Supplementary Methods

Media and growth conditions

Yeast strain S288C was grown at 30 C to exponential phase ($4x10^7$ cells/ml) in YPD medium.

RNA preparation

Total RNA was extracted from cells using a slightly modified protocol by using hot, acid phenol (Sigma) (*1*). Poly(A) RNA was obtained by two rounds of selection using the Poly(A) purist Kit according to manufacturer's instructions (Ambion). A diagram outlininging the PARS protocol is provided in Supplementary Fig. 2.

Preparation of RNA transcripts in vitro

RNA transcripts of P4P6, P9-9.2, HOTAIR, and HOTAIR fragments are generated by in vitro transcription using RiboMAX Large Scale RNA production Systems Kit according to the manufacturer's instructions (Promega). The RNA was purified using denaturing polyacrylamide gel electrophoresis (PAGE). The composition of the gel is 8% 19:1 acrylamide:bisacrylamide, 7 M urea, 90 mM Tris-borate and 1 mM EDTA. The gel mix is polymerized using 10% APS and 1% Temed. The RNA bands are visualized by UV shadowing and are excised out of the gel. The RNA is recovered by freeze-thawing the gel slices three times in dry ice, followed by passive diffusion into Rnase free water overnight at 4C with continuous rocking. The RNA is then ethanol precipitated (0.3M Sodium Acetate, 1% glycogen and 3 volumes of cold 100% ethanol) and resuspended in water.

Full length YKL185W (Ash1), a fragment of YNL229C, YLR110C (CCW12), YDL184C (RPL41A) are obtained by PCR using primers against the yeast genome followed by in vitro transcription using RiboMAX Large Scale RNA production Systems Kit according to the manufacturer's instructions (Promega). The RNAs are

purified using RNeasy Mini kit (Qiagen) following manufacturer's instructions.

Enzymatic Structure Probing

5 prime labeling of the RNA with gamma p32 for structure probing:

In vitro transcribed RNA was treated with 5 units of Antarctic Phosphatase (NEB) at 37C for 30min, followed by heat inactivation at 65C for 7min. T4 polynucleotide kinase (PNK) was then used to add [γ-^{32}P]ATP to the 5'end of RNA by incubating at 37C for 30min. An equal volume of RNA loading dye (95% Formamide, 18 mM EDTA, 0.025% SDS, 0.025% Xylene Cyanol, 0.025% Bromophenol Blue) was added, and the RNA was denatured at 70C for 5min before it was run on an 8% denaturing PAGE gel with 7M urea. Bands corresponding to the right size were excised out of the gel. The gel slices were freeze-thawed on dry ice three times, and RNA was recovered by immersion of the gel slice in 100ul of water, at 4C, overnight. The amount of radioactivity present was measured by scintillation spectroscopy.

Structure probing of the labeled RNA:

Prior to structure mapping, the labeled RNA (50000 units per lane) was added to 1ug of total yeast RNA and was renatured by heating to 90C, cooled on ice, and slowly brought to room temperature in structure buffer (10mM Tris pH7, 100mM KCl, 10mM MgCl2). Structure determination was obtained by digesting with dilutions of RNase V1 (Ambion) and RNase S1 (Fermentas) at room temperature for 15min. The reaction was stopped by using inactivation and precipitation buffer (Ambion), the RNA was recovered using ethanol precipitation and was dissolved in RNA loading dye. The RNA was resolved by running a 8% denaturing PAGE gel and visualized by exposing a phosphoimager plate.

T1 urea sequencing ladder was obtained by incubating labeled RNA, mixed with 1ug of total RNA, in sequencing buffer (20mM sodium citrate pH 5, 1 mM EDTA, 7 M urea) at 50C for 5min. The samples were cooled to room temperature and cleaved using 10-100 fold dilutions of RNase T1 for 15min. The reaction was stopped by

adding inactivation and precipitation buffer (Ambion), and the RNA was recovered using ethanol precipitation and dissolved in RNA loading dye. The RNA was resolved by running a 8% denaturing PAGE gel.

Alkaline hydrolysis ladder was obtained by incubating labeled RNA in alkaline hydrolysis buffer (50mM Sodium Carbonate [$NaHCO_3/Na_2Co_3$] pH 9.2, 1 mM EDTA) at 95C for 5-10min. An equal volume of the RNA loading dye was added to the fragmented RNA and resolved using 8% denaturing PAGE gel.

Quantification of band intensities

Band intensities on the sequencing gel are quantified using the program semi automated footprinting analysis (SAFA) (*2*).

SOLiD™ Library construction

P4P6, P9-9.2, HOTAIR, YKL185W and fragment of YNL229C were doped into the double-selected poly(A)+ mRNA as controls. The RNA pool was then folded and probed for structure using 0.01U of RNase V1 (Ambion), or 1000U of S1 nuclease (Fermentas), in a 100ul reaction volume, as described above. To capture the cleaved fragments and convert them into a library for Solid sequencing, we used the SOLiD™ Small RNA Expression Kit (Ambion) and modified the manufacturor's instructions as follows.

Briefly: RNase V1 and S1 nuclease cleaved RNA pool was fragmented using alkaline hydrolysis buffer at 95C for 3min. The fragments were resolved on a 6% denaturing PAGE gel and a band corresponding to 75-200bases of RNA size was excised out of the PAGE gel. The gel slice was frozen and thawed three times and crushed. RNA was recovered by passive diffusion into water at 4C, overnight, followed by ethanol precipitation. The RNAs were ligated to 5'adaptors by adding T4 RNA ligase2 and adaptor mixA (SOLiD™ Small RNA Expression Kit) and incubating at 16C,

overnight. The RNA was then treated with Antarctic Phosphatase (NEB), 37C for 1 hour, and heat inactivated at 65C for 7min. Adaptor mixA was re-added to the RNA to maximize ligation to the 3' end of the RNA and incubated at 16C for 6hours. Reverse transcription was carried out using ArrayScript reverse transcriptase (Ambion) and RNA was removed using RNase H. 18-20 rounds of PCR were carried out using SOLiD PCR primers provided in the kit.

SOLiD™ Sequencing

cDNA libraries were amplified on beads by emulsion PCR, and the resulting beads were deposited onto the surface of a glass slide according to the standard protocol described in the SOLiD Library Preparation Guide (Applied Biosystems). 35-50 bp sequences were generated on a SOLiD™ System sequencing platform according to the standard protocol described in the SOLiD Instrument Operation Guide (Applied Biosystems). The sequences generated were further analyzed according to the methods described elsewhere in this manuscript.

Sequence mapping

Obtained sequences were truncated to 35bp before mapping, and required to map uniquely to either the yeast genome or transcriptome, allowing up to one mismatch and no insertions or deletions. Mapping results are provided in Supplementary Table 2.

Mapping of the short reads to the yeast transcriptome was done using version 1.1.0 of SHRiMP (*3*) downloaded from http://compbio.cs.toronto.edu/shrimp/. We required the alignment to start from the first base of the read, as PARS relies on the first base to recover a valid enzyme cleavage point. Reads that were not uniquely mapped were discarded and all genomic locations to which those reads mapped were marked as 'unmappable' due to ambiguity. In addition, genomic locations from which no reads were obtained in any of the replicates were also marked 'unmappable'.

Genome and transcriptome assembly

The yeast genome was downloaded from The Saccharomyces Genome Database (SGD, http://www.yeastgenome.org/) on June 2008. The yeast transcriptome was assembled by SGD annotations (downloaded June 2008). Untranslated regions (UTR) lengths were taken from Nagalkshmi *et al* (*4*). The set of genes predicted to encode secretory proteins is based on Emanuelsson et al (*5*).

Quantifying cleavage data

For each nucleotide along a transcript, we count the number of reads whose first mapped base was one base 3' of the inspected nucleotide.

The *load* of a transcript is defined as the total number of reads that mapped to the transcript, divided by the effective transcript length, which is the annotated transcript length minus the number of unmappable locations (see “sequence mapping” above). This measure is a proxy to the transcript's abundance in the sample. The *ratio* score of a nucleotide is defined as the ratio between the number of reads obtained for that nucleotide and the *load* of that transcript.

Detecting overlapping peaks

We compute the abovementioned *ratio* score separately for the V1- and S1-treated samples. Nucleotides for which this computed ratio is greater than one are defined as “peaks”. An “overlapping peak” nucleotide is a nucleotide for which both the V1 and S1-computed ratios exceed 1. Statistics about those nucleotides are provided in Supplementary Table 4; the full list of those nucleotides can be downloaded from our website (see “Online Resources” below).

Computing the PARS Score

For each nucleotide, we compute the logarithm of the ratio between the number of reads obtained for that nucleotide in the V1-treated sample and that obtained in the S1-treated sample.

Specifically, the PARS Score is defined as the $\log_2$ of the ratio between the number of times the nucleotide immediately downstream to the inspected nucleotide was observed as the first base when treated with RNase V1 and the number of times it was observed in the RNase S1 treated sample. The score of base *i* is thus defined as:

$$Score_i = \log_2\left(\frac{|V1_{i+1}|+1}{|S1_{i+1}|+1}\right)$$

where $|V1_{i+1}|$ and $|S1_{i+1}|$ are the number of times the nucleotide immediately downstream to the inspected nucleotide was observed as the first base of a sequence read in the V1- and S1- treated samples, respectively.

To account for differences in overall sequencing depth between the V1- and S1-treated samples, the number of reads for each nucleotide is normalized prior to the computation of the ratio:

$$|V1_i| = k_V \cdot |RawV1_i|$$
$$|S1_i| = k_S \cdot |RawS1_i|$$

Where $RawS1_i$ and $RawV1_i$ are the raw number of reads observed for nucleotide *i* in the V1 and S1 treated samples, respectively, and the normalizing constants k_v and k_s are computed as follows:

$$k_v = \frac{\left(|V1| + |S1|\right)/2}{|V1|}$$

$$k_s = \frac{\left(|V1| + |S1|\right)/2}{|S1|}$$

Higher PARS (and positive) scores indicate higher double stranded propensity and lower (and negative) scores indicate that the base was less likely to be in a double-stranded conformation. We cap the PARS score to ±7. Nucleotides with zero evidence counts on both lanes have a zero PARS score and are excluded from all subsequent analysis.

Enrichment of Gene Ontology annotations in over- and under-structured genes

For each gene, we separately computed the average PARS score of its 5' UTR, CDS, and 3' UTR, and used the Wilcoxon rank sum test to ask whether genes with similar Gene Ontology (GO) annotations tend to have similar average PARS scores in any of the inspected regions. Multiple-hypothesis correction was done by FDR with a cutoff of 0.05. The Wilcoxon rank sum test results obtained for each gene set are listed in Supplementary Table 5.

Predicted structure data

We used the Vienna package (*6*) to fold transcripts, calculate the partition function of the structures ensemble and base pairing probabilities. We examined global and local (folding in selected short sliding windows) folding schemes. To compute the pairing probability of a nucleotide we re-fold the transcript for every window, moving the window a single basepair at a time, and take the average pairability reported for that nucleotide across all windows that cover it.

Comparison between the PARS readout and the predicted pairability (Fig. 4a and Supplementary Fig. 13) was done by binning nucleotides showing similar PARS scores (±0.5) and computing the average pairability assigned by the prediction

algorithm. The predicted pairability scores shown in Fig. 4a relate to a 200-bp local folding window.

Periodicity and codon signature

Periodicity analysis was done by a straightforward application of Discrete Fourier Transform to the average PARS score collected from the following genomic features: last 100 bases of the 5' UTR, first 200 bases of the coding sequence, 100 first bases of the 3' UTR.

The codon signature shown in the inset of Fig. 3b was computed by separately averaging the PARS score reported for each codon position, collected from the entire coding sequence of each of the 3000 mRNAs that went into our analysis. The reported *p*-values are computed by applying a t-test on the distribution of PARS scores of the different codon positions.

Clustering structure profiles

We applied *k*-means clustering to the structural profiles of all genes whose 5' UTR is at least 50 bases long. To bring all profiles to the same baseline we use a relative PARS score, which is obtained by subtracting the average PARS score of the gene from each nucleotide. To account for missing values in the clustering, we first smooth the profile by interpolating neighboring data (±10 window average) to assign a PARS score to bases that were unmappable. We require no missing values for further analysis.

Online Resources

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (*7*) and are accessible through GEO Series accession number GSE22393. Nucleotide-resolution raw reads and PARS scores for the 3000 genes included in our analysis can be visualized and downloaded at http://genie.weizmann.ac.il/pubs/PARS10

References

1. A. Lee, K. D. Hansen, J. Bullard, S. Dudoit, G. Sherlock, *PLoS Genet* **4**, e1000299 (Dec, 2008).

2. R. Das *et al.*, *RNA* **11**, 344-354 (2005).

3. S. M. Rumble *et al.*, *PLoS Comput Biol* **5**, e1000386 (May, 2009).

4. U. Nagalakshmi *et al.*, in *Science*. (2008), vol. 320, pp. 1344-9.

5. O. Emanuelsson, S. Brunak, G. von Heijne, H. Nielsen, *Nat Protoc* **2**, 953 (2007).

6. I. L. Hofacker, M. Fekete, P. F. Stadler, *J Mol Biol* **319**, 1059 (Jun 21, 2002).

7. R. Edgar, M. Domrachev, A. E. Lash, *Nucleic Acids Res* **30**, 207 (Jan 1, 2002).

Supplementary Note: Current methods for measuring RNA structure

Experimentally, advanced methods for measuring RNA structure such as X-ray crystallography[1], NMR[2], and cryo-electron microscopy[3], provide detailed three-dimensional descriptions of the probed RNA. However, these methods can only probe a single RNA structure per experiment, and are limited in the length of the probed RNA. Indeed, only ~750 structures from various organisms were collectively solved by these methods in the past three decades[4], the vast majority of which being very short RNAs (<50 nucleotides).

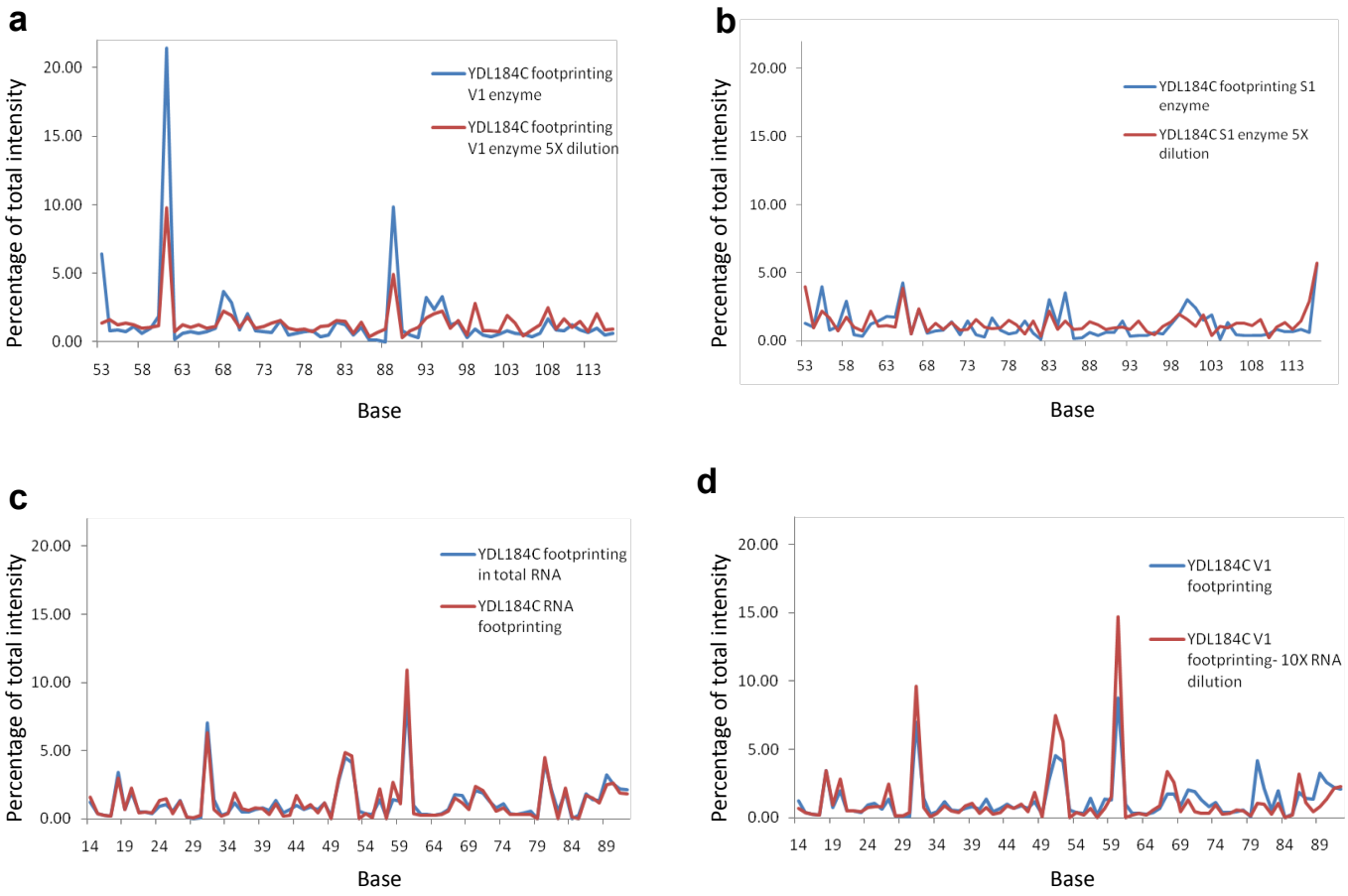
As they are easier to implement, chemical and enzymatic probing methods have become widely used for RNA secondary structure analysis[5-7]. For example, the analyzed RNA can be radiolabelled at one end and digested with an RNase that preferentially cuts double-stranded nucleotides. The length distribution of the resulting RNA fragments is then used to infer which nucleotides of the original RNA molecule were in a double-stranded conformation. Enzymatic probing, however, is also limited to the measurement of one RNA structure per experiment, and depending on whether the enzymatic activity is assayed using standard gel or capillary electrophoresis, only ~100-600 nucleotides can be analyzed at a time[8,9]. Although there has been considerable success in probing RNA structures of increasing lengths[10-12], no genome-scale collection of RNA structures currently exists.

Given the experimental difficulties in measuring RNA structure, algorithms for predicting RNA structure from primary sequence have been developed and applied in many settings[13-19]. Although prediction algorithms achieve accuracies of ~40-70% [20,21], their predictive power is limited by the complexity of modeling important factors such as long-distance intramolecular connections or pseudoknots. More importantly, since there is little experimental data regarding how environmental factors such as changes in pH, temperature, or interactions with metabolites and RNA binding proteins affect RNA structure, these effects cannot be predicted reliably with existing algorithms.

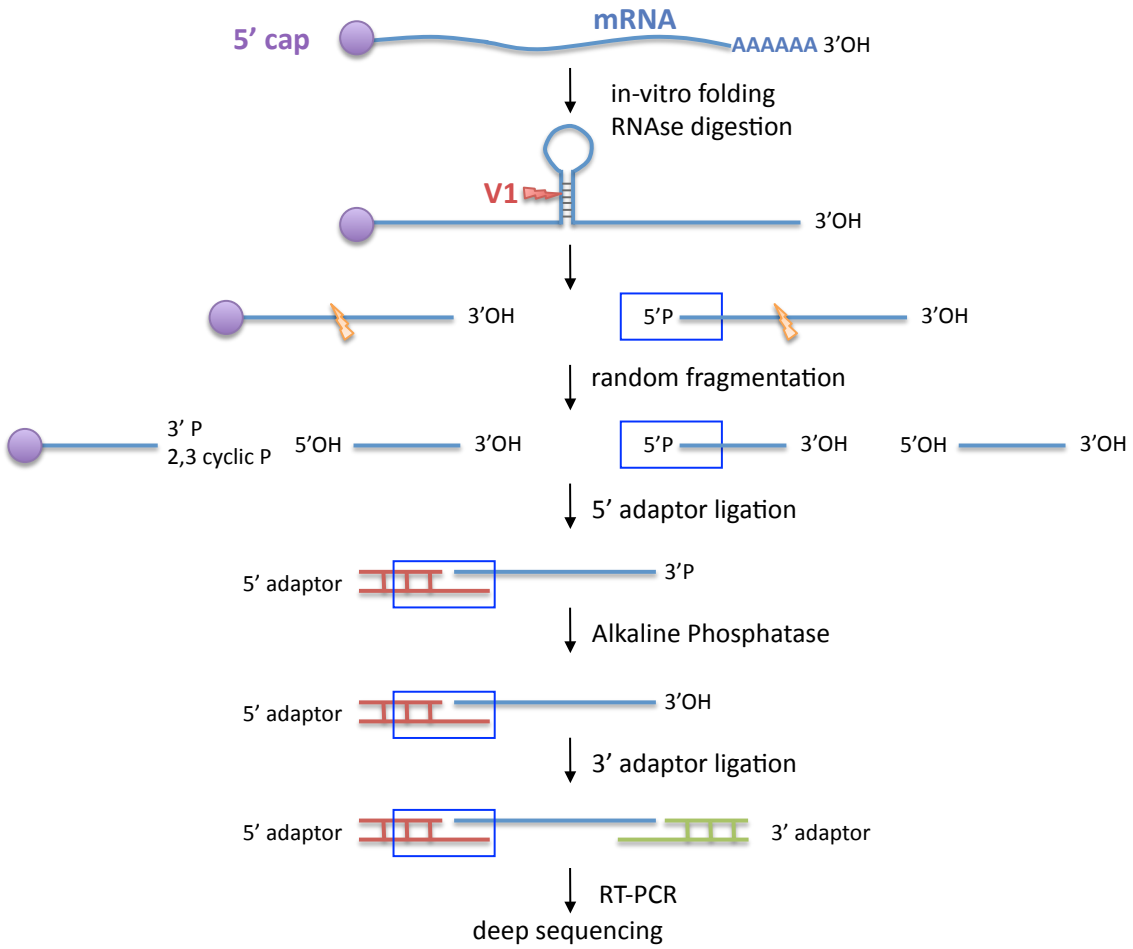
References

1. Guo, F., Gooding, A.R. & Cech, T.R. Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* **16**, 351-62 (2004).
2. Latham, M.P., Brown, D.J., McCallum, S.A. & Pardi, A. NMR methods for studying the structure and dynamics of RNA. *Chembiochem* **6**, 1492-505 (2005).
3. Mueller, F. et al. The 3D arrangement of the 23 S and 5 S rRNA in the Escherichia coli 50 S ribosomal subunit based on a cryo-electron microscopic reconstruction at 7.5 A resolution. *J Mol Biol* **298**, 35-59 (2000).
4. Bernstein, F.C. et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* **112**, 535-42 (1977).
5. Brenowitz, M., Chance, M.R., Dhavan, G. & Takamoto, K. Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical "footprinting". *Curr Opin Struct Biol* **12**, 648-53 (2002).
6. Alkemar, G. & Nygard, O. Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. *Biochemistry* **45**, 8067-78 (2006).
7. Romaniuk, P.J., de Stevenson, I.L., Ehresmann, C., Romby, P. & Ehresmann, B. A comparison of the solution structures and conformational properties of the somatic and oocyte 5S rRNAs of Xenopus laevis. *Nucleic Acids Res* **16**, 2295-312 (1988).
8. Deigan, K.E., Li, T.W., Mathews, D.H. & Weeks, K.M. Accurate SHAPE-directed RNA structure determination. in *Proc Natl Acad Sci USA* Vol. 106 97-102 (2009).
9. Das, R. et al. Structural inference of native and partially folded RNA by high-throughput contact mapping. *Proc Natl Acad Sci U S A* **105**, 4144-9 (2008).
10. Mitra, S., Shcherbakova, I.V., Altman, R.B., Brenowitz, M. & Laederach, A. High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res* **36**, e63 (2008).
11. Wilkinson, K.A. et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6**, e96 (2008).
12. Watts, J.M. et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-6 (2009).
13. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-15 (2003).
14. Hofacker, I.L., Fekete, M. & Stadler, P.F. Secondary structure prediction for aligned RNA sequences. *J Mol Biol* **319**, 1059-66 (2002).
15. Do, C.B., Woods, D.A. & Batzoglou, S. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* **22**, e90-8 (2006).
16. Mathews, D.H., Sabina, J., Zuker, M. & Turner, D.H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**, 911-40 (1999).

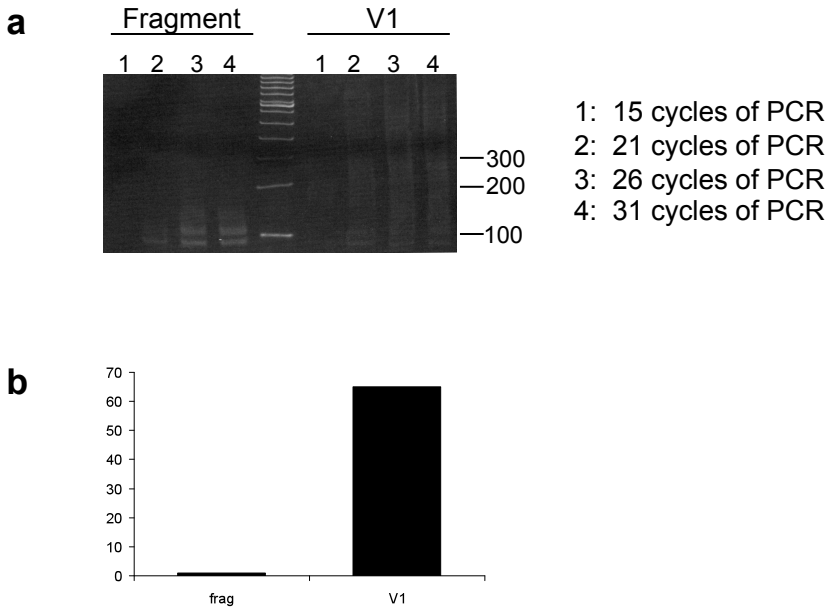
17. Mathews, D.H. Revolutions in RNA secondary structure prediction. in *J Mol Biol* Vol. 359 526-32 (2006).
18. Rabani, M., Kertesz, M. & Segal, E. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. in *Proc Natl Acad Sci USA* Vol. 105 14885-90 (2008).
19. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. in *Nat Genet* Vol. 39 1278-84 (2007).
20. Dowell, R.D. & Eddy, S.R. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinformatics* **5**, 71 (2004).
21. Doshi, K.J., Cannone, J.J., Cobaugh, C.W. & Gutell, R.R. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* **5**, 105 (2004).



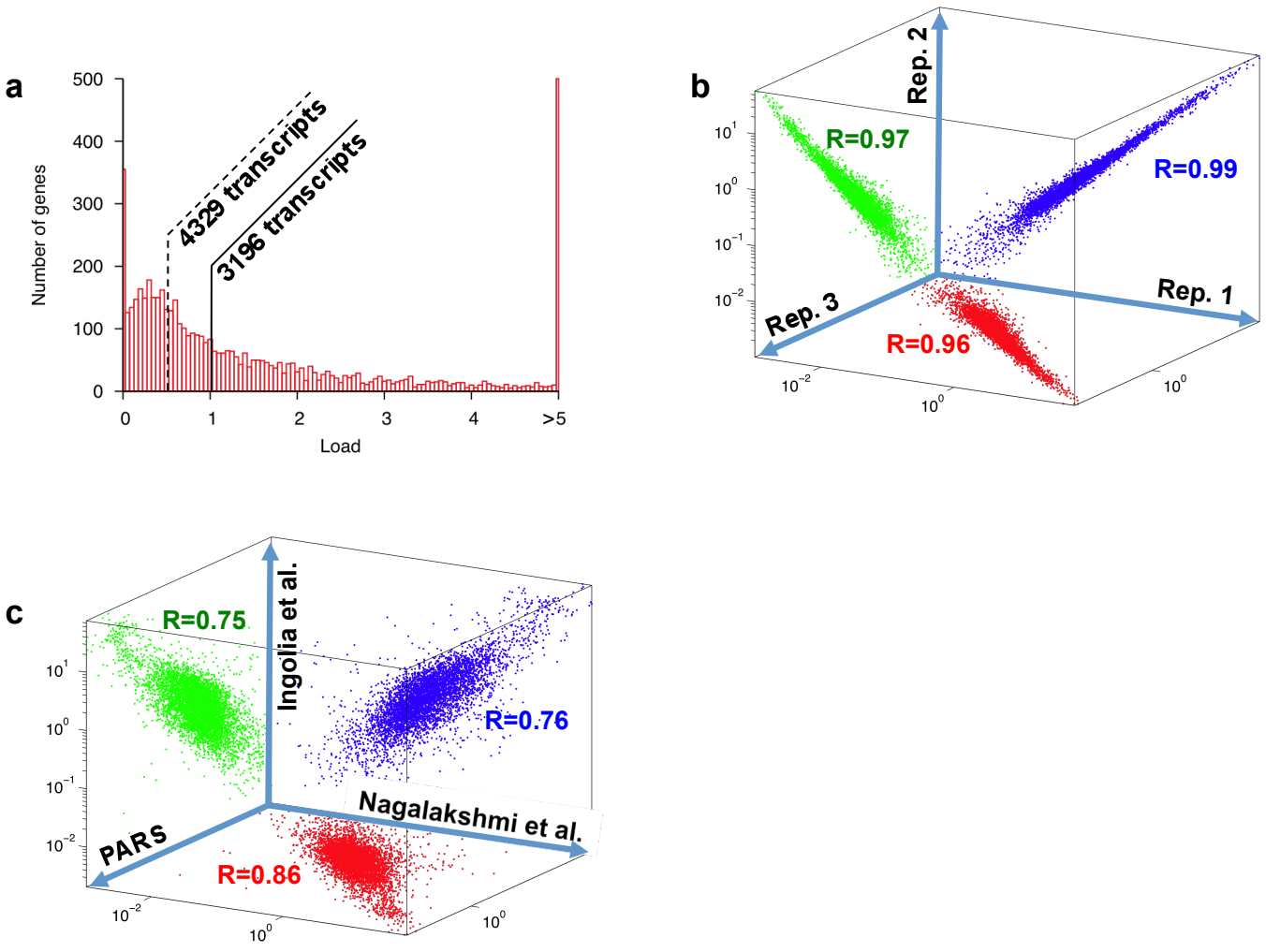
Supplementary Figure 1. The enzyme concentration used in PARS cuts RNA with single hit kinetics and occur at regions resulting from intra-molecular interactions. **(a)** Shown are traces indicating footprinting intensities of P32-labeled *in vitro* transcribed YDR184C that were quantified using SAFA. The footprint of YDR184C obtained with the RNase V1 concentration used in PARS matches very well with the footprint obtained with a 5-fold dilution of RNase V1 (correlation=0.93). **(b)** The footprint of YDR184C obtained with the RNase S1 concentration used in PARS matches well with the footprint obtained with a 5 fold dilution of RNase S1 (correlation=0.65). **(c)** P32-labeled RNA is folded and cleaved either by itself or is folded and cleaved in a population of mRNAs. YDR184C folds into a similar structure when it is alone in solution or when it is in the presence of other RNAs (correlation=0.97); **(d)** P32 RNA mixed with yeast mRNAs is either folded at 10ul or folded at 100ul (10X) before being cleaved by RNase V1. YDR184C folds into a similar conformation with or without 10X dilution, indicating that most of the folding is driven by intra-molecular interactions (correlation=0.9). Y-axis shows the intensity of cleavage per base, calculated using the percentage of intensity at the base over the total intensity of the area probed.



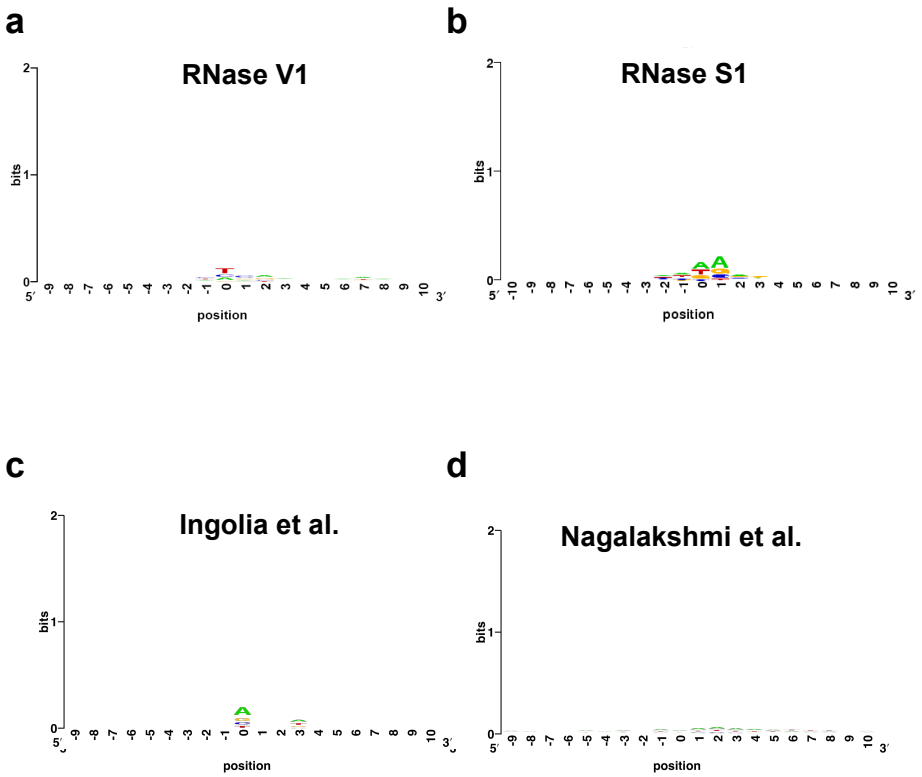
Supplementary Figure 2. Detailed schematic of RNA structure probing by high throughput sequencing. Total RNA is isolated from yeast, selected for poly(A)+ transcripts and renatured in vitro. The folded RNA is then cut by RNase V1 (shown), or RNase S1, resulting in 5'P. The RNA then undergoes fragmentation. Although fragmentation generates different products, only the RNA that has been cut by RNase V1 or S1 contains 5'P that is ligation competent. The RNA then undergoes size selection, followed by 5'adaptor ligation. Fragmentation products with 3'P are converted to 3'OH by alkaline phosphatase making them able to ligate to 3' adaptors. This is followed by reverse transcription, PCR and another size selection to make a library that is suitable for high throughput sequencing.



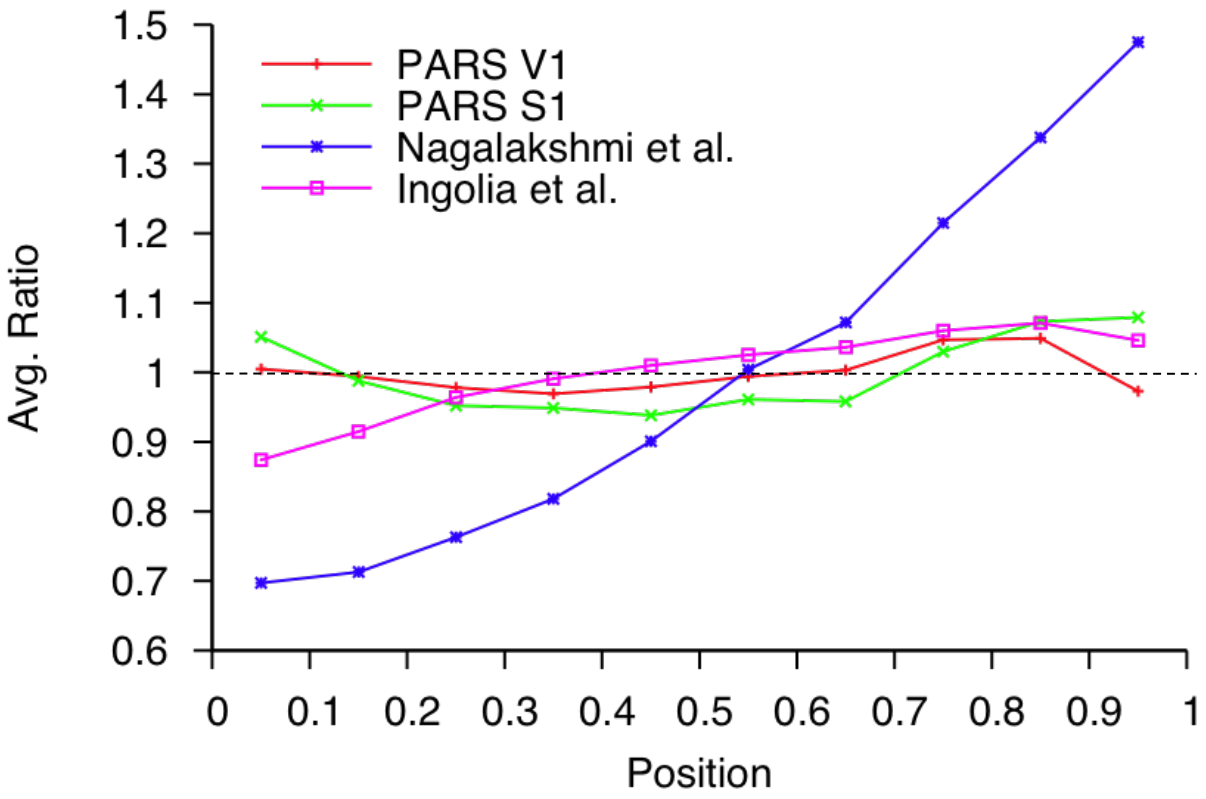
Supplementary Figure 3. PARS captures fragments generated from V1 cleavages and not random fragmentation products from alkaline hydrolysis. (a) RNA libraries ran on 5% native PAGE and stained using ethidium bromide. Fragments above 120 bases indicate yeast RNA fragments that were ligated to adaptors and cloned into a library. The RNAs are either treated ("V1") or not treated ("Fragment") before they are fragmented at 95C for 3.5min. They are then ligated to 5' and 3' adaptors. Lanes 1, 2, 3 and 4 refer to the amount of library that is amplified with 15, 21, 26, and 31 cycles of PCR. We typically excise between 150 bases to 250 bases from the native PAGE gel for high throughput sequencing. **(b)** qPCR quantification of the library after 18 cycles of PCR amplification and size selection between 150-250 bases using native PAGE.



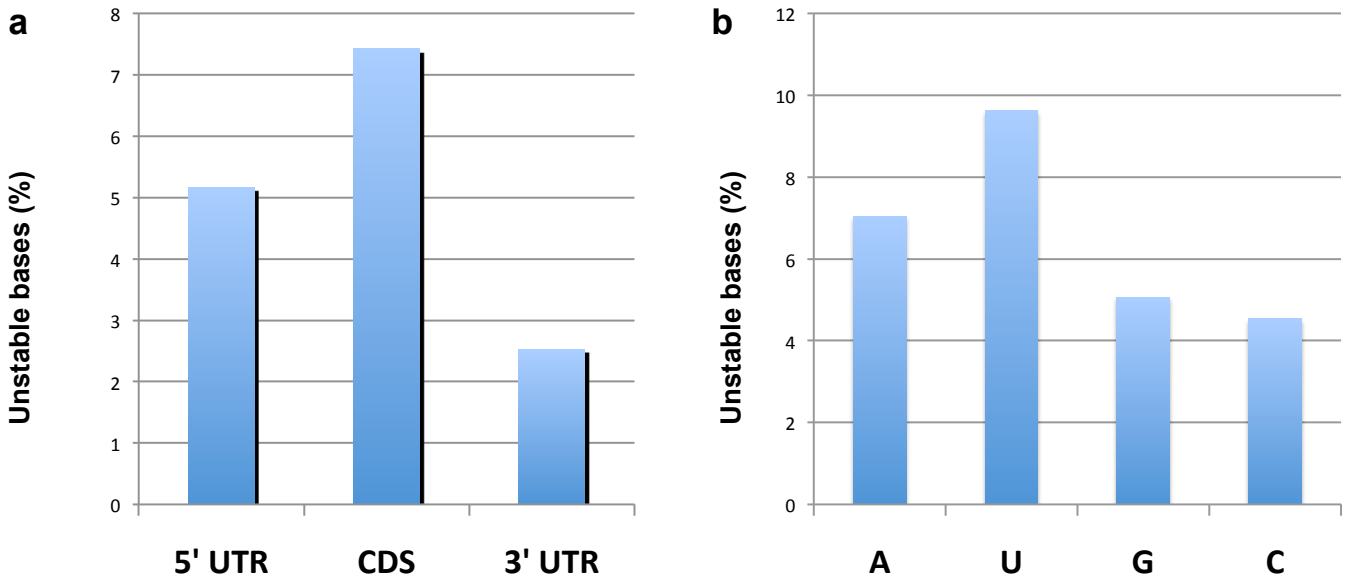
Supplementary Figure 4. PARS samples cleaved RNA fragments in proportion to their abundance. (a) Histogram showing for the number of transcripts as a function of load obtained by merging the readout of all replicates of our PARS experiments. Applying a threshold of load > 1, we obtain structural information for 3196 transcripts (solid black line). By performing more sequencing runs, better coverage can be obtained, allowing PARS to obtain structural information for many more transcripts. For example, it is likely that structural information regarding ~1100 additional transcripts will be obtained by doubling the number of sequencing runs (dashed line) **(b)** Comparison of mRNA abundance levels per transcript between three biological replicates of our samples treated by the double-stranded cutter V1. The abundance level of each transcript is computed as the total number of reads mapped to the transcript divided by the transcript length. **(c)** Same as (b), but when comparing our abundance levels and those of the ribosomal profiling method[16] and RNA-Seq method[21]. The high correlation between mRNA coverage measurements across biological replicates (correlation>0.96), as well as among our measurements and the previous sequencing-based approaches (correlation=0.86 and 0.75 respectively) suggests that our protocol cleaves and captures RNA fragments in proportion to their abundance in the initial pool.



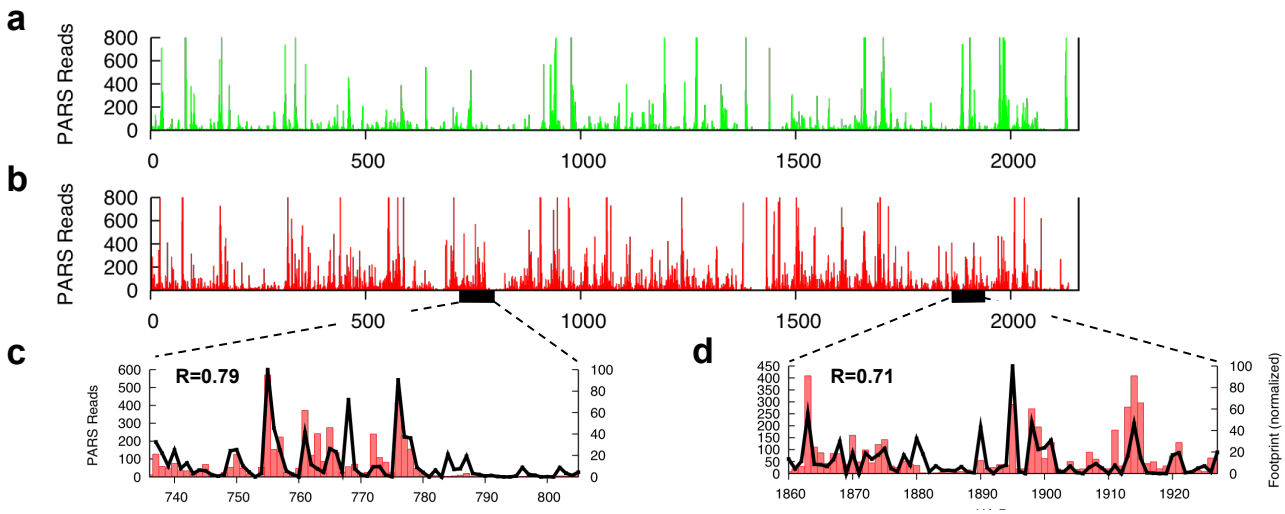
Supplementary Figure 5. PARS has minimal sequence-dependent bias. (a) To determine whether there is a bias towards RNA fragments with particular sequences, we examined the nucleotide distribution over the first bases of the sequenced fragments. Shown is the sequence specificity across all sequence reads that we uniquely mapped to the genome from our generated V1 libraries. The specificity was derived from an alignment of the 20 nucleotides in the genome that surround the first mapped base of each sequence read and are shown as a standard position specific scoring matrix (PSSM), which displays the information content of the nucleotide distribution at each position of the alignment. The sequence composition at these bases does not show a strong sequence bias at the first base or around it, suggesting that RNase cleavage, adaptor ligation, and cDNA conversion does not introduce significant sequence biases. **(b)** Same as (a), for the data obtained from our S1 libraries. **(c,d)** Same as (a), for the data obtained in the RNA-Seq study[21] and the ribosomal profiling study[16].



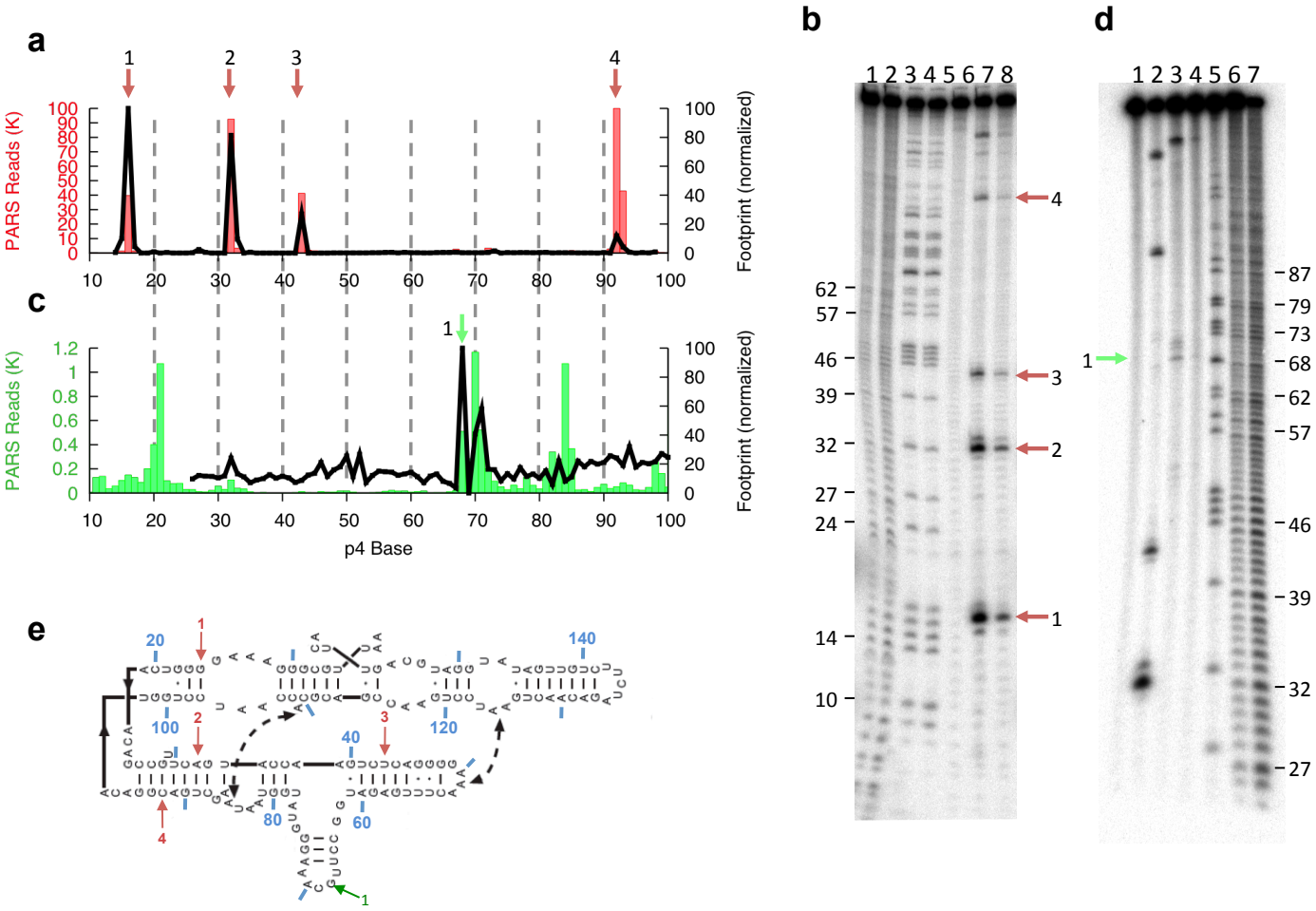
Supplementary Figure 6. PARS has minimal bias towards particular regions of the transcript. As we are interested in structural information across the entire transcript, it was important to know whether we obtained reads uniformly from both the 5' and the 3' end of the transcripts. Shown is the number of sequence reads along each nucleotide of the annotated coding region of each transcript, averaged across all transcripts. The number of sequence reads are shown after normalizing for the abundance of each transcript, by dividing the number of sequence reads at each nucleotide with the total number of reads for its embedding transcript. Since transcripts vary in length, the position of each normalized read is then projected onto a 0-1 range denoting the 5' to 3' end of the coding region of each transcript. Data is shown for our double-stranded (red) and single-stranded (green) cutters, and for the RNA-Seq data[21] (blue) and ribosomal profiling data[16] (pink). Positions that exhibited the largest deviation from the mean coverage in our method had 8% more reads than the mean coverage, suggesting that our protocol has a relatively small bias towards particular regions along the transcript.



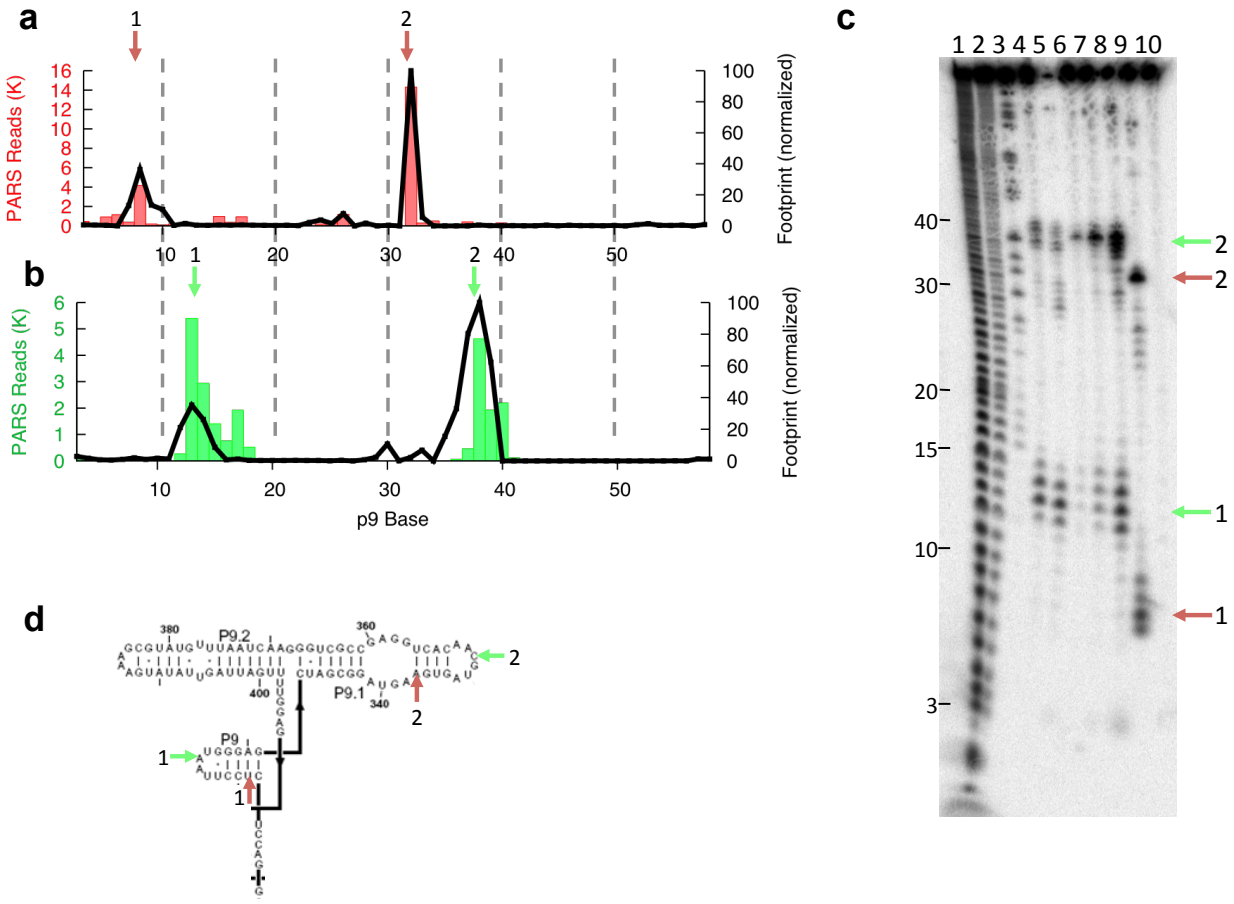
Supplementary Figure 7. Nucleotide composition and location properties of unstable nucleotides within transcripts. Occurrence (in percent) of nucleotides for which a high readout in both V1- and S1-treated lanes was obtained. Those joint peaks could represent areas of the transcript that fold into more than one conformation. Alternatively, they could also represent single-stranded helical regions that can be cleaved by RNase V1. **(a)** Untranslated regions show significantly lower levels of overlapping readout, suggesting that RNA within those areas (and within the 3' UTR in particular) tend to fold into a single conformation, whereas the coding region is more structurally dynamic. **(b)** A high readout in both V1- and S1-treated lanes is more common for Adenine and Uracil than Guanine or Cytosine, in line with the fact A:U pairing is weaker, and therefore more dynamic than G:C pairing. Interestingly, Uracil shows more instability than Adenine and Guanine more instability than Cytosine, suggesting that G:U pairing might be another cause of unstable base pairing.



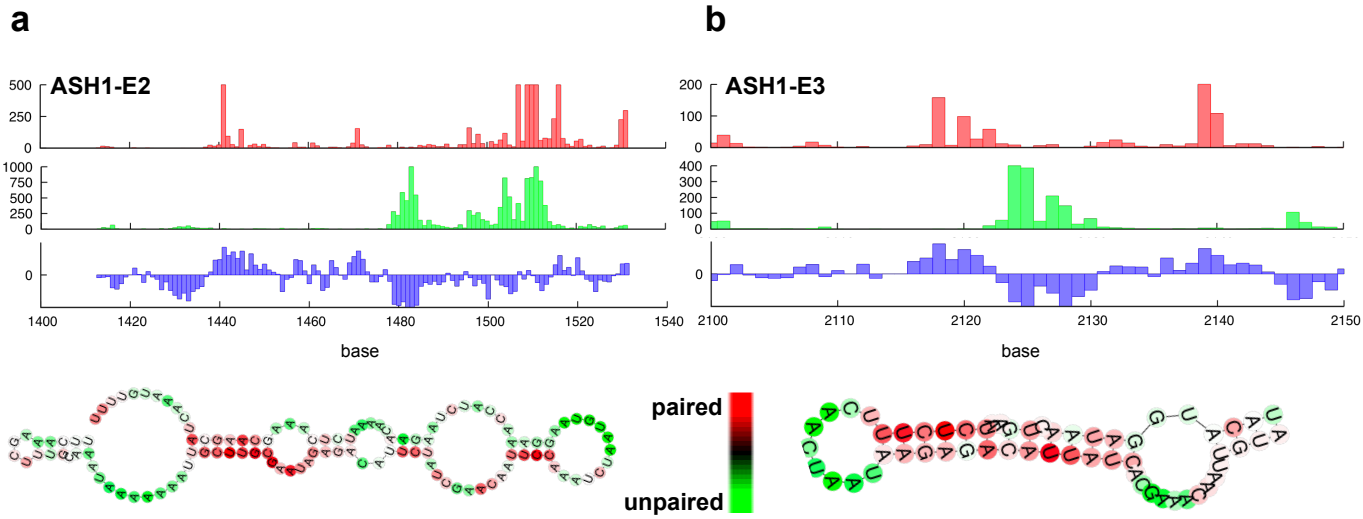
Supplementary Figure 8. PARS is able to probe the structure of HOTAIR, a long non-coding RNA, in a length independent way. (a,b) Single-stranded and double-stranded signal of PARS obtained using the RNase S1 (green bars) and RNase V1 (red bars) across the 2.2kb HOTAIR[8] transcript which we added to our samples and whose structure was previously unknown. **(c,d)** Detailed view of the PARS V1 signal from (b) across two domains from the full transcript. For each domain, shown is the signal obtained when subjecting this domain to traditional footprinting (black line). The correlations between PARS and traditional footprinting are indicated.



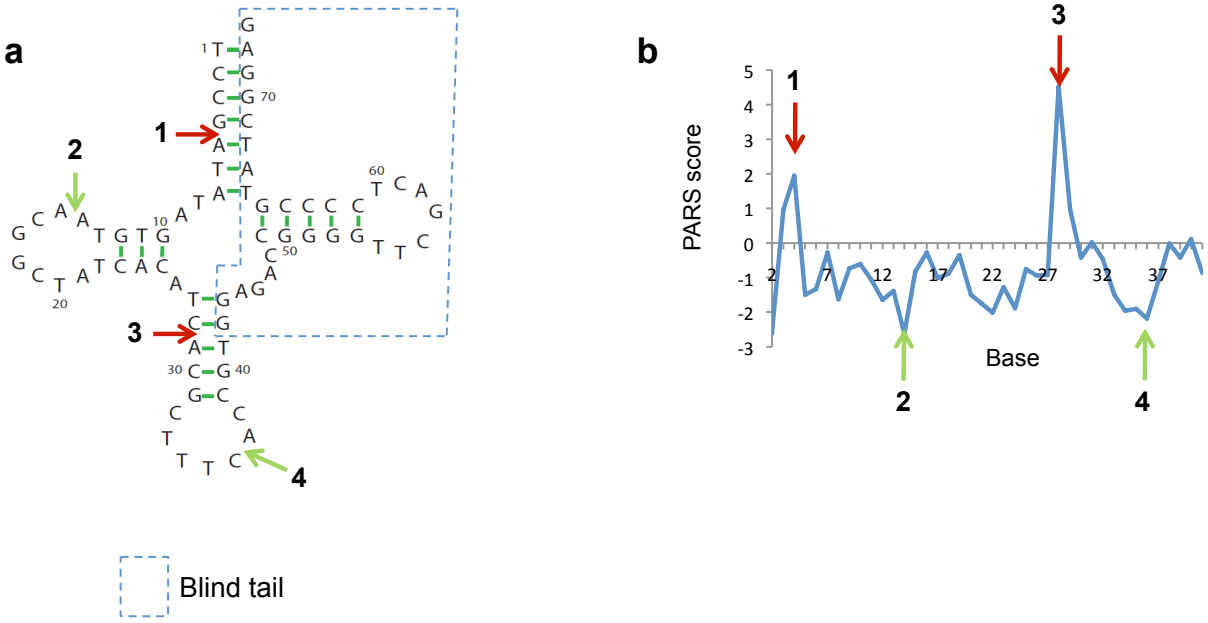
Supplementary Figure 9. PARS correctly recapitulates results of RNA footprinting for the p4p6 domain of the *Tetrahymena* Ribozyme. (a) RNase V1 cleaves the folded p4p6 domain of the *Tetrahymena* ribozyme at four distinct sites, which are accurately captured by PARS. The double-stranded signal of PARS, obtained using the double-stranded cutter RNase V1 (red bars), is shown as the number of sequence reads mapped along each nucleotide of the p4p6 domain. Also shown is the signal obtained on the p4p6 domain using traditional footprinting (black line) and automated quantification of the RNase V1 lane shown in (b). Red arrows indicate cleavages that are seen in gel (b). **(b)** The gel resulting from RNase V1 (Lanes 7,8) enzymatic probing of the p4p6 domain. Alkaline hydrolysis (Lanes 1,2), RNase T1 ladder (Lanes 3,4) and no RNase treatment (Lane 6) are also shown. **(c)** Single-stranded signal of PARS obtained using the single-stranded cutter RNase S1 (green bars), compared to the signal obtained using traditional footprinting (black line). Green arrows indicate cleavages that are seen in gel (d). **(d)** The gel resulting from RNase V1 (Lane 2) and RNase S1 (Lane 3) enzymatic probing of the p4p6 domain. Alkaline hydrolysis (Lanes 6,7), RNase T1 ladder (Lane 5) and no RNase treatment (Lane 4) are also shown. **(e)** Known secondary structure of the p4p6 domain[22]. Arrows mark nucleotides that were identified by both PARS and enzymatic probing as double-stranded (red arrows) or single-stranded (green arrow).



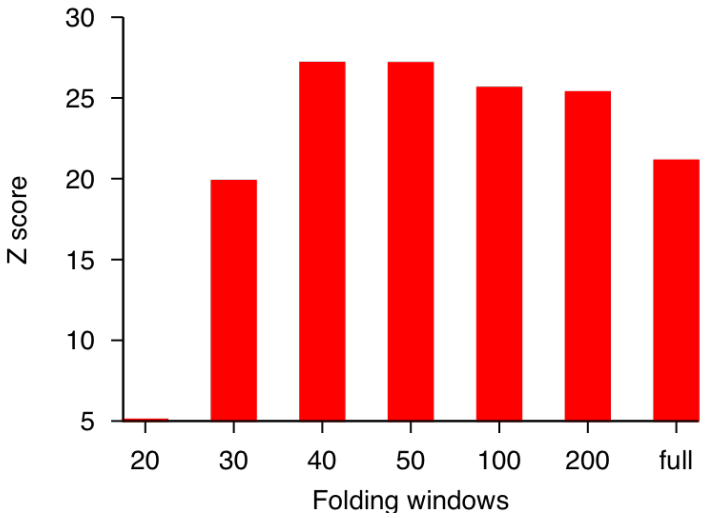
Supplementary Figure 10. PARS correctly recapitulates results of RNA footprinting for the p9-9.2 domain of the *Tetrahymena* ribozyme. (a) RNase V1 cleaves the folded p9-9.2 domain of the *Tetrahymena* ribozyme at two distinct sites, which are accurately captured by PARS. The double-stranded signal of PARS obtained using the double-stranded cutter RNase V1 (red bars) is shown as the number of sequence reads mapped along each nucleotide of the p4p6 domain. Also shown is the signal obtained on the p4p6 domain using traditional footprinting (black line) and automated quantification of the RNase V1 lane shown in (c). Red arrows indicate cleavages that are seen in gel (c). **(b)** Single-stranded signal of PARS obtained using the single-stranded cutter RNase S1 (green bars), compared to the signal obtained using traditional footprinting (black line). Green arrows indicate cleavages that are seen in gel (c). **(c)** The gel resulting from RNase V1 (Lane 9) and RNase S1 (Lanes 7,8,9 at pH7 and Lanes 5,6 at pH4.5). Alkaline hydrolysis (Lanes 1,2), RNase T1 ladder (Lane 3) and no RNase treatment (Lane 10) are also shown. **(d)** Known secondary structure of the p9-9.2 domain[22]. Arrows mark nucleotides that were identified by both PARS and enzymatic probing as double-stranded (red arrows) or single-stranded (green arrows).



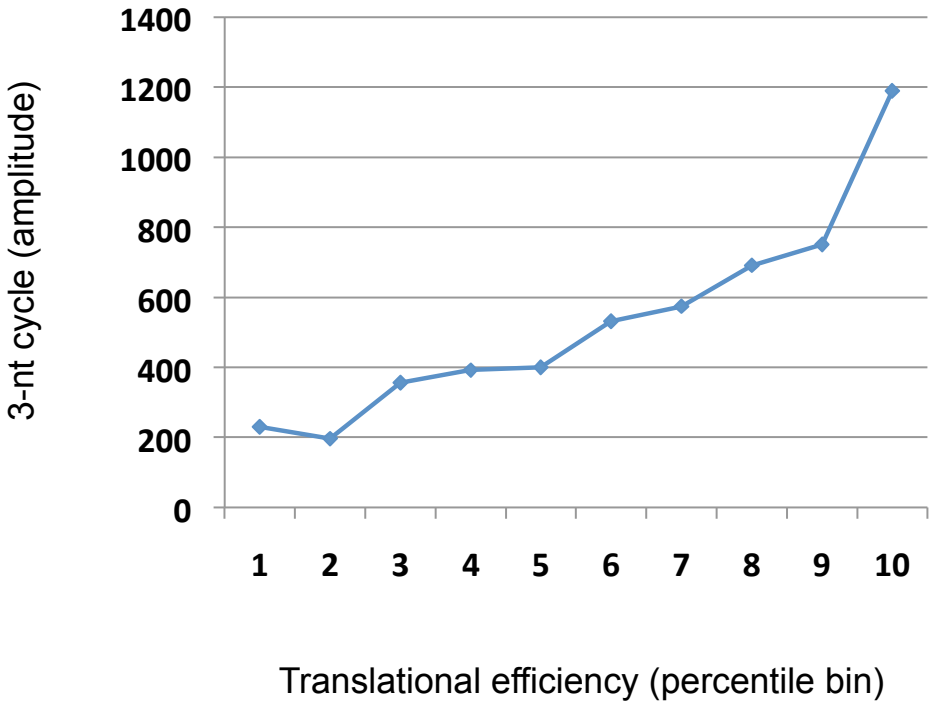
Supplementary Figure 11. PARS correctly recapitulates known RNA structures. Raw number of reads obtained using RNase V1 (red bars) or RNase S1 (green bars) and the resulting PARS score (blue bars) along the inspected domain of ASH1-E2 **(a)** and ASH1-E3 (**b**). Also shown are the known structures of the inspected domains. Nucleotides are color-coded according to their computed PARS score (paired nucleotides are marked in red, unpaired nucleotides are marked in green).



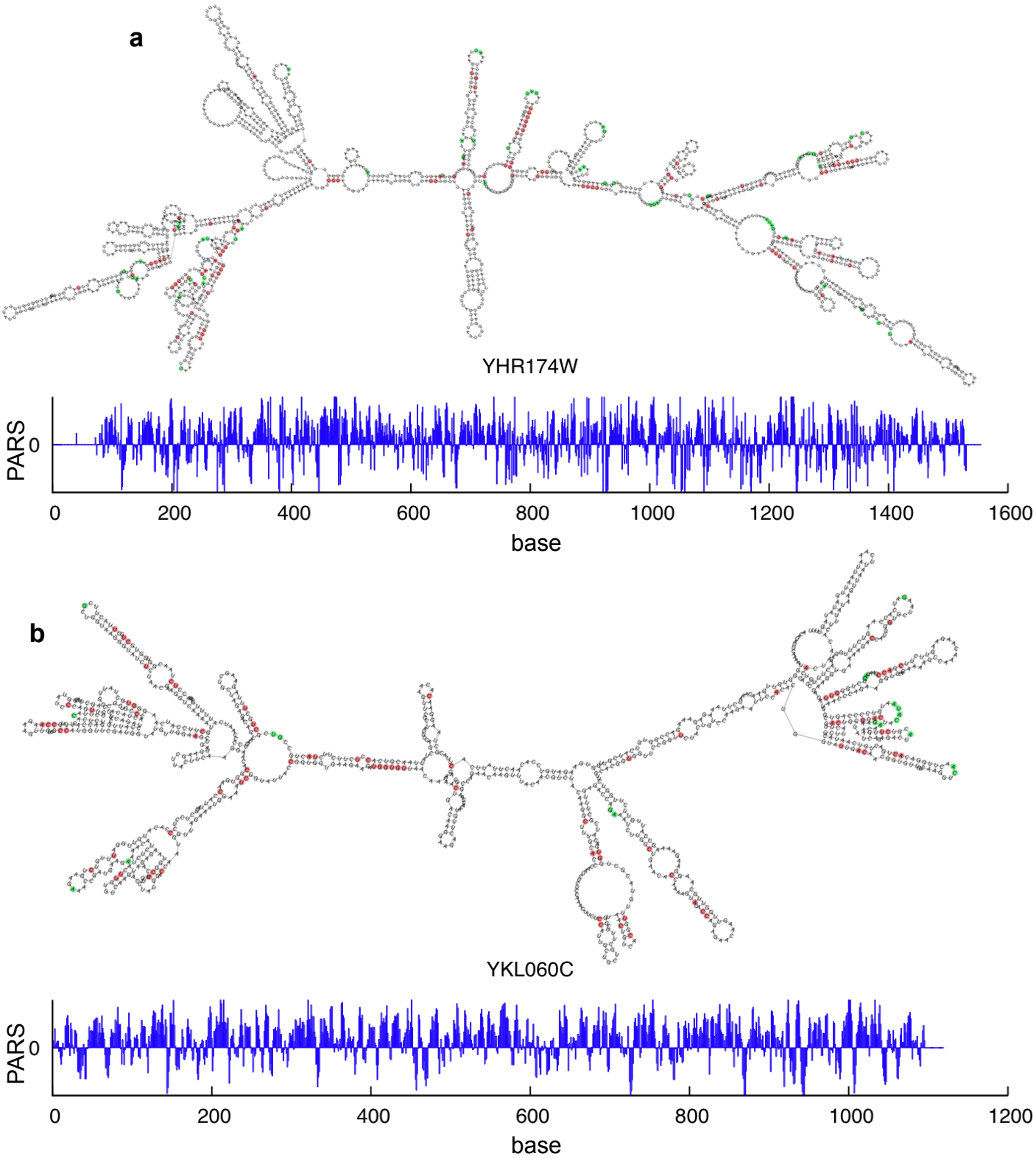
Supplementary Figure 12. PARS correctly recapitulates the secondary structure of glu-tRNA. (a) Known secondary structure of the yeast glu-tRNA. The boxed area in gray is not detectable by sequencing. Green arrows indicate nucleotides with low PARS score (single-stranded) and red arrows indicate nucleotides having a high PARS score (double-stranded). **(b)** The PARS score along this non-coding transcript.



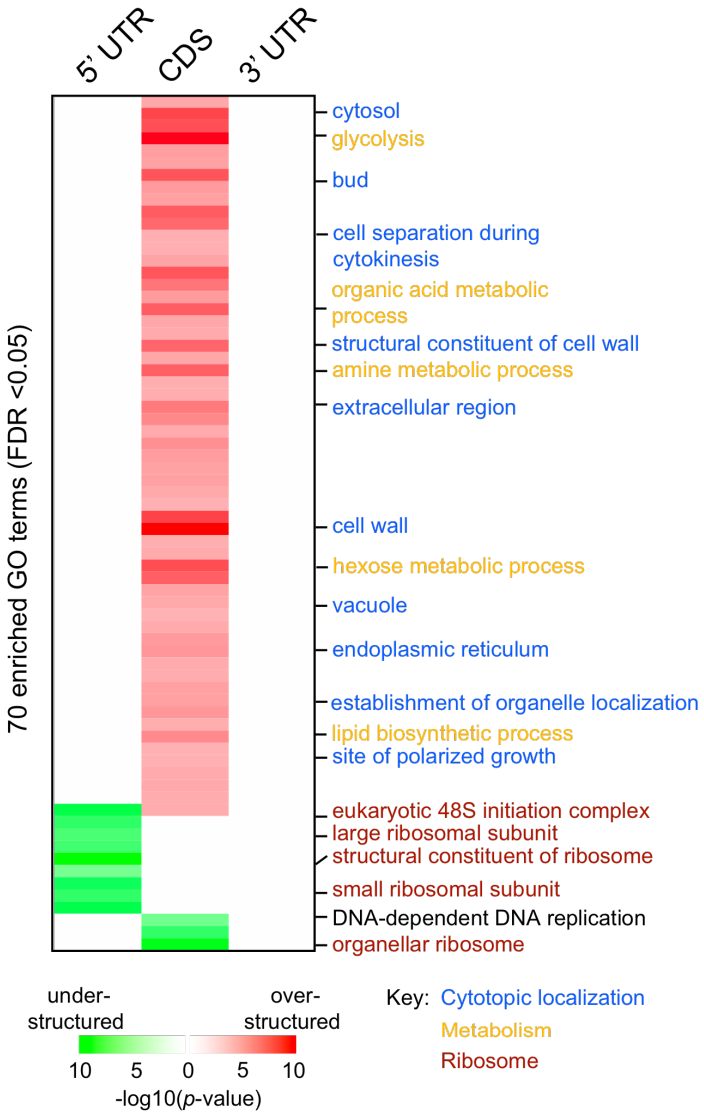
Supplementary Figure 13. The effect of folding window size in computational predictions of RNA structure on correspondence to PARS. Shown is the z-score obtained by comparing the average prediction score for bases which obtained a high PARS score (≥7) to random shuffle. Similar results are obtained when computationally folding the yeast transcriptome in windows ranging from ~40 nucleotides up to windows that cover the entire transcript, suggesting that folding algorithms correctly capture local interactions but do not improve in accuracy using the entire transcript.



Supplementary Figure 14. The coding region of genes with high translational efficiency tends to show a stronger three-nucleotide periodicity. Genes were sorted by their measured ribosome occupancy[16] (a proxy for translational efficiency) and divided into ten equally-sized sets. We applied a discrete Fourier transform analysis to the average PARS signal of each of those sets separately, and plotted the amplitude of the three-nucleotide periodic signal. While the three-nucleotide periodic signal is apparent in all sets, it significantly increases with translational efficiency (R^2=0.93, p=1.1x10^{-5}). Transcript abundance is also significantly associated with translational efficiency. Multivariate regression analysis with both periodicity and transcript abundance shows that periodicity is still a significant predictor of translational efficiency (p=0.02). The association between periodicity and translational efficiency independent of transcript abundance is strongest for the top 10 percentile of most efficiently translated transcripts (p=0.004).



Supplementary Figure 15. Examples of PARS-assisted structure predictions for long mRNAs. Shown is the PARS score along the transcript (bottom) and the predicted secondary structure obtained by providing the prediction algorithm[12] constraints extracted from our experimental measurements. Nucleotides having a high PARS score (>7, marked in red) are constrained to be double-stranded whereas nucleotides having low PARS scores (<-7, marked in green) are constrained to be single-stranded.



Supplementary Figure 16. Distinct patterns of secondary structures in mRNA are associated with cytotopic localization and protein function. For each gene, we separately computed the average PARS score of its 5' UTR, CDS, and 3' UTR. For each of these three regions, we then used the Wilcoxon rank sum test to compute a p-value for whether genes with similar Gene Ontology (GO) annotations have PARS scores that are higher or lower than expected. Multiple-hypothesis correction was done by FDR with a cutoff of 0.05. The Wilcoxon rank sum test results for each GO category are listed in Supplementary Table 5. As can be seen, mRNAs encoding proteins with specific sub-cellular localizations (blue) or function in several metabolic pathways (yellow) tend to have excess secondary structure in the coding regions, while mRNAs encoding ribosomal proteins (dark red) tend to have less secondary structure than expected in their 5'UTR and CDS.

Replicates		Correlation
RNase S1 rep. 1	RNase S1 rep. 2	**0.93**
RNase S1 rep. 1	RNase S1 rep. 3	**0.76**
RNase S1 rep. 2	RNase S1 rep. 3	**0.60**
RNase V1 rep. 1	RNase V1 rep. 2	**0.75**
RNase V1 rep. 1	RNase V1 rep. 3	**0.73**
RNase V1 rep. 1	RNase V1 rep. 4	**0.62**
RNase V1 rep. 2	RNase V1 rep. 3	**0.91**
RNase V1 rep. 2	RNase V1 rep. 4	**0.61**
RNase V1 rep. 3	RNase V1 rep. 4	**0.64**

Supplementary Table 1: Reproducibility of PARS signal at single nucleotide resolution. For each pair of replicates, the table shows the Pearson correlation coefficient when comparing the raw number of reads obtained for each nucleotide of the two replicates.

Lane	Input reads	Mapped to genome	%	Mapped to transcript	%	Uniquely mapped	%
V1 (1)	11863204	9279439	78.22	9011318	75.96	6629996	55.89
V1 (2)	7160815	4600352	64.24	4473520	62.47	3341250	46.66
V1 (3)	6528001	4356121	66.73	4250412	65.11	3097273	47.45
V1 (4)	29180169	20033105	68.65	19508366	66.85	12976611	44.47
S1 (1)	23115630	16076977	69.55	15783615	68.28	11086282	47.96
S1 (2)	23023964	15713013	68.25	15426287	67.00	10668210	46.34
S1 (3)	23749516	16537687	69.63	16223036	68.31	10896861	45.88
TOTAL	124621299	86596694	69.49	84676554	67.95	58696483	47.10

Supplementary Table 2: Mapping statistics of the obtained sequences to the yeast genome and transcriptome for all our sequencing runs. Table columns show, for each replicate ("lane"), the number of raw sequences obtained ("input reads"), the number of sequences that mapped to the yeast genome or transcriptome ("mapped to genome", "mapped to transcriptome" respectively), and the number of reads which mapped uniquely within the transcriptome.

	Total	Unstable	Unstable (%)
5' UTR	205084	10579	5.16
CDS	3722310	276273	7.42
3' UTR	357679	9035	2.53
Total	4285073	295887	6.91

	Total	Unstable	Unstable (%)
A	1362616	95883	7.04
U	1234260	118864	9.63
G	869515	43984	5.06
C	818682	37156	4.54
Total	4285073	295887	6.91

Supplementary Table 4: List of nucleotides showing strong peaks in both V1- and S1-treated samples. Table lists the genomic locations (coding sequence, 3' UTR or 5' UTR) of the nucleotides for which a strong signal was obtained in both V1- and S1-treated samples, as well as the base composition of those joint peaks. A list containing the exact position of each of those ~300,000 nucleotides can be downloaded from our website (Methods).