Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing

Yue Wan[1,2], Kun Qu[1,2], Zhengqing Ouyang[3,4] & Howard Y Chang[1,2]

[1]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California, USA. [2]Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California, USA. [3]The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA. [4]Department of Biomedical Engineering, University of Connecticut, Storrs, Connecticut, USA. Correspondence should be addressed to H.Y.C. (howchang@stanford.edu).

Published online 4 April 2013; doi:10.1038/nprot.2013.045

RNA structure is important for RNA function and regulation, and there is growing interest in determining the RNA structure of many transcripts. Here we provide a detailed protocol for the parallel analysis of RNA structure (PARS) for probing RNA secondary structures genome-wide. In this method, enzymatic footprinting is coupled to high-throughput sequencing to provide secondary structure data for thousands of RNAs simultaneously. The entire experimental protocol takes ~5 d to complete, and sequencing and data analysis take an additional 6–8 d. PARS was developed using the yeast genome as proof of principle, but its approach should be applicable to probing RNA structures from different transcriptomes and structural dynamics under diverse solution conditions.

INTRODUCTION

RNA has an important role in regulating cellular gene expression[1]. In addition to its primary sequence content, RNA folds into extensive secondary and tertiary structures that can influence its function. The structure of RNA is involved in every step of its life cycle, including cellular processes such as transcription, splicing, localization, translation and decay[2]. The secondary and tertiary conformations of RNA can also govern the ability of RNA molecules to catalyze reactions, as well as to bind to specific ligands and cellular factors to regulate gene expression under different environmental stimuli[3,4]. Therefore, studying RNA structures and their organization in the cell is crucial to our understanding of cellular function.

Traditionally, the secondary structure of an RNA molecule of interest can be probed in solution, one molecule at a time, using chemicals and enzymes that either cleave or modify single- or double-stranded regions[5–7]. The cleaved sites can be identified through size fractionation by running a sequencing gel on the cleaved RNA pool. Alternatively, the modification and cut sites can be identified by reverse transcription (RT), whereby the RT enzyme is either blocked by the modification site or falls off the cleaved fragment. The exact location of modification sites can be determined by capillary electrophoresis (CE). Hundreds of nucleotides can be read out in one CE run, and recently multiplexing has increased the throughput of structure probing[8,9]. However, probing the secondary structures of long RNAs and/or large numbers of RNAs is still a relatively slow and tedious process. Furthermore, the identity and sequence content of the RNA needs to be known in order to design primers for RT, making *de novo* probing of RNAs of unknown sequence impossible with this approach. Large-scale RNA structural information on most of the RNA in the cell is needed to fully understand the effect of RNA structure on the biology of cells. To this end, we and others have developed high-throughput methods for coupling structure determination with deep sequencing in order to probe the secondary structures of thousands of RNAs simultaneously[10–14].

Overview of PARS

We recently developed PARS, which enables high-throughput experimental determination of the secondary structure of a solution of thousands of yeast-derived RNAs[10]. Double- and single-stranded bases that are present in a pool of isolated cellular RNAs are cleaved using RNase V1 and S1 nuclease, two nucleases that are double- and single-strand specific, respectively (**Fig. 1**). Instead of detecting the cleavage sites by running a sequencing gel or CE experiment, the cleavage sites are identified by converting the RNA into a cDNA library and performing high-throughput sequencing on this library. The secondary structures of thousands of RNAs can be determined simultaneously because the deep sequencing reads can be mapped specifically to the yeast transcriptome, thus enabling researchers to identify where the cleavage event has occurred. The extent of enzymatic cleavage at a base is estimated by counting the number of reads that start at that particular base. Depending on the sequencing depth of the libraries, structural information on thousands of RNAs can be obtained. This information can provide insight into the structural properties of most mRNAs in the cell, as well as into how structural organization of groups of genes can affect cellular biology.

Here we describe a detailed experimental and analysis protocol for genome-wide RNA structure probing in yeast using PARS. The PARS protocol begins with the isolation of total RNA from log-phase growing yeast and ends with the generation of PARS scores per base and the development of PARS-guided secondary structure models. After partially digesting the RNA with RNase V1 or S1 nuclease for structure probing, the RNA is fragmented randomly to a size of ~200 nt in order to enable cloning (**Fig. 1**). As the RNase V1 and S1 nucleases cut RNA, leaving behind 5′-phosphate (5′P) and 3′-hydroxyl (3′OH) groups, the base immediately after a paired (in the case of RNase V1) or unpaired (in the case of S1 nuclease) base can be ligated to a 5′ adapter. By contrast, random fragmentation products contain 5′-hydroxyl (5′OH) groups that cannot be ligated. Subsequent 3′-end treatment, 3′ adapter ligation, RT and PCR steps result in selective amplification of nuclease-cleaved fragments because they contain both 5′ and 3′ adapters. Upon cDNA library construction and deep sequencing, the sequencing reads are mapped against the yeast transcriptome (**Fig. 2**). As the two nucleases cut at the 3′ end of either paired or unpaired bases, the actual cleavage site is 1 nt in front of the base to which the double- and single-stranded sequencing

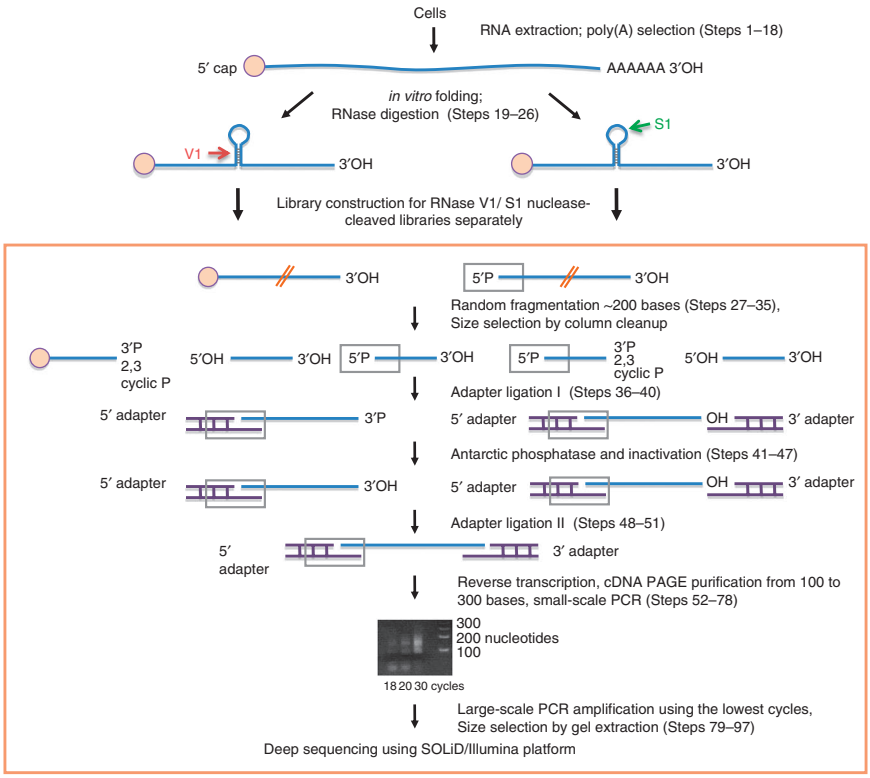

Figure 1 | Detailed experimental schematic of PARS. Total RNA is isolated from cells, enriched for poly(A)$^+$ transcripts and renatured *in vitro*. The folded RNA is then cut by RNase V1 and S1 nuclease separately, resulting in 5′P overhangs. For each of the two pools of cleaved RNA, the RNA is fragmented and made into a cDNA library. However, only the sites cleaved by RNase V1 or S1 nuclease contain 5′Ps that are ligation-competent (the ends of RNA fragments where these overhangs are located are highlighted by gray rectangles). The RNA is then size-selected, followed by 5′ adapter ligation. Fragmentation products with 3′P groups are converted to 3′OH groups by Antarctic phosphatase, enabling these products to be ligated to 3′ adapters. This step is followed by reverse transcription, size selection and PCR to produce a cDNA library that is suitable for high-throughput sequencing. This figure is partially reproduced from Kertesz *et al.*[10].

reads are mapped. The ratio of the number of double- and single-stranded reads per base enables the experimenter to calculate a structure accessibility score (PARS score) that indicates whether a base is likely to be paired or unpaired, thus becoming part of a double- or single-stranded region of an RNA, respectively. PARS information can be integrated into SeqFold software to provide experimentally guided secondary structure models[15]. In total, the experimental procedure takes 5 d, sequencing takes 4–5 d and the analysis takes ~2–3 d to complete.

Applications and limitations

PARS has also been modified and applied to the study of changes in RNA structure across a temperature gradient (parallel analysis of RNA structure with temperature elevation, PARTE)[16] to address questions on the melting landscape of the yeast transcriptome, including (i) whether regions exist with different degrees of structural stability, (ii) where these regions are and (iii) how differences in structural stability can affect RNA function. We used RNase V1 digestion to probe for local RNA melting in double-stranded regions in the yeast transcriptome from 23 to 75 °C (ref. 16) and calculated a melting temperature per base. We observed that regions near the start codon of an mRNA tend to have low melting temperatures, whereas regions near the stop codon have high melting temperatures. Furthermore, low melting temperatures of certain RNAs also seem to facilitate their decay during heat shock.

In addition to yeast, the PARS approach should be applicable to probing *in vitro* the structures of RNAs isolated from many other organisms, including those with larger and more complex transcriptomes; it should also be useful for studying structural RNA changes under different solution conditions. However, performing PARS on organisms with highly repetitive genomes may be challenging because of the difficulty in mapping the sequencing reads to the individual transcripts accurately. Similarly to an RNA-sequencing library, paired-end sequencing of a PARS library can enhance mapping of sequencing reads, as information from the two ends of the read can extend beyond the repetitive region and reside in a region with unique sequence content. This enables the accurate assignment of the read to a specific transcript.

Alternative methods for probing RNA structures

In addition to PARS, several groups have also developed high-throughput methods to scale-up RNA structure probing. These methods include SHAPE-CE, dsRNA-seq, FragSeq and SHAPE-seq[9,11–13]. These methods use enzymes or small molecules to probe the RNA structure and use either capillary sequencing or deep sequencing to read out the cleavage or modification sites. The structural information of RNA can also be integrated into prediction algorithms to generate accurate secondary structure models. However, there are several limitations associated with the different methods, preventing them from being utilized in a truly genome-wide manner. In SHAPE-CE and SHAPE-seq, flexible bases on RNAs are selectively acylated by chemicals such as NMIA and 1M7 (refs. 9,13). Modifications in the RNA can then be read out as RT stop sites because reverse transcriptase is blocked by the modification. In SHAPE-CE, the structural features of individual RNA sequences several hundreds of bases long are detected at each step of the procedure, thus making genome-wide structure probing for thousands of RNA molecules tedious. In SHAPE-seq, the cDNA RT stop sites are read out by deep sequencing, enabling a more complex pool of RNAs to be probed simultaneously. As each gene in the pool is first individually cloned in front of a structure cassette, which consists of the reverse-priming site[13], the generation of the initial starting pool of RNA for structure probing is time consuming. Furthermore, the sequence of the RNA also needs to be known to design primers and/or clone the transcript, thereby making structural probing of novel, undiscovered RNAs in the cell by SHAPE-CE and SHAPE-seq impossible.

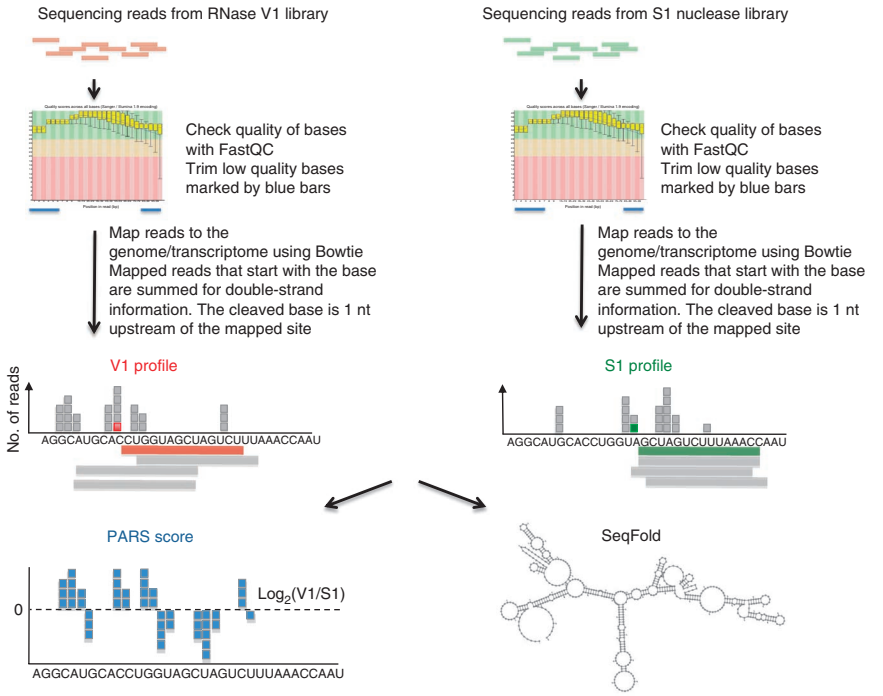

Figure 2 | Detailed analysis pipeline of PARS. Deep sequencing reads from SOLiD or Illumina sequencing can be mapped to the transcriptome using the software Bowtie. (Optional) For Illumina reads, the quality of the sequenced reads can be determined using the program FastQC. Bases with poor-quality scores at the ends of the reads (shown in blue bars) can be trimmed before aligning the bases to the transcriptome to enable more accurate mapping. For every base along a transcript, the number of double- and single-stranded reads that start with the base is read. Structural information is then inferred through the base 1 nt upstream of the mapped base because RNase V1 cleaves RNA cuts after a paired base, whereas S1 nuclease cleaves it after an unpaired base. The PARS score, log ratio of V1/S1, provides information on whether a base is paired or unpaired. The larger the PARS score, the more likely a base is to be paired (and part of a nucleotide that 'terminates' a double-stranded stretch of RNA). Information on double- and single-stranded regions of the RNA can also be integrated into SeqFold, which is a structure prediction program that uses experimental PARS data to guide computationally predicted structure ensembles to provide more accurate RNA secondary structure models. This figure is partially reproduced from Kertesz *et al.*[10].

In FragSeq and dsRNA-seq, enzymatic structure probing has been carried out using single-strand-specific nucleases to probe the structures of the mouse and *Arabidopsis* transcriptomes, respectively, using deep sequencing. FragSeq uses the enrichment of single-stranded reads via P1 nuclease digestion (versus background) to identify single-stranded bases in an RNA[12]. Owing to a lack of information on the location of double-stranded sections, the number of structurally informative bases identified by FragSeq is smaller than in PARS. As there is no random fragmentation step in the FragSeq procedure, FragSeq misses many of the longer RNAs in the transcriptome, resulting in a library that is enriched with shorter RNAs, such as small nucleolar RNAs (snoRNAs), which fall within the cloning range. dsRNA-seq, in contrast, uses RNase I to probe for the presence of paired bases[11] (corresponding to double-stranded RNA regions). However, RNase I and random fragmentation both result in 5′OH and 2,3 cyclic phosphates, making it difficult to distinguish real nuclease cleavages from random fragmentation sites.

Despite the caveats to using SHAPE-CE to obtain genome-scale structure data for thousands of RNAs, SHAPE-CE provides excellent quantification and accurate determination of RNA structure because of a lack of signal amplification that can be introduced in the PCR step in library generation. Furthermore, NMIA and 1M7 are small chemicals that are less sterically hindered than RNA nucleases. As such, structural information for most of the bases along the RNA can be determined[17]. This information can provide very accurate RNA secondary structure models[18] when it is integrated into prediction algorithms such as RNAStructure. Although all high-throughput RNA structure data, including data from SHAPE-seq, FragSeq and PARS, can be integrated into structure prediction algorithms to generate structure models, SHAPE-CE is likely to provide more informative and quantitative data compared with sequencing-based methods when structural information about only a few RNAs is required. When secondary structures of several highly similar RNA sequences are of interest, SHAPE-seq can be a good structure-probing strategy, as the unique barcode on each RNA enables the accurate assignment of sequencing reads to each transcript[13]. In PARS and FragSeq, it may be difficult to assign sequencing reads to the correct transcript in a group of highly similar sequences, as the same read can potentially be mapped to many different places.

Experimental design

Choice of nucleases. One of the main reasons for our decision to use RNase V1 and S1 nuclease as our structural probes, despite the large repertoire of chemical and enzymatic probes used in the literature, is that these enzymes cleave RNA, leaving behind 5′P and 3′OH overhangs, which enables oligonucleotide ligation to a 5′ adapter[7,19,20]. To confirm that the 5′Ps cloned in the procedure are mostly generated from enzymatic cleavage, and not from endogenous 5′Ps already present in the cell, alongside our normal PARS library we built a library of RNA fragments obtained without adding the nucleases to the RNA pool[10]. We observed that we needed many more cycles of PCR amplification (27–28 cycles as opposed to 18) to generate enough DNA material for sequencing when the enzymes were not added. This evidence indicates that the amount of endogenous 5′P is small and that in a PARS library the vast majority of captured 5′Ps comes from enzymatic cleavage.

Many other chemicals and enzymes identify paired bases by cleaving the RNA and leaving behind a 5′OH overhang[6]. These 5′OH overhangs need to be treated with T4 polynucleotide kinase (PNK) in the presence of ATP to generate 5′Ps before they can be ligated. This phosphate addition step can introduce noise into the sequencing data because degraded RNA fragments that are

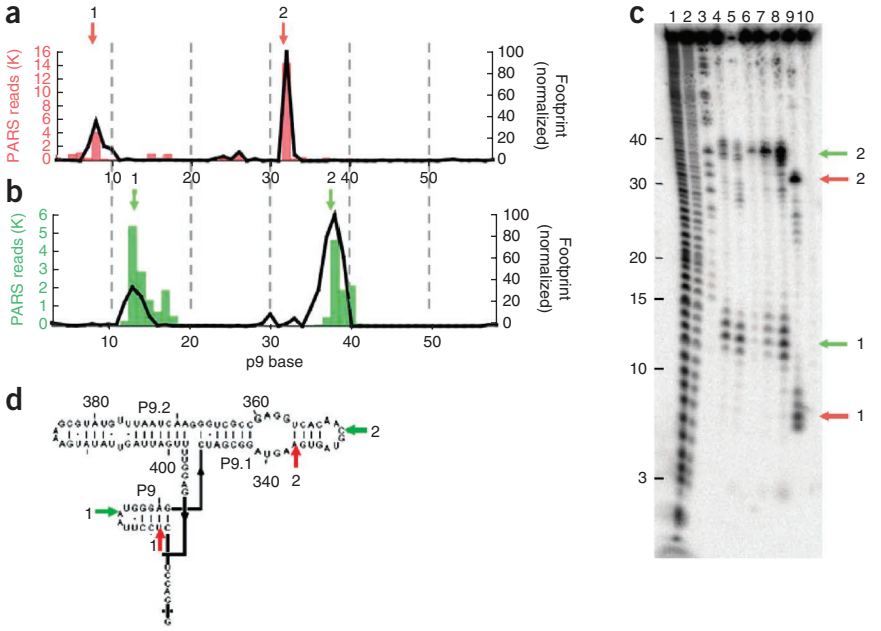

Figure 3 | PARS correctly recapitulates results of RNA footprinting for the P9-9.2 domain of the Tetrahymena ribozyme[10]. (**a**–**c**) RNase V1 cleaves the folded p9-9.2 domain of the tetrahymena ribozyme at two distinct sites, which are accurately captured by PARS[10]. (**a**) The double-stranded 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 P9-9.2 domain. Also shown is the signal obtained on the P9-9.2 domain using traditional footprinting (black line) and a semiautomated quantification of the RNase V1 lane shown in panel (**c**). Red arrows indicate cleavages that are seen in **c**. (**b**) Single-stranded signal of PARS obtained using the single-stranded cutter RNase S1 (green bars), compared with the signal obtained using traditional footprinting (black line). Green arrows indicate cleavages that are seen in **c**. (**c**) The gel resulting from RNase V1 (lane 9) and RNase S1 (lanes 6, 7 and 8 at pH 7 and lanes 4 and 5 at pH 4.5) digestions. Alkaline hydrolysis (lanes 1 and 2), RNase T1 ladder (lane 3) and no RNase treatment (lane 10) are also shown. Arrows mark nucleotides that were identified by PARS as double (red arrows) or single stranded (green arrows). (**d**) The known secondary structure of the p9-9.2 domain. Arrows mark nucleotides that were identified by both PARS and enzymatic probing as double (red arrows) or single stranded (green arrows). This figure is reproduced from Kertesz *et al.*[10].

originally present in the cell are also captured and sequenced. One potential way to distinguish between endogenous breakage and RNA structure-probing cleavage is to clone the endogenous 5′OH groups in the absence of nuclease cleavage. This approach involves the use of T4 PNK to convert endogenous 5′OH into 5′P groups before fragmenting to construct a cDNA library, which could potentially serve as a background lane to subtract noisy signals from the structurally probed RNAs.

As shown in **Figure 3**, according to the model of the secondary structure of the P9-9.2 domain of the tetrahymena ribozyme, there are many more bases located in single-stranded and double-stranded regions than are cleaved by RNase V1 and S1 nuclease. Typically, a strong nuclease cleavage is highly indicative of the double- or single-stranded nature of the base. As both enzymes are comparatively large, they can be sterically hindered such that not all the bases are accessible to them[21]. For example, the S1 nuclease can detect loops at the end of stems better than bulges, as well as mismatches along the stem; also, RNase V1 may only cleave at one side of a stem.

Titration of nucleases for single-hit kinetics. During structure probing, an RNA molecule should be cleaved, on average, once per molecule (single-hit kinetics should be enforced) so that conformational changes that result from the first enzymatic cleavage are not captured because of a second cleavage event. In general, we titrate the amount of enzyme used such that 10–20% of the RNA is cleaved (**Box 1**). As enzyme reactivity can vary from batch to batch, in this protocol RNase V1 and S1 nuclease reaction efficiencies are tested by incubating serial dilutions of each enzyme with a radiolabeled RNA substrate of known secondary structure, such as the tetrahymena ribozyme, for each new batch of enzyme. This titration step is particularly important for experiments that probe RNA structural changes genome-wide under different solution conditions. This is because nucleases can cleave at different rates or become inactive under different reaction conditions, such as changing temperature, in the presence of low salt concentration or in the presence of detergents. For example, enzymes typically work faster at higher temperatures, unless they become denatured. In PARTE, the concentration of RNase V1 used at each temperature is decreased as the temperature increases from 23 to 75 °C, such that a similar amount of RNA is cleaved at each temperature[16]. This approach enables the experimenter to systematically compare RNA structures across different temperatures to identify regions along the RNA—and in the transcriptome—that selectively melt at each temperature. PARTE was used to probe the melting profile of the small cytoplasmic RNA 1 (*SCR1*) RNA to show that different regions of the RNA melt at different temperatures (**Fig. 4** and ref. 16). Furthermore, it is also crucial to ensure that the enzyme retains its double- or single-stranded specificity under different reaction conditions. This can be done by structurally probing an RNA species with a known secondary structure that is not expected to change under different reaction conditions. Similar structural profiles obtained using the same nuclease confirm that the nuclease is functioning properly.

In this protocol, enzymatic titrations are set up in the same reaction conditions as the PARS experiment. A radiolabeled tetrahymena ribozyme is cleaved using different dilutions of RNase V1 and S1 nuclease in the presence of 2 μg of poly(A)$^+$ RNA in a 100-μl reaction solution (**Box 1**). When the resulting RNA pool is run on a sequencing gel, the pattern of enzymatic cleavages should match expectations based on the known secondary structure model[4]. The extent of RNA digestion can be determined by quantifying the amount of full-length, uncut radiolabeled RNA before and after the enzymatic reaction using ImageQuant software**.** The enzyme concentration that results in 10–20% cleavage is then used for downstream experiments to generate PARS sequencing libraries.

Box 1 | Determination of the conditions for single-hit kinetics for RNase V1 and S1 nuclease ● TIMING 3 d

1. 5′-end-label 100 pmoles of the P4P6 domain of the tetrahymena ribozyme RNA using T4 PNK and γ-^{32}P–ATP.
! CAUTION ^{32}P is radioactive. Use safety goggles and lab wear, and follow all relevant radiation safety guidelines.
2. Add 2 µg of poly(A)$^+$-enriched RNA and 50,000 counts of radioactively labeled P4P6 RNA to 80 µl of nuclease-free water in a 0.2-ml PCR tube. Prepare five reaction mixtures each for the RNase V1 and S1 nuclease titration.
3. Heat the RNA at 90 °C for 2 min in a thermal cycler with heated lid on; then immediately place the tubes on ice for 2 min.
4. Add 10 µl of ice-cold 10× RNA structure buffer to the tube and mix by pipetting up and down several times. Transfer the tubes from ice to the thermal cycler; program the thermal cycler such that the temperature increases slowly from 4 to 23 °C over 20 min.
5. Prepare three serial 5× dilutions of S1 nuclease from the stock solution (100 U µl^{-1}) to generate 1:5, 1:25 and 1:125 dilutions of S1 nuclease in nuclease-free water.
6. Prepare a 1:10 dilution of RNase V1 from the stock solution (0.1 U µl^{-1}). Prepare three serial 5× dilutions from the 1:10 dilution to generate 1:50, 1:250 and 1:1250 dilutions of RNase V1 in nuclease-free water.
7. Add 10 µl each of the undiluted, 1:5, 1:25 and 1:125 S1 nuclease solutions to different tubes containing 90 µl of folded RNA kept at 23 °C. As a negative control, add 10 µl of nuclease-free water to a fifth tube containing 90 µl of folded RNA. Mix the solution by pipetting. Incubate the tubes at 23 °C for 15 min.
8. Add 10 µl each of the 1:10, 1:50, 1:250, 1:1250 RNase V1 dilutions to different tubes containing 90 µl of folded RNA kept at 23 °C. As a negative control, add 10 µl of nuclease-free water to a fifth tube containing 90 µl of folded RNA. Mix by pipetting. Incubate the tubes at 23 °C for 15 min.
9. Transfer each reaction mixture to a 1.5-ml microcentrifuge tube containing 100 µl of phenol:chloroform:isoamyl alcohol. Vortex the tube vigorously. Spin the tubes in a microcentrifuge at 4 °C at 13,000*g* for 10 min.
10. Remove the top aqueous layer carefully from each reaction mixture and transfer it to a new 1.5-ml microcentrifuge tube. Add 10 µl of 3 M sodium acetate, 1 µl of glycogen and 300 µl of 100% cold ethanol (−20 °C) to the aqueous layer. Precipitate the RNA by incubating the tube at −80 °C for 1 h or at −20 °C overnight.
■ PAUSE POINT The RNA can be stored at −80 °C indefinitely.
11. Spin the tube at 13,000*g* at 4 °C for 20 min in a centrifuge from step 10. Remove the supernatant and add 1 ml of 70% (vol/vol) ethanol to the RNA pellet.
12. Spin the tubes at 13,000*g* at 4 °C for 20 min in a centrifuge. Remove the supernatant; then quick spin the tubes for a few seconds and remove the residual 70% (vol/vol) ethanol using a p10 pipette tip. Add 5 µl of gel loading dye II to each tube and resuspend it by pipetting.
13. Heat the samples at 70 °C for 5 min and snap-cool on ice. Load the samples onto an 8 M TBE urea 20-cm × 52-cm sequencing gel. Run the sequencing gel at 50 W for about 1 h.
14. Remove the gel from the glass plates and place it onto a piece of filter paper. Cover the top of the gel with a piece of plastic lab wrap. Dry the gel piece in a gel dryer at 80 °C for 1 h.
15. Expose a storage phosphor screen with the dried gel in a storage phosphor cassette overnight.
16. Visualize the bands on the storage phosphor screen using a Typhoon scanner.
17. Quantify the amount of uncut, full-length radioactive RNA in each lane using ImageQuant software. The percentage of uncut RNA at each dilution is the amount of intact full-length RNA in each digested lane divided by the amount of full-length RNA in the undigested lane. Typically choose the nuclease dilution that results in 80–90% of intact, full-length RNA.

Starting pool of RNA. In the experiment we describe here to illustrate the protocol, we grew ~500 ml of yeast to log phase for each biological replicate and used a hot acidic phenol solution to isolate the total RNA from cellular proteins and DNA[22]. We then performed poly(A)$^+$ selection, which uses beads that contain oligo(dT)s on their surface to bind to RNAs with poly(A)$^+$ tails, such as mRNAs. This step removes most of the rRNA and other highly abundant nonpolyadenylated RNAs, including tRNAs, snoRNAs and small nuclear RNAs (snRNAs), from the total RNA pool so that the majority of our deep-sequencing reads comes from a diverse population of mRNAs in the cell.

Recently, it has become increasingly clear that the genome is extensively transcribed[23–25]. Many RNAs in the cell, including non-poly(A)$^+$ RNAs, may have biologically important functions and are missing from the poly(A)$^+$ pool. However, PARS use is not restricted to poly(A)$^+$ RNA as input material. Diverse populations of RNAs including total RNA, RNAs from rRNA depletion protocols (such as Ribominus or Ribozero), poly(A)$^-$ RNAs and RNAs from cellular fractions can all be probed with PARS. Depending on the composition of the RNA pool and the presence of dominant RNA species, however, deeper sequencing may be needed. For example, in ribosomal depletion protocols, rRNAs are removed from total RNA, but other highly abundant RNAs, such as tRNAs, snoRNAs and snRNAs, are present in the starting RNA pool. These libraries typically require deeper sequencing to obtain structural information for the rest of the RNAs, as most of the sequencing reads are dominated by a few distinct populations of RNAs.

RNA renaturation. As yeast RNA extraction with hot acidic phenol and its poly(A)$^+$ selection result in the denaturation of RNA structures, we introduced an RNA renaturation step to refold the RNA for structural probing. RNA folding can be sensitive to various factors, including pH and the concentration of divalent and monovalent cations. Therefore, most of the standard folding buffers include a high concentration of magnesium (for example 5 mM Mg^{2+}) and sodium or potassium ions (for example 100 mM K^+)

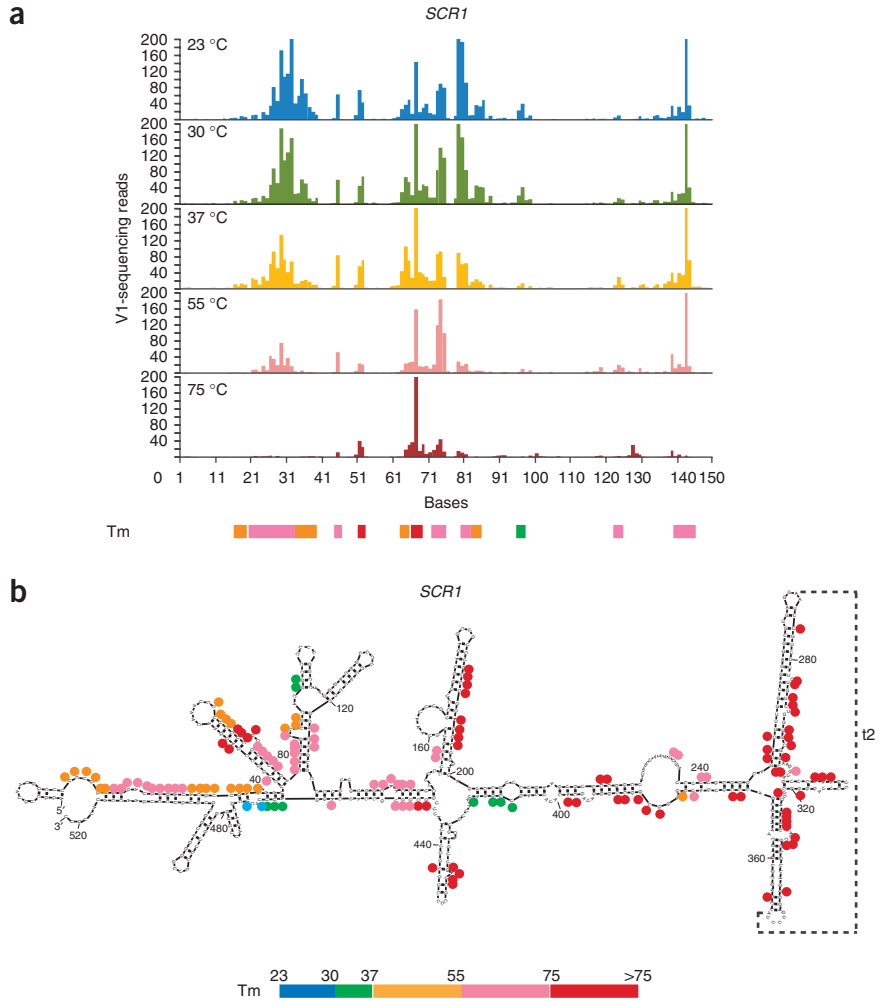

Figure 4 | Probing melting dynamics of the RNA structure across different temperatures. (**a**) Raw RNase V1 sequencing reads for the first 150 nt of the *SCR1* RNA at 23 °C, 30 °C, 37 °C, 55 °C and 75 °C. Bases that melt at the respective temperatures are shown as colored bars at the bottom of the graph, and the color indicates the highest temperature at which the structure was found to be stable. (**b**) RNA secondary structure model of the full-length *SCR1* mRNA[35]. A tertiary interaction is indicated in gray dotted lines. The melting transitions obtained from PARTE are indicated as colored dots. This figure is reproduced from Wan *et al.*[16].

at pH 7. One approach to refolding RNA after denaturation is to heat the RNA to 90 °C for 2 min before cooling it on ice, and then to allow the RNA to warm up to room temperature (23 °C) slowly in RNA structure buffer[17]. Another way of refolding RNA involves heating the RNA and cooling it slowly to room temperature. As different RNAs have different folding kinetics, which may be affected differently by the renaturation process, it is difficult to identify the best renaturation procedure for a population of RNAs. In this protocol, RNA is heated, cooled over ice and finally allowed to warm to room temperature. We chose this renaturation procedure because it has been used successfully in other RNA structure probing experiments, such as SHAPE[17]. By way of this approach, we were also able to capture RNA secondary structures known to exist in yeast, indicating that the folding is likely to be accurate at least for many RNAs in the refolded yeast population[10].

RNA fragmentation and size selection. As each RNA molecule is cleaved, on average, only once, long RNAs that are cleaved can still be too large to be cloned into a cDNA library. In order to capture most of the cleaved RNAs in the population, after nuclease digestion, RNAs are fragmented to a size of ~200–300 bases. Alkaline hydrolysis at 95 °C is performed to shear the RNAs more uniformly along its length. Depending on the average physical length of the transcriptome, different fragmentation times may be needed to fragment RNAs in a population to a specific size. Hence, the best fragmentation time must be determined in a titration performed at the start of an experiment so that ~200-base-long RNAs are obtained for single-end sequencing (**Box 2**). For complex and highly repetitive transcriptomes, paired-end sequencing may be needed to facilitate accurate mapping of the reads to the genome and/or the transcriptome. To generate paired-end sequencing libraries, RNAs can be fragmented to slightly longer lengths (e.g., 300–400 bases) to span exon junctions and highly repetitive regions.

In the present approach, a cleanup step is performed after RNA fragmentation because many very short RNA fragments (<10 bases) may be generated at this stage and some of them can end up being cloned into the cDNA library. To exclude these very short RNA fragments, size selection can be achieved by passing the RNAs either through a PAGE gel or through a column (RiboMinus concentration module). In PAGE purifications, RNA fragments between 50 and 300 bases in size are isolated and collected. In passing the fragmented RNA through a column, RNAs <30 bases in size are positively selected. We have been using column purification routinely, as opposed to PAGE purification, because it is easy to use and it allows higher RNA recovery. This RNA cleanup step, as well as size selection steps during cDNA purification and large-scale PCR purification, is not expected to result in erroneous structure calls along the body of a transcript. This is because, under conditions of single-hit kinetics, it is unlikely that a second cut will occur consistently next to the first cut (<50 bases away). The situation changes if a cut occurs within 30–50 bases of the end of an RNA molecule, because the fragment will be lost through size selection. Hence, this protocol does not enable the identification of RNA structural features close to the 3′ ends of transcripts. This 'blind tail' of 30–50 bases affects the researcher's ability to obtain structure information from poly(A)$^-$ transcripts, but not from poly(A)$^+$ transcripts, as poly(A)$^+$ mRNAs typically have a poly(A)$^+$ tail longer than 60 nt. This blind tail is likely to exist in all protocols that use size fractionation to separate ligated products from unligated adapters or to separate PCR products from primer dimers. One possible way to detect the structure at the 3′ end of transcripts is to perform SHAPE-CE, which directly detects the structural modifications by RT.

Box 2 | Determination of the optimal time for alkaline hydrolytic fragmentation ● TIMING 1 d

1. Dissolve 2 µg of RNA in 4 µl of nuclease-free water. Transfer the RNA to a 0.2-ml PCR tube and preheat the RNA at 95 °C for 15 s. Preheat 100 µl of 1× alkaline hydrolysis buffer at 95 °C for 1 min. Prepare five samples of the RNA for five titrations.
2. Add 16 µl of heated 1× alkaline hydrolysis buffer to 4 µl of RNA heated at 95 °C.
3. Incubate each tube for one of five different lengths of time, 3 min, 4 min, 5 min, 6 min or 7 min.
4. Place the tubes on ice and add 2 µl of 3 M sodium acetate to stop the reaction. Transfer the fragmented RNA to a 1.5-ml microcentrifuge tube. Add 1 µl of glycogen and 70 µl of ethanol and precipitate the RNA at −80 °C for 1 h.
5. Spin the tubes at 13,000*g* at 4 °C for 20 min in a centrifuge. Remove the supernatant and add 0.2 ml of 70% (vol/vol) ethanol to the RNA pellet.
6. Spin the tubes at 13,000*g* at 4 °C for 20 min in a centrifuge. Remove the supernatant and spin the tubes again for a few seconds; remove the residual 70% (vol/vol) ethanol using a p10 pipette tip. Resuspend the RNA in 5 µl of nuclease-free water.
7. Quantify the average size of the fragmented RNA using the Agilent Bioanalyzer. To generate the single-end PARS library in yeast, typically use the fragmentation time that results in the maximal amount of product that is 200 nt long.

cDNA library construction and TA cloning. To determine where the paired and unpaired bases are located in the transcript, the bases are cloned immediately downstream of the cleavage event by adapter ligation, as these bases contain 5′P groups. We have tried single-strand RNA ligation as well as splint-end ligation, and we have found that the splint-end ligation strategy in the SOLiD total RNA-seq kit (for sequencing on the SOLiD platform) and the Ambion RNA-seq library construction kit (for sequencing on the Illumina platform) provide more uniform capture of the cleaved RNA products. In contrast to making an RNA-seq library characterized by a single adapter ligation step, two such steps are implemented in PARS library construction (**Fig. 1**). The first ligation step adds a 5′ adapter to the 5′P group generated by nuclease cleavage and a 3′ adapter to the 3′OH group at the end of the RNA fragment. However, as many of the 3′ ends contain 3′P or 2,3 cyclic phosphates, the 3′ ends need to be converted to 3′OH groups for them to be cloned. In the present approach, Antarctic phosphatase is used to generate 3′OH groups after the first adapter ligation step. This timing is chosen because adding Antarctic phosphatase before that stage would remove 5′Ps that mark the structure cleavage sites. The second ligation step then ligates 3′ adapters to the Antarctic phosphatase–generated 3′OH group ends. Only the nuclease-digested fragments will contain both 5′ and 3′ adapters, which enables them to be selectively amplified during the PCR amplification step.

To preserve the relative abundance of the different RNA molecules initially extracted, a small-scale PCR is performed to identify the lowest number of PCR cycles needed to construct the library (**Fig. 1**). A large-scale PCR reaction is then performed using the number of PCR cycles identified in the small-scale PCR to amplify enough material for deep sequencing. Eighteen PCR cycles or fewer are optimal because the amplification process is more likely to be linear. Typically, cDNA from PCR cycles past the 20th is not utilized, because of the greater potential for overamplification of the template. In fact, overamplified PARS libraries tend to show a low correlation between sequencing data and traditional footprinting data. After the cDNA library is generated, some of the products are cloned and sequenced by traditional CE sequencing to confirm that the cDNA library consists of diverse sequences and is not dominated by primer dimers or an overamplified single product. As the Phusion DNA polymerase used in this protocol generates cDNA with blunt ends, the cDNA products are cloned using a Zero Blunt TOPO PCR cloning kit, according to the manufacturer's instructions. Typically, about 20 colonies are sequenced to determine whether the library is dominated by a single species, such as rRNA or linker dimers.

Controls. To ensure that the sequencing data obtained from PARS is of good quality and accurately captures RNA structure, our protocol calls for the comparison of the RNA sequencing profile of a few RNAs with known secondary structures with their manual footprinting data. In every PARS reaction, control RNA, such as different fragments from the tetrahymena ribozyme, is 'doped' into our starting pool of poly(A)$^+$-selected RNAs, so that 1–2% of the sequencing reads consist of these controls. Upon mapping the reads to the sequence of the doped-in controls, the double- and single-stranded sequencing profiles of the RNA are compared with low-throughput double- and single-strand structure probing that is visualized by running on a sequencing gel (**Fig. 3**). A high correlation ($R > 0.6$) indicates that the PARS experiment has been successful.

Data analysis. Many of the trimming, mapping and statistical tools that have been developed to analyze RNA sequencing are also applicable to PARS data. After obtaining the raw sequencing reads, the reads can be mapped to the genome and/or transcriptome using a variety of different programs, including Shrimp, BWA, TopHat and Bowtie[26–29]. We have used many of these programs to directly map the sequencing data to a well-annotated transcriptome and have obtained similar structural profiles for RNAs, irrespective of the software used. However, depending on the size and complexity of the genomes, different software packages may perform better in mapping RNA structure across junctions or to accommodate indel mutations and mismatches. In this protocol, we chose to use Bowtie to map both SOLiD and Illumina reads to larger transcriptomes primarily because of Bowtie's ability to align reads at a rate of millions per CPU hour with considerable accuracy.

One of the major differences in downstream analysis between an RNA sequencing experiment and a PARS experiment is in how the reads are counted across the transcript. In an RNA sequencing experiment, the total number of reads that are mapped to a whole transcript provides an estimate of RNA abundance (such as RPKM) for that transcript[10,30]. We found that by summing the

total number of RNase V1 reads and S1 nuclease reads across a transcript, we obtained a very high correlation with RNA abundance in RNA sequencing. To obtain RNA structure information per base on an RNA molecule, the total number of reads that start with that base for both V1 and S1 libraries are counted in the protocol (**Fig. 2**). This base lies immediately downstream of the structural evidence point, in front of the nuclease cleavage sites. The double- and single-strand reads are assigned to 1 nt upstream of the base that they are mapped to along an RNA molecule. Bases with high RNase V1 cleavage counts and low S1 nuclease cleavage counts are likely to represent paired nucleotides, whereas bases with low RNase V1 cleavage counts and high S1 nuclease cleavage counts are likely to represent unpaired nucleotides.

To illustrate how the analysis is performed, we provide a sample of RNase V1 SOLiD sequencing data (V1.csfasta) and perl analysis scripts from a typical yeast PARS experiment. The perl scripts, sam2tab.pl, normalize.pl and calculate_PARS.pl, convert the sequencing data to PARS score per base for each transcript to obtain RNA secondary structure information. We used the SeqFold tool[15] to incorporate S1 and V1 sequencing count data into computational modeling to reconstruct the RNA secondary structures of the entire transcriptome.

MATERIALS

REAGENTS

▲ **CRITICAL** All reagents need to be RNase-free; use reagents of the highest quality whenever possible.

- Yeast cells (S288C, ATCC, cat. no. 26108) or other sources of RNA to be investigated
- Nuclease-free water (Life Technologies, cat. no. AM9930)
- RNase V1 (Life Technologies, cat. no. AM2275)
- S1 nuclease (Fermentas, cat. no. EN0321)
- RNA structure buffer, 10× (Life Technologies, cat. no. AM2275)
- Alkaline hydrolysis buffer (Life Technologies, cat. no. AM2275)
- TBE-urea gels, 6% (wt/vol), 1.0 mm, ten well (Life Technologies, cat. no. EC6865BOX)
- Superasin RNase inhibitor (Life Technologies, cat. no. AM2696)
- TBE, 10× (Life Technologies, cat. no. 15581-044)
- Gel loading dye II (Life Technologies, cat. no. AM8546G)
- Novex TBE-urea sample buffer (Life Technologies, cat. no. LC6876)
- BlueJuice gel loading buffer, 10× (Life Technologies, cat. no. 10816-015)
- RNA century markers (Life Technologies, cat. no. AM7140)
- Phusion master mix with HF buffer (NEB, cat. no. F-531S)
- Ambion RNA-seq library construction kit (Life Technologies, cat. no. 4454073) for Illumina sequencing or SOLiD total RNA-seq kit (Life Technologies, cat. no. 4445374) for SOLiD sequencing
- Zero Blunt TOPO PCR cloning kit with One Shot TOP10 chemically competent *E. coli*. (Life Technologies, cat. no. K2800-20)
- RiboMinus concentration module (Life Technologies, cat. no. K1550-05)
- MinElute PCR purification kit (Qiagen, cat. no. 28004)
- MinElute gel extraction kit (Qiagen, cat. no. 28604)
- MicroPoly(A)Purist kit (Life Technologies, cat. no. AM1919)
- Costar Spin-X centrifuge tube filters (Sigma-Aldrich, cat. no. CLS8162-96EA)
- Sterile scalpels (Fisher Scientific, cat. no. 14-827-71)
- Corning bottle-top vacuum filters, 0.45 µm, 250 ml (Sigma-Aldrich, cat. no. CLS430768)
- Polypropylene conical tubes, 50 ml (BD Falcon, cat. no. 352070)
- Glycogen (Life Technologies, cat. no. AM9510)
- Ammonium acetate, 5 M (Life Technologies, cat. no. AM9070G)
- Sodium acetate, 3 M (Life Technologies, cat. no. AM9740)
- Tris, pH 7 (Life Technologies, cat. no. AM9850G)
- Tris, pH 8 (Life Technologies, cat. no. AM9856)
- Ethidium bromide solution, 10 mg ml^{-1} ! **CAUTION** Ethidium bromide is a potential carcinogen. Wear personal protection and dispose waste according to institutional regulations.
- SYBR Gold nucleic acid gel stain (Life Technologies, cat. no. S11494) ! **CAUTION** Gel staining dyes are potential carcinogens. Wear personal protection and dispose waste according to institutional regulations.
- NuSieve GTG agarose (Lonza, cat. no. 50084)
- Antarctic phosphatase (NEB, cat. no. M0289)
- Ethanol ! **CAUTION** Ethanol is volatile and flammable.
- Isopropanol (Fisher Scientific, cat. no. BP2618-500) ! **CAUTION** Isopropanol is volatile and flammable.
- Phenol solution, pH 4.3 (Sigma, cat. no. P4682) ! **CAUTION** Phenol is toxic. It can also cause skin burns when it comes in contact with bare skin. Wear proper protection (gloves) when using phenol and dispose it according to institutional regulations.
- Phenol:chloroform:isoamyl alcohol (25:24:1, vol/vol; Life Technologies, cat. no. 15593-031) ! **CAUTION** Phenol and chloroform are both toxic. Wear proper protection and dispose the chemicals according to institutional regulations.
- Chloroform (MG Scientific, cat. no. 9182-01) ! **CAUTION** Chloroform is toxic. Wear proper protection and dispose it according to institutional regulations.
- Microcentrifuge tubes, 1.5 ml (Sigma-Aldrich, cat. no. Z606340)
- Microcentrifuge tubes, 0.6 ml (Sigma-Aldrich, cat. no. T5149)
- Thin-walled PCR tubes, 200 µl (Sigma-Aldrich, cat. no. Z316121)
- Needle, 18.5 G (BD, cat. no. 305196)
- γ-^{32}P–ATP (Perkin Elmer, cat. no. BLU502Z250UC) ! **CAUTION** ^{32}P is radioactive. Use safety goggles and lab wear and follow radiation safety guidelines.
- T4 polynucleotide kinase (PNK; New England Biolabs, cat. no. M0201S)
- Nucleic acid sequencing unit, 20 cm × 52 cm (Sigma-Aldrich, cat. no. Z351857)
- Glass plates, 20 cm × 52 cm (Sigma-Aldrich, cat. no. Z352551)
- Plastic wrap (Fisher Scientific, cat. no. 22305654)
- RNaseOUT recombinant ribonuclease inhibitor (Life Technologies, cat. no. 10777-019)
- Gel elution buffer (Reagent Setup)
- Yeast lysis buffer (Reagent Setup)

P4P6 domain of tetrahymena ribozyme

- 5′-GGAAUUGCGGGAAAGGGGUCAACAGCCGUUCAGUAC CAAGUCUCAGGGGAAACUUUGAGAUGGCCUUGCAAAG GGUAUGGUAAUAAGCUGACGGACAUGGUCCUAACCACG CAGCCAAGUCCUAAGUCAACAGAUCUUCUGUUGAUAUG GAUGCAGUUC-3′ (This gene can be synthesized as a MiniGene from Integrated DNA Technologies and then *in vitro* transcribed to generate the RNA)

P9-9.2 domain of tetrahymena ribozyme

- 5′-GGACCUCUCCUUAAUGGGAGCUAGCGGAUGAAGUGAUGCAA CACUGGAGCCGCUGGGAACUAAUUUGUAUGCGAAAGUAUAUU GAUUAGUUUUGG-3′ (This gene can be synthesized as a MiniGene from Integrated DNA Technologies and then *in vitro* transcribed to generate the RNA)

Data and software

- Data file: https://s3.amazonaws.com/changbackup/ywan/PARS_Nature_Protocols/sample_data.tar.gz, which contains the following: V1.csfasta: sample SOLiD sequencing data for the RNase V1 library; sam2tab.pl: script to sum the V1 and S1 reads per base after mapping to the transcriptome; normalize.pl: script for normalizing mapped reads across libraries; and calculate_PARS.pl: script to calculate PARS score from mapped V1 and S1 reads

Box 3 | Downloading and installing software ● TIMING variable; a few hours

1. The example files, downloaded software and analysis files should all exist in a single directory (e.g., 'my_PARS_seq'). Create a PARS directory to store all of the executable programs used in this protocol:

```
$ mkdir $HOME/opt
```

2. Download sample data and perl scripts under 'my_PARS_seq' folder.
3. Install Bowtie using the latest binary package (http://bowtie-bio.sourceforge.net/index.shtml)
4. Unpack the Bowtie zip archive and move the Bowtie to the $HOME/opt directory:

```
$ unzip bowtie-0.12.7-macos-10.5-x86_64.zip
```

5. Move the unzipped Bowtie folder to the directory /opt/: $ mv bowtie-0.12.7 /opt/
6. Add Bowtie executables directory to your PATH environment variable:

```
$ PATH = ' $PATH:$HOME/opt/bowtie-0.12.7
```

7. Update $PATH so that Bowtie can be called anywhere in a terminal:

```
$ export PATH
```

8. Request a license from the Sfold website and follow the online instructions to install the program.
9. Download SeqFold: http://www.stanford.edu/~zouyang/seqfold/seqfold.tar.gz
10. Unzip the SeqFold folder and move it under $HOME/opt/:

```
$ tar -zxvf seqfold.tar.gz
```

11. Move the SeqFold folder under $HOME/ope/ directory:

```
$ mv seqfold $HOME/opt/
```

12. Download sample S1 sequencing count data (http://www.stanford.edu/~zouyang/seqfold/sce_S1.tab.tar.gz) and V1 sequencing count data (http://www.stanford.edu/~zouyang/seqfold/sce_V1.tab.tar.gz) from a typical yeast PARS experiment. Each file contains the S1 (or V1) read counts with one transcript per row in tab-delimited format: transcript name (column 1), counts (column 2, semicolon separated, the raw number of reads obtained for each base)
13. Unzip the S1 and V1 count files (tar -zxvf sce_S1.tab.tar.gz and tar -zxvf sce_V1.tab.tar.gz)

- Bowtie software (http://bowtie-bio.sourceforge.net/index.shtml/)[26]
- SeqFold software (http://www.stanford.edu/~zouyang/seqfold/)[15]
- Sfold software (http://sfold.wadsworth.org/cgi-bin/index.pl)[31]
- Treeview software (http://jtreeview.sourceforge.net/)[32]
- VARNA software (http://varna.lri.fr/)[33]
- FASTQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
- Perl and Python software (http://www.activestate.com/activeperl, http://www.python.org/getit/)
- ImageQuant software (GE healthcare)

EQUIPMENT

- Tabletop microcentrifuge
- XCell SureLock mini cell electrophoresis system (Life Technologies, cat. no. EI0001)
- Power supply (Bio-Rad, cat. no. 164-5056)
- Thermal cycler
- Heat block or water bath
- Rocking shaker
- Vacuum centrifuge
- NanoDrop Spectrophotometer or blue-light transilluminator (Life Technologies, cat. no. G6600)
- Freezers, −80 and −20 °C
- Liquid nitrogen tank
- Agilent Bioanalyzer
- Qubit fluorometer (Life Technologies, cat. no. Q32866)
- Gel dryer
- Storage phosphor screen (GE healthcare)
- Typhoon scanner

Computer hardware

- 64-bit computer running either Linux or Mac OS X (10.6 Snow Leopard or more recent); 8 GB of RAM (16 GB preferred); mutiple CPU cores are preferred (we recommend >100 CPU cores for PARS-guided RNA secondary structure reconstruction for the yeast transcriptome using SeqFold)
- Genome Analyzer II or HiSeq 2000 (Illumina)

REAGENT SETUP

Gel elution buffer To prepare gel elution buffer, add 5 ml of TE buffer (pH 8; 10 mM Tris-HCl (pH 8), 1 mM EDTA) to 5 ml of 5 M ammonium acetate. It can be stored at room temperature for 1 year.

Yeast lysis buffer To prepare yeast lysis buffer, add 2 ml of 0.5 M EDTA (pH 8), 5 ml of 10% (wt/vol) SDS and 1 ml of 1 M Tris-HCl (pH 7.5) to 92 ml of nuclease-free water. It can be stored at room temperature for 1 year.

EQUIPMENT SETUP

Downloading and installing software See directions in **Box 3**.

PROCEDURE

Total RNA extraction from yeast and Poly(A)$^+$ RNA selection ● TIMING ~2 d

1| Filter 500 ml of log-phase yeast cells (4 × 10^7 cells per ml) using two 0.45-µm, 250-ml Corning bottle-top vacuum filters to collect the yeast.

2| Remove the yeast-containing filter paper from the rest of the bottle-top filter. Fill a 50-ml Falcon polypropylene conical tube with liquid nitrogen and snap-freeze the yeast with the filter paper in the liquid nitrogen. Store the yeast cells at −80 °C until they are ready to be lysed for RNA extraction.

■ **PAUSE POINT** Yeast cells can be stored at −80 °C indefinitely.

3| Loosen the cap of a bottle of phenol solution (pH 4.3) and warm up the phenol at 65 °C for 20 min. Add 10 ml of yeast lysis buffer and 10 ml of phenol solution (pH 4.3) to the 50-ml Falcon tube and vortex the solution vigorously. Lyse the yeast at 65 °C for 1 h, vortexing the solution vigorously every 20 min.

4| Spin the 50-ml Falcon polypropylene tube for 15 min at 4 °C. Transfer the upper aqueous layer to a new 50-ml Falcon polypropylene tube, taking care not to transfer the white intermediate layer. Add 10 ml of phenol solution (pH 4.3) to the aqueous solution in the new tube. Mix well by vortexing the solution vigorously.

5| Repeat Step 4 for a total of two phenol extractions.

6| Spin the 50-ml Falcon polypropylene tube at 10,000*g* for 15 min at 4 °C. Carefully transfer the upper aqueous layer to a new 50-ml Falcon polypropylene tube and add 10 ml of chloroform to the aqueous solution in the new tube. Vortex the solution vigorously.

7| Spin the 50-ml Falcon polypropylene tube at 10,000*g* for 15 min.

8| Transfer the upper aqueous layer to a new 50-ml Falcon polypropylene tube. Add 1 ml of 3 M sodium acetate and 30 ml of 100% ethanol. Mix well by inverting the tubes several times and precipitate the RNA at −20 °C overnight or at −80 °C for 1 h.
■ **PAUSE POINT** The RNA can be stored at −80 °C indefinitely.

9| Spin the 50-ml Falcon polypropylene tube at 10,000*g* for 30 min at 4 °C to pellet the RNA. Decant the supernatant from the tube.

10| Add 40 ml of 70% (vol/vol) ethanol to the Falcon tube to wash the RNA. Spin the tube at 10,000*g* for 30 min at 4 °C. Decant the supernatant from the tube.

11| Spin the tube at 5,000*g* for 1 min at 4 °C to collect residual ethanol from the walls of the Falcon tube. Remove excess ethanol using a pipette.

12| Add 10 ml of nuclease-free water to the RNA. Place the Falcon tube on a rocking shaker at room temperature for 15 min to completely resuspend the RNA. Measure RNA concentration using a NanoDrop spectrophotometer. If the RNA concentration is >1.6 μg μl^{-1}, dilute it to 1.6 μg μl^{-1}.

13| Add 250 μl of the 2× binding solution from the MicroPoly(A)Purist kit to 250 μl of the RNA solution (concentration ≤1.6 μg μl^{-1}) and mix thoroughly. Follow the manufacturer's instructions according to the MicroPoly(A)Purist kit manual, starting by adding the 500-μl mixture to a tube of cellulose beads (provided in the kit) for binding, washing and elution of poly(A)$^+$ RNA.

14| To further enrich for poly(A)$^+$ RNA, repeat the poly(A)$^+$ enrichment step using the MicroPoly(A)Purist kit. Heat the RNA storage solution at 75 °C for 10 min. Elude the RNA from the beads using 200 μl of 75 °C RNA storage solution. Add 10 μl of 5 M ammonium acetate, 1 μl of glycogen and 550 μl of 100% ethanol to the RNA. Mix the reagents well by inverting the tubes several times. Precipitate the mixture at −20 °C overnight or at −80 °C for 1 h. Typically, we retrieve about 1% of the total RNA after two rounds of selection.
■ **PAUSE POINT** The RNA can be stored at −80 °C indefinitely.

15| Spin the microcentrifuge tube containing the RNA at 13,000*g* for 20 min, at 4 °C, to pellet the RNA precipitate.

16| Add 1 ml of 70% (vol/vol) ethanol to wash the RNA. Mix thoroughly by vortexing. Spin the microcentrifuge tube at 13,000*g* for 15 min at 4 °C to re-pellet the RNA. Carefully remove the supernatant.

17| Spin the microcentrifuge tube briefly to collect the residual ethanol at the bottom of the tube. Remove the excess ethanol carefully using a 10-μl pipette tip. Add 40 μl of nuclease-free water to the RNA to resuspend it. Measure the concentration of the RNA using a NanoDrop spectrophotometer.
■ **PAUSE POINT** The RNA can be stored at −80 °C indefinitely.

18| Run 60 ng of total RNA, 1× poly(A)$^+$-enriched RNA and 2× poly(A)$^+$-enriched RNA on the Agilent Bioanalyzer. The rRNA bands (for yeast, two predominant bright bands due to rRNA are present, one at 2 kb and the other at 3.8 kb) should be progressively fainter with poly(A)$^+$ selection, and there should be an increasing smear indicating the presence of other RNAs in the cell.
? TROUBLESHOOTING

RNA folding and enzymatic cleavage ● TIMING ~4 h
19| Add 2 µg of 2× poly(A)$^+$-enriched RNA in 80 µl of nuclease-free water to each of two 200-µl thin-walled PCR tubes for two reactions (one for RNase V1 digestion and another for S1 nuclease digestion). Make sure to include a doped-in positive control with a known RNA secondary structure to reach ~1% of the total moles in each tube.
▲ CRITICAL STEP Including a doped-in control with a known secondary structure (e.g., tetrahymena ribozyme) in the PARS library enables you to determine whether the secondary structure obtained from sequencing matches the known secondary structure.

20| Heat the tubes containing RNA at 90 °C for 2 min in a thermal cycler with a heated lid and then immediately place the tubes on ice for 2 min.

21| Add 10 µl of ice-cold 10× RNA structure buffer to the tubes and mix the contents by pipetting up and down several times. Transfer the tubes from ice to the thermal cycler; program the thermal cycler so that the temperature increases slowly from 4 to 23 °C over 20 min.
▲ CRITICAL STEP Bringing RNA to room temperature slowly enables proper folding.

22| Add 10 µl of S1 nuclease at the dilution determined as optimal in the titration described in **Box 1** to one tube and 10 µl of RNase V1 also at the dilution determined as optimal in the titration described in **Box 1** to the second tube and mix by pipetting. Incubate the samples at 23 °C for 15 min.
▲ CRITICAL STEP Nuclease digestion time has to be tightly controlled to prevent overdigestion.

23| Transfer the two reaction mixtures from Step 22 to 1.5-ml microcentrifuge tubes containing 100 µl of phenol:chloroform: isoamyl alcohol. Vortex the tubes vigorously. Spin the tubes in a microcentrifuge at 4 °C at 13,000*g* for 10 min.

24| Remove the top aqueous layers carefully and transfer them to new 1.5-ml microcentrifuge tubes. Add 10 µl of 3 M sodium acetate, 1 µl of glycogen and 300 µl of 100% cold ethanol (−20 °C) to the aqueous layers. Mix well by inverting the tubes several times. Precipitate the RNA by incubating the tubes at −80 °C for 1 h or at −20 °C overnight.
■ PAUSE POINT The RNA can be stored at −80 °C indefinitely.

25| Spin the tubes at 13,000*g* at 4 °C for 20 min in a centrifuge. Remove the supernatants and add 1 ml of 70% (vol/vol) ethanol to the RNA pellets.

26| Spin the tubes at 13,000*g* at 4 °C for 20 min in a centrifuge. Remove the supernatants and quick spin for a few seconds and remove the residual 70% (vol/vol) ethanol using a p10 pipette tip. Dissolve the RNA pellets in 4 µl of nuclease-free water and transfer the resuspended RNA to new PCR tubes. Keep the tubes on ice.

RNA fragmentation and cleanup ● TIMING ~2 h
27| Add 40 µl of 1× alkaline hydrolysis buffer to a clean PCR tube and heat it to 95 °C. After 45 s, transfer the two PCR tubes containing 4 µl of RNA each from ice to a thermal cycler set at 95 °C and heat them for 15 s. These preheating steps prevent the RNA from being fragmented before it is fully denatured.

28| Add 16 µl of the heated 1× alkaline hydrolysis buffer to each heated RNA solution and mix by pipetting several times. Incubate the tubes at 95 °C for the fragmentation time determined in the experiment described in **Box 2** and then immediately place the RNA on ice. Add 2 µl of 3 M sodium acetate to stop the fragmentation reaction.
▲ CRITICAL STEP The fragmentation time is crucial for controlling the size of the RNA fragments.

29| Add 6 ml of 100% ethanol to 1.5 ml of wash buffer (W5) from the RiboMinus concentration module.

30| Add 78 µl of 10 mM Tris, pH 7.0, to each of the two fragmentation RNA reaction mixtures. Mix the reagents by pipetting.

31| Add 100 µl of binding buffer (L3) from the RiboMinus concentration module to each of the 100-µl RNA fragmentation reaction mixtures, and then add each mixture to 250 µl of 100% ethanol. Mix by pipetting.

32| For each of the two fragmentation mixtures, place a spin column in a new 1.5-ml wash tube and transfer 450 µl of the RNA sample onto the spin column. Spin at 12,000*g* for 1 min and discard the flow-through. Put the spin column back into the empty wash tube.

33| Add 500 µl of wash buffer (W5, with ethanol added) to each spin column. Spin at 12,000*g* for 1 min and discard the flow-through. Place both spin columns back into empty wash tubes and spin the tubes at maximum speed for 2 min to dry the column.

34| Discard the wash tubes that contain the flow-through and place the spin columns in clean 1.5-ml recovery tubes. Add 12 µl of nuclease-free water to the center of each spin column. Wait for 1 min at room temperature. Spin the columns at maximum speed for 1 min at 4 °C. Run 1 µl of each of the two fragmented RNA mixtures on the Agilent Bioanalyzer to determine the fragmentation size.
? TROUBLESHOOTING

35| Dry the fragmented RNA in each of the two tubes in a vacuum centrifuge at low heat (<40 °C) for ~10 min. Suspend the procedure once the RNA is dry or has a volume lower than 3 µl.
▲ **CRITICAL STEP** Overdrying of RNA can cause the RNA to be very difficult to resuspend.

Adapter ligation (I) ● **TIMING ~17 h**
▲ **CRITICAL** The library preparation steps given below are modifications made to the Ambion RNA-seq library construction kit for sequencing on the Illumina platform and to the SOLiD Total RNA-seq kit for sequencing on the SOLiD platform. For simplicity, the components used in adapter ligation I are from the Ambion RNA-seq library construction kit unless specified otherwise.

36| Resuspend the dried RNA in a 1.5-ml microcentrifuge tube by adding nuclease-free water to a final volume of 3 µl. Mix by pipetting.

37| Add 2 µl of adapter mix A and 3 µl of hybridization buffer to the RNA and mix by pipetting.

38| Heat the tube to 65 °C in a thermal cycler for 10 min, and then incubate it at 16 °C for 5 min.

39| Add 10 µl of 2× ligation buffer slowly, as the 2× ligation buffer is very viscous. Add 2 µl of the ligation enzyme mix. Mix well by flicking the tube such that the sample looks homogenous. Briefly spin the tubes to collect the sample at the bottom of the tubes.

40| Incubate the sample in a thermal cycler at 16 °C overnight, with the heated lid off.

3′end treatment and inactivation of Antarctic phosphatase ● **TIMING ~3 h**
41| Add 20 µl of nuclease-free water to the sample and mix by pipetting. Add 5 µl of 10× Antarctic phosphatase buffer to the sample, followed by 2.5 µl of Superasin RNase inhibitor and 2.5 µl of Antarctic phosphatase enzyme. Mix the sample by flicking the tube. Briefly spin the tube to collect the contents at the bottom of the tube.

42| Incubate the sample at 37 °C in a water bath or heat block for 1 h.

43| Transfer the reaction mixture to a 1.5-ml microcentrifuge tube. Add 150 µl of nuclease-free water, followed by 100 µl of 100 mM Tris (pH 8); mix by pipetting.

44| Add 300 µl of phenol:chloroform:isoamyl alcohol mix and vortex to mix well. Spin the tube at 13,000*g* for 10 min at 4 °C.

45| Transfer the top aqueous layer of the sample to a new 1.5-ml tube and add 30 µl of 3 M sodium acetate, followed by 3 µl of glycogen and 900 µl of 100% ethanol. Incubate the sample at −20 °C overnight or at −80 °C for 1 h.
■ **PAUSE POINT** The RNA can be stored at −80 °C indefinitely.

46| Spin the tube at 13,000*g* for 20 min at 4 °C. Remove the supernatant and add 1 ml of 70% (vol/vol) ethanol to wash the pellet.

47| Spin the tube again at 13,000*g* for 20 min at 4 °C. Remove the supernatant; spin briefly to collect the excess 70% (vol/vol) ethanol at the bottom of the tube. Remove the residual ethanol using a 10-µl pipette tip. Dissolve the RNA pellet in 3 µl of nuclease-free water.

Adapter ligation (II) ● **TIMING ~17 h**

▲ **CRITICAL** For simplicity, the components used in adapter ligation II are from the Ambion RNA-seq library construction kit unless specified otherwise.

48| Add 2 µl of adapter mix A and 3 µl of hybridization buffer to the RNA in a 1.5-ml microcentrifuge tube. Mix the reagents by pipetting and transfer the contents to a 200-µl PCR tube.

49| Heat the PCR tube to 65 °C in a thermal cycler for 10 min and then incubate it at 16 °C for 5 min.

50| Add 10 µl of 2× ligation buffer slowly. Add 2 µl of the ligation enzyme mix. Mix well by flicking the tube until the sample looks homogenous. Briefly spin the tubes to collect the sample at the bottom of the tubes.

51| Incubate the sample in a thermal cycler at 16 °C overnight, with the heated lid off.

RT and cleanup ● **TIMING ~1.5 h**

52| Prepare the RT master mix on ice:

Component	Amount (µl)
Ligated RNA sample	20
Nuclease-free water	9
RT buffer	4
dNTP mix	2
RT primer	4
Total	39

53| Incubate the PCR tube containing the RNA in a thermal cycler with a heated lid at 70 °C for 5 min. Snap-cool the tube on ice for 2 min.

54| Add 1 µl of ArrayScript RT enzyme to each sample. Flick the tubes to mix the sample well and briefly spin the tubes to collect the sample at the bottom of the tube. Incubate the tubes at 42 °C for 30 min in a thermal cycler.

■ **PAUSE POINT** The resulting cDNA can be stored at −20 °C for several months.

55| Add 60 µl of nuclease-free water to the 40 µl of cDNA and transfer the 100 µl of sample to a clean 1.5-ml microcentrifuge tube. Add 500 µl of buffer PB from the Qiagen MinElute PCR purification kit to the sample and transfer the sample to a spin column.

56| Spin the column at 13,000*g* for 1 min at room temperature and discard the flow-through. Place the column back into the tube. Add 750 µl of buffer PE (with ethanol added) to the spin column. Spin the column for 1 min at 13,000*g*. Discard the flow-through.

57| Spin the column at 13,000*g* for 1 min to completely dry the column.

58| Add 10 µl of nuclease-free water to the column in a clean 1.5-ml microcentrifuge tube. Wait for 1 min before spinning the column at 13,000*g* for 1 min.

Size selection of cDNA ● TIMING ~1 d

59| Add 24 µl of nuclease-free water to 1 µl of 50-bp DNA ladder. Transfer 5 µl of the diluted DNA into a new 1.5-ml microcentrifuge tube and add 5 µl of 2× Novex TBE-urea sample buffer. Mix by pipetting.

60| Add 5 µl (out of 10 µl) of cDNA to a clean 1.5-ml microcentrifuge tube. Add 5 µl of 2× Novex TBE-urea sample buffer to the cDNA. Mix by pipetting.

61| Heat the 10 µl of DNA ladder and 10 µl of cDNA that contains the 1× Novex TBE-urea sample buffer from Step 60 at 95 °C for 3 min. Snap-cool on ice.

62| Load each sample into a well in a 6% (wt/vol) Novex TBE-urea PAGE gel (1 mm). Separate the samples from each other by at least two lanes.

63| Run the PAGE gel in 1× TBE at 180 V until the leading blue dye is about 1 cm below the center of the gel (~25 min).

64| Add 1 µl of SYBR Gold to 10 ml of 1× TBE. Mix and incubate the solution with the gel for 5 min in the dark. Visualize the size distribution of the ladder and the sample by using either UV or blue light.

65| Puncture two holes at the bottom of a clean 0.6-ml microcentrifuge tube using a clean 18.5-G needle and place the 0.6-ml tube inside a 1.5-ml microcentrifuge tube.

66| Cut the gel using a sterile scalpel from 100–300 nt (corresponding to insert sizes of 50–250 nt). Transfer the gel slice to the punctured 0.6-ml tube. Spin the tubes at 12,000*g* for 1 min at room temperature. Make sure that all the gel slices are at the bottom of the 1.5-ml tube and discard the 0.6-ml tube.

67| Add 300 µl of gel elution buffer to the shredded gel and incubate it at room temperature overnight.

68| Transfer 300 µl of the gel elution buffer into a new 1.5-ml tube on ice. Add another 300 µl of gel elution buffer to the shredded gel and incubate it at room temperature for another 2 h.

69| Cut the end of a 1-ml pipette tip to increase the surface area. Use it to transfer the elution buffer from the first overnight incubation into a Costar Spin-X centrifuge tube filter. Spin at 10,000*g* for 2 min at room temperature in a centrifuge. Add the second elute to the same column using the cut 1-ml pipette tip and spin the column at 10,000*g* for 2 min.

70| Add 6 µl of glycogen and 450 µl of isopropanol to the elution buffer from the Spin-X centrifuge filter. Incubate the mixture at room temperature for 10 min.

71| Spin the column at 13,000*g* for 20 min. Remove the supernatant and add 1 ml of 70% (vol/vol) ethanol. Gently pipette the solution without disturbing the pellet. Remove the 70% (vol/vol) ethanol and repeat this washing procedure three times.

72| Add 10 µl of water to the DNA pellet. Transfer 1 µl of cDNA to a clean 0.2-ml PCR tube to conduct a small-scale PCR.

PCR amplification (small-scale PCR) ● TIMING 4–5 h

73| Prepare the PCR master mix on ice using reagents from the Ambion RNA-seq library construction kit.

Component	Amount (µl)
cDNA library	1
Nuclease-free water	22
Ambion 5′ PCR primer	1
Ambion 3′ PCR primer	1
2× Phusion high-fidelity PCR master mix	
Total	25

74| To determine the fewest PCR cycles needed to amplify the sample, perform the PCR reaction from Step 73 using the cycling conditions in the in-text table below. After 15 cycles, pause the program at 72 °C and transfer 10 µl of the PCR reaction to a clean 1.5-ml tube. Place the tube on ice. Un-pause the program and repeat the same sample collection step after 20, 25 and 30 cycles. At the end of the PCR, the researcher should have four tubes, each with a sample that has undergone 15, 20, 25 or 30 PCR cycles.

Cycle number	Denature	Anneal	Extend
1	98 °C, 2 min		
2–31	98 °C, 30 s	65 °C, 30 s	72 °C, 30 s
32			72 °C, 5 min

75| Load 10 µl of 100-bp ladder and load 10 µl of 1:10 diluted 50-bp ladder into two separate wells.

76| Add 1 µl of 10× BlueJuice gel loading buffer to each of the 10-µl PCR products at 15, 20, 25 and 30 cycles; then load each sample into a well in a 2% (wt/vol) agarose gel made with 1× TBE. Skip at least one lane between the sample and the ladders.

77| Run at about 120 V in 1× TBE until the running dye for the ladder is near the bottom of the gel.

78| Visualize the gel under UV light. The amplified samples should have a smear at around 250 bases all the way to 300 bases. Choose the lowest-numbered PCR cycle showing such a smear in order to amplify the samples for large-scale PCR amplification.
? TROUBLESHOOTING

Large-scale PCR amplification and cleanup ● TIMING 3–4 h
79| Prepare the PCR master mix on ice in a new 1.5-ml microcentrifuge tube. The PCR is set up to a total volume of 200 µl, to be split into two 0.2-ml PCR tubes.

Component	Amount (µl)
cDNA library	4
Nuclease-free water	88
Ambion 5′ PCR primer	4
Ambion 3′ PCR primer	4
2× Phusion high-fidelity PCR master mix	100
Total	200

80| Perform the PCR using the cycling conditions described in the in-text table below. *X* is the lowest number of PCR cycles necessary to see amplification in the small-scale PCR reaction, as determined in Step 78.

Cycle number	Denature	Anneal	Extend
1	98 °C, 2 min		
X	98 °C, 30 s	65 °C, 30 s	72 °C, 30 s
32			72 °C, 5 min

81| Pool replicate large-scale PCR reactions from the same sample together. Add 5 volumes of buffer PB from the MinElute gel extraction kit to the sample and mix by pipetting several times. If the volume is too large to be contained in a 1.5-ml tube, split the reaction into several tubes.

82| Add 750 µl of the sample to a MinElute column. Spin the column at 13,000*g* for 1 min at room temperature. Load the sample multiple times through the column until all the samples have passed through the column.

83| Add 750 µl of buffer PE (part of the kit; add ethanol before using PE) to the column. Spin the column at 13,000*g* for 1 min at room temperature.

84| Spin the column at 13,000*g* for 1 min at room temperature to dry the column.

85| Add 20 µl of nuclease-free water to the center of the column in a new 1.5-ml tube to elute the DNA. Repeat this elution step.

Size selection of PCR products from large-scale amplification ● TIMING 4–5 h

86| Add 5 µl of 10× BlueJuice gel loading buffer to 40 µl of eluted PCR product.

87| Load 10 µl of 100-bp ladder and 10 µl of 1:10 diluted 50-bp ladder into two separate wells in a 3% (wt/vol) NuSieve GTG agarose gel (made with 1× TBE).

88| Skip at least two lanes between the sample and the ladders and then load 45 µl of the PCR product into one well in the agarose gel. Run the agarose gel at 100 V for 2–3 h in TBE until the leading dye front from the ladder is at the end of the gel.

89| Visualize the gel in blue light (blue light is preferred, although it has lower sensitivity) or under UV light. Cut out a gel slice containing the PCR product between 150 and 300 nt using a clean, sterile scalpel. Transfer the gel slice into a clean 2-ml microcentrifuge tube.

90| Add 1 ml of QG buffer from the MinElute gel extraction kit to the tube. Place the tube containing the gel slice and QG buffer at room temperature in a rocking shaker until the gel slice is dissolved.
▲ CRITICAL STEP Dissolving the gel slices at room temperature, and not at 50 °C, reduces local melting of AT-rich cDNA and prevents these cDNAs from being preferentially lost through the column[34].

91| Add 300 µl of isopropanol to the dissolved gel slice. Mix by pipetting.

92| Transfer 700 µl of the dissolved gel slice onto a clean MinElute column. Spin the MinElute column at 13,000*g* for 1 min at room temperature. Discard the flow-through. Repeat this step until all of the dissolved gel slice has been loaded onto the column.

93| Add 750 µl of the PE buffer (with ethanol added) to the column. Spin the column at 13,000*g* for 1 min. Discard the flow-through. Place the column back into the empty tube.

94| Spin the column at 13,000*g* for 1 min at room temperature to dry the column. Place the column in a clean 1.5-ml microcentrifuge tube.

95| Add 15 µl of nuclease-free water to the center of the column. Let the column sit for 1 min at room temperature. Spin the column at 13,000*g* for 1 min at room temperature.

96| Measure the concentration of the eluted PARS cDNA library using a Qubit fluorometer. Use 2 µl of the PARS library to check the size and concentration of the library using the Agilent Bioanalyzer.

97| Clone the cDNA products using the Zero Blunt TOPO PCR cloning kit and pick 20 colonies for capillary sequencing. Blast the reads against the yeast genome to identify the sequences cloned.
? TROUBLESHOOTING

High-throughput sequencing ● TIMING 4–5 d

98| Sequence the libraries (from Step 96) using Illumina's GA II or Hi-seq machine, according to the manufacturer's protocol.

Data analysis ● TIMING 1–2 d

▲ **CRITICAL** Most of the commands detailed below can be run using UNIX shell prompt and are intended to run from the example working directory. All executable commands are prefixed with a '$' character in the UNIX shell (e.g., bash or csh).

99| Download the raw sequencing reads (.fastq files for Illumina or .csfasta files for SOLiD) from the sequencers.

100| For Illumina sequencing, check the quality of the sequencing reads from the fastq files using the program FastQC. The low-quality reads are trimmed to enable accurate mapping.

101| Create a transcriptome index file using yeast RNA sequences to enable mapping of the sequencing reads to the yeast transcriptome using Bowtie, using either the Illumina platform (option A) if the Ambion RNA-seq library construction kit is used, or the SOLiD platform (option B) if the SOLiD Total RNA-seq kit is used.

(A) Generating an index for mapping using the Illumina platform

(i) Type the following command line:

```
$ bowtie-build sce_genes.fa sce_genes
```

(B) Generating an index for mapping using the SOLiD platform

(i) Type the following command line:

```
$ bowtie-build -C sce_genes.fa sce_genes_c
```

102| Map the sequencing reads to the yeast transcriptome by aligning the raw reads to the transcriptome indexes using the Bowtie software. This action creates a SAM format file that indicates the positions along the transcriptome where cleavage by the nuclease has occurred. Depending on whether the libraries were sequenced using Illumina single-end sequencing (option A), Illumina paired-end sequencing (option B), SOLiD single-end sequencing (option C) or SOLiD paired-end sequencing (option D), map the sequencing reads to the transcriptome using the following command lines. Please note that the meanings of the parameters and suggested initial values in the command lines below are clarified in the in-text table at the end of this step.

(A) Trimming and mapping of single-end reads generated from the Illumina platform

(i) Type the following command lines:

```
$ bowtie -5 NumA -3 NumB -p NumC -S sce_genes S1.fastq S1.sam
$ bowtie -5 NumA -3 NumB -p NumC -S sce_genes V1.fastq V1.sam
```

(B) Trimming and mapping of paired-end reads generated from the Illumina platform

(i) Type the following command lines:

```
$ bowtie -5 NumA -3 NumB -p NumC -S sce_genes -1 S1_R1.fastq -2 S1_R2.fastq
S1.sam
$ bowtie -5 NumA -3 NumB -p NumC -S sce_genes -1 V1_R1.fastq -2 V1_R2.fastq
V1.sam
```

(C) Trimming and mapping of single-end reads generated from the SOLiD platform.

(i) Type the following command lines:

```
$ bowtie -5 NumA -3 NumB -p NumC -S -C -f sce_genes_c S1.csfasta S1.sam
$ bowtie -5 NumB -3 NumB -p NumC -S -C -f sce_genes_c V1.csfasta V1.sam
```

(D) Trimming and mapping of paired-end reads generated from the SOLiD platform

(i) Type the following command lines:

```
$ bowtie -5 NumA -3 NumB -p NumC -S -C -f sce_genes_c -1 S1_R1.csfasta -2
S1_R2.csfasta S1.sam
$ bowtie -5 NumA -3 NumB -p NumC -S -C -f sce_genes_c -1 V1_R1.csfasta -2
V1_R2.csfasta V1.sam
```

Parameter	Description
NumA	The number of bases you would like to trim from the 5′ end of the raw read, depending on the raw reads' quality. If no trimming is needed, the suggested initial value is 0
NumB	The number of bases you would like to trim from the 3′ end of the raw read, depending on the raw reads' quality. If no trimming is needed, the suggested initial value is 0
NumC	The number of alignment threads to launch, e.g., 8 for an eight-core processor

103| Sum the number of V1 and S1 reads at each base for all the transcripts, using the 'sam2tab.pl' script. This script generates a table whereby each row consists of the name and structural information of the transcript, which is the number of RNase V1 or S1 nuclease reads mapped to each base of the transcript. The name and data are tab-separated, whereas the number of reads at each base is separated with semicolons.

104| To calculate the number of RNase V1 cleavages that have occurred at each base from the SAM file after alignment, use the following command line (the parameters *x*, *y* and *z* are defined in the in-text table in Step 105):

```
$ perl sam2tab.pl x y z V1.sam V1.tab
```

105| To calculate the number of S1 nuclease cleavages that have occurred at each base from the SAM file after alignment, use the following command line:

```
$ perl sam2tab.pl x y z S1.sam S1.tab
```

Parameter	Input	Description
x	0	The read is mapped onto the forward strand of the transcriptome
y	1	Only the number of mapped reads from the first read of a pair-end sequence, or single-end sequence, is counted
z	NumA+1	NumA was defined as the number of bases that were trimmed from the 5′ end of the read. *z* = NumA+1. By correcting for *z*, we identify the actual position where the cleavage occurred. For example, a cleavage event could have occurred at base 10, and the bases 11–61 are sequenced as a read. After we trimmed off two (i.e., NumA = 2) low-quality reads from the 5′ end, the nucleotide position that the read will map to on the RNA is base 13. However, if we subtract 2 + 1 bases from the mapped position, we will end up mapping the read at base 10, which is the base where the cleavage occurred

106| Normalize for sequencing depth across different PARS samples using the total mapped reads for the V1 and S1 libraries using the following command line:

```
$ perl normalize.pl Sample1_S1.tab Sample1_V1.tab Sample2_S1.tab Sample2_V1.tab …
SampleN_S1.tab SampleN_V1.tab
```

? TROUBLESHOOTING

107| Calculate the PARS score at each base, in a PARS experiment, by taking the log ratio of V1 reads over S1 reads at every base using a script 'calculate_PARS.pl' with the following command line:

```
$ perl calculate_PARS.pl norm.S1.tab norm.V1.tab sample.pars.txt
```

Using PARS data to guide secondary structure models in SeqFold ● TIMING 1 d

108| Generate structure preference profiles using the following command line:

```
$ python $HOME/opt/seqfold/pars2spp.py sce_S1.tab sce_V1.tab sce
```

The resulting file sce.spp contains the structure preference profiles with one transcript per row in tab-delimited format: transcript name (column 1), structure preferences (column 2, semicolon separated).

109| Generate sample structures and clusters for each transcript using the following command line:

```
$ perl $HOME/opt/seqfold/sfold_wrapper.pl sfold_executable_file sce_genes.fa
sfold_output_directory
```

The sfold_executable_file is the path to the executable file of Sfold. The sfold_output_directory is the directory in which each transcript has a folder with structure sampling results. By default, 1,000 sample structures are generated and clustered into distinct groups for each transcript.

110| (Optional) In a parallel computing environment, the runtime can be sped up by modifying the value of $para in sfold_wrapper.pl. For example:

```
$para = 'bsub -M 3072000 -W 6:00';
```

or

```
$para = 'qsub -cwd -V -l h_vmem=3G -l h_rt=6:00:00 -m ea -w e -b y';
```

Under these settings, each transcript will be processed in one CPU core independently and hundreds to thousands of transcripts can be processed in parallel.

111| Generate RNA secondary structure predictions and base-level accessibilities using the following command line and by choosing the optional parameters in the in-text table below:

```
$ python $HOME/opt/seqfold/seqfold.py sfold_output_directory sce.spp
```

Optional parameter	Description
-d	SeqFold output directory. Default: ./
-o	Prefix of output summary files. Default: out
-f	Cutoff to filter transcripts with the fraction of sites having experimental data <=cutoff_frac. Default: 0

Two files, A and B, are generated under the SeqFold output directory. A is '*.seqfold.ct.' Please note that each *.seqfold.ct file contains a predicted secondary structure for a transcript where * represents the name of the transcript. The structures are in the CT format. File B is 'out.acc.' Each acc file contains the estimated accessibility of each base of each transcript with one transcript per row in tab-delimited format: transcript name (column 1), accessibilities (column 2, semicolon separated, the accessibility of each base).

112| Visualize the predicted RNA secondary structures using the program VARNA. Please note that each *.seqfold.ct file can be visualized by VARNA for each transcript individually.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
18	Poly(A)$^+$ RNA looks degraded	Nuclease contamination (such as from RNase A)	Keep all reagents RNase-free. Wipe the workspace and all pipettes with RNaseOUT before starting
34	RNA fragments are either too long or too short to be cloned	The input RNAs are either very short or very long	Reduce fragmentation time from 3.5 min to 2 min if there is too much fragmentation. Alternatively, increase the fragmentation time to 4–5 min if the RNA is too long

(continued)

TABLE 1 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
78	No amplification products in small-scale PCR	RNase V1 or S1 nuclease is not working under the buffer conditions	Perform manual footprinting with increasing concentration (fivefold more) of RNase V1 or S1 nuclease to determine whether the enzymes are active and to identify the concentration needed in the new buffer conditions
97	TOPO cloning products are mostly adapter linkers	The gel slice is cut too closely to 100 bases during size selection after a large-scale PCR	Run a 3% (wt/vol) NuSieve GTG agarose gel for a longer time (3–4 h at 100 V) to separate the PCR products from the adapter dimers better. Cut the gel slice from 150 to 250 bases. Make sure that the gel does not heat up during electrophoresis, as heating may denature the double-stranded cDNA to cause changes in migration patterns
	TOPO cloning products are mostly of the same RNA	The library may be of low complexity, so the same product is cloned repeatedly	Repeat the protocol with twice as much starting 2× poly(A)$^+$ RNA
106	V1 structure profile and S1 structure profile look the same	RNase V1 or S1 nuclease is not working properly	Perform manual footprinting of a known RNA substrate with increasing concentration (fivefold increase) of RNase V1 or S1 nuclease to ensure that the enzymes are cleaving as they should

● TIMING

Steps 1–18, total RNA extraction from yeast and PolyA selection + RNA selection: ~2 d
Steps 19–26, RNA folding and enzymatic cleavage: ~4 h
Steps 27–35, RNA fragmentation and cleanup: ~2 h
Steps 36–40, adapter ligation (I): ~17 h
Steps 41–47, 3′-end treatment and inactivation of Antarctic phosphatase: ~3 h
Steps 48–51, adapter ligation (II): ~17 h
Steps 52–58, RT and cleanup: ~1.5 h
Steps 59–72, size selection of cDNA: ~1 d
Steps 73–78, PCR amplification (small-scale PCR): ~4–5 h
Steps 79–85, large-scale PCR amplification and cleanup: ~3–4 h
Steps 86–97, size selection of PCR products from large-scale amplification: ~4–5 h
Step 98, high-throughput sequencing: 4–5 d
Steps 99–107, data analysis: ~1–2 d
Steps 108–112, using PARS data to guide secondary structure models in SeqFold: ~1 d
Box 1, determination of the conditions for single-hit kinetics for RNase V1 and S1 nuclease: 3 d
Box 2, determination of the optimal time for alkaline hydrolytic fragmentation: 1 d
Box 3, downloading and installing software: a few hours

ANTICIPATED RESULTS

The poly(A)$^+$ selection typically leads to the isolation of about 1–2% of the total RNA. We can typically see a smear of cDNA products ranging from 100 to 300 bases using SYBR Gold stain on 8 M TBE urea gel. After 18 cycles of PCR, we typically see PCR products from 150 to 300 bases. Quantification of the PCR products after size selection generally results in at least 200 ng of DNA ready for deep sequencing.

In the process of structural probing of the yeast transcriptome, we sequenced about 90 million reads to obtain structural information for more than 3,000 genes. We first map the raw reads to the sequence of the positive control to ensure that we can capture the secondary structure of P9-9.2 accurately using PARS. As shown in **Figure 3**, the structural profile of P9-9.2 RNA from a PARS experiment correlates well with the structural information obtained by traditional footprinting, with the S1 nuclease cleaving in regions expected to be single stranded and RNase V1 cleaving in regions expected to be double stranded, according to the secondary structure model[7]. After checking that the PARS experiment captures the secondary structure of P9-9.2 RNA accurately, we map the sequencing reads to the rest of the yeast transcriptome by following the step-by-step protocol in the PROCEDURE section. Typically, more than 60% of the reads map uniquely to the yeast transcriptome, providing RNA structural information for thousands of genes.

ACKNOWLEDGMENTS This work was supported by National Institutes of Health (R01-HG004361). Y.W. is funded by the Agency of Science, Technology and Research of Singapore. H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS Y.W. and H.Y.C. developed the protocol and designed the experiments; Y.W. performed the experiments; K.Q. analyzed the data; Z.O. developed the SeqFold pipeline; Y.W. and H.Y.C. wrote the paper with contributions from all authors.

COMPETING FINANCIAL INTERESTS The authors declare competing financial interests: details are available in the online version of the paper.

1. Sharp, P.A. The centrality of RNA. *Cell* **136**, 577–580 (2009).
2. Wan, Y., Kertesz, M., Spitale, R.C., Segal, E. & Chang, H.Y. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* **12**, 641–655 (2011).
3. Breaker, R.R. Prospects for riboswitch discovery and analysis. *Mol. Cell* **43**, 867–879 (2011).
4. 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**, 351–362 (2004).
5. Deigan, K.E., Li, T.W., Mathews, D.H. & Weeks, K.M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. USA* **106**, 97–102 (2009).
6. Gornicki, P. *et al.* Use of lead(II) to probe the structure of large RNA's. Conformation of the 3′ terminal domain of *E. coli* 16S rRNA and its involvement in building the tRNA binding sites. *J. Biomol. Struct. Dyn.* **6**, 971–984 (1989).
7. Auron, P.E., Weber, L.D. & Rich, A. Comparison of transfer ribonucleic acid structures using cobra venom and S1 endonucleases. *Biochemistry* **21**, 4700–4706 (1982).
8. Watts, J.M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2009).
9. 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).
10. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103–107 (2010).
11. Zheng, Q. *et al.* Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in *Arabidopsis*. *PLoS Genet.* **6**, e1001141 (2010).
12. Underwood, J.G. *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7**, 995–1001 (2010).
13. Lucks, J.B. *et al.* Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. USA* **108**, 11063–11068 (2011).
14. Li, F. *et al.* Global analysis of RNA secondary structure in two metazoans. *Cell Rep.* **1**, 69–82 (2012).
15. Ouyang, Z., Snyder, M.P. & Chang, H.Y. SeqFold: Genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res.* (2012).
16. Wan, Y. *et al.* Genome-wide measurement of RNA folding energies. *Mol. Cell* **48**, 169–181 (2012).
17. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).
18. Low, J.T. & Weeks, K.M. SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150–158 (2010).
19. Wurst, R.M., Vournakis, J.N. & Maxam, A.M. Structure mapping of 5′-32P-labeled RNA with S1 nuclease. *Biochemistry* **17**, 4493–4499 (1978).
20. Lowman, H.B. & Draper, D.E. On the recognition of helical RNA by cobra venom V1 nuclease. *J. Biol. Chem.* **261**, 5396–5403 (1986).
21. Ehresmann, C. *et al.* Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 9109–9128 (1987).
22. Lee, A., Hansen, K.D., Bullard, J., Dudoit, S. & Sherlock, G. Novel low abundance and transient RNAs in yeast revealed by tiling microarrays and ultra high-throughput sequencing are not conserved across closely related yeast species. *PLoS Genet.* **4**, e1000299 (2008).
23. Neil, H. *et al.* Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**, 1038–1042 (2009).
24. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037 (2009).
25. ENCODE Project Consortium *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
26. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
27. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
28. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
29. Rumble, S.M. *et al.* SHRiMP: accurate mapping of short color-space reads. *PLoS Comput. Biol.* **5**, e1000386 (2009).
30. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* **5**, 621–628 (2008).
31. Ding, Y., Chan, C.Y. & Lawrence, C.E. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* **11**, 1157–1166 (2005).
32. Saldanha, A.J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
33. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
34. Quail, M.A. *et al.* A large genome center's improvements to the Illumina sequencing system. *Nat. Methods* **5**, 1005–1010 (2008).
35. Zwieb, C., van Nues, R.W., Rosenblad, M.A., Brown, J.D. & Samuelsson, T.A nomenclature for all signal recognition particle RNAs. *RNA* **11**, 7–13 (2005).