondary structures of ncRNA

The secondary structure of rRNA was derived from two sources. The co-variance structure was downloaded from the Comparative RNA Web (CRW, http://www.rna.icmb.utexas.edu) and was derived from evolutionary constraints. The crystal structure was converted from the three-dimensional crystal structure of the ribosome [23] using RNAView [24]. The co-variance structure may contain some base pairings *in vivo* that are not found in the crystal structure. The secondary structure of eukaryotic-specific expansion segment ES6S in the co-variance structure is left as a large loop because of insufficient evolutionary constraints. The secondary structure of ES6S was obtained from the crystal structure of the ribosome using RNAView.

The numbers of contacts were calculated from the crystal structure of the ribosome. We used a coarse-grained approach to model RNA and protein molecules: each RNA nucleotide is represented by three pseudoatoms corresponding to the base, sugar and phosphate groups, and each amino acid residue is represented by two pseudoatoms corresponding to the backbone and side chain groups. We then calculated the number of pseudoatoms corresponding to side chain groups around each pseudoatom corresponding to the base group within a sphere of 10 angstrom radius to derive the number of contacts with the RNA nucleotide.

The identities and coordinates of tRNA and snoRNA structural elements were obtained from the tRNA database (http://trnadb.bioinf.uni-leipzig.de) [25] and the snoRNA Orthological Gene Database (http://snoopy.med.miyazaki-u.ac.jp), respectively. The secondary structures of ncRNAs were visualized using VARNA [26].

1.4 Grouping of UTRs by localization, decay rate and stress-specific binding

We grouped the RNA sequences by decay rates, cellular localization and specific binding patterns under stresses, and used RNApromo to detect the common structural motifs in each group. RBP binding sites of UTRs were extended to the same length of 70 nt. Genome-wide decay rates of mRNA in yeast were downloaded from [27]. Genes were grouped according to their mRNAs' long (⩾60 min) or short (⩽6 min) half-lives. The cellular localization annotations of yeast genes were obtained from the SGD. Genes with mRNAs having the same cellular localization were also grouped together.

In UTRs, RBP binding sites that overlapped with each other across wild type (WT) and stress conditions were merged together into larger blocks. We then calculated averaged RBP binding affinity scores for each block. All RBP binding blocks in the 3′ UTRs could be divided into three groups according to their averaged binding affinity: (i) specifically bound under non-stress conditions, (ii) specifically bound under stress conditions (i.e., glucose or nitrogen starvation) and (iii) bound under both conditions. Condition-specific binding blocks were defined according to the following criterion: the average non-stress binding affinity/average stress condition binding affinity is smaller than 0.25, and the inverse is larger than 0.5. We focused mainly on the condition-specific binding blocks in subsequent analysis.

1.5 Prediction of RNA structural motifs in different groups of UTRs

For each functional annotation of yeast genes, we calculated a RBP binding enrichment score as follows:

$$\frac{No.\ of\ genes\ containing\ binding\ sites\ with\ GOi\ /No.\ of\ genes\ with\ GOi}{No.\ of\ genes\ with\ GOi\ /Total\ No.\ of\ genes},$$

where *GOi* is the Gene Ontology term in our analysis. We then focused on those functional annotations with high enrichment scores to detect their structural motifs. Before running RNApromo, we filtered each set of RNA sequences to exclude the sequences with >90% sequence similarity, because high sequence similarity may result in higher AUC scores even when a functional motif does not exist. We then used RNApromo to predict the consensus structural motif for each non-redundant sequence set (parameters: -fold 70, 35 -shuffle 5,70,35 -bg 0.01 -n 5). For each structural motif, an AUC score and *P*-value were given to examine the per-

formance of the result. An AUC score close to 0.5 indicates that no significant motif is detected. If the structural motif's *P*-value is larger than the threshold ('bg' in the parameters), RNApromo will not report any motif results.

1.6 Assembly of novel ncRNA transcripts bound by RBPs

We used poly(A) RNA sequencing data (downloaded from GSE43747) from non-stressed cells (WT) to assemble novel ncRNA transcripts. First, we stripped off 3′ adaptor sequences and discarded short reads (<18 nt) using FASTX. We then assembled novel ncRNA transcripts using TopHat and Cufflinks as described in [28]. The gap threshold in Cufflinks to join non-overlapping reads is 50 nt. Assembled transcripts with lengths shorter than 50 nt were discarded; the average and maximum lengths of these transcripts were 879 and 30337 nt, respectively. Assembled transcripts were annotated with genomic elements (annotations from SGD, including annotations of UTRs from [22]). In parallel, we used gPAR-CLIP light-total-RNA libraries to identify RBP binding sites in these novel ncRNA transcripts. We then mainly focused on novel ncRNA transcripts containing RBP binding sites.

We downloaded the seven-way alignments of yeast species from UCSC (http://genome.ucsc.edu). In order to filter novel ncRNA transcripts with low evolutionary conservation, we then overlapped these novel ncRNA transcripts with the regions of multiple alignment to obtain more credible ncRNA transcripts. Only those located entirely in regions of multiple alignment were retained for further analysis. Next, we used RNAz (version 2.1) [29] to detect secondary structures for those novel ncRNA transcripts. For each secondary structure, two parameters, the *z*-score and SCI (structure conservation index), were calculated to evaluate its thermodynamic stability and structure conservation, respectively. Negative *z*-scores indicate that the RNA sequence is more stable than the random background. The SCI will be around 1 if the structure is quite conserved, and it will be around 0 if a consensus structure is not found. To identify novel ncRNA transcripts with high confidence, we set the probability cutoff as 0.9 in RNAz. For these newly identified long ncRNAs (lncRNAs), we also calculated their coding potential scores using CPC [30]. In addition, experimentally identified lncRNAs in budding yeast were manually collected from lncRNAdb (http://www.lncrnadb.org) [31].

2 Results

2.1 A computational framework to identify RBP ensemble binding patterns in non-coding regions of the yeast genome

We developed a computational framework to describe the global signatures of RBP binding sites in structured non-coding regions in yeast (Figure 1). First, the global set of RBP binding sites was identified using PARalyzer [18] (Figure 1A) based on 10 gPAR-CLIP datasets (Table S1 in Supporting Information; see Materials and methods). The gPAR-CLIP data includes information about the RNA binding of various RBPs. A nucleotide with a signal value (U to C) higher than background value (U to U) was defined to be bound by RBPs. For each single nucleotide in the RBP binding sites, we calculated a binding affinity score to represent the binding strength (Figure 1A; see Materials and methods).

We then overlapped the RBP binding sites with annotated yeast genomic elements (annotations from SGD, including annotations of untranslated regions (UTRs) from [22]). The RBP binding to three RNA elements was profiled: (i) UTRs, (ii) structured canonical ncRNAs, and (iii) novel ncRNAs (Figure 1B). (i) We detected various structural motifs affecting mRNA decay rates and localization in 3′ and 5′ UTR sequences, respectively, using RNApromo [32]. We also examined the secondary structural dynamics of RBP binding sites in 3′ UTRs upon glucose and nitrogen deprivation. (ii) We generated RBP binding maps on canonical ncRNAs (including rRNA, tRNA, snoRNA, snRNA and known lncRNAs), where the protein binding sites were mapped to secondary structures. (iii) In addition to the annotated ncRNAs and UTRs, we also described 663 novel ncRNAs, including 196 intergenic and 467 antisense ncRNAs, which were derived from a set of total RNA sequencing data (Table S1 in Supporting Information) [19]. We also analyzed the evolutionary covariance signatures of the RBP binding regions in novel ncRNA transcripts.

2.2 Global distribution and coverage of RBP binding sites in non-coding regions

We obtained a global picture of protein binding sites in the non-coding regions of yeast RNAs from the gPAR-CLIP data. It provides a high-resolution profile of RBP ensemble binding signatures (Figure 2A). We assayed the binding signals from three gPAR-CLIP libraries: poly-A+ RNAs, heavy RNAs and light RNAs (see Materials and methods). In mRNAs (from poly-A+ RNA library), 71% of the RBP binding sites were distributed in coding sequences (CDSs), 22% in 3′ UTRs, and 7% in 5′ UTRs. In rRNAs (from the heavy RNA library), most binding sites were observed in 18S rRNA (29%) and 25S rRNA (37%). Many binding sites are in pre-rRNAs (27%), indicating the processing of pre-rRNA by RBPs. In other ncRNAs (from the light RNA library), most binding sites are located in tRNAs (27%), snoRNAs (26%) and newly identified antisense ncRNAs (26%).

We also calculated the exact fraction of each RNA transcript covered by RBPs at the single nucleotide level, because of the high resolution of gPAR-CLIP data. The total

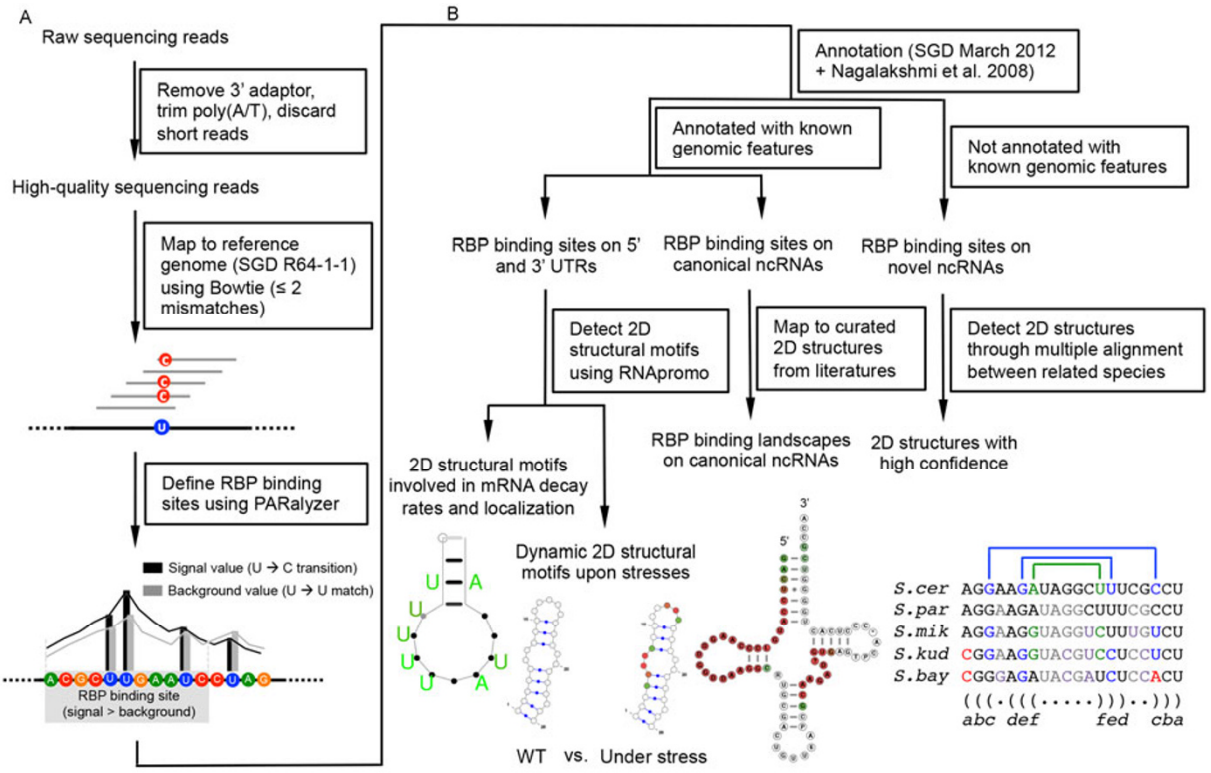


Figure 1 Computational framework to identify RBP ensemble binding sites on non-coding regions of yeast genome. A, Pipeline for data pre-processing and identification of RBP binding sites. Substitutions of U to C between the reference genome and sequencing reads indicate potential RBP binding events. A continuous region in which the signal values (U-to-C mutation) are larger than the background values (U-to-U mapping) can be defined as an RBP binding site using PARalyzer. B, Pipeline for the functional analyses of RBP binding sites in non-coding regions. Novel ncRNAs can be defined from RNA sequencing data using TopHat and Cufflinks.

coverage for each type of RNA element is summarized as the total number of nucleotides in RBP binding sites divided by the total RNA length (Figure 2B). We found that 3′ UTRs, in total, had higher coverage (~15%) by RBPs than 5′ UTRs (~9%) did, and CDSs had the lowest coverage (~3%). This suggests that UTRs, especially 3′ UTRs, may be more important in post-transcriptional regulation. A previous study also reported that different RBPs might have different preferences among 3′ UTRs, 5′ UTRs and coding sequences [5]. Among ncRNAs, we found three of them (tRNA, snoRNA and snRNA) with greater coverage by RBPs than the remainder.

2.3 RBPs bind preferentially to single-stranded loops and conserved structures

We inquired whether RBPs display any preference for binding to particular secondary structures of non-coding regions. We first derived base-pairing information from a genome-wide structure measurement in yeast, where the base-pairing probability was measured using an enzymatic probing technology, creating a so-called PARS score [33]. We found no preference for single-strandedness or base-pairing for all nucleotides in non-coding regions (Figure 2C, left panel), but nucleotides bound by RBPs (binding affinity larger than 0.5) comprised nearly twice as many unpaired nucleotides as base-paired nucleotides (P=1.43×10^{-4}; Fisher's exact test) (Figure 2C, right panel). This genome-wide observation is consistent with a previous biophysical analysis, showing that RBP binding to RNAs is preferentially to unpaired nucleotides [10]. We also examined the distribution of binding affinities for the conserved secondary structures among seven yeast species [34], and found that the RBP binding affinities for conserved, structured regions are higher for CDSs, 5′ UTRs, 3′ UTRs, tRNAs, snoRNAs and snRNAs (Figure 2D). This preference might be caused by positive structural selection in RBP binding sites. Our results suggest that the RBP binding sites are concentrated in the regions conserved at both the sequence [19] and structural levels.

2.4 RBP binding facilitates RNA secondary structure conformations *in vivo* and their cellular functions

The observation that RBP binding has a preference for unpaired nucleotides suggested a mechanism of structural regulation by RBP binding. To understand how RBP binding influences the secondary and tertiary structures of ncRNAs, we carefully examined RBP binding in relation to the structures of U6 snRNA, RNase P/MRP (Figures S1 and S2 in

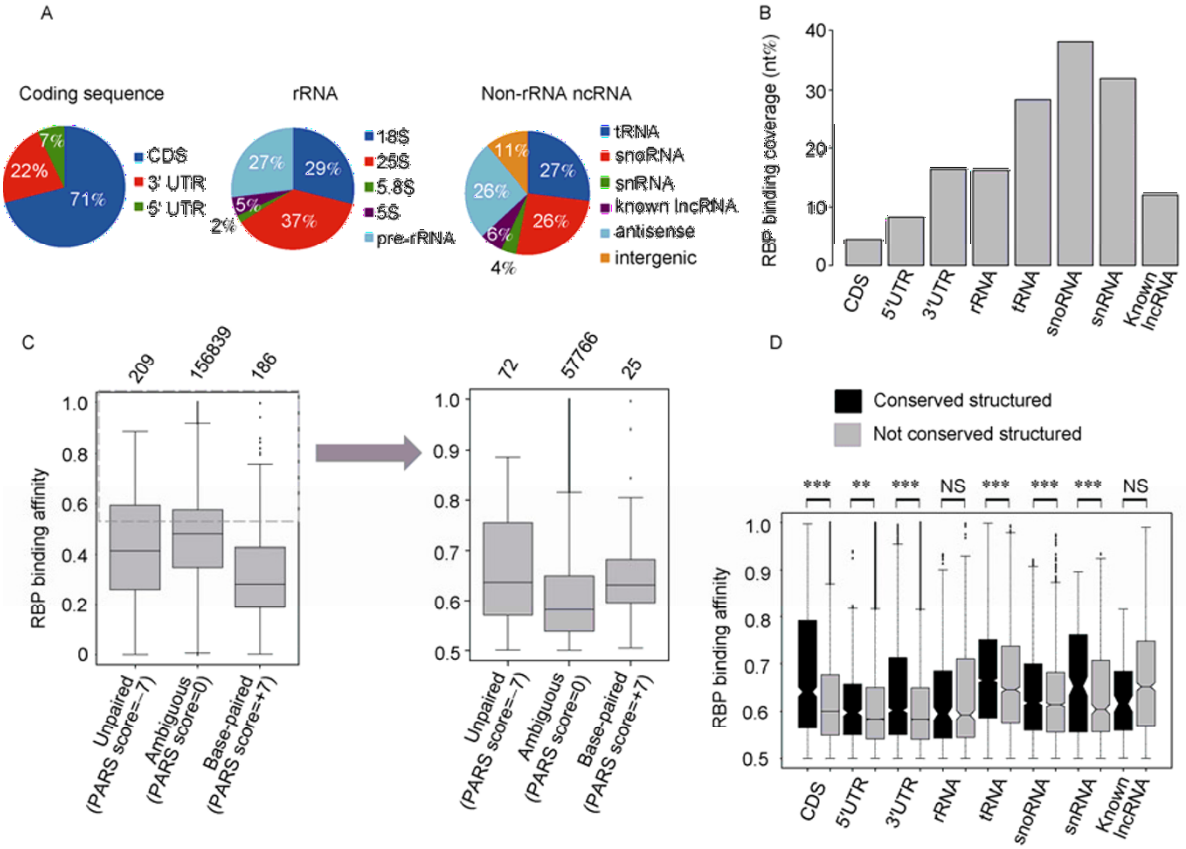


Figure 2 Overview of RBP binding in non-coding regions. A, Distribution of RBP binding sites in non-coding regions in three different RNA libraries. B, RBP binding coverage in different genomic elements. C, RBP binding in non-coding regions has a preference for unpaired nucleotides. RBP binding affinity of unpaired, ambiguous and base-paired nucleotides for all nucleotides (209 unpaired and 186 base-paired) (left) and nucleotides for which binding affinity is larger than 0.5 (72 unpaired and 25 base-paired) (right) in non-coding regions. The PARS scores range from −7 to +7; −7 indicates a single-stranded conformation, +7 indicates a double-stranded conformation and 0 indicates an ambiguous conformation. D, Comparison of RBP binding affinities of nucleotides located in conserved structured regions and otherwise for different genomic elements. **, $P<0.01$; ***, $P<10^{-3}$ (Wilcox test); NS, not significant at a threshold of 0.01.

Supporting Information), box C/D snoRNA (Figure S3 in Supporting Information) and eukaryotic-specific expansion segments in 18S rRNA.

The secondary structures of naked U6 snRNA (a small nuclear RNA) and U6 snRNA incorporated into the U6 snRNP (small nuclear ribonucleoprotein, snRNA-protein complex) were recently shown to be dramatically different [35]. The 3′ stem-loop in the naked U6 snRNA is more compact than it is in the U6 snRNP (Figure S4 in Supporting Information). Four known U6 snRNP protein-binding sites on the U6 snRNA were confirmed by our work (Figure 3A). For example, a binding site on the large 5′ loop of Prp8 was also recently revealed by CLIP experiments [36]. In addition to these, we also identified a novel binding site with an open conformation on the 3′ stem-loop (Figure 3A). The bound protein might recognize and stabilize the open conformation of the U6 snRNA.

The RNA components of RNase P and RNase MRP are highly conserved and similar to each other, and most of the protein components of RNase P and RNase MRP are shared [37]. Almost all known RBP binding sites on RNase P and RNase MRP were identified in our data. We also identified novel binding sites in the RNAs of both RNase P and RNase MRP (Figures S1 and S2 in Supporting Information). Phylogenetic analyses uncovered a well-conserved GARAR element in RNase MRP. Others have suggested that this conserved GARAR element may form a pentaloop, P8 [38]. However, the P8 structure was not detected by holoenzyme footprinting analyses *in vitro* [39]. In our analysis, the P8 region was found to interact with proteins, suggesting that the conformation of P8 may change when it is bound by an RBP *in vivo*.

We also identified novel binding sites in rRNAs (Figure S5 in Supporting Information). The major difference between bacterial and eukaryotic rRNA is the presence of eukaryotic-specific expansion segments (ESs) [40–42] (Table S2 in Supporting Information). Information about the expansion segments of eukaryotic rRNA is quite limited [40]. The largest expansion segment in small 18S rRNA is ES6S, which is about 200 nucleotides in length. The secondary structure of ES6S cannot be determined using a co-variance model without phylogenetic data (Figure S5 in Supporting

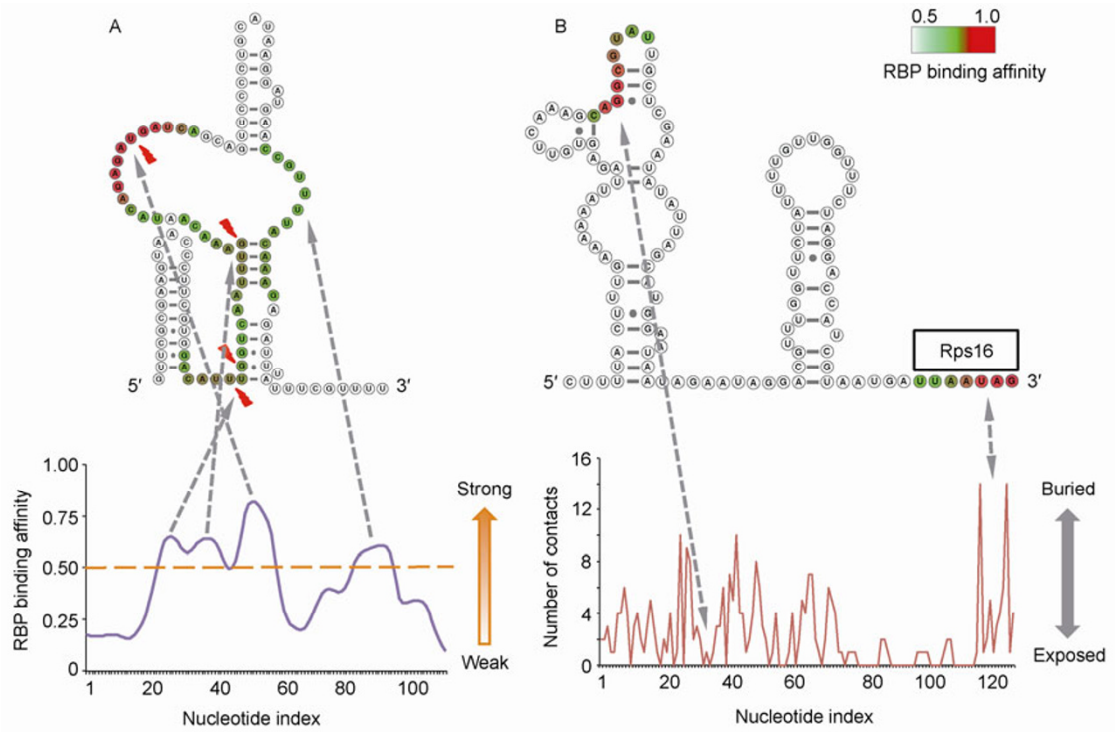


Figure 3 RBP binding facilitates structure conformations of ncRNAs *in vivo*. A, RBP binding landscape on secondary structure of U6 snRNA *in vivo* (top). The red arrows indicate known RBP binding sites in U6 snRNA. RBP binding affinity across the U6 snRNA (bottom). B, RBP binding sites in the secondary structure of the 3′ half of ES6S (C741-G866) (top). The structure is based on the crystal structure of the ribosome. A previously known binding site of ribosomal protein Rps16 is indicated. The numbers of contacts across the 3′ half of ES6S (bottom) reflect the accessibility of neighboring ribosomal proteins in the fragment.

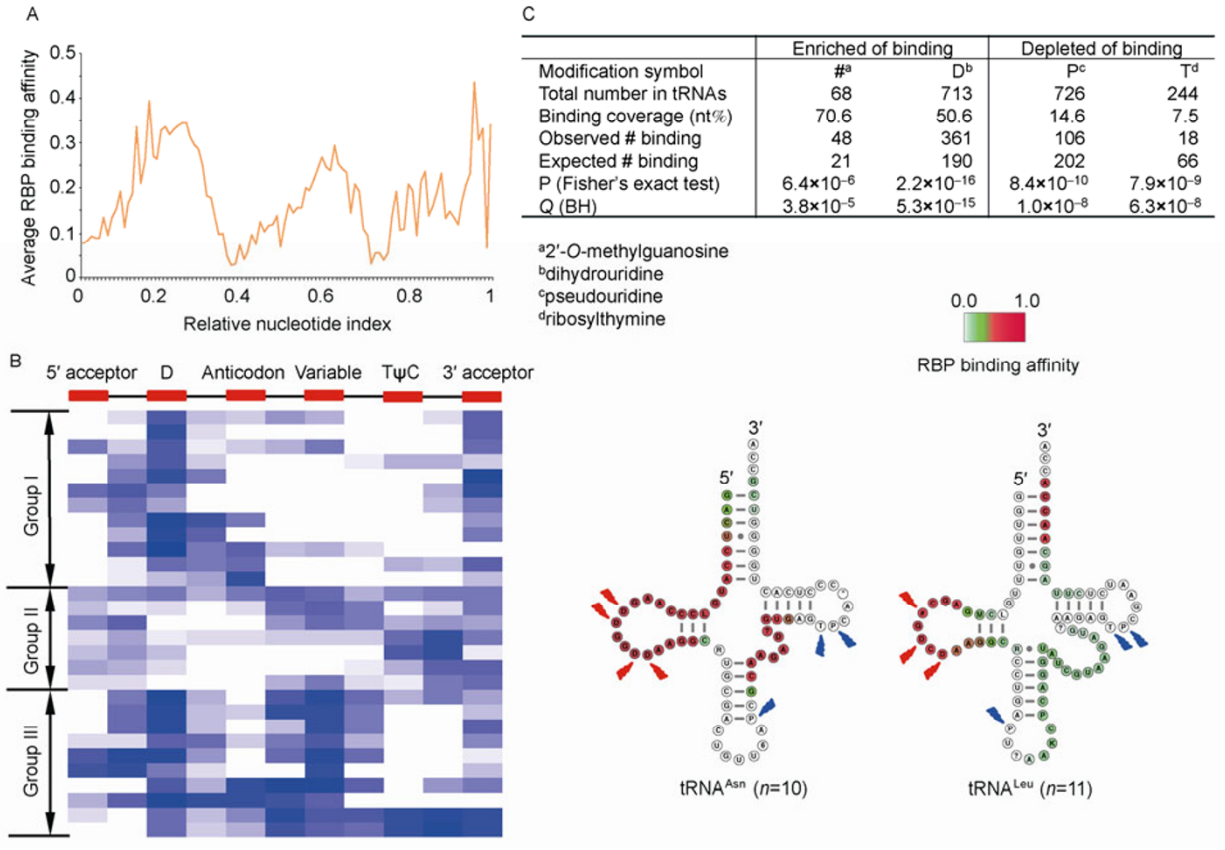
Information), but two long helices in ES6S can be derived from crystal structures [23,43] (Figure 3B). In the crystal structure, these ES6S helices are exposed on the ribosome surface [40]. The 3′-end of ES6S is known to be tightly associated with ribosomal protein Rps16 [44], and we also observed this interaction. Moreover, we identified another RBP footprint 5′ to the known binding site (Figure 3B). ES12S also has an obvious RBP binding site on the long helix. We then compared the patterns of RBP binding to the conserved core components and to the expansion segments of rRNA, but found no significant differences in affinity or coverage.

To assess the ribosomal protein accessibility of these RBP binding sites, we calculated the number of contacts for each nucleotide in ES6S and ES12S using 3D structure models derived from crystals [23]. Usually, nucleotides with a small number of contacts are far from ribosomal proteins in space, whereas nucleotides with a large number of contacts are closer to ribosomal proteins. In ES6S, there are a large number of contacts across the Rps16 binding site, suggesting that this site may be closely associated with the ribosomal protein Rps16. However, there are a small number of contacts across the newly identified site, indicating that it is quite distant from ribosomal proteins (Figure 3B). Similarly, few contacts were observed in the long helix of ES12S, suggesting that the novel binding sites found *in vivo* could interact with proteins other than ribosomal proteins (Figure S6 in Supporting Information). The non-ribosomal proteins binding to the expansion segments may play important roles in the specific regulation of eukaryotic translation.

2.5 RBP binding is associated with RNA modifications

In almost all organisms, tRNAs are heavily modified (http://rnamdb.cas.albany.edu/RNAmods). A few mechanistic studies have indicated that specific base modification of tRNAs influences translation efficiency and accuracy, tRNA stability and other cellular processes [45–48]. Here, we performed a global analysis of the relationship between tRNA modification and RBP binding (see results in Figure 4).

We first calculated the average RBP binding affinity along tRNA sequences. Usually, a tRNA sequence consists of five structural elements: the 5′-acceptor stem, D loop, anticodon loop, variable loop, TψC loop and 3′-acceptor stem [25]. RBP binding was obviously enriched at the D loop, variable loop and 3′ acceptor stem (Figure 4A). RBP binding to the 3′-acceptor stem has been reported [49]. tRNAs can be classified (K-means classification, see Materials and methods) into three different categories according to their RBP binding signatures at the D loop and variable



	Enriched of binding		Depleted of binding	
Modification symbol	#[a]	D[b]	P[c]	T[d]
Total number in tRNAs	68	713	726	244
Binding coverage (nt%)	70.6	50.6	14.6	7.5
Observed # binding	48	361	106	18
Expected # binding	21	190	202	66
P (Fisher's exact test)	6.4×10^{-6}	2.2×10^{-16}	8.4×10^{-10}	7.9×10^{-9}
Q (BH)	3.8×10^{-5}	5.3×10^{-15}	1.0×10^{-8}	6.3×10^{-8}

Figure 4 RBP binding associated with tRNA nucleotide modifications. A, RBP binding affinity along tRNAs, averaged across all transcripts used in our analysis. The length of different transcripts is normalized to a scale ranging from 0 to 1. B, Heatmap of RBP binding affinity along tRNA structural elements. The positions of the acceptor stem, D loop, anticodon loop, variable loop and TψC loop are indicated at the top. All tRNAs can be grouped into three categories according to their binding patterns. C, tRNA modifications that are concentrated or scarce in RBP binding sites. Two examples of RBP binding to secondary structures of type I tRNA ($tRNA^{Asn}$, bottom left) and type II tRNA ($tRNA^{Leu}$, bottom right). Nucleotide modifications that are concentrated or scarce in RBP binding sites are indicated by red and blue arrows, respectively. The list of abbreviations for modified nucleotides in tRNAs can be found in tRNAdb (http://trnadb.bioinf.uni-leipzig.de).

loop (Figure 4B). Because dihydrouridine is a feature of the D loop, we inquired whether RBP binding to different structural elements of tRNAs is related to specific nucleotide modifications.

Furthermore, we identified two tRNA modifications that were significantly concentrated in RBP binding regions: 2′-*O*-methylguanosine ($P=3.82\times10^{-5}$; Fisher's exact test) and dihydrouridine ($P=5.28\times10^{-15}$; Fisher's exact test) (Figure 4C; Table S3 in Supporting Information). In addition, we also identified two modified nucleotides that were significantly depleted in RBP binding regions: pseudouridine ($P=1.01\times10^{-8}$; Fisher's exact test) and ribosylthymine ($P=6.29\times10^{-8}$; Fisher's exact test) (Figure 4C). Of note, we found that the associations between RBP binding and the above four modifications are robust upon increasing RBP binding affinity thresholds (Table S4 in Supporting Information). The depletion of dihydrouridine in RNA stems may be caused by its conformational flexibility [49]. This conformational flexibility may also contribute to the strong RBP binding in the D loop. On the other hand, the pseudouridine and ribosylthymine of the TψC loop provide increased structure stability, through tighter base stacking and the interaction between pseudouridine and RNA backbone phosphates through a bridging water molecule [49]. The paucity of RBP binding to pseudouridine and ribosylthymine of the TψC loop may be attributed to the stacking effect between pseudouridine and backbone phosphates. Indeed, the dynamics of tRNA structures are regulated by the interactions between modified nucleotides, proteins and ions in tRNA structures.

2.6 RBPs recognize structural motifs to regulate mRNA at the post-transcriptional level

We averaged the RBP binding affinities along the entire lengths of mRNAs, and found that coding regions exhibited significantly low binding affinity than 5′ and 3′ UTRs ($P<2.2\times10^{-16}$; Wilcox test) (Figure 5A). Notably, both the translational start and stop sites exhibit a local peak of protein binding affinity, indicating increased accessibility for RBPs. These findings agree with those of a previous study showing that the start and stop codons tend to have single-

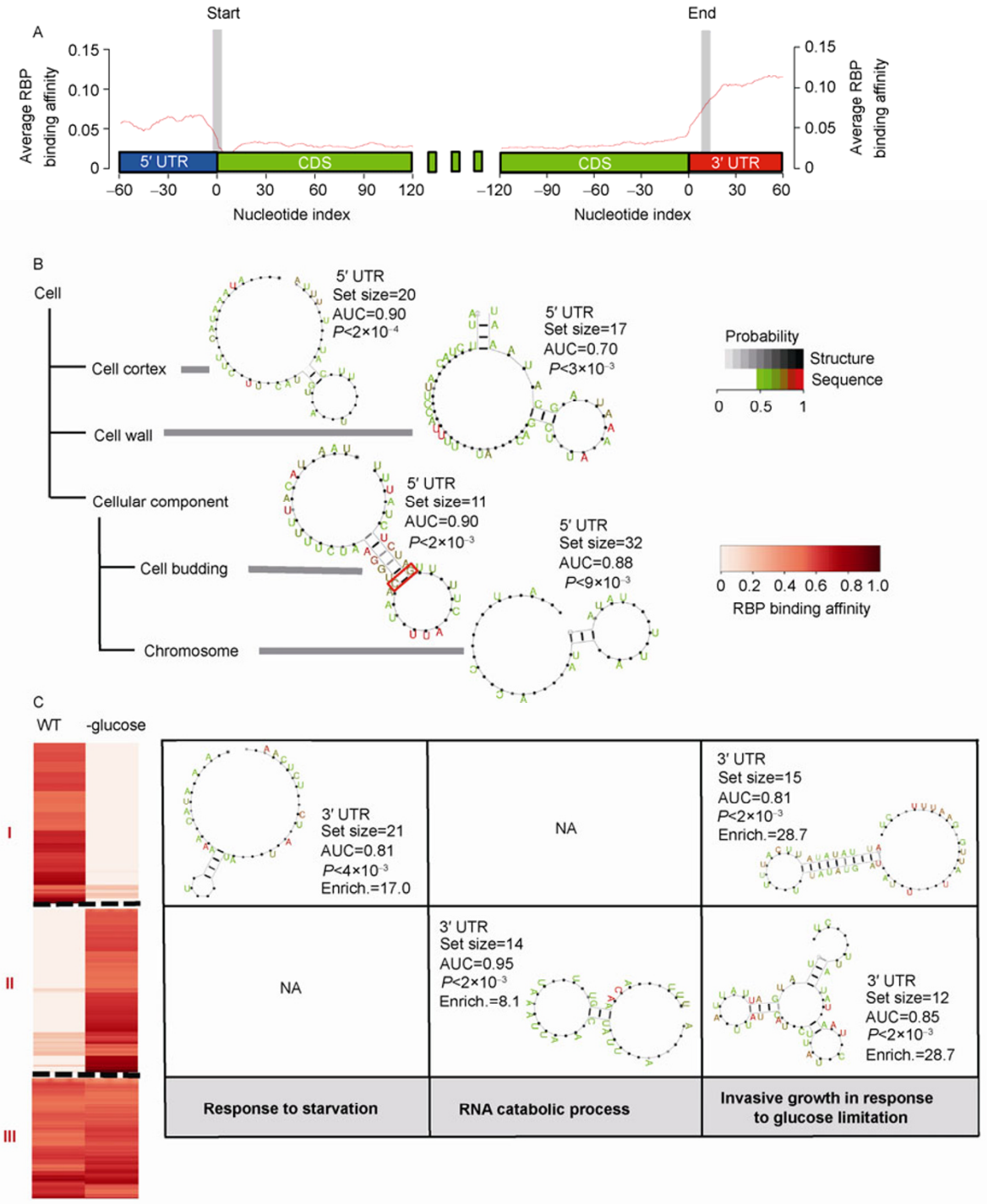


Figure 5 Structural motif prediction from the UTR sequences bound by RBPs. A, RBP binding affinity across the 5′ UTR, the coding region and the 3′ UTR, averaged across all transcripts used in our analysis. Translational start (left panel) and termination (right panel) sites are indicated by gray bars. B, Predicted structural motifs involved in mRNA localization. The location of each set is indicated in their localization annotation hierarchy. C, RBP binding dynamics on 3′ UTRs under glucose starvation. The UTRs differentially bound by RBPs were grouped together (left). Predicted structural motifs from differentially bound regions, which are enriched with specific GOs (gene ontologies) (right).

stranded conformation [33].

Some secondary structural motifs in UTRs involved in regulating mRNA decay rates and localization in yeast have been identified recently [32,50,51]. We grouped the 5′ UTR and 3′ UTR sequences according to cellular localization and decay rates, respectively, and predicted the structural motifs in the UTRs' RBP binding regions using RNApromo [32]. We detected eight specific structural motifs regulating mRNAs' distinct cellular localizations (Figure 5B; more examples in Figure S7 in Supporting Information). For ex-

ample, we found a structural motif enriched in mRNAs localized to the cell bud (AUC=0.90; $P<2\times10^{-3}$), which is consistent with previous experiments that found a core CG dinucleotide in the stem of the She2p/3p mRNA [51].

In addition to structural motifs regulating localization, we found a large AU-rich loop in the 3′ UTRs of mRNAs having long half-lives (AUC=0.91; $P<4\times10^{-3}$), and a stem-loop structure enriched with AU in the 3′ UTRs of mRNAs with short half-lives (AUC=0.95; $P<4\times10^{-3}$) (Figure S8 in Supporting Information). Additional predicted 3′ UTR structural motifs are consistent with previous results [32] (Figure S8 in Supporting Information). We also used MEME [52] to discover whether the same regions are enriched with any particular sequence motifs and found none. This suggests that secondary structure, rather than primary sequence, is conserved in the RBP binding sites to modulate mRNA stability and decay rates.

Another function of RBP binding to UTRs is to mediate global gene expression in response to stresses [53–55]. Using the gPAR-CLIP data from different stress conditions (i.e., glucose and nitrogen starvation), we calculated the changes in RBP binding affinities (see Materials and methods). In 3′ UTRs, we found 154 structural motifs in RBP binding sites that were specifically bound under different stresses (Figure 5C; Figure S9 in Supporting Information). For example, mRNAs could exhibit a distinct structural motif in their binding regions to promote invasive growth in response to glucose starvation (Figure 5C). Besides, mRNAs with other cellular functions (e.g., response to starvation and RNA catabolism) contain specific structural motifs that only appear upon either normal or stress condition (Figure 5C). These results reveal that RBP binding to UTRs is dynamic and may alter the conformation of the RNA, ultimately inducing different post-transcriptional regulation in response to stress.

2.7 Novel lncRNAs bound by RBPs

We curated 15 lncRNAs recently identified in yeast [31]. Six of them exhibit RBP binding signals in the gPAR-CLIP data (Table S5 in Supporting Information). In addition to these, we assembled 663 novel ncRNA transcripts from newly published poly(A) RNA sequencing data [19] using TopHat and Cufflinks. These ncRNAs have a minimum length of 50 nt and an average length of 301 nt (Figure S10 in Supporting Information). The newly identified ncRNA transcripts can be divided into two groups: 467 are antisense transcripts and 196 are intergenic transcripts. We observed that RBP binding coverage on these newly identified ncRNA transcripts is relatively low. This is probably due to the low level of expression of lncRNAs, resulting in a paucity of sequencing reads from binding sites (Figure S11 in Supporting Information).

Although the average binding coverage is low on novel lncRNAs compared with that on canonical ncRNAs, we still found 182 (38.9%) antisense and 59 (30.1%) intergenic ncRNA transcripts containing RBP binding signatures. Among them, 101 (82 antisense and 19 intergenic ncRNA transcripts) are conserved in a seven-way alignment of yeast genomes (Tables S6 and S7 in Supporting Information). Most of these novel lncRNA transcripts exhibit low coding potential scores, confirming their identity as ncRNAs. Furthermore, we used RNAz to predict the secondary structures of these ncRNA transcripts. Finally, we discovered 20 novel ncRNA transcripts (8 antisense and 12 intergenic) with high confidence levels (Figure 6A). For example, a novel intergenic ncRNA transcript (chrXV, + strand, from 978470 to 978868) is highly expressed and evolutionarily conserved (Figure 6B). The predicted secondary structure of the intergenic transcript is quite stable (the z-score is −27.58), and exhibits strong RBP binding sites (Figure 6C). The novel ncRNA transcript candidates we identified are supported by multiple lines of evidence, including RNA expression (RNA-seq), protein binding (gPAR-CLIP) and evolutionary constraints on their secondary structures.

3 Discussion

Here we presented the first transcriptome-wide RBP binding landscape in non-coding regions in a eukaryote. We developed a computational framework that identifies transcriptome-wide RBP binding sites and analyzes their functions and dynamics in non-coding regions in yeast. We included the entire ensemble of RBP binding sites in the analysis by using gPAR-CLIP, and focused on protein binding in the context of RNA secondary structure, rather than primary sequence.

Although PAR-CLIP has proved its effectiveness in RBP binding site identification, its limitations cannot be neglected. RBP binding sequences, which are devoid of U, cannot be captured by 4sU-based crosslinking [14]. Therefore, it is important to be aware that the nuceotide components of the RBPs' footprints could infulence the crosslinking efficiency, although U-less sites are extremely rare. Only about 1.7% of the sequencing reads are U-less, with the proportion of Us in the reads less than 10%. In addition, it is noteworthy that photoactivable nucleotides, such as 4sU, are toxic for yeast, as mentioned in [14] as well.

We found many novel binding sites facilitating RNA secondary structures *in vivo*, and discovered four tRNA modifications associated with protein binding. Many novel structural motifs were found to be associated with mRNA decay rates and localization. Notably, the results we obtained by analyzing structural motifs associated with mRNA decay are consistent with a previous study [32]. In addition, it is noteworthy that most of the mRNAs in our analysis are non-translating ones and do not represent the spectra of RBPs bound to mRNAs upon translation. Finally, we discovered many novel lncRNAs whose existence is supported

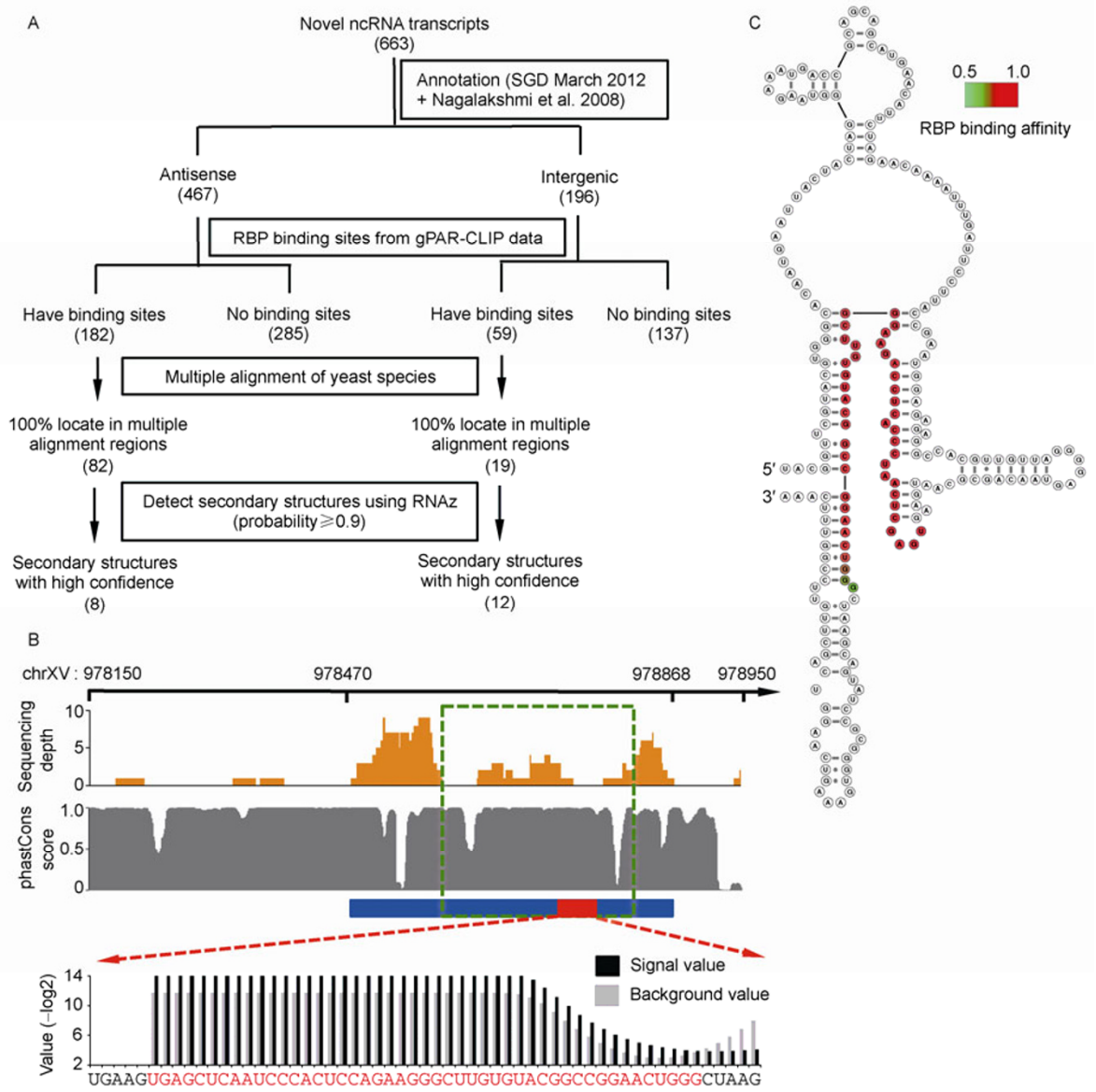


Figure 6 Novel ncRNAs bound by RBPs. A, Summary of novel ncRNA transcripts that contain RBP binding sites. The novel ncRNA transcripts were annotated with antisense and intergenic genomic elements, and ignored in the subsequent analysis if they did not contain RBP binding sites. B, RNA sequencing depth and phastCons scores across a novel intergenic ncRNA transcript (chrXV, + strand, from 978470 to 978868), which is indicated by the blue bar (top). The RBP binding site is indicated by the red bar, on which the signal and background values are zoomed in (down). C, The secondary structure around the RBP binding sites (chrXV, + strand, from 978584 to 978819; indicated by the green box in Figure 6B) in the novel intergenic ncRNA transcript. The structure's z-score and SCI are −27.58 and 0.49, respectively.

by secondary structure, RNA expression and/or protein binding data. These results would be a good starting point for further functional studies.

gPAR-CLIP may generate false positives because the crosslinking can cause nonspecific binding effects, and many transient and dynamic binding interactions may not be detected in limited CLIP-seq experiments. In addition, because of their low expression levels, many lncRNAs [56] may be omitted during transcript assembly using RNA sequencing data and RBP binding site identification using gPAR-CLIP data. Due to the limitations of RNA sequencing and gPAR-CLIP data in uncovering RBP binding signatures in novel lncRNAs, further experiments are needed to validate our initial results and systematically discover more RBP binding signatures in lncRNAs.

RBP binding to RNA has proven to be a complex event in post-transcriptional regulation. RNA primary sequence and, more importantly, secondary structure are now considered to be involved in regulating protein recognition. Our results suggest some biological functions of RNA structures bound by RBPs in a eukaryotic genome. This study is a step toward the long-term goal of understanding how RNA structures are regulated and recognized by proteins at the post-transcriptional level.

This work was supported by the National Natural Science Foundation of China (31271402 and 31100601) and the National Key Basic Research Program (2012CB316503). We thank Ting Han, Mallory Freeberg and John Kim in Department of Human Genetics, University of Michigan for the helpful advice and sharing the gPAR-CLIP data. We also thank Yunjiang Qiu, Yuchuan Wang and Yifang Liu for their advice and help during data analysis. Raw sequence data are available through Gene Expression Omnibus using series entry GSE48888.

1 Keene JD. RNA regulons: coordination of post-transcriptional events. Nat Rev Genet, 2007, 8: 533–543
2 Moore MJ. From birth to death: the complex lives of eukaryotic mRNAs. Science, 2005, 309: 1514–1518
3 Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol, 2007, 8: 479–490
4 Kedde M, Strasser MJ, Boldajipour B, Oude Vrielink JA, Slanchev K, le Sage C, Nagel R, Voorhoeve PM, van Duijse J, Orom UA, Lund AH, Perrakis A, Raz E, Agami R. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell, 2007, 131: 1273–1286
5 Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol, 2008, 6: e255
6 Jamonnak N, Creamer TJ, Darby MM, Schaughency P, Wheelan SJ, Corden JL. Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. RNA, 2011, 17: 2011–2025
7 Granneman S, Petfalski E, Swiatkowska A, Tollervey D. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. EMBO J, 2010, 29: 2026–2036
8 Joshi A, Van de Peer Y, Michoel T. Structural and functional organization of RNA regulons in the post-transcriptional regulatory network of yeast. Nucleic Acids Res, 2011, 39: 9108–9117
9 Lukong KE, Chang KW, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. Trends Genet, 2008, 24: 416–425
10 Iwakiri J, Tateishi H, Chakraborty A, Patil P, Kenmochi N. Dissecting the protein-RNA interface: the role of protein surface shapes and RNA secondary structures in protein-RNA recognition. Nucleic Acids Res, 2012, 40: 3299–3306
11 Valley CT, Porter DF, Qiu C, Campbell ZT, Hall TM, Wickens M. Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. Proc Natl Acad Sci USA, 2012, 109: 6054–6059
12 Oberstrass FC, Lee A, Stefl R, Janis M, Chanfreau G, Allain FH. Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. Nat Struct Mol Biol, 2006, 13: 160–167
13 Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, Darnell JC, Darnell RB. Hits-clip yields genome-wide insights into brain alternative RNA processing. Nature, 2008, 456: 464–469
14 Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr., Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell, 2010, 141: 129–141
15 Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol, 2010, 17: 909–915
16 Uren PJ, Bahrami-Samani E, Burns SC, Qiao M, Karginov FV, Hodges E, Hannon GJ, Sanford JR, Penalva LO, Smith AD. Site identification in high-throughput RNA-protein interaction data. Bioinformatics, 2012, 28: 3013–3020
17 Zhang C, Darnell RB. Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechol, 2011, 29: 607–614
18 Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U. Paralyzer: Definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol, 2011, 12: R79
19 Freeberg MA, Han T, Moresco JJ, Kong A, Yang YC, Lu ZJ, Yates JR, Kim JK. Pervasive and dynamic protein binding sites of the mRNA transcriptome in *Saccharomyces cerevisiae*. Genome Biol, 2013, 14: R13
20 Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 2009, 10: R25
21 Quinlan AR, Hall IM. Bedtools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 2010, 26: 841–842
22 Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science, 2008, 320: 1344–1349
23 Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M. The structure of the eukaryotic ribosome at 3.0 Å resolution. Science, 2011, 334: 1524–1529
24 Yang H, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, Westhof E. Tools for the automatic identification and classification of RNA base pairs. Nucleic Acids Res, 2003, 31: 3450–3460
25 Juhling F, Morl M, Hartmann RK, Sprinzl M, Stadler PF, Putz J. tRNAdb 2009: Compilation of tRNA sequences and tRNA genes. Nucleic Acids Res, 2009, 37: D159–162
26 Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics, 2009, 25: 1974–1975
27 Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO. Precision and functional specificity in mRNA decay. Proc Natl Acad Sci USA, 2002, 99: 5860–5865
28 Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nat Protocols, 2012, 7: 562–578
29 Gruber AR, Findeiss S, Washietl S, Hofacker IL, Stadler PF. Rnaz 2.0: Improved noncoding RNA detection. In: Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing, 2010. 69–79
30 Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res, 2007, 35: W345–349
31 Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res, 2011, 39: D146–151
32 Rabani M, Kertesz M, Segal E. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. Proc Natl Acad Sci USA, 2008, 105: 14885–14890
33 Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E. Genome-wide measurement of RNA secondary structure in yeast. Nature, 2010, 467: 103–107
34 Steigele S, Huber W, Stocsits C, Stadler PF, Nieselt K. Comparative analysis of structured RNAs in *S. cerevisiae* indicates a multitude of different functions. BMC Biol, 2007, 5: 25
35 Karaduman R, Fabrizio P, Hartmuth K, Urlaub H, Luhrmann R. RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. J Mol Biol, 2006, 356: 1248–1262
36 Li X, Zhang W, Xu T, Ramsey J, Zhang L, Hill R, Hansen KC, Hesselberth JR, Zhao R. Comprehensive *in vivo* RNA-binding site analyses reveal a role of Prp8 in spliceosomal assembly. Nucleic Acids Res, 2013, 41: 3805–3818
37 Esakova O, Krasilnikov AS. Of proteins and RNA: the RNase P/MRP family. RNA, 2010, 16: 1725–1747
38 Davila Lopez M, Rosenblad MA, Samuelsson T. Conserved and variable domains of RNase MRP RNA. RNA Biol, 2009, 6: 208–220
39 Esakova O, Perederina A, Quan C, Schmitt ME, Krasilnikov AS. Footprinting analysis demonstrates extensive similarity between eukaryotic RNase P and RNase MRP holoenzymes. RNA, 2008, 14: 1558–1567
40 Melnikov S, Ben-Shem A, Garreau de Loubresse N, Jenner L, Yusupova G, Yusupov M. One core, two shells: bacterial and eukaryotic ribosomes. Nat Struct Mol Biol, 2012, 19: 560–567
41 Wilson DN, Doudna Cate JH. The structure and function of the eu-

karyotic ribosome. Cold Spring Harbor Perspect Biol, 2012, 4, pii: a011536
42 Klinge S, Voigts-Hoffmann F, Leibundgut M, Ban N. Atomic structures of the eukaryotic ribosome. Trends Biochem Sci, 2012, 37: 189–198
43 Ben-Shem A, Jenner L, Yusupova G, Yusupov M. Crystal structure of the eukaryotic ribosome. Science, 2010, 330: 1203–1209
44 Leshin JA, Heselpoth R, Belew AT, Dinman J. High throughput structural analysis of yeast ribosomes using hSHAPE. RNA Biol, 2011, 8: 478–487
45 Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, Frye M, Helm M, Stoecklin G, Lyko F. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. Nat Struct Mol Biol, 2012, 19: 900–905
46 Novoa EM, Pavon-Eternod M, Pan T, Ribas de Pouplana L. A role for tRNA modifications in genome structure and codon usage. Cell, 2012, 149: 202–213
47 Waas WF, Druzina Z, Hanan M, Schimmel P. Role of a tRNA base modification and its precursors in frameshifting in eukaryotes. J Biol Chem, 2007, 282: 26026–26034
48 Persson BC. Modification of tRNA as a regulatory device. Mol Microbiol, 1993, 8: 1011–1016
49 Alexander RW, Eargle J, Luthey-Schulten Z. Experimental and computational determination of tRNA dynamics. FEBS Lett, 2010, 584: 376–386
50 Olivier C, Poirier G, Gendron P, Boisgontier A, Major F, Chartrand P. Identification of a conserved RNA motif essential for She2p recognition and mRNA localization to the yeast bud. Mol Cell Biol, 2005, 25: 4752–4766
51 Jambhekar A, McDermott K, Sorber K, Shepard KA, Vale RD, Takizawa PA, DeRisi JL. Unbiased selection of localization elements reveals *cis*-acting determinants of mRNA bud localization in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA, 2005, 102: 18005–18010
52 Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. Meme suite: Tools for motif discovery and searching. Nucleic Acids Res, 2009, 37: W202–208
53 Rodriguez-Gabriel MA, Burns G, McDonald WH, Martin V, Yates JR 3rd, Bahler J, Russell P. RNA-binding protein Csx1 mediates global control of gene expression in response to oxidative stress. EMBO J, 2003, 22: 6256–6266
54 Satoh R, Tanaka A, Kita A, Morita T, Matsumura Y, Umeda N, Takada M, Hayashi S, Tani T, Shinmyozu K, Sugiura R. Role of the RNA-binding protein Nrd1 in stress granule formation and its implication in the stress response in fission yeast. PLoS ONE, 2012, 7: e29683
55 Jung HJ, Kim MK, Kang H. An ABA-regulated putative RNA-binding protein affects seed germination of *Arabidopsis* under ABA or abiotic stress conditions. J Plant Physiol, 2013, 170: 179–184
56 Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature, 2012, 482: 339–346

Supporting Information

Figure S1 RBP binding landscape on the secondary structure of RNase P. The black arrows indicate known RBP binding sites based on [1].

Figure S2 RBP binding landscape on the secondary structure of RNase MRP. The black arrows indicate known RBP binding sites based on [1]. The evolutionary conserved GARAR element is indicated by black triangles [2].

Figure S3 RBP binding affinity along box C/D snoRNA structural elements. RBP binding affinity along box C/D snoRNAs, averaged across all transcripts used in our analysis (top). The length of different transcripts is normalized to a scale ranging from 0 to 1. The positions of C, D, C' and D' boxes (red), and the two guide regions (green) are indicated at the top. All box C/D snoRNAs can be generally grouped into two distinct categories according to whether containing RBP binding signatures in the C' box motif (bottom).

Figure S4 Secondary structure of U6 snRNA *in vitro*. The diagram is based on [3]. The naked secondary structure of U6 snRNA was obtained by chemical structure probing. It differs significantly from that of U6 snRNP, which exhibits an open conformation.

Figure S5 RBP binding landscape on the secondary structure of 18S rRNA. The diagram is based on co-variance model (http://www.rna.icmb.utexas.edu/). Dashed boxes indicate ribosomal protein interaction sites on 18S rRNA, based on various experimental data (adopted from [4]).

Figure S6 RBP binding sites on the secondary structure of ES12S. The secondary structure of ES12S (top). The 5′ and 3′ coordinate of ES12S in 18S rRNA is 1224U and 1259A, respectively. The numbers of contacts across the ES12S (bottom).

Figure S7 RNA structural motifs predicted from whole 3′ UTRs and RBP binding regions with different decay rates. The structural motifs are derived from fast- and slow-decaying mRNAs. The 5′ end of the motifs is circled.

Figure S8 RNA structural motifs predicted from whole 3′ UTRs with different decay rates. The structural motifs are derived from fast- (left) and slow-decaying (right) mRNAs. The 5′ end of the motifs is circled.

Figure S9 RBP binding dynamics on 3′ UTRs under nitrogen starvation. All 3′ UTRs can be separated into three distinct groups according to their RBP binding affinities: exhibition of specific binding under WT and nitrogen starvation condition, respectively; and both under the two conditions (heatmap in the left). Predicted structural motifs of specifically bound regions, which are enriched with specific GOs (box in the right).

Figure S10 Length distribution of novel ncRNA transcripts. Number of novel ncRNA transcripts with different length distributions (top). Number of anti-sense and intergenic ncRNA transcripts with different length distributions, respectively (bottom).

Figure S11 RBP binding coverage on different genomic elements, including novel ncRNAs.

Table S1 Description of gPAR-CLIP and RNA-seq libraries

Table S2 Summary of expansion segments (ESs) in ribosomal RNA

Table S3 RBP binding enrichment on all nucleotide modifications in tRNA

Table S4 RBP binding enrichment on four tRNA modifications upon increasing RBP binding affinity thresholds

Table S5 RBP binding sites on known long ncRNAs in budding yeast

Table S6 Novel antisense ncRNA transcripts containing RBP binding sites with 100% located in multiple alignment regions

Table S7 Novel intergenic ncRNA transcripts containing RBP binding sites with 100% located in multiple alignment regions

References

1 Khanova E, Esakova O, Perederina A, Berezin I, Krasilnikov AS. Structural organizations of yeast RNase P and RNase MRP holoenzymes as revealed by UV-crosslinking studies of RNA-protein interactions. RNA, 2012, 18: 720–728
2 Esakova O, Krasilnikov AS. Of proteins and RNA: the RNase P/MRP family. RNA, 2010, 16: 1725–1747
3 Karaduman R, Fabrizio P, Hartmuth K, Urlaub H, Luhrmann R. RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. J Mol Biol, 2006, 356: 1248–1262
4 Granneman S, Petfalski E, Swiatkowska A, Tollervey D. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. EMBO J, 2010, 29: 2026–2036