Published in final edited form as:
Nat Rev Genet. ; 12(9): . doi:10.1038/nrg3049.

Understanding the transcriptome through RNA structure

Yue Wan[1], **Michael Kertesz**[2], **Robert C. Spitale**[1], **Eran Segal**[3], and **Howard Chang**[1]

[1]Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

[2]Howard Hughes Medical Institute and Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

[3]Department of Computer Science and Applied Mathematics and, Weizmann Institute of Science, Rehovet 76100, Israel

Abstract

RNA structure is critical for gene regulation and function. In the past, transcriptomes have been largely parsed by primary sequences and expression levels, but it is now becoming feasible to annotate and compare transcriptomes based on RNA structure. In addition to computational prediction methods, the recent advent of experimental techniques to probe RNA structure by deep sequencing has enabled genome-wide measurements of RNA structure, and provided the first picture of the structural organization of an eukaryotic transcriptome—the "RNA structurome". With additional advances in method refinement and interpretation, structural views of the transcriptome should help to identify and validate regulatory RNA motifs that are involved in diverse cellular processes, and thereby increase understanding of RNA function.

Introduction

RNA is a unique informational molecule. In addition to carrying information in their linear sequences of nucleotides (primary structure), RNA molecule fold into intricate shapes. Pairing of local nucleotides can create secondary structures such as hairpins and stem loops, and interaction among distantly located sequences can further create tertiary structures. In every step of its life cycle, RNA structures influence the transcription, splicing, cellular localization, translation, and turnover of the RNA (Fig. 1). The topic of RNA structures in different cellular processes have been covered in several excellent reviews[1–5]. Although the structures of multiple RNAs have been studied in detail, structural information for most RNAs in cell, such as mRNAs, is missing due to the low throughput nature of RNA structure probing and the difficulty in probing long RNAs. Classic techniques require individually cloned RNA sequences, and only a few hundred bases can be interrogated per experiment. As most of the RNA structures are studied on a case-by-case basis, it is difficult to determine what the full impact an RNA's structure has on cellular biology. To close this gap, genome-wide RNA structure determination has relied heavily on computational predictions to create structural models for hypothesis testing. Computational RNA prediction algorithms have advanced greatly in their ability to predict more accurate secondary structures from both primary sequences and sequence covariation. However, these predicted structures are typically confirmed by secondary structure probing, which still serves as the gold standard of RNA structure determination.

Correspondence to: H.Y.C. at howchang@stanford.edu.

The advent of ultra high throughput sequencing technologies has enabled the sequencing of hundreds of millions of bases at a time, and greatly increased the speed and precision of genomic data. High throughput sequencing has been applied successfully in many applications, including genome discovery, transcriptome annotation, and global mapping of DNA-protein interactions[6–8]. Coupling RNA structure probing to high throughput sequencing yields genome-scale RNA structural information, providing insights to the secondary structures of thousands of transcripts in the cell. Here, we briefly summarize the importance of RNA structure in various cellular processes by highlighting a few recently discovered examples, review advances in computational structure predictions, focus on experimental approaches to large-scale RNA structure maps, and discuss the potential impact of this new kind of transcriptomic information.

Biological relevance of RNA structures

RNA secondary and tertiary structures influence the function of almost all classes of RNAs, including mRNAs, non-coding RNAs such as riboswitches, ribozymes, long non-coding RNAs (lncRNA) and microRNAs (miRNA). RNA structures play roles in nearly every step of gene expression from transcription, mRNA processing, RNA localization, translation, to RNA decay (Table 1). RNA structures enable RNA to interact with itself, with other RNAs, with ligands and with RNA binding proteins. Many of these structures can exert their influence by helping to provide specific binding sites for RNA binding proteins (RBP) as well as restricting protein binding by altering accessibility. Identifying RBP binding sites and RBP consensus motifs is an area of intense study (Box 1).

BOX 1

RNA binding proteins: motif identification and prediction

RNA binding proteins (RBP) interact with RNAs to regulate diverse cellular processes. While many of these interactions are mediated by linear sequence motifs, RNA structural motifs as well as the structure context in which linear motifs are embedded also influence RBP binding. Different strategies have been developed to identify RNA consensus motifs. Transcripts associated with RBPs can be computationally searched for consensus nucleotide sequences that are selectively enriched in bound versus un-bound transcripts using programs such as MEME, FIRE and REFINE[141–143]. Experimentally, Selex and RNAcompete enable the determination of RNA consensus motifs experimentally by incubating an RBP with a complex pool of randomized short RNA sequences to selectively identify the sequences that have stronger binding affinities to the RBP[142, 144]. The development of new methods such as High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) and Photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation (PAR-CLIP) allow the identification of both RBP bound transcripts as well as the protein binding site, greatly reducing the search space for consensus motif finding in RBP bound targets[24, 25]. Importantly, incorporation of predicted RNA secondary structure can substantially increase the explanatory power of some linear RBP binding motifs; for instance, several motifs are shown to bind RBP only when the motif occurs in the context of a single-stranded, accessible region of mRNA[145]. Combined with an increased amount of available RNA structure data, it would be possible to predict consensus RNA structural motifs and assess the impact of RNA structures in RNA protein interactions.

Multiple RNA structures can potentially be formed from a long linear sequence. RNA structures are frequently dynamic and RNAs can undergo different conformational changes based on their solvent conditions. RNAs can react to various inputs including differences in protein binding, changes in ligand and salt concentrations, and varying temperatures to result in gene expression changes, providing an additional layer of complexity to gene regulation. This role of RNA as a molecular sensor requires that RNA structures are highly specific, so that distinct RNA structures can respond to specific cellular stimuli, and that RNA structures are dynamic, so that the cellular response is fairly rapid. Below we elaborate on some examples that illustrate the specificity and dynamic character of RNA structures and how identifying such structures in a transcriptome-wide manner can enhance our understanding on RNA function.

The specificity and dynamics of riboswitches

One of the best examples that illustrates the specificity and the dynamics of RNA structures is a riboswitch. Riboswitches are RNAs sensors that can detect changes in cellular stimuli in the absence of other cofactors such as proteins[5, 9]. As such, some of the first riboswitches were discovered based on changes in RNA structure induced by specific ligands[5, 10]; sequence alignment with established riboswitches allowed subsequent identification of riboswitch families[11, 12].

A riboswitch typically consists of two domains, an aptamer domain that recognizes its specific ligand and an expression domain. Upon interacting with a ligand in its ligand binding domain, the riboswitch undergoes a conformational change that results in gene expression changes. Multiple classes of riboswitches exist that respond to a wide range of cellular stimuli including amino acids, nucleotides, metal ions, coenzymes and temperature to regulate processes such as transcription termination, changes in translation rate, splicing and mRNA decay[13–16]. Although first discovered in bacteria, riboswitches have been found in other organisms such as yeast, algae and plants, indicating the prevalence of this important regulatory mechanism in multiple kingdoms of life[17, 18]. However, only the thiamine pyrophosphate (TPP) riboswitch has been found outside eubacteria, and none has been found in mammals[12].

Because the aptamer domains of riboswitches form multiple Watson and crick bases with their ligands, riboswitches are typically very specific for their ligands and can discriminate between their true ligands and other similarly structured molecules[5]. This specificity of its metabolite enables a riboswitch to serve as a cellular sensor. An example of this is the adenine riboswitch whereby a single base pair change from U to C in the ligand binding site changes the affinity of the riboswitch for adenine to guanine[19]. This riboswitch is found in the 5′UTR of the ydhL mRNA and forms a secondary structure upon binding to adenine that prevents the formation of the terminator loop and transcription termination. High levels of adenine hence result in high protein levels of ydhL, which is a purine efflux pump, to pump purines out of the cells. Another example is the SAM riboswitch. Distinct classes of the SAM riboswitches can bind to S-adenosylmethione (SAM), a coenzyme for methylation, or S-adenosylcytosine (SAH), a byproduct of the methylation reaction, even though SAM and SAH are highly similar in structure (Fig. 1A). This distinction is important to prevent the accumulation of toxic SAH and to recycle SAH to form SAM[5]. The diversity of SAM riboswitches also illustrates the possibility of multiple RNA structural solutions to the same biochemical challenge, raising the need to experimentally probe RNA structural dynamics rather than relying purely on sequence conservation.

Dynamics of RNA structures in higher eukaryotes- mammals

The dynamics of RNA structure is also a recurring theme in mammalian RNAs. While the binding of protein factors to specific RNA elements has been extensively studied, it is recently emerging that this binding can result in a corresponding change in RNA structure, which affects gene expression. The *VEGFA* mRNA contains a 125 base, hypoxia stability region, in its 3′UTR and the structure of this region changes depending on whether the cell is exposed to normoxic or hypoxic conditions in the presence of interferon gamma[20]. During normoxia, the presence of the GAIT complex results in the *VEGFA* mRNA to form a structure that is not permissive to translation. However during hypoxia, the binding of HNRNPL results in the RNA conformation to switch to a different structure that permits protein translation.

MicroRNAs (miRNAs) are ~23-nt short RNAs that modulate gene expression in normal development and disease pathogenesis. Recently, RNA conformations within a transcript have also been found to be one of the determinants of whether a transcript is targeted by specific miRNAs. The interaction between miRNAs and 3′ UTRs of their targets can lead to mRNA destabilization and/or translation inhibition. Accessibility of miRNA target sites can influence miRNA binding, as target sites that are buried in secondary structures may sterically hinder their interaction with miRNAs[21]. Interestingly, accessibility of miRNA target sites can change in different biological states indicating an additional layer of gene regulation[22]. One prime example is the regulation of levels of p27, a cyclin dependent kinase inhibitor, during different stages of the cell cycle. p27 protein level is low in dividing cells but high in non-dividing, quiescent cells. Upon growth factor stimulation, Pumilio-1 protein is activated, binds to the p27 mRNA 3′ UTR, and results in a RNA structural change. This structural change exposes the microRNA target sites in the 3′ UTR of p27, allowing miR-221 and miR-222 to interact with the p27 3′ UTR, causing translation repression and a reduction in p27 protein levels (Fig. 1B).

There is an increasing amount of genome-wide datasets on RNA binding proteins and their targets, as well as where these proteins bind to their mRNA targets[23–25]. Probing RNA structures in a genome-wide manner both in-vitro and in-vivo would enable us to study both the structural context that determines protein binding to RNAs as well as identify regions of RNA structural changes that occur in the presence and absence of protein binding. As many of such structural changes result in meaningful functional outputs, such as changes in translation or decay, this would enrich our mechanistic understanding of how RNA structures impact cellular function.

Computational approaches to RNA secondary structures

Given the experimental difficulties in measuring RNA structure, algorithms for predicting RNA structure from primary sequence have been developed and applied in many settings[26–31]. When accurate, these approaches have clear advantages, as they do not require experimentation, and can also be used to predict the structure of any arbitrary transcript, including hypothetical transcripts with designed mutations. Indeed, approaches based on computational predictions have led to many biological discoveries and insights. For example, for specific classes of ncRNAs whose members share structural properties essential for their function, computational methods utilizing secondary structure predictions were successfully used to annotate new members of that ncRNA class. Examples include methods for predicting tRNAs[32, 33], snoRNAs[32] and microRNAs[34]. By combining RNA structure predictions with comparative genomic analysis, the more general task of identifying novel ncRNAs from a genome sequence has also been addressed in many organisms[35–37]. Finally, several methods have been developed for identifying structural motifs that are common to multiple RNAs, and that may have a role in the subcellular localization, stability, or the function of the RNA in which they are embedded[30, 38–41] (Fig. 2a,b).

Computational RNA structure prediction—covariation

Several different approaches exist for predicting RNA secondary structure. Methods based on comparative sequence analysis rely on the fact that many of the known functional RNA structures are conserved in evolution. Examples include tRNAs, rRNAs, and group I and group II introns[42, 43]. Covariation methods determine secondary structure by examining conservation patterns of basepairs among homologous or paralogous genes. Such covariation methods search for two distinct genomic sequences in which evolutionary sequence changes in one sequence are accompanied by compensatory sequence changes in the other sequence that preserve RNA structure[42]. For example, the pairing of G-C nucleotides between two distinct genomic sequences can be maintained at the structure level in another species if the G-C nucleotides have changed to A-U nucleotides (Fig. 2c). The structure can be determined directly from the pattern of conserved pairings when enough homologous sequences are available, and several methods exist for this[27, 44–47]. In other cases, a combined thermodynamic-covariation method can be used[48].

Computational RNA structure prediction—thermodynamic modeling

When only a single sequence is available, an accurate and popular method is thermodynamic computation of the minimal free energy structure. This method uses efficient dynamic programming algorithms in conjunction with experimentally-derived energy parameters to scan the entire landscape of possible secondary structure configurations and identify the most thermodynamically stable structure[26, 49, 50]. For sequences that are shorter than 700bp, ~70% of the known basepairs are correctly predicted by these methods. However for longer sequences, the accuracy drops to ~20–60% when the predicted structures are compared to high resolution crystal structures and structural predictions obtained using comparative analysis[51, 52]. As an alternative to free energy minimization methods, algorithms based on probabilistic modeling using stochastic context-free grammars (SCFGs) were also developed, but since their accuracy is lower, thus far they have not replaced free energy minimization methods[28]. Another recent improved strategy was developed using both thermodynamic modeling and machine learning methods, and the strategy was based on choosing the nucleotide set with the maximal sum of pairing probabilities[53, 54]. An interesting application of thermodynamic modeling techniques is the evaluation of potential RNA structural changes caused by noncoding single nucleotide polymorphisms associated with human diseases. Laederach and colleagues identified multiple disease-associated mutations in UTRs that alter the mRNA structural ensemble of the associated gene, providing new hypotheses for causes of human disease and variation[55].

Computational RNA structure prediction—Incorporating experimental data

Another successful approach has been to incorporate experimentally derived structural information into computational predictions. This approach has been in use since the first prediction algorithms became available and has been further developed throughout the years[29, 56–59]. When the experiment can only derive binary information for each nucleotide, namely whether the nucleotide was paired or unpaired, the dynamic programming algorithm can be modified such that large positive free energy terms are added to nucleotides that are known to be unpaired, thus restricting the algorithm from marking them as paired[57]. More recently, methods that use quantitative, nucleotide resolution experimental data (discussed below) to direct the prediction of a folding algorithm have been introduced, by integrating an additional per nucleotide pseudo-free energy term into the dynamic programming algorithm[59]. This method was shown to significantly increase the accuracy of structure prediction.

Ongoing challenges

Despite their many successes, current prediction algorithms have several limitations. First, RNA molecules in solution may adopt secondary structures that are only partially determined by thermodynamics, as RNA molecules can undergo conformational changes upon interaction with other RNAs and RNA-binding proteins. These context-dependent RNA protein interactions are extremely complex to model and are thus excluded from all prediction algorithms. Second, although our knowledge of thermodynamic rules and parameters has greatly improved, it is far from being complete[29, 57, 60, 61]. Finally, most folding algorithms use approximations in order to efficiently scan the vast landscape of possible secondary structures. Important limitations are the difficulty to predict pseudoknots (RNA topologies that contain non-nested nucleotide pairings) or take into account long-range model and tertiary structure interactions. Pseudoknots have been observed in a number of functional RNA sequences, such as ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) or the genomes of viral RNAs[62], where they have been shown to be involved in unique mechanisms of viral translation initiation and elongation[63]. Thus, ignoring pseudoknots results in inaccurate structure predictions[62, 64]. In contrast to the prediction of nested structures (free of pseudoknots), which can be efficiently solved using dynamic programming, predicting structures that contain pseudoknots is very challenging computationally. Pseudoknot prediction has proven to be a class of computational problems with no fast solutions, termed “NP-complete”, for a large class of models of pseudoknots[65]. As a result, several methods have been developed that focus on specific types of pseudoknots[66–68], or employ heuristics[69–73], to bring running time to down. Nonetheless, computational prediction of pseudoknot still scales exponentially with the length of the RNA [on the order of O(n^4)-O(n^6) where n is the length of the RNA sequence].

Thus, although the extensive research and development of RNA structure prediction tools has led to many successes and discoveries, the applicability of existing tools is still limited and further experimental data is needed to bridge the knowledge gap. However, the accumulation of additional experimental data should lead to better optimization of existing algorithms and to the development of new strategies, some of which may combine experimental and computational approaches.

RNA structure maps—the first steps

Probing RNA structures in solution by RNA footprinting

RNA footprinting is a method that probes RNA in solution using a variety of chemical and enzymatic probes[74]. With *in vitro* footprinting, an RNA of interest is typically transcribed *in vitro* and folded in solution before being subjected to a battery of different structural probes that determine which of the bases are single stranded, double stranded, or solvent exposed[74, 75]. Chemicals including dimethyl sulfide (DMS), 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate (CMCT), kethoxal, lead (Pb2+) and N-methylisatoic anhydride (NMIA), and nucleases including RNase I, T1, A and S1 nuclease, interact with single stranded or flexible bases to modify or cleave them[76–80]; enzymes such as RNase V1 recognize and cleave at double stranded bases[81]; hydroxyl radicals cleave at RNA bases that are solvent exposed[82, 83]. The combinatorial usage of the above probes provides structural information on most bases in the RNA. Upon cleavage or modification, the reaction sites can be detected by autoradiography, or alternatively reverse transcription, followed by gel or capillary electrophoresis (Fig. 3). The location of the cleavage is determined from the migration pattern of the bands and the intensity of the bands can be quantified using image processing tools, such as the program semiautomated footprinting analysis (SAFA)[84].

RNA footprinting can also be performed *in vivo*[85, 86]. Because some RNAs are able to fold into alternative conformations *in vitro* that do not reflect their *in vivo* biological conformations, structure probing *in vivo* may provide more accurate information on biologically relevant RNA structures[87]. RNA footprinting can be carried out inside the cells using chemicals that can penetrate the cell membrane such as lead and DMS, or with high energy X-rays[82, 86, 88]. Lead probing has been successfully applied to *in vivo* structure probing in bacteria while DMS has been applied to both prokaryotic and eukaryotic cells[86, 88]. However, *in vivo* RNA footprinting may not be able to interrogate all regions of a RNA of interest due steric hindrance from protein interactions. The dynamic cellular environment also presents RNA in heterogeneous states: RNA in different stages of its lifecycle during transcription, translation and decay are all present. Averaging the structural signal from heterogeneous states may also prove to be inaccurate. As such, structural probing *in vitro* and *in vivo* provide complementary information about RNA structures. In all footprinting experiments, it is important to titrate the amount of structural probe used to single hit kinetics such that on average, the RNA of interest is only cleaved once per molecule. This ensures that the footprinting is performed on the original folded RNA, instead of on RNA that has refolded incorrectly after it has been cleaved.

Application of capillary electrophoresis to RNA structure probing is an important step in increasing the throughput of RNA structure data. Although RNA probing in solution can be readily implemented for short RNAs, probing of long RNAs can be challenging. Gel electrophoresis typically resolve about a hundred bases of RNA at a time and hence probing an RNA of several kilobases long would require running tens to hundreds of gels. Capillary electrophoresis allows the resolution of 300–650 bases from a structure probing experiment and multiple lanes can be run at the same time to increase its throughput of RNA structure probing[89, 90]. The readout of the probing experiment is typically through the reverse transcription of a 5′ fluorescently labeled DNA primer that anneals specifically to the RNA of interest. If the RNA is several kilobases long, multiple primers are designed to anneal along the length of the transcript. Modification or cleavage of the RNA template results in premature stops in the primer extension reaction, leading to different lengths of the cDNA product which are resolved by capillary electrophoresis. Software tools such as CAFA and Shapefinder can automate the data acquisition from capillary electrophoresis and further improve speed and accuracy[89, 90] (Fig. 3).

SHAPE and its applications to long RNAs

The method SHAPE uses the chemical NMIA and its derivatives to interrogate flexible regions in RNA secondary structure[80]. The 2′ OHs of flexible bases are able to orient themselves more readily for attack by the electrophile NMIA, resulting in the formation of 2-O adducts. These 2-O adducts can be detected by reverse transcription and capillary electrophoresis. As every ribonucleotide contains a 2′ OH, SHAPE has the advantage of being able to probe most bases in an RNA. With the coupling of SHAPE to capillary sequencing, SHAPE has been applied to interrogate the secondary structures of long RNAs, such as the 16S rRNA and the RNA genome of the human immunodeficiency virus (HIV)[59, 91, 92].

The construction of the secondary structure of the HIV genome using SHAPE was a landmark that demonstrates the substantial value of comprehensive RNA structure analysis[91]. The HIV genome is a 9kb long single stranded RNA that encodes nine open reading frames that are translated into fifteen proteins important for HIV infection and replication. Initial probing of the first 900 bases of the HIV genome across four different biological states showed highly similar secondary structures *in virio* and *ex virio*[92]. Regulatory regions within the 900 bases are found to be more structured than protein coding regions, and multiple regions within the RNA are found to interact with the nucleocapsid proteins. Structure probing of the entire 9kb HIV genome *ex virio* by SHAPE further found numerous regions within the genome that have functional roles in HIV replication[91]. These structured RNA domains provide insights into Gag-Pol frame-shifting, hyper-variable domains, and translocation of the Env protein. Interestingly, the nucleotides that encode for loops between independently folded protein domains are more structured than their surrounding bases, and are able to fold into secondary structures that retard the mobility of ribosomes for co-translational protein folding of modular domains[91].

Coupling RNA footprinting, such as SHAPE, to capillary sequencing has opened the door to structure probing of large RNAs, and it is likely that more RNA genomes, such as the polio virus and HCV virus, will be structurally probed to understand the role of RNA structures in viral replication. Furthermore, RNA structure probing is likely to extend beyond the probing of a single viral genome to families of viral genomes, to discover conserved or rapidly evolving structural elements that are likely to be functionally important in viral biology or pathogenicity. To facilitate this, the throughput of RNA structure probing can be greatly enhanced by coupling RNA footprinting to high throughput sequencing, which provides orders of magnitude of more sequencing information than capillary sequencing.

Genome-wide RNA structure maps—the next generation

Parallel analysis of RNA structure (PARS) and Fragmentation sequencing (Frag-seq)

The application of next-generation sequencing allowed the next major advance in genome-wide measurements of RNA structure, since millions of sequence reads can be obtained in a single experiment (Fig. 4). Cleavages or modifications at double or single stranded bases from structure probing can be captured and converted into cDNA libraries that are sequencing compatible. These sequencing reads are mapped back to the genome or the transcriptome to identify the transcript and the locations along the transcript that the cleavages occurred. The intensity of the cleavage at a base can also be calculated by summing the reads that are mapped to the base. This strategy allows the simultaneous identification of double or single stranded/flexible bases in thousands of RNAs in one experiment. In a strategy termed Parallel Analysis of RNA structure (PARS), deep sequencing reads of double- or single-stranded regions of RNAs generated by RNase V1 and S1 nuclease respectively are compared[21]. An alternative strategy, named Fragmentation sequencing (Frag-seq), quantifies deep sequencing reads generated specifically by RNase P1, a single-strand specific nuclease[93].

Using PARS, Kertesz et al. measured the secondary structure of the yeast transcriptome, generating structural information on ~4.2 million bases in over 3000 yeast transcripts[21]. Mapping PARS scores to known structures of regulatory motifs, such as Ash1 localization elements (required to properly localize *Ash1* mRNAs to the yeast bud tip) and the internal ribosomal entry site of *URE2* mRNA, indicates that PARS is able to capture the structural information in these elements, demonstrating the utility of this high throughput data. The large amount of PARS data provides insights into the global structural organization of mRNAs, including the presence of more secondary structure in coding regions as compared to untranslated regions, a three-nucleotide periodicity of secondary structure along the coding regions and an anti-correlation between mRNA translation efficiency and structure over mRNA translation start site (Fig. 5). Using Frag-seq, Underwood et al. correctly reconstructed the secondary structure of snoRNAs in mouse cells[93]. Both Frag-seq and PARS data can be integrated into structure prediction programs for more accurate RNA secondary structure prediction. PARS data was used to constrain a thermodynamic RNA structure prediction algorithm as binary inputs (paired vs. unpaired), while custom algorithm was developed to accommodate Frag-seq data. In essence, the nature, number, and location of structured regions in the transcriptome can be rapidly discovered, leading to many hypotheses and potential insights into gene regulation.

Comparison of PARS and Frag-seq reveals the complementary nature of the information that they both provide. First, because Frag-seq isolates RNAs between 20 to 100 bases after P1 cleavage without an additional fragmentation step, many sequence reads came from small nuclear RNAs, such as snoRNAs, while larger RNAs may be under-represented. Second, structured regions appear as “blanks” on Fraq-seq data, and other information is thus necessary to ensure that these regions are not missed due to mapping or cloning difficulties. Third, while PARS compares the cleavage sites of a single- vs. double-strand specific enzymes, Fraq-seq uses as background the endogenous 5′ OH and 5′ P within the transcriptome. This latter control can also identify regions that vary in their ability to be cloned and amplified during library production. Thus, by combining features from PARS and Frag-seq, future experiments can exploit the strengths of each to improve the accuracy of genome-scale measurements of RNA structure.

Recently, SHAPE has also been coupled to deep sequencing[94]. Lucks et al. in vitro transcribed seven short RNAs, each appended with a unique sequence tag (a barcode). After reacting with the SHAPE chemical 1M7 to acylate flexible bases, the reacted bases are indirectly detected by their ability to terminate the reverse transcription reaction and read out by sequencing the cDNAs. Because of the bar code, multiple sequences, even those with extensive sequence similarity, can be probed simultaneously. SHAPE-seq data correlate well with SHAPE followed capillary sequencing data for RNase P and pT181 attenuator, showing that sequencing largely captured similar structural information as capillary sequencing. This approach is likely useful for studying multiple mutants of one RNA or multiple members of closely related RNA family. Comparison of SHAPE-seq with PARS or Frag-seq illustrates several trade-offs in experimental design. The use of individual barcodes to assign identity to RNAs enables studies of highly related RNAs, but limits the ability to scale the same procedure up genome-wide, particularly when RNA sequences are not known a priori. Also, the choice to measure the cDNA product in SHAPE-seq, rather than directly clone the RNA fragments in PARS and Frag-seq, means that the processivity of reverse transcription becomes a dominant factor in SHAPE-seq data processing and the modeling of RNA secondary structure. SHAPE-seq signal progressively decays from 3′ to 5′ of the RNA template, the direction of reverse transcription, and a detailed mathematical model has been developed to correct for this signal decay[95]. Such models and the use of many more internal primers may allow full length mRNAs to be assessed by SHAPE-seq.

Advances relative to prior methods

The genome-scale RNA structure maps have three important advantages over prior methods. The first advantage is the amount of data measured by deep sequencing, which in itself is rapidly developing. While RNA footprinting with capillary sequencing is still very much directed at interrogating a single RNA of interest, PARS and Frag-seq have the power of probing structures of entire transcriptomes, comprised of tens of thousands of transcripts. Second, the degree of parallel multiplexing is much enhanced in the new methods. Capillary sequencing is typically performed with one purified RNA product and one primer per well. Thus, to study multiple genes, an investigator needs to clone each of these genes as well as prepare unique primers that span the length of the transcripts. In contrast, due to the massively parallel nature of deep sequencing technology, thousands of distinct RNAs of multiple kilobases long can be probed easily with high throughput sequencing, as long as the RNAs are fragmented to a size that is captured by the library preparation. This genome-wide approach allows biologists to compare the structural profile of one transcript to another in the transcriptome easily, enabling them to classify transcripts according to specific structural features.

Finally, PARS and Frag-Seq can also perform de novo transcript discovery and probe the structures of RNAs that were either not known to be present previously or underwent post-transcriptional modifications such as alternative splicing or RNA editing. In contrast, for capillary sequencing (or SHAPE-seq as it is currently practiced), the nucleotide sequence, as well as how the RNA is spliced, needs to be known so as to design primers along the length of the RNA to identify the bases that reacted with structural probes. This process is not only tedious but also restricts capillary sequencing to be used on structure probing of transcripts that are well annotated in the transcriptome.

Despite potential advantages, care and thoughtful controls are necessary to design and interpret genome-scale RNA structure maps, as has been done with RNA footprinting by capillary sequencing[96]. Key considerations include replicates to examine reproducibility, titration of structural probes to maintain single-hit kinetics, and controls to assess various biases that may arise from library preparation, deep sequencing, or mapping[97]. The use of positive control RNAs with well known structuresthat are doped into the genome-scale reactions is a useful measure to assess the quality of structural information generated by deep sequencing.

Using RNA structure maps to understand the transcriptome—the future

Much remains to be done and learned from genome-wide maps of RNA structure. First, it is likely that multiple technical advances will improve the quality of the maps. With classic RNA footprinting, multiple enzymes and chemical reagents are used to generate a consensus picture of RNA structure, and it is likely that multiple reagents, including DMS, lead and others, will be adapted to deep sequencing readouts. The use of third generation, single molecule sequencing platforms that do not require amplification, and are capable of reading hundreds to thousands of nucleotides, may also expand the range of questions that can be addressed. For instance, long-range structural impacts of alternative splicing of exons located hundreds or thousands of bases apart can be more simply evaluated.

Second, *in vivo* and dynamic RNA structure maps will yield critical understanding of how RNA structures may change and help regulate different biological states. Currently, both PARS and Frag-seq have probed the structures of RNAs that are isolated from cells and renatured in vitro, but these techniques can be readily applied to native RNA isolated without denaturation. Several chemical probes such as lead, DMS, NMIA and hydroxyl radicals, have been used successfully to probe RNA structures in vivo by penetrating cellular membranes[79, 82, 88, 92]. RNA footprinting can also occur under diverse conditions, such as alterations in temperature, the presence of specific proteins, or small molecule ligands, to probe the impact of these perturbations on RNA structure[10, 22, 98, 99].

Third, new computational strategies are emerging to better integrate experimental and computational RNA structures and delineate the impact on RNA function[58, 59]. The challenges are to predict the accurate structure of an RNA given its profile in the genomic RNA structure map, and further predict impacts of changes in the RNA structure (due to single nucleotide polymorphism, changes in biological state, or drug) on biological outcome. It is likely that cross comparison of genomic RNA structure maps with high resolution maps of RNA-protein interactions will be one immediate avenue whereby such integrative analyses can yield useful biological insights[24, 25].

Acknowledgments

We gratefully acknowledge the support of NIH (R01- HG004361), Agency of Science, Technology and Research of Singapore (Y.W.), and A.P. Giannini Foundation (R.C.S.). E.S. is the incumbent of the Soretta and Henry Shapiro career development chair. H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

References

1. Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. Nat Rev Mol Cell Biol. 2007; 8:113–126. [PubMed: 17245413]
2. Warf MB, Berglund JA. Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem Sci. 2010; 35:169–178. [PubMed: 19959365]
3. Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. Cell. 2009; 136:719–730. [PubMed: 19239891]
4. Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene. 2005; 361:13–37. [PubMed: 16213112]
5. Breaker RR. Riboswitches and the RNA World. Cold Spring Harb Perspect Biol. 2010
6. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet. 2009; 10:669–680. [PubMed: 19736561]
7. Nagalakshmi U, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008; 320:1344–1349. [PubMed: 18451266]
8. Blencowe BJ, Ahmad S, Lee LJ. Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. Genes Dev. 2009; 23:1379–1386. [PubMed: 19528315]
9. Henkin TM. Riboswitch RNAs: using RNA to sense cellular metabolism. Genes Dev. 2008; 22:3383–3390. [PubMed: 19141470]
10. Winkler W, Nahvi A, Breaker RR. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature. 2002; 419:952–956. [PubMed: 12410317]
11. Weinberg Z, et al. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biol. 2010; 11:R31. [PubMed: 20230605]
12. Barrick JE, Breaker RR. The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biol. 2007; 8:R239. [PubMed: 17997835]
13. Dann CE 3rd, et al. Structure and mechanism of a metal-sensing regulatory RNA. Cell. 2007; 130:878–892. [PubMed: 17803910]
14. Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR. Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. Cell. 2003; 113:577–586. [PubMed: 12787499]
15. Mandal M, et al. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science. 2004; 306:275–279. [PubMed: 15472076]
16. Nahvi A, Barrick JE, Breaker RR. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. Nucleic Acids Res. 2004; 32:143–150. [PubMed: 14704351]
17. Croft MT, Moulin M, Webb ME, Smith AG. Thiamine biosynthesis in algae is regulated by riboswitches. Proc Natl Acad Sci U S A. 2007; 104:20770–20775. [PubMed: 18093957]
18. Sudarsan N, Barrick JE, Breaker RR. Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA. 2003; 9:644–647. [PubMed: 12756322]
19. Mandal M, Breaker RR. Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat Struct Mol Biol. 2004; 11:29–35. [PubMed: 14718920] A single base in the adenine riboswitch determines the affinity of the aptamer for adenine versus guanine and illustrates the specificity of RNA structures in binding to their substrates.
20. Ray PS, et al. A stress-responsive RNA switch regulates VEGFA expression. Nature. 2009; 457:915–919. [PubMed: 19098893]
21. Kertesz M, et al. Genome-wide measurement of RNA secondary structure in yeast. Nature. 2010; 467:103–107. [PubMed: 20811459] First genome-wide RNA structure probing to probe double and single stranded regions in yeast, *in vitro*, using RNase V1 and S1 nuclease.
22. Kedde M, et al. A Pumilio-induced RNA structure switch in p27-3′ UTR controls miR-221 and miR-222 accessibility. Nat Cell Biol. 2010; 12:1014–1020. [PubMed: 20818387] Pumilio-1 protein binding results in a conformational change that allows miRNA binding sites in p27 to be accessible for regulation, illustrating that dynamics of RNA structure contributes to the complexity of gene regulation.

23. Zhao J, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell. 2010; 40:939–953. [PubMed: 21172659]

24. Licatalosi DD, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature. 2008; 456:464–469. [PubMed: 18978773]

25. Hafner M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010; 141:129–141. [PubMed: 20371350]

26. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003; 31:3406–3415. [PubMed: 12824337]

27. Hofacker IL, Fekete M, Stadler PF. Secondary structure prediction for aligned RNA sequences. J Mol Biol. 2002; 319:1059–1066. [PubMed: 12079347]

28. Do CB, Woods DA, Batzoglou S. CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics. 2006; 22:e90–8. [PubMed: 16873527]

29. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol. 1999; 288:911–940. [PubMed: 10329189] The paper reports the measurement of thermodynamic parameters for the stability of secondary structure motifs. Those have been extensively used by various prediction algorithms.

30. Rabani M, Kertesz M, Segal E. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. Proc Natl Acad Sci U S A. 2008; 105:14885–14890. [PubMed: 18815376]

31. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet. 2007; 39:1278–1284. [PubMed: 17893677]

32. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 2005; 33:W686–9. [PubMed: 15980563]

33. Chan PP, Lowe TM. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. 2009; 37:D93–7. [PubMed: 18984615]

34. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res. 2008; 36:D154–8. [PubMed: 17991681]

35. Rivas E, Eddy SR. Noncoding RNA gene detection using comparative sequence analysis. BMC Bioinformatics. 2001; 2:8. [PubMed: 11801179]

36. Lu ZJ, et al. Prediction and characterization of noncoding RNAs in C. elegans by integrating conservation, secondary structure, and high-throughput sequencing and array data. Genome Res. 2011; 21:276–285. [PubMed: 21177971]

37. Pedersen JS, et al. Identification and classification of conserved RNA secondary structures in the human genome. PLoS Comput Biol. 2006; 2:e33. [PubMed: 16628248]

38. Gautheret D, Lambert A. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. J Mol Biol. 2001; 313:1003–1011. [PubMed: 11700055]

39. Hochsmann M, Toller T, Giegerich R, Kurtz S. Local similarity in RNA secondary structures. Proc IEEE Comput Soc Bioinform Conf. 2003; 2:159–168. [PubMed: 16452790]

40. Pavesi G, Mauri G, Stefani M, Pesole G. RNAProfile: an algorithm for finding conserved secondary structure motifs in unaligned RNA sequences. Nucleic Acids Res. 2004; 32:3258–3269. [PubMed: 15199174]

41. Gorodkin J, Heyer LJ, Stormo GD. Finding the most significant common sequence and structure motifs in a set of RNA sequences. Nucleic Acids Res. 1997; 25:3724–3732. [PubMed: 9278497]

42. Eddy SR, Durbin R. RNA sequence analysis using covariance models. Nucleic Acids Res. 1994; 22:2079–2088. [PubMed: 8029015]

43. Sun FJ, Caetano-Anolles G. The origin and evolution of tRNA inferred from phylogenetic analysis of structure. J Mol Evol. 2008; 66:21–35. [PubMed: 18058157]

44. Washietl S, Hofacker IL, Stadler PF. Fast and reliable prediction of noncoding RNAs. Proc Natl Acad Sci U S A. 2005; 102:2454–2459. [PubMed: 15665081]

45. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. Nucleic Acids Res. 2003; 31:439–441. [PubMed: 12520045]

46. Yao Z, Weinberg Z, Ruzzo WL. CMfinder–a covariance model based RNA motif finding algorithm. Bioinformatics. 2006; 22:445–452. [PubMed: 16357030]

47. Knudsen B, Hein J. Pfold: RNA secondary structure prediction using stochastic context-free grammars. Nucleic Acids Res. 2003; 31:3423–3428. [PubMed: 12824339]

48. Seemann SE, Gorodkin J, Backofen R. Unifying evolutionary and thermodynamic information for RNA folding of multiple alignments. Nucleic Acids Res. 2008; 36:6355–6362. [PubMed: 18836192]

49. Mathews DH, Turner DH. Prediction of RNA secondary structure by free energy minimization. Curr Opin Struct Biol. 2006; 16:270–278. [PubMed: 16713706]

50. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res. 2008; 36:W70–4. [PubMed: 18424795] The Vienna RNA package is one of the most commonly used software suites for folding single and aligned sequences and predicting RNA-RNA interactions.

51. Dowell RD, Eddy SR. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. BMC Bioinformatics. 2004; 5:71. [PubMed: 15180907]

52. Doshi KJ, Cannone JJ, Cobaugh CW, Gutell RR. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. BMC Bioinformatics. 2004; 5:105. [PubMed: 15296519]

53. Hamada M, Kiryu H, Sato K, Mituyama T, Asai K. Prediction of RNA secondary structure using generalized centroid estimators. Bioinformatics. 2009; 25:465–473. [PubMed: 19095700]

54. Lu ZJ, Gloor JW, Mathews DH. Improved RNA secondary structure prediction by maximizing expected pair accuracy. RNA. 2009; 15:1805–1813. [PubMed: 19703939]

55. Halvorsen M, Martin JS, Broadaway S, Laederach A. Disease-associated mutations that alter the RNA structural ensemble. PLoS Genet. 2010; 6:e1001074. [PubMed: 20808897] Single nucleotide polymorphisms (SNPs) present in different disease states result in different predicted RNA structural conformations in UTRs.

56. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 1981; 9:133–148. [PubMed: 6163133]

57. Mathews DH, et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A. 2004; 101:7287–7292. [PubMed: 15123812]

58. Quarrier S, Martin JS, Davis-Neulander L, Beauregard A, Laederach A. Evaluation of the information content of RNA structure mapping data for secondary structure prediction. RNA. 2010; 16:1108–1117. [PubMed: 20413617]

59. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. Proc Natl Acad Sci U S A. 2009; 106:97–102. [PubMed: 19109441]

60. Xia T, et al. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry. 1998; 37:14719–14735. [PubMed: 9778347]

61. Liu B, Diamond JM, Mathews DH, Turner DH. Fluorescence competition and optical melting measurements of RNA three-way multibranch loops provide a revised model for thermodynamic parameters. Biochemistry. 2011; 50:640–653. [PubMed: 21133351]

62. van Batenburg FH, Gultyaev AP, Pleij CW. PseudoBase: structural information on RNA pseudoknots. Nucleic Acids Res. 2001; 29:194–195. [PubMed: 11125088]

63. Brierley I, Pennell S, Gilbert RJ. Viral RNA pseudoknots: versatile motifs in gene expression and replication. Nat Rev Microbiol. 2007; 5:598–610. [PubMed: 17632571]

64. Chen HL, Condon A, Jabbari H. An O(n(5)) algorithm for MFE prediction of kissing hairpins and 4-chains in nucleic acids. J Comput Biol. 2009; 16:803–815. [PubMed: 19522664]

65. Lyngso RB, Pedersen CN. RNA pseudoknot prediction in energy-based models. J Comput Biol. 2000; 7:409–427. [PubMed: 11108471]

66. Rivas E, Eddy SR. A dynamic programming algorithm for RNA structure prediction including pseudoknots. J Mol Biol. 1999; 285:2053–2068. [PubMed: 9925784]

67. Dirks RM, Pierce NA. An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. J Comput Chem. 2004; 25:1295–1304. [PubMed: 15139042]

68. Reeder J, Giegerich R. Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. BMC Bioinformatics. 2004; 5:104. [PubMed: 15294028]

69. Ruan J, Stormo GD, Zhang W. An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. Bioinformatics. 2004; 20:58–66. [PubMed: 14693809]

70. Ren J, Rastegari B, Condon A, Hoos HH. HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. RNA. 2005; 11:1494–1504. [PubMed: 16199760]

71. Chen X, et al. FlexStem: improving predictions of RNA secondary structures with pseudoknots by reducing the search space. Bioinformatics. 2008; 24:1994–2001. [PubMed: 18586700]

72. Bellaousov S, Mathews DH. ProbKnot: fast prediction of RNA secondary structure including pseudoknots. RNA. 2010; 16:1870–1880. [PubMed: 20699301]

73. Sato K, Kato Y, Hamada M, Akutsu T, Asai K. IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. Bioinformatics. 2011; 27:i85–i93. [PubMed: 21685106]

74. Weeks KM. Advances in RNA structure analysis by chemical probing. Curr Opin Struct Biol. 2010; 20:295–304. [PubMed: 20447823]

75. Ehresmann C, et al. Probing the structure of RNAs in solution. Nucleic Acids Res. 1987; 15:9109–9128. [PubMed: 2446263]

76. Romby P, et al. Ribosomal 5S RNA from Xenopus laevis oocytes: conformation and interaction with transcription factor IIIA. Biochimie. 1990; 72:437–452. [PubMed: 2124147]

77. Wurst RM, Vournakis JN, Maxam AM. Structure mapping of 5′-32P-labeled RNA with S1 nuclease. Biochemistry. 1978; 17:4493–4499. [PubMed: 363143]

78. 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. 1989; 6:971–984. [PubMed: 2686708]

79. Wells SE, Hughes JM, Igel AH, Ares M Jr. Use of dimethyl sulfate to probe RNA structure in vivo. Methods Enzymol. 2000; 318:479–493. [PubMed: 10890007]

80. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc. 2005; 127:4223–4231. [PubMed: 15783204] SHAPE reagents react with 2′OH of all four RNA bases to probe for flexible structural regions in an RNA and have emerged to be excellent chemical probes.

81. Lowman HB, Draper DE. On the recognition of helical RNA by cobra venom V1 nuclease. J Biol Chem. 1986; 261:5396–5403. [PubMed: 2420800]

82. Adilakshmi T, Lease RA, Woodson SA. Hydroxyl radical footprinting in vivo: mapping macromolecular structures with synchrotron radiation. Nucleic Acids Res. 2006; 34:e64. [PubMed: 16682443]

83. Shcherbakova I, Mitra S. Hydroxyl-radical footprinting to probe equilibrium changes in RNA tertiary structure. Methods Enzymol. 2009; 468:31–46. [PubMed: 20946763]

84. Das R, Laederach A, Pearlman SM, Herschlag D, Altman RB. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. RNA. 2005; 11:344–354. [PubMed: 15701734]

85. Zemora G, Waldsich C. RNA folding in living cells. RNA Biol. 2010; 7:634–641. [PubMed: 21045541]

86. Liebeg A, Waldsich C. Probing RNA structure within living cells. Methods Enzymol. 2009; 468:219–238. [PubMed: 20946772]

87. Russell R. RNA misfolding and the action of chaperones. Front Biosci. 2008; 13:1–20. [PubMed: 17981525]

88. Lindell M, Romby P, Wagner EG. Lead(II) as a probe for investigating RNA structure in vivo. RNA. 2002; 8:534–541. [PubMed: 11991646]

89. Mitra S, Shcherbakova IV, Altman RB, Brenowitz M, Laederach A. High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. Nucleic Acids Res. 2008; 36:e63. [PubMed: 18477638]

90. Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA. 2008; 14:1979–1990. [PubMed: 18772246]

91. Watts JM, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 2009; 460:711–716. [PubMed: 19661910] RNA structure probing of the entire 9kb HIV RNA genome provided many insights into differentially structured regions and their biological functions.

92. Wilkinson KA, et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol. 2008; 6:e96. [PubMed: 18447581]

93. Underwood JG, et al. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. Nat Methods. 2010; 7:995–1001. [PubMed: 21057495] Genome-wide RNA structure probing in using P1 nuclease to probe for single stranded regions in mouse cells, *in vitro*.

94. Lucks JB, et al. Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc Natl Acad Sci U S A. 2011; 108:11063–11068. [PubMed: 21642531] Chemical structure probing is coupled to high throughput sequencing by using a SHAPE reagent to probe flexible structural regions of seven RNAs *in vitro*.

95. Aviran S, et al. From the Cover: Modeling and automation of sequencing-based characterization of RNA structure. Proc Natl Acad Sci U S A. 2011; 108:11069–11074. [PubMed: 21642536]

96. Weeks KM. RNA structure probing dash seq. Proc Natl Acad Sci U S A. 2011; 108:10933–10934. [PubMed: 21700884]

97. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 2011; 12:R22. [PubMed: 21410973]

98. Leipply D, Draper DE. Dependence of RNA tertiary structural stability on Mg2+ concentration: interpretation of the Hill equation and coefficient. Biochemistry. 2010; 49:1843–1853. [PubMed: 20112919]

99. Chowdhury S, Maris C, Allain FH, Narberhaus F. Molecular basis for temperature sensing by an RNA thermometer. EMBO J. 2006; 25:2487–2497. [PubMed: 16710302]

100. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008; 322:750–756. [PubMed: 18974356]

101. Kanhere A, et al. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol Cell. 2010; 38:675–688. [PubMed: 20542000]

102. Tsai MC, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010; 329:689–693. [PubMed: 20616235]

103. Kaneko S, et al. Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. Genes Dev. 2010; 24:2615–2620. [PubMed: 21123648]

104. Kotake Y, et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene. 2011; 30:1956–1962. [PubMed: 21151178]

105. Wanrooij PH, Uhler JP, Simonsson T, Falkenberg M, Gustafsson CM. G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. Proc Natl Acad Sci U S A. 2010; 107:16072–16077. [PubMed: 20798345]

106. Henkin TM, Grundy FJ. Sensing metabolic signals with nascent RNA transcripts: the T box and S box riboswitches as paradigms. Cold Spring Harb Symp Quant Biol. 2006; 71:231–237. [PubMed: 17381302]

107. Wang J, Nikonowicz EP. Solution structure of the K-turn and Specifier Loop domains from the Bacillus subtilis tyrS T-box leader RNA. J Mol Biol. 2011; 408:99–117. [PubMed: 21333656]

108. Lu C, et al. SAM recognition and conformational switching mechanism in the Bacillus subtilis yitJ S box/SAM-I riboswitch. J Mol Biol. 2010; 404:803–818. [PubMed: 20951706]

109. Deikus G, Bechhofer DH. Bacillus subtilis trp Leader RNA: RNase J1 endonuclease cleavage specificity and PNPase processing. J Biol Chem. 2009; 284:26394–26401. [PubMed: 19638340]

110. Butler EB, Xiong Y, Wang J, Strobel SA. Structural basis of cooperative ligand binding by the glycine riboswitch. Chem Biol. 2011; 18:293–298. [PubMed: 21439473]

111. Zhang Q, Kang M, Peterson RD, Feigon J. Comparison of solution and crystal structures of preQ1 riboswitch reveals calcium-induced changes in conformation and dynamics. J Am Chem Soc. 2011; 133:5190–5193. [PubMed: 21410253]

112. Kar A, et al. RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5′ splice site. Mol Cell Biol. 2011; 31:1812–1821. [PubMed: 21343338]

113. Warf MB, Diegel JV, von Hippel PH, Berglund JA. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. Proc Natl Acad Sci U S A. 2009; 106:9203–9208. [PubMed: 19470458]

114. Oikawa D, Tokuda M, Hosoda A, Iwawaki T. Identification of a consensus element recognized and cleaved by IRE1 alpha. Nucleic Acids Res. 2010; 38:6265–6273. [PubMed: 20507909]

115. Yang Y, et al. RNA secondary structure in mutually exclusive splicing. Nat Struct Mol Biol. 2011; 18:159–168. [PubMed: 21217700]

116. Cheah MT, Wachter A, Sudarsan N, Breaker RR. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature. 2007; 447:497–500. [PubMed: 17468745]

117. Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. Science. 2010; 329:845–848. [PubMed: 20705859]

118. Aragon T, et al. Messenger RNA targeting to endoplasmic reticulum stress signalling sites. Nature. 2009; 457:736–740. [PubMed: 19079237]

119. Gonsalvez GB, Urbinati CR, Long RM. RNA localization in yeast: moving towards a mechanism. Biol Cell. 2005; 97:75–86. [PubMed: 15601259]

120. Bullock SL, Ringel I, Ish-Horowicz D, Lukavsky PJ. A′-form RNA helices are required for cytoplasmic mRNA transport in Drosophila. Nat Struct Mol Biol. 2010; 17:703–709. [PubMed: 20473315]

121. Subramanian M, et al. G-quadruplex RNA structure as a signal for neurite mRNA targeting. EMBO Rep. 2011; 12:697–704. [PubMed: 21566646]

122. Chao JA, et al. ZBP1 recognition of beta-actin zipcode induces RNA looping. Genes Dev. 2010; 24:148–158. [PubMed: 20080952]

123. Van Etten RA, et al. The COOH terminus of the c-Abl tyrosine kinase contains distinct F- and G-actin binding domains with bundling activity. J Cell Biol. 1994; 124:325–340. [PubMed: 8294516]

124. Mayer C, Neubert M, Grummt I. The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. EMBO Rep. 2008; 9:774–780. [PubMed: 18600236]

125. Parsons CJ, et al. Mutation of the 5′-untranslated region stem-loop structure inhibits alpha1(I) collagen expression in vivo. J Biol Chem. 2011; 286:8609–8619. [PubMed: 21193410]

126. Cho HH, et al. Selective translational control of the Alzheimer amyloid precursor protein transcript by iron regulatory protein-1. J Biol Chem. 2010; 285:31217–31232. [PubMed: 20558735]

127. Goforth JB, Anderson SA, Nizzi CP, Eisenstein RS. Multiple determinants within iron-responsive elements dictate iron regulatory protein binding and regulatory hierarchy. RNA. 2010; 16:154–169. [PubMed: 19939970]

128. Shahid R, Bugaut A, Balasubramanian S. The BCL-2 5′ untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. Biochemistry. 2010; 49:8300–8306. [PubMed: 20726580]

129. Derecka K, et al. Occurrence of a quadruplex motif in a unique insert within exon C of the bovine estrogen receptor alpha gene (ESR1). Biochemistry. 2010; 49:7625–7633. [PubMed: 20715834]

130. Gomez D, et al. A G-quadruplex structure within the 5′-UTR of TRF2 mRNA represses translation in human cells. Nucleic Acids Res. 2010; 38:7187–7198. [PubMed: 20571083]

131. Reineke LC, Komar AA, Caprara MG, Merrick WC. A small stem loop element directs internal initiation of the URE2 internal ribosome entry site in Saccharomyces cerevisiae. J Biol Chem. 2008; 283:19011–19025. [PubMed: 18460470]

132. Feng S, et al. Alternate rRNA secondary structures as regulators of translation. Nat Struct Mol Biol. 2011; 18:169–176. [PubMed: 21217697]

133. Waldminghaus T, Heidrich N, Brantl S, Narberhaus F. FourU: a novel type of RNA thermometer in Salmonella. Mol Microbiol. 2007; 65:413–424. [PubMed: 17630972]

134. Giuliodori AM, et al. The cspA mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. Mol Cell. 2010; 37:21–33. [PubMed: 20129052]

135. Kortmann J, Sczodrok S, Rinnenthal J, Schwalbe H, Narberhaus F. Translation on demand by a simple RNA-based thermosensor. Nucleic Acids Res. 2011; 39:2855–2868. [PubMed: 21131278]

136. Badis G, Saveanu C, Fromont-Racine M, Jacquier A. Targeted mRNA degradation by deadenylation-independent decapping. Mol Cell. 2004; 15:5–15. [PubMed: 15225544]

137. Prouteau M, Daugeron MC, Seraphin B. Regulation of ARE transcript 3′ end processing by the yeast Cth2 mRNA decay factor. EMBO J. 2008; 27:2966–2976. [PubMed: 18923425]

138. Fukuchi M, Tsuda M. Involvement of the 3′-untranslated region of the brain-derived neurotrophic factor gene in activity-dependent mRNA stabilization. J Neurochem. 2010; 115:1222–1233. [PubMed: 20874756]

139. Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. Control of gene expression by a natural metabolite-responsive ribozyme. Nature. 2004; 428:281–286. [PubMed: 15029187]

140. McCown PJ, Roth A, Breaker RR. An expanded collection and refined consensus model of glmS ribozymes. RNA. 2011; 17:728–736. [PubMed: 21367971]

141. Elemento O, Slonim N, Tavazoie S. A universal framework for regulatory element discovery across all genomes and data types. Mol Cell. 2007; 28:337–350. [PubMed: 17964271]

142. Riordan DP, Herschlag D, Brown PO. Identification of RNA recognition elements in the Saccharomyces cerevisiae transcriptome. Nucleic Acids Res. 2011; 39:1501–1509. [PubMed: 20959291]

143. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994; 2:28–36. [PubMed: 7584402]

144. Ray D, et al. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nat Biotechnol. 2009; 27:667–670. [PubMed: 19561594]

145. Li X, Quon G, Lipshitz HD, Morris Q. Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. RNA. 2010; 16:1096–1107. [PubMed: 20418358]

146. Montange RK, Batey RT. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. Nature. 2006; 441:1172–1175. [PubMed: 16810258]

147. Gilbert SD, Rambo RP, Van Tyne D, Batey RT. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. Nat Struct Mol Biol. 2008; 15:177–182. [PubMed: 18204466]

148. Lu C, et al. Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism. Nat Struct Mol Biol. 2008; 15:1076–1083. [PubMed: 18806797]

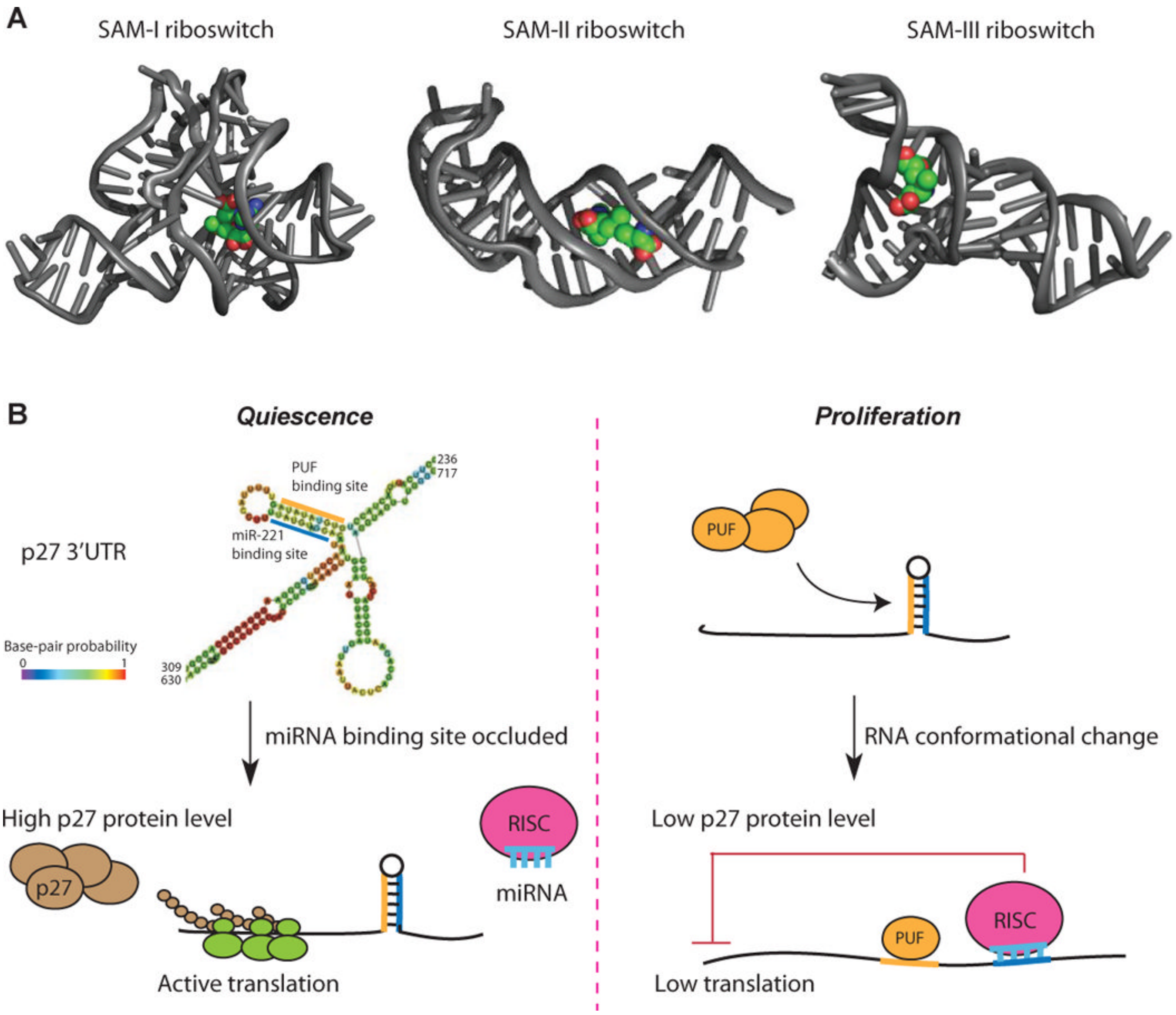

Figure 1.

Diversity and dynamics of RNA structures. **A**, Different classes of SAM riboswitches bind specifically to SAM. The backbone of the riboswitch is in grey while the SAM molecule is colored. SAM-I[146] [PDB number: 2GIS]; SAM-II[147] [PDB number:2QWY]; S(MK) riboswitch[148] [PDB number3E5C]. Images are generated using Pymol. **B**, Dynamic changes in p27 mRNA structure upon Puf binding results in changes in p27 gene expression. Left panel: During quiescence, the miRNA binding site in 3′UTR of p27 is in a folded structure and is not accessible to miRNA. Translation of p27 mRNA results in high p27 protein levels to maintain quiescence. Right panel: During cellular proliferation, binding of Puf proteins to p27 mRNA causes a structural change that allows miRNA binding sites to be accessible to miR-221 and miR-222, resulting in translation repression of p27. Low p27 levels allow the cells to exit cellular quiescence and enter the cell cycle. [Figure modified from Kedde M. et al, 2010]

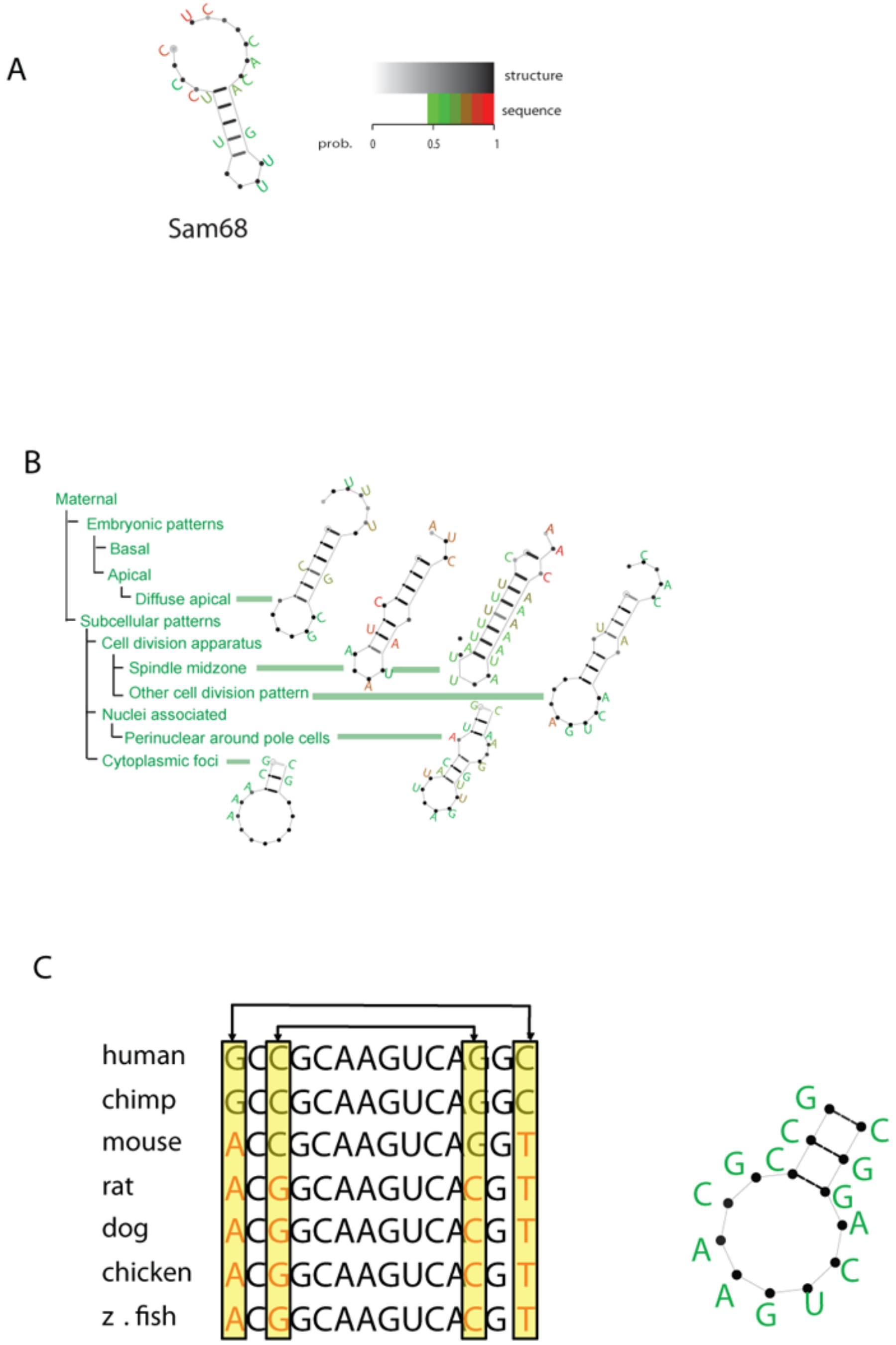

Figure 2.

Predicting structural motifs for RNA binding protein (RPB) targets in mRNAs from different organisms. **A**, By applying an RNA motif finder, investigators were able to identify a significant stem-loop structural motif by analyzing the eight known targets of the human RBP Sam68. **B**, The same RNA motif finder was used to analyze data collected from a large study on mRNA localization during fly embryonic development, to predict significant motifs in six sets of colocalized maternal transcripts. Shown is the structural motif enriched in each set of mRNAs. **C**, Conservation of base pairs in homologous sequences directs structure prediction. Sequence covariation found at aligned positions. Shown is an example alignment of seven RNA sequences. In the example, sequence covariation in between the two sets of marked columns hint at interacting bases.

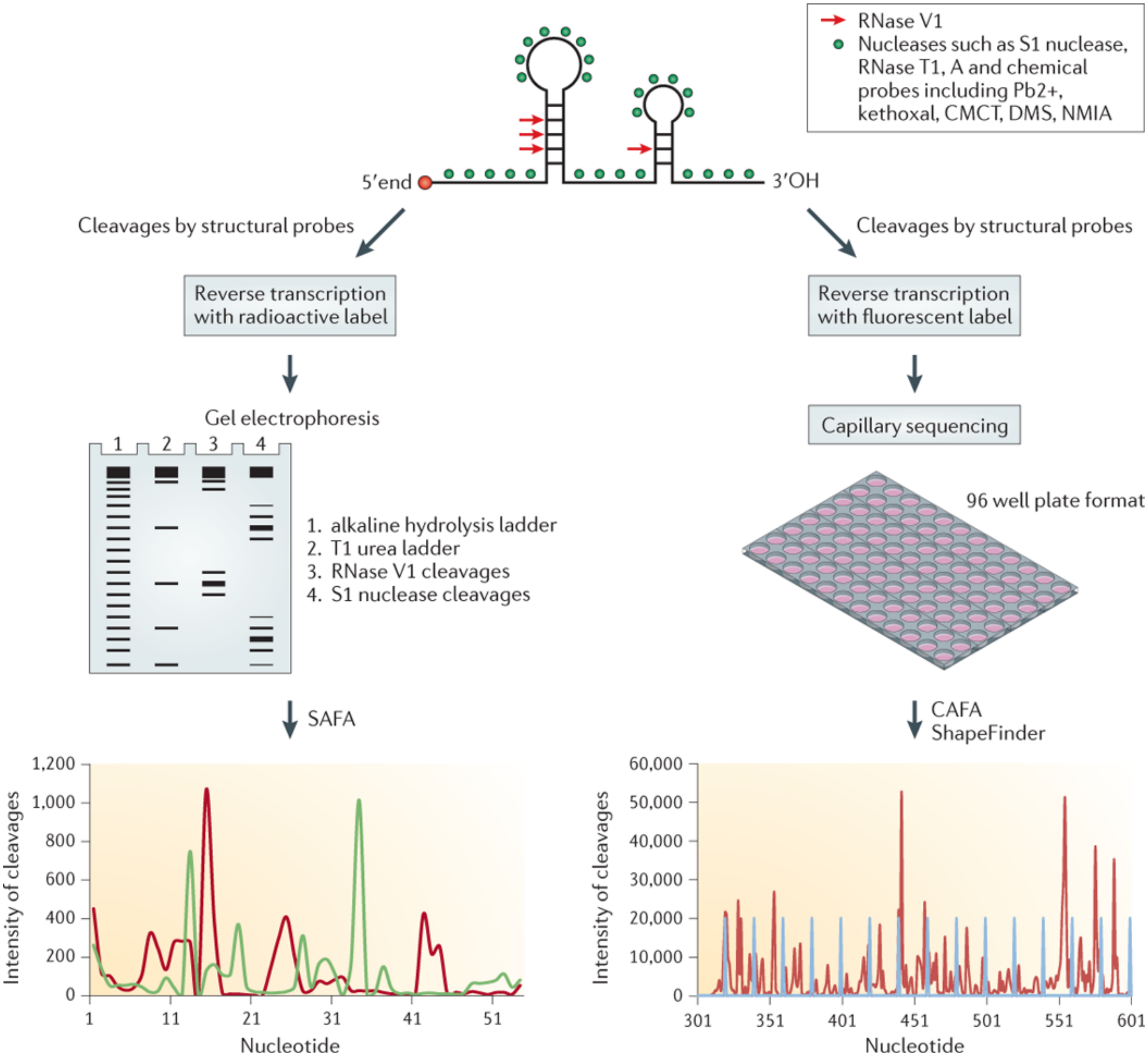

Figure 3.

Structure probing by RNA footprinting followed by gel and capillary electrophoresis. An RNA of interest is typically in-vitro transcribed, folded and subjected to a combination of single and double stranded structural probes in solution. Cleavages in double or single stranded regions can either be identified by running a gel electrophoresis (RNA needs to be radioactively labelled at one end) or be identified via primer extension followed by capillary electrophoresis (primer needs to be florescently labelled). The bands from gel electrophoresis can be quantitated using a program called SAFA, while bands in capillary electrophoresis are identified and quantitated using CAFA or ShapeFinder. In gel electrophoresis, the SAFA quantitated green lines refer to the intensity of S1 nuclease cleavages while the red lines refer to the intensity of RNase V1 cleavages. The positions of these cleaved bases are determined from the RNase T1 ladder and alkaline hydrolysis ladder. In capillary electrophoresis, the red line indicates the intensity of structure probing sites that are detected by reverse transcription, while the grey line corresponds to a ladder that positions the RNA bases.

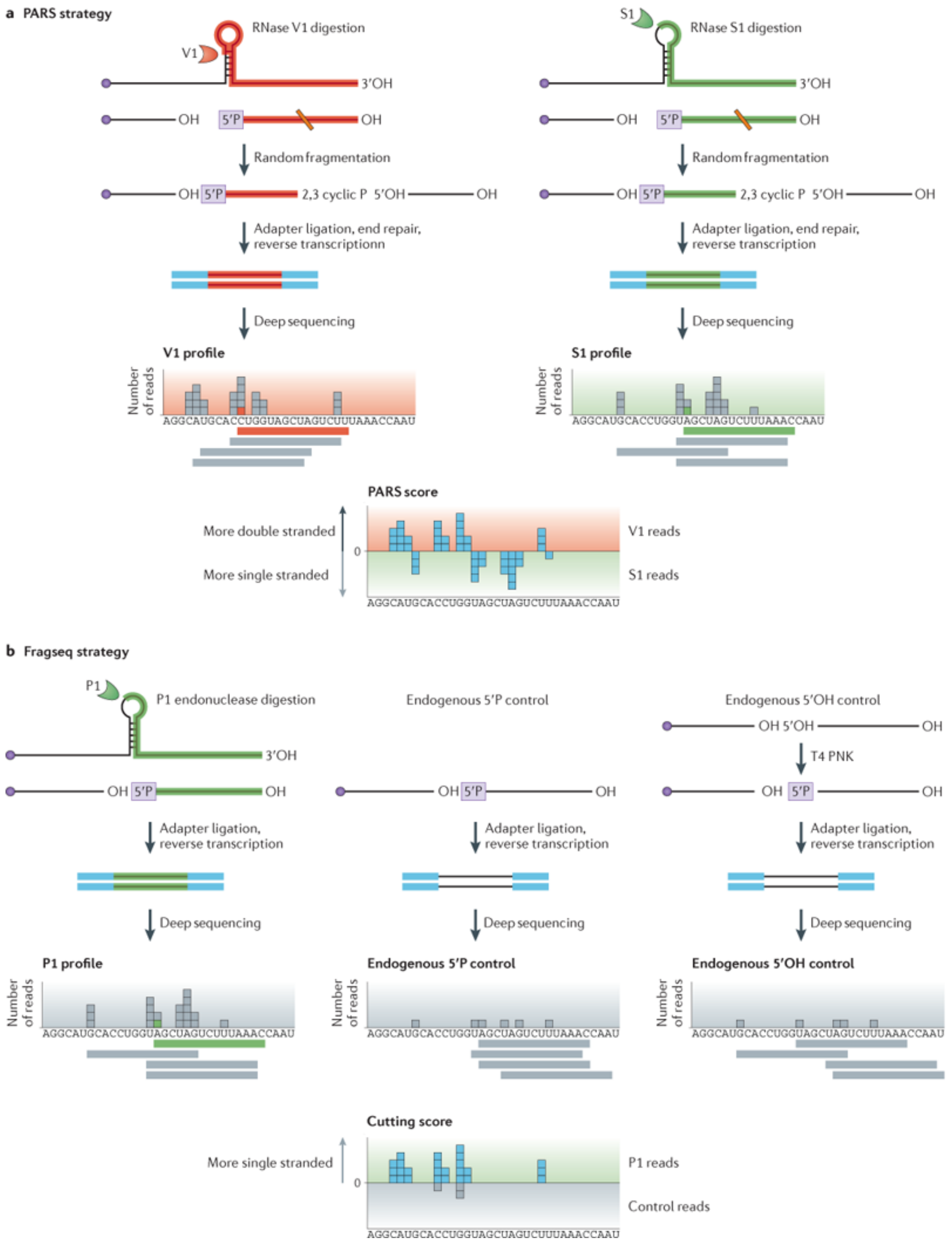

Figure 4.

PARS and Frag-seq methods. **A**, PARS strategy. In PARS, polyA selected RNA is folded in vitro and incubated with either RNase V1 or S1 nuclease to probe for double and single stranded regions respectively. RNase V1 and S1 nuclease cleave resulting in a 5′P leaving group. The enzymatically probed RNA is then fragmented. As enzymatic cleavage products contain 5′P whereas fragmentation and degradation products have 5′OH, only true structure probing sites can be ligated to adapters and reverse transcribed. The cDNA library is sequenced using high throughput sequencing and the resulting reads are mapped to the genome to identify double/single stranded regions in the transcriptome. A PARS score can be calculated at each base whereby a positive PARS score indicates that a base is double stranded and a negative PARS score indicates that a base is single stranded. **B**, Fragseq stratgy. Nuclear RNA is folded in vitro and probed in solution with P1 endonuclease. P1 cleaves at single stranded regions, resulting in a 5′P leaving group. This 5′P can be captured by adapter ligation, followed by reverse transcription and high throughput sequencing. Sequencing reads are mapped back to the genome to identify where single stranded bases are located in the transcriptome. Fragseq also contains controls which include sequencing of endogenous 5′P and 5′OH that are originally present in the untreated RNA samples. A cutting score can be calculated at each base which incorporates reads from P1 nuclease and reads from endogenous degradation or fragmentation products. A positive cutting score indicates that the base is single stranded.

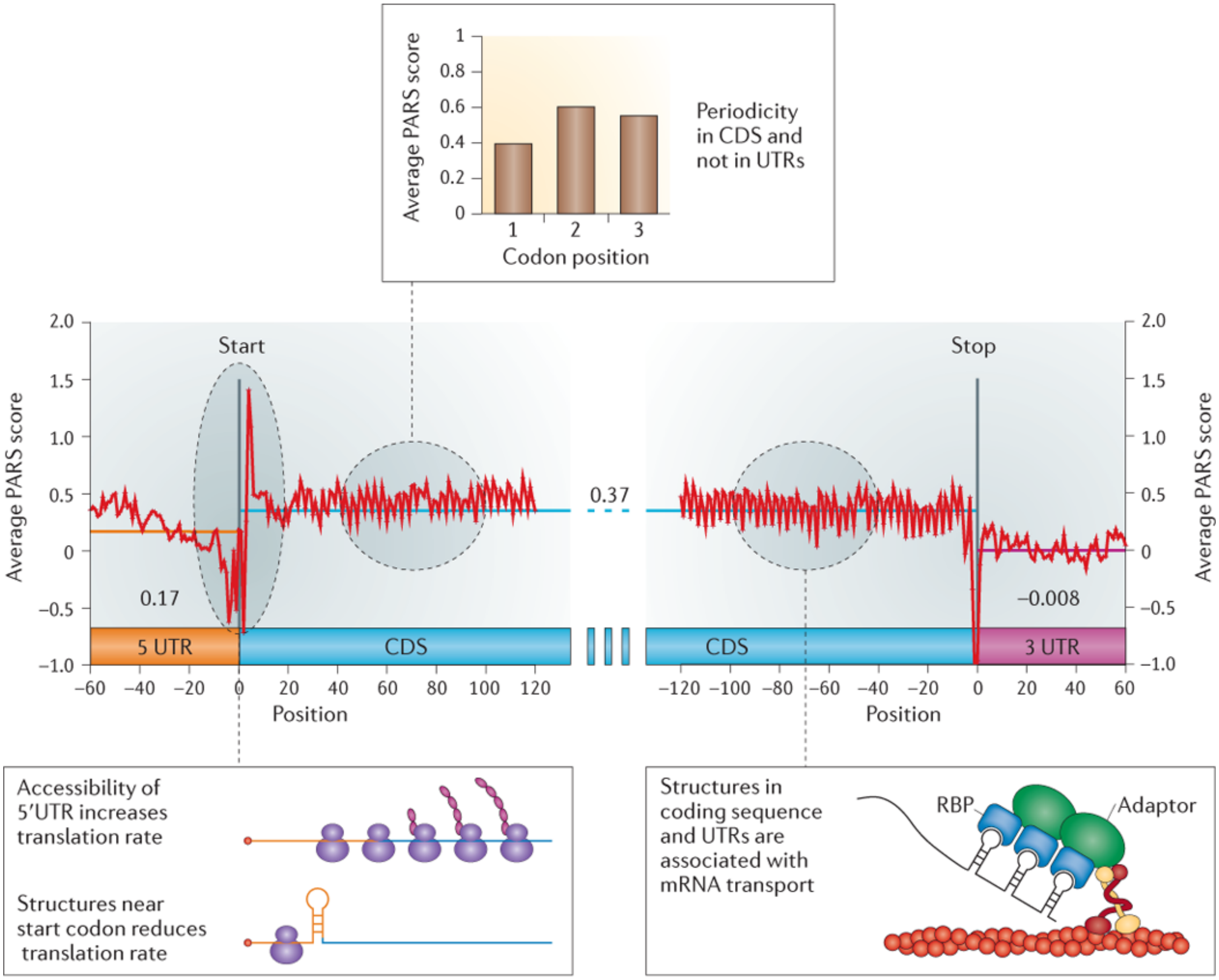

Figure 5.

Structural organization of the mRNA transcriptome. Thousands of yeast mRNAs are structurally probed in PARS and aligned according to their start and stop codons. The average PARS score of the coding sequence (CDS) is shown in blue; 5′ untranslated region (UTR) in yellow; 3′ UTR in red. The organization of secondary structures within the transcriptome revealed an increased accessibility of RNA structure near the start codon important for translation efficiency, shown by the negative spike. The coding sequence is more structured than the UTRs, as shown by the higher average (blue line), compared to the UTRs (orange and red lines). Some of these structures are important for cellular processes such as mRNA transport. A three nucleotide periodicity (inset box) in RNA is also seen in the coding region and is absent from the UTRs.

Table 1

Diverse roles of RNA structure in gene expression. Several recent examples are highlighted.

Process	RNA type	Examples	Role of RNA structures	Refs
Transcription	LncRNAs	Xist, HOTAIR, ANRIL, promoter associated RNAs	Double stem loop and other structural motifs recruit Polycomb complex for gene silencing.	100–104
	ncRNA	Mitochondria RNA	G-quadruplex structures cause transcription termination-mammals	105
	Riboswitch	Adenine, guanine, lysine, glycine, T-box, trp, SAM, preQ1	Structure change upon ligand binding results in either transcription termination or activation- bacteria	5, 106–111
Splicing	mRNAs	Tau, cardia troponin	Protein binding to stem loop at cause alternative splicing-mammals	112, 113
		CD59, XBP1	IRE1α recognizes stem loop for splicing- mammals	114
		14-3-3ξ	Inter-intronic RNA pairing results in mutually exclusive splicing-Drosophila	115
	Riboswitch	Group I ribozyme, TPP	Binding to metabolites alters splicing- bacteria, fungi, plants	116, 117
RNA Localization	mRNAs	Hac1	Localization to yeast ER membrane	118
		ATP2, ATM1	Localization to yeast mitochondria	119
		K10	A form helix causes localization to anterior of *Drosophila* oocyte	120
		PSD-95/CaMKIIa	G-quadruplex in 3′UTR targeting to neurites–mammals	121
		Beta-actin	Localizes to the leading edge of fibroblasts/neurons–mammals	122, 123
	ncRNA	Promoter RNA	Stem loop results in nucleoli localization–mammals	124
Translation	mRNAs	p27, VEGFA	Protein binding causes structural changes in mammals	20, 22
		Collagen genes, amyloid precursor protein, ferritin	Stem loop at 5′UTR	125–127
		BCL-2, ERα, TRF2	G-quadruplex in 5′UTR affects translation-mammals	128–130
		URE2	Stem loop as internal ribosomal entry site- yeast	131
	ncRNA	rRNA	Binding of Z-DNA-binding domain to rRNA structures block translation-bacteria/mammals	132
	Riboswitch	FourU, ROSE element, cspA	RNA thermometers respond to temperature changes	99, 133–135
RNA Decay	mRNAs	Rps28b	Structure recruits decapping proteins for decay-yeast	136
		Cth2	adenosine/uridine-rich (ARE) elements in 3′UTR-yeast	137
		BDNF	Stem loop in 3′UTR prevents decay in presence of Ca2+ –mammals	138
	Riboswitch	GlcN6P riboswitches	Ligand binding results in cleavage of RNA	139, 140