Understanding the transcriptome through RNA structure

Yue Wan, Michael Kertesz‡, Robert C. Spitale*, Eran Segal§ and Howard Y. Chang**

Abstract | RNA structure is crucial for gene regulation and function. In the past, transcriptomes have largely been 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 high-throughput sequencing has enabled genome-wide measurements of RNA structure and has provided the first picture of the structural organization of a 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.

Sequence covariation
Nucleotide substitutions that differ between two or more homologous genes but retain the potential for Watson–Crick base pairing in an RNA molecule in each sequence, thus suggesting a selective pressure to retain those base pairings.

**Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA.*
‡Howard Hughes Medical Institute and Department of Bioengineering, Stanford University, Stanford, California 94305, USA.
§Department of Computer Science and Applied Mathematics and Weizmann Institute of Science, Rehovot 76100, Israel.
Correspondence to H.Y.C.
e-mail:
howchang@stanford.edu
doi:10.1038/nrg3049

RNA is a unique informational molecule. In addition to carrying information in their linear sequences of nucleotides (that is, their primary structure), RNA molecules fold into intricate shapes. Pairing of local nucleotides can create secondary structures, such as hairpins and stem–loops, and interactions among distantly located sequences can create tertiary structures. RNA structures influence the transcription, splicing, cellular localization, translation and turnover of the RNA. The topic of RNA structures in different cellular processes has been covered in several excellent reviews[1–5]. Although the structures of multiple RNAs have been studied in detail, structural information for most RNAs — such as mRNAs — is missing owing 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 RNA structures are studied on a case-by-case basis, it is difficult to determine the full impact of an RNA's structure 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 greatly advanced in their ability to predict more accurate secondary structures from both primary sequences and sequence covariation. These predicted structures are typically confirmed by secondary-structure probing, which still serves as the gold standard of RNA structure determination.

The advent of high-throughput sequencing technologies has enabled the sequencing of hundreds of millions of bases at a time and has greatly increased the speed of acquisition and precision of genomic data. High-throughput sequencing has been successfully applied 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 into 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, and we review advances in computational structure predictions. We then focus on experimental approaches to derive 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 and non-coding RNAs (ncRNAs), such as riboswitches, ribozymes, long non-coding RNAs (lncRNAs) and microRNAs (miRNAs). RNA structures have roles in nearly every step of gene expression (TABLE 1). RNA structures enable an RNA to interact with itself, with other RNAs, with ligands and with RNA-binding proteins (RBPs). Many of

Table 1 | **Examples of the diverse roles of RNA structure in gene expression**

RNA type	Examples	Example roles of RNA structures	Refs
Transcription			
Long and short ncRNAs	*Xist*, *HOTAIR*, *ANRIL*, promoter associated RNAs	Double stem–loop and other structural motifs recruit Polycomb complex for gene silencing (mammals)	111–115
Mitochondrial RNA		G-quadruplex structures cause transcription termination (mammals)	116
Riboswitch	Adenine, guanine, lysine, glycine, T box, TPP, SAM, pre-Q1	Structure change upon ligand binding results in either transcription termination or activation (bacteria)	5,117–122
Splicing			
mRNAs	Tau, cardiac troponin	Protein binding to stem–loop regulates alternative splicing (mammals)	123,124
	CD59, *XBP1*	IRE1α recognizes stem–loop for splicing (mammals)	125
	14-3-3ξ	Inter-intronic RNA pairing results in mutually exclusive splicing (*Drosophila melanogaster*)	126
Riboswitch	Group I ribozyme, TPP	Binding to metabolites alters splicing (bacteria, fungi, plants)	127,128
RNA localization			
mRNAs	*Hac1*	Localization to yeast endoplasmic reticulum membrane	129
	ATP2, *ATM1*	Localization to yeast mitochondria	130
	fs(1)K10	A-form helix causes localization to anterior of *Drosophila* oocyte	131
	PSD95/CaMKIIa	G-quadruplex in 3′UTR targeting to neuritis (mammals)	132
	β-actin	Localizes to the leading edge of fibroblasts or neurons (mammals)	133,134
ncRNA	Promoter RNA	Stem–loop results in nucleoli localization (mammals)	135
Translation			
mRNAs	*p27*, *VEGFA*	Protein binding causes structural changes (mammals)	20,22
	Collagen genes, amyloid precursor protein, ferritin	Stem–loop at 5′UTR (mammals)	136–138
	BCL2, *ERA*, *TRF2*	G-quadruplex in 5′UTR affects translation (mammals)	139–141
	URE2	Stem–loop as internal ribosomal entry site (yeast)	142
ncRNA	rRNA	Binding of Z-DNA-binding domain to rRNA structures block translation (bacteria and mammals)	143
Riboswitch	FourU, ROSE element, CSPA, TPP, SAM	Structure change on ligand binding and temperature variance changes accessibility of Shine–Dalgarno sequence for ribosomal recognition (bacteria)	9,102, 144–146
RNA decay			
mRNAs	*RPS28B*	Structure recruits decapping proteins for decay (yeast)	147
	Cth2	Adenosine or uridine-rich (ARE) elements in 3′UTR (yeast)	148
	BDNF	Stem–loop in 3′UTR prevents decay in presence of Ca^{2+} (mammals)	149
Riboswitch	GlcN6P riboswitches	Ligand binding results in cleavage of RNA (bacteria)	150,151

ANRIL, antisense non-coding RNA in the INK locus; *ATM1*, ABC transporter mitochondrial; *ATP2*, ATP-synthase subunit β-mitochondrial; *BCL2*, B-cell lymphoma 2; *BDNF*, brain-derived neurotrophic factor; CSPA, cold shock protein A; *Cth2*, cysteine-three-histidine protein 2; *CaMKIIa*, calmodulin-dependent protein kinase II; *ERA*, oestrogen receptor-α; *fs(1)K10*, female sterile (1) K10; GlcN6P, glucosamine-6-phosphate; *Hac1*, homologous to ATF/CREB1; *HOTAIR*, Hox transcript antisense RNA; IRE1α, serine/threonine-protein kinase/endoribonuclease IRE1; ncRNA, non-coding RNA; pre-Q1, pre-queuosine 1; *PSD95*, post-synaptic density protein 95; ROSE, repression of heat shock gene expression; *RPS28B*, ribosomal protein S28B; rRNA, ribosomal RNA; SAM, S-adenosylmethione; TPP, thiamine pyrophosphate; *TRF2*. telomeric repeat-binding factor 2; *VEGFA*, vascular endothelial growth factor A; *XBP1*, X box binding protein 1; *Xist*, X-inactive specific transcript.

Box 1 | **RNA-binding proteins: motif identification and prediction**

RNA-binding proteins (RBPs) interact with RNAs to regulate diverse cellular processes. Although 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 that are associated with RBPs can be computationally searched for consensus nucleotide sequences that are selectively enriched in bound versus unbound transcripts using programs such as MEME, FIRE and REFINE[103–105]. Experimentally, SELEX and RNAcompete enable the determination of RNA consensus motifs 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[104,106]. 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), allows the identification of both RBP-bound transcripts, as well as the protein binding site, thus 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[107]. 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.

these structures can exert their influence by helping to provide specific binding sites for RBPs, 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).

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 variance in temperature — that result in changes in gene expression, which adds 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 can be rapid. Below we elaborate on a few examples that demonstrate the specificity and dynamic character of RNA structures and how identifying such structures in a transcriptome-wide manner can enhance our understanding of RNA function.

The specificity and dynamics of riboswitches. One of the best examples demonstrating the specificity and dynamics of RNA structures is a riboswitch. Riboswitches are RNA 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 that had been 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, the riboswitch undergoes a conformational change that results in gene-expression changes. Multiple classes of riboswitch exist that respond to a wide range of cellular stimuli, including amino acids, nucleotides, metal ions, coenzymes and temperature, which allows them 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 riboswitch has been found outside eubacteria, and no riboswitches have been found in mammals[12].

The aptamer domain of a riboswitch binds to a ligand through multiple interactions, such as hydrogen bonding and electrostatic interactions. Riboswitches that bind metabolites are typically very specific for certain metabolites and can discriminate between their true ligands and other similarly structured molecules[5]. This specificity for 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 from adenine to guanine[19]. The aptamer is found in the 5′UTR of the bacteria *ydhL* mRNA and forms a secondary structure when bound to adenine. This prevents the formation of a transcription terminator loop and so transcription occurs. High levels of adenine hence result in high levels of YdhL, which is a purine efflux pump, and so adenine can be pumped out of the cell. Another example is the *S*-adenosylmethione (SAM) riboswitch. Distinct classes of the SAM riboswitches can distinguish between SAM — a coenzyme for methylation reactions — and *S*-adenosylcytosine (SAH) — a by-product of the methylation reaction — even though SAM and SAH are highly similar in structure. This distinction is important to prevent the accumulation of toxic SAH and to recycle SAH to form SAM[5]. The diversity of SAM riboswitches, as shown by high-resolution crystal structures (FIG. 1a), also illustrates the possibility of multiple RNA structural solutions to the same biochemical challenge. The presence of diverse RNA structures also shows the need to experimentally probe RNA structural dynamics rather than relying purely on sequence conservation.

Dynamics of RNA structures in mammals. The dynamics of RNA structure is also a recurring theme in mammalian RNAs. The binding of protein factors to specific RNA elements has been extensively studied, and it is emerging that this binding can result in a corresponding change in RNA structure, which affects gene expression. The human vascular endothelial growth factor A (*VEGFA*) mRNA contains a 125-base hypoxia-stability region in its 3′UTR. The structure of this region changes depending on whether the cell is exposed to normoxic conditions or whether it

SELEX
(Systematic evolution of ligands by exponential enrichment). In the context of RNA, this is a method for identifying consensus protein binding sequences on RNA substrates by *in vitro* selection of short RNAs that bind preferentially to RNA-binding proteins.

RNAcompete
An *in vitro* method to identify the structural and linear sequence motifs of RNAs that interact strongly with RNA-binding proteins in a complex pool of k-mer RNAs.

PUF
(Pumilio family). This is a family of evolutionarily conserved RNA-binding proteins. They preferentially bind to the 3′UTR of mRNAs to regulate gene expression.

is exposed to hypoxic conditions in the presence of interferon-γ (IFNγ)[20]. During normoxia, the presence of the IFNγ-activated inhibitor of translation (GAIT) complex causes the *VEGFA* mRNA to form a structure that is not permissive to translation. However, during hypoxia, the binding of heterogenous nuclear ribonucleoprotein L (HNRNPL) results in the RNA conformation switching to a different structure that permits protein translation.

An miRNA is a ~23 nt short RNA that modulates gene expression in normal development and disease pathogenesis. The interaction between miRNAs and 3′UTRs of their target mRNAs can lead to mRNA destabilization and/or translation inhibition. Recently, RNA conformations within a transcript have been found to be one of the determinants of whether a transcript is targeted by specific miRNAs; target sites that are buried in secondary structures may sterically hinder

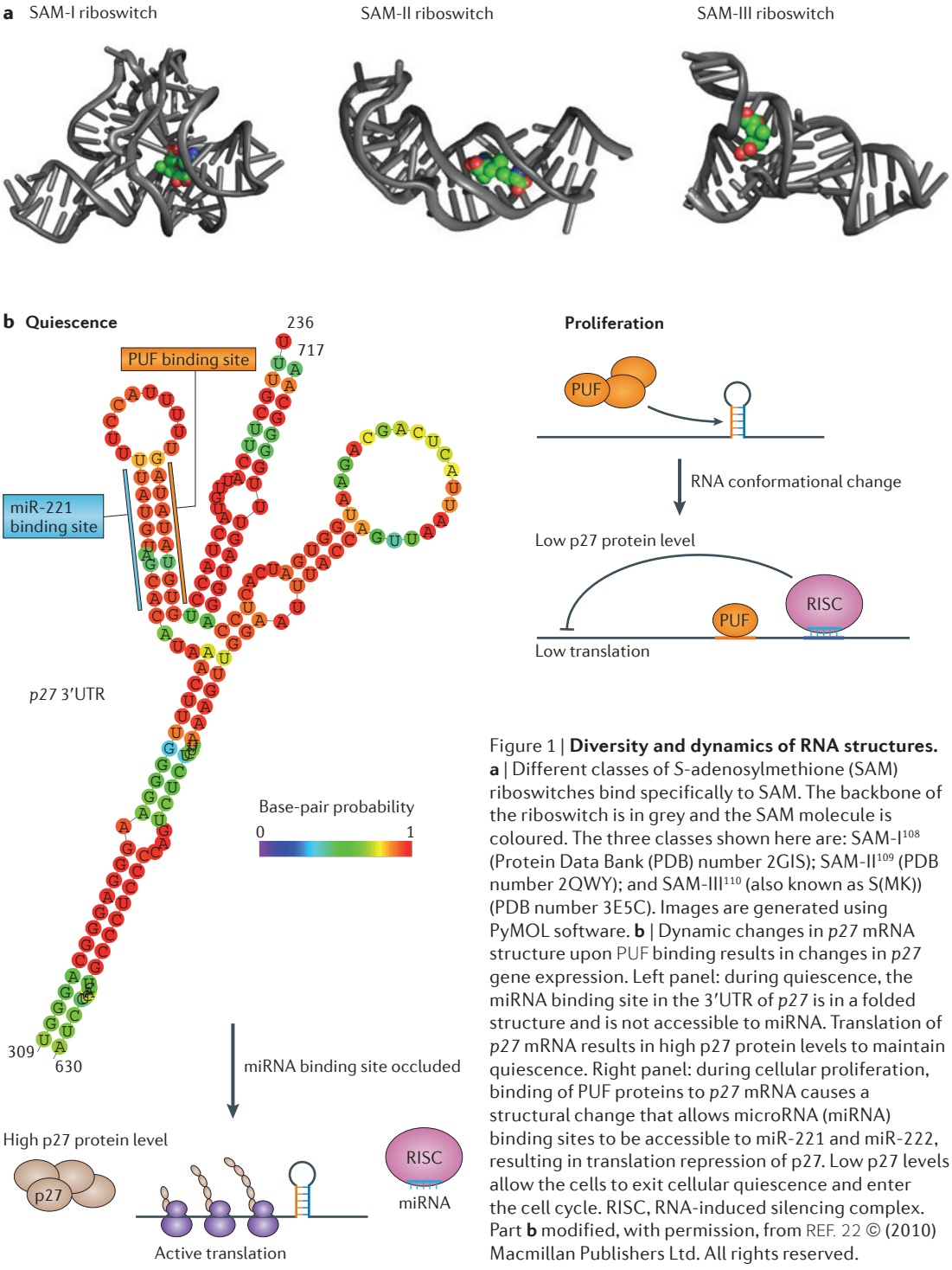


Figure 1 | **Diversity and dynamics of RNA structures.** **a** | Different classes of *S*-adenosylmethione (SAM) riboswitches bind specifically to SAM. The backbone of the riboswitch is in grey and the SAM molecule is coloured. The three classes shown here are: SAM-I[108] (Protein Data Bank (PDB) number 2GIS); SAM-II[109] (PDB number 2QWY); and SAM-III[110] (also known as S(MK)) (PDB number 3E5C). Images are generated using PyMOL software. **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 the 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 microRNA (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. RISC, RNA-induced silencing complex. Part **b** modified, with permission, from REF. 22 © (2010) Macmillan Publishers Ltd. All rights reserved.

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. Levels of p27 are low in dividing cells but high in non-dividing, quiescent cells. Upon growth-factor stimulation, Pumilio 1 is activated, binds to the *p27* mRNA 3′UTR and results in an RNA structural change. This structural change exposes the miRNA 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 levels (FIG. 1b).

Genome-wide data sets that identify RBP targets and where these proteins bind to mRNAs are increasing[23–25]. Probing RNA structures in a genome-wide manner both *in vitro* and *in vivo* would enable us to study the structural context that determines protein binding to RNAs and would also enable us to identify regions of RNA structural changes that occur in the presence and absence of protein binding. As many 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 have an impact on cellular function.

RNA structure — computational approaches

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 that are essential for their function, computational methods using secondary-structure predictions were successfully used to annotate new members of that ncRNA class. Examples include methods for predicting tRNAs[32,33], small nucleolar RNAs (snoRNAs)[32] and miRNAs[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 function of the RNA in which they are embedded[30,38–41] (FIG. 2a,b).

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, ribosomal RNAs (rRNAs) and group I and group II introns[42,43]. Covariation methods determine secondary structure by examining conservation patterns of base pairs among orthologous or paralogous genes. These 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, thus preserving RNA structure[42]. For example, the pairing of G-C nucleotides between two distinct genomic sequences can be maintained at the structural level in another species if the G-C nucleotides have changed to A-U nucleotides (FIG. 2c). The structure can be determined 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 combination of a covariation method and a thermodynamic method (discussed below) can be used[48].

Thermodynamic modelling. 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 to identify the most thermodynamically stable structure[26,49,50]. For sequences that are shorter than 700 bp, ~70% of the base pairs that are determined by experimental methods are correctly predicted by these methods. However, for longer sequences, the accuracy drops to ~20–60% when the predicted structures are compared with 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 modelling using stochastic context-free grammars (SCFGs) were also developed, but as their accuracy is lower, they have not thus far replaced free-energy minimization methods[28]. Another recent, improved strategy was developed using both thermodynamic modelling 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 modelling techniques is the evaluation of potential RNA structural changes caused by non-coding SNPs that are associated with human diseases. Laederach and colleagues[55] identified multiple disease-associated mutations in UTRs that alter the mRNA structure, providing new hypotheses for causes of human disease and variation.

Incorporating experimental data. Another successful approach has been to incorporate experimentally derived structural information into computational predictions. This approach has been developed over many years since the first prediction algorithms became available[29,56–59]. In cases in which the experiment can only derive binary information for each nucleotide — namely whether the nucleotide was paired or unpaired — the dynamic programming

Small nucleolar RNAs (snoRNAs). RNAs that are involved in guiding the modification of other RNAs, such as ribosomal RNAs, tRNAs and small nuclear RNAs.

Dynamic programming
A method for solving complex problems by breaking them down into simpler 'sub-problems'. This method is used by most RNA structure-prediction algorithms to efficiently scan the entire landscape of possible secondary structures.

Stochastic context-free grammars
(SCFGs). Mathematical models in which base pairings in an RNA molecule are described as a set of production rules, each augmented with a probability.

Machine learning methods
Algorithms that use empirical data (called the training set) to capture characteristics of unknown underlying phenomena and improve predictions about new data (called the test set).

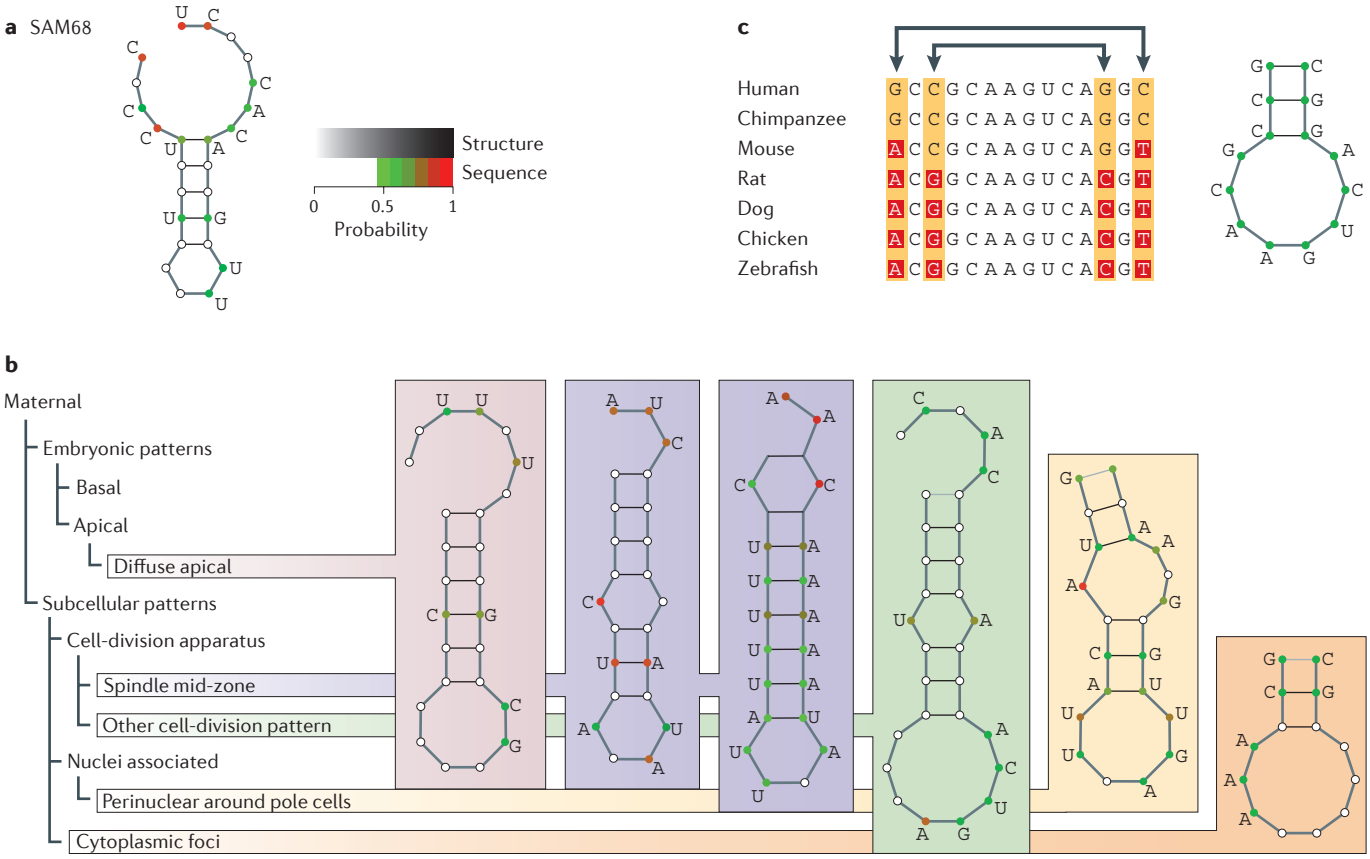


Figure 2 | **Predicting structural motifs for RNA-binding-protein targets in mRNAs from different organisms.** **a** | A stem–loop structural motif that was found in the eight known targets of the human RNA-binding protein (RBP) SAM68 (also known as KHDRBS1; not to be confused with *S*-adenosylmethione) using an RNA motif finder (RNApromo)[30]. **b** | The same RNA motif finder was used to analyse data collected from a large study of mRNA localization during fly embryonic development to predict significant motifs in six sets of colocalized maternal transcripts[30]. Shown is the structural motif enriched in each set of mRNAs. **c** | Conservation of base pairs in homologous sequences directs structure prediction. Sequence covariation is 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 suggests that these bases might interact, which could give the motif shown on the right. Parts **a** and **b** are modified, with permission, from REF. 30 © (2008) National Academy of Sciences, USA.

algorithm can be modified such that large, positive free-energy terms are added to nucleotides that are known to be unpaired. This thereby restricts 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[59]. By integrating an additional per-nucleotide pseudo-free-energy term into the dynamic programming algorithm, 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 RBPs. These environment-dependent interactions are extremely complex to model and are thus excluded from 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 encountered in predicting pseudoknots or in taking into account long-range and tertiary-structure interactions. Although those more complicated aspects of RNA structure are computationally hard to predict, some algorithms have been developed that allow for pseudoknots or are able to predict tertiary structure[62–64]. Pseudoknots have been observed in a number of functional RNA sequences, such as rRNAs, tRNAs or the genomes of viral RNAs[65], in which they have been shown to be involved in unique mechanisms of viral translation initiation and elongation[66]. Thus, ignoring pseudoknots results in inaccurate structure predictions[65,67]. In contrast to the prediction of

Pseudoknots
RNA topologies that contain non-nested nucleotide pairings.

nested structures (that is, structures that are free of pseudoknots), which can be efficiently solved using dynamic programming, predicting structures that contain pseudoknots is very challenging computationally. There are no efficient algorithms for the prediction of most classes of pseudoknot[68]. As a result, several methods have been developed that focus on specific types of pseudoknots[69–71] or use heuristics[72–76] to bring down running time. Nonetheless, computational prediction of pseudoknots still scales exponentially with the length of the RNA on the order of n^4 to 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, further experimental data are needed. Furthermore, 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[77]. 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[77,78]. Numerous reagents interact with single-stranded or flexible bases to modify or cleave them[79–83], including chemicals such as dimethyl sulphide (DMS), 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulphonate (CMCT), kethoxal, Pb^{2+} and *N*-methylisatoic anhydride (NMIA), as well as nucleases such as RNases I, T1 and A and S1 nuclease. Enzymes such as RNase V1 recognize and cleave at double-stranded bases[84], and hydroxyl radicals cleave at RNA bases that are solvent-exposed[85,86]. Chemical and enzymatic probes are frequently used 'hand-in-hand' to provide structural information. The enzyme RNase V1 is the only structural probe that recognizes and preferentially cleaves at double-stranded regions. Chemical probes are frequently used to provide structural information on single-stranded regions. Because they are much smaller in size than enzymes, chemical probes encounter less steric hindrance and are able to probe more bases in a folded RNA. Different chemicals can also react with different moieties on unpaired bases; thus, comparing the patterns of reactivity of multiple chemical probes can provide additional insight into RNA structure. Upon cleavage or modification, the reaction sites can be detected by autoradiography or reverse transcription followed by gel or capillary electrophoresis (FIG. 3). In the case of gel electrophoresis, 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 semi-automated footprinting analysis (SAFA)[87].

The 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 resolves about a hundred bases of RNA at a time, and hence probing an RNA of several kilobases in length 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[88,89]. The readout of the probing experiment is typically achieved through the reverse transcription of a 5′ fluorescently labelled DNA primer that specifically anneals 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 capillary-automated footprinting analysis (CAFA) and ShapeFinder can automate the data acquisition from capillary electrophoresis and can further improve speed and accuracy[88,89] (FIG. 3).

RNA footprinting can also be performed *in vivo*[90,91]. Because some RNAs are able to fold into 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[92]. RNA footprinting can be carried out inside cells using chemicals that can penetrate the cell membrane, such as Pb^{2+} and DMS, or with high-energy X-rays[85,91,93]. Lead probing has been successfully applied to *in vivo* structure probing in bacteria, whereas DMS has been applied to both prokaryotic and eukaryotic cells[91,93]. However, *in vivo* RNA footprinting may not be able to interrogate all regions of an RNA of interest owing to steric hindrance from protein interactions. The dynamic cellular environment also presents RNA in heterogeneous states: RNA is present in all of the different stages of its life cycle, including transcription, translation and decay. Averaging the structural signal from heterogeneous states provides inaccurate models. As such, structural probing *in vitro* and *in vivo* provides complementary information about RNA structures. In all footprinting experiments, it is important to titrate the correct amount of structural probe to achieve '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.

SHAPE and its applications to long RNAs. The selective 2-hydroxyl acylation analysed by primer extension (SHAPE) method uses the chemical NMIA and its derivatives to interrogate flexible regions in RNA secondary structure[83]. The 2′OHs of flexible bases are

Heuristics
An experience-based method of problem solving that is used in cases in which an exhaustive search is impractical to speed up the process of finding a solution. There is usually some loss of accuracy.

Single-hit kinetics
The kinetics of reactions involving chemical and enzymatic probes that react with RNA, such that, on average, there is only one cut per molecule.

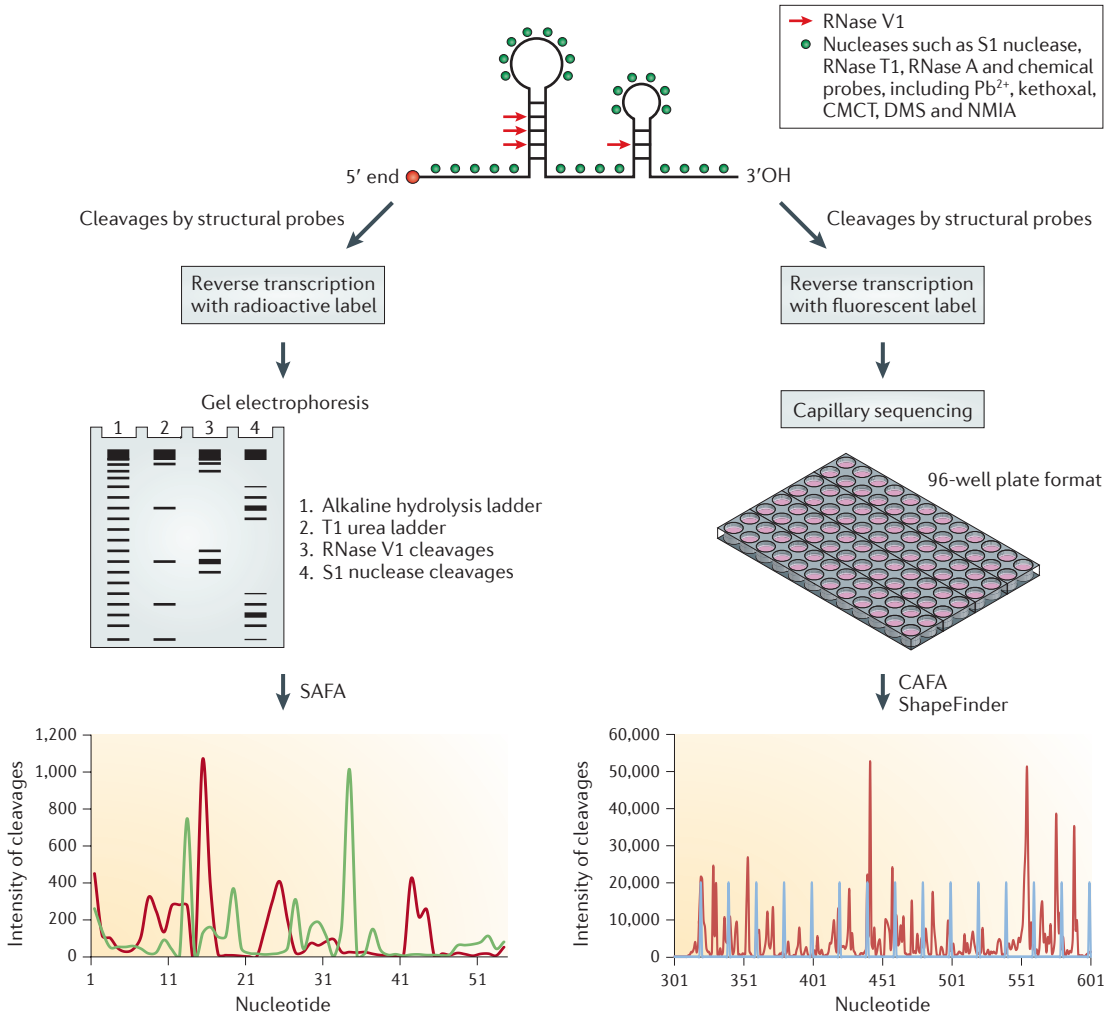


Figure 3 | **Structure probing by RNA footprinting followed by gel or capillary electrophoresis.** An RNA of interest is typically transcribed *in vitro*, folded and subjected to a combination of single- and double-stranded structural probes in solution. Cleavages in double- or single-stranded regions can either be identified by running a gel electrophoresis (RNA needs to be radioactively labelled at one end) or by primer extension followed by capillary electrophoresis (the primer needs to be fluorescently labelled). The bands from gel electrophoresis can be quantified using a program called semi-automated footprinting analysis (SAFA), and bands in capillary electrophoresis are identified and quantified using capillary-automated footprinting analysis (CAFA) or ShapeFinder. In the illustrated example of data from gel electrophoresis, the green lines on the graph refer to the intensity (quantified by SAFA) of S1 nuclease cleavages and 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, which are run alongside the samples. In the example of capillary electrophoresis, the red line of the graph indicates the intensity of structure probing sites that are detected by reverse transcription, and the blue line corresponds to a ladder that positions the RNA bases. CMCT, 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulphonate; DMS, dimethyl sulphide; NMIA, *N*-methylisatoic anhydride.

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 followed by capillary electrophoresis, because they cause termination of the reverse transcription reaction. As every ribonucleotide contains a 2′OH, SHAPE has the advantage of being able to probe most bases in an RNA. With its coupling 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 HIV[59,94,95].

The construction of the secondary structure of the HIV genome using SHAPE was a landmark that demonstrated the substantial value of comprehensive RNA structure analysis[94]. The HIV genome is a 9 kb long single-stranded RNA that encodes nine ORFs, which are translated into 15 proteins that are 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*[95]. 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 9 kb HIV genome *ex virio* by SHAPE further found numerous regions within the genome that have functional roles in HIV replication[94]. These structured RNA domains provide insights into Gag-Pol frame shifting, hyper-variable domains and translocation of the Env protein. Interestingly, the nucleotides between independently folded protein domains are more structured than surrounding bases and are able to fold into secondary structures that retard ribosome progression to facilitate co-translational protein folding of modular domains[94].

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 the hepatitis C virus (HCV), will be structurally probed to understand the role of RNA structures in viral replication. Furthermore, RNA structure probing is likely to be extended to families of viral genomes in order to discover conserved or rapidly evolving structural elements that have the potential 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 more sequencing information than capillary sequencing.

Genome-wide RNA structure maps

Parallel analysis of RNA structure, fragmentation sequencing and SHAPE sequencing. Next-generation sequencing has enabled the next major advance in genome-wide measurements of RNA structure, because 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 at which the cleavages or modifications occurred. The frequency of the cleavage or modification at a base can also be estimated by summing the reads that are mapped to the base. This strategy allows the simultaneous identification of double- or single-stranded (or 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[96]. Recently, SHAPE has also been coupled to deep sequencing (SHAPE-seq)[97].

PARS was used to measure the secondary structure of the yeast transcriptome through generating structural information on ~4.2 million bases in over 3,000 yeast transcripts[21]. Mapping PARS scores to known structures of regulatory motifs, such as *Ash1* localization elements and the internal ribosomal entry site of *URE2* mRNA, indicates that PARS is able to capture the structural information in these elements. 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 than in UTRs, a three-nucleotide periodicity of secondary structure along the coding regions and an anti-correlation between mRNA translation efficiency and structure over the mRNA translation start site (FIG. 5). Frag-seq was used to reconstruct the secondary structure of snoRNAs in mouse cells[96]. Both Frag-seq and PARS data can be integrated into structure-prediction programs for more accurate RNA secondary-structure prediction. For example, PARS data were used to constrain a thermodynamic RNA structure-prediction algorithm as binary inputs (paired versus unpaired), and a custom algorithm was developed to accommodate Frag-seq data.

Comparison of PARS and Frag-seq reveals the complementary nature of the information that they provide. First, because Frag-seq isolates RNAs that are 20–100 bases long after P1 nuclease cleavage without an additional fragmentation step, many sequence reads come from small nuclear RNAs (snRNAs), such as snoRNAs, and larger RNAs may be under-represented. Second, structured regions appear as 'blanks' in Fraq-seq data, and other information is thus necessary to ensure that these regions are not being missed owing to mapping or cloning difficulties. Third, PARS compares the cleavage sites of single- versus double-strand-specific enzymes, whereas Fraq-seq uses the endogenous 5′OH and 5′P within the transcriptome as a background. Capturing the endogenous 5′P allows filtering of reads that result from molecule ends that are already present in the transcriptome and are not a result of P1 nuclease cleavage. The 5′OH control identifies 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.

In a SHAPE-seq experiment, seven short RNAs were transcribed *in vitro*, and then each was appended with a unique sequence tag (a barcode)[97]. These RNAs were then treated with the chemical 1-methyl-7-nitroisatoic anhydride (1M7) — a derivative of NMIA — to acylate flexible bases. The reacted bases were indirectly detected by their ability to terminate the reverse transcription reaction, as assessed by sequencing the cDNAs. Because of the barcode, multiple sequences — even those with extensive sequence similarity — can be probed simultaneously. For RNAs such as RNase P and pT181 attenuator, SHAPE-seq data correlate well with data obtained from SHAPE followed by capillary sequencing. SHAPE-seq is likely to be useful for studying multiple mutants of one RNA or multiple members of a closely related RNA family. Comparison of SHAPE-seq with PARS or Frag-seq illustrates several

Nucleocapsid
A coat of proteins that surrounds the genomic content of a virus.

Gag-Pol
The Gag polyprotein is processed into several proteins including the matrix, capsid, spacer peptides, p6 and nucleocapsid proteins. Pol includes reverse transcriptase, integrase and protease.

Env
(Envelope protein). This is found on the surface of the retroviruses and contains glycoproteins that enable the virus to recognize and enter host cells.

Ash1 localization elements
Sequences that are required to properly localize *Ash1* mRNAs to the yeast bud tip.

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 achieve genome-wide scale, particularly when RNA sequences are not known *a priori*. Because SHAPE-seq measures the cDNA product rather than cloned RNA fragments (as in PARS and Frag-seq), the processivity of reverse transcription needs to be taken into account in SHAPE-seq data processing and RNA secondary-structure modelling. The SHAPE-seq signal progressively decays from the 3′ end to the 5′ end of the RNA template (the direction of reverse transcription), so a detailed mathematical model has been developed to correct for this signal decay[98]. 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 that can be measured by deep sequencing, an amount that is still rapidly increasing. Although RNA footprinting with capillary sequencing is still very much directed at interrogating a single RNA of interest, methods based on high-throughput sequencing have the power to probe entire transcriptomes. 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. By contrast, owing to the massively parallel nature of deep-sequencing technology, thousands of distinct RNAs that are many kilobases long can easily be probed by 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 easily compare the structural profile of one transcript with another in the transcriptome, 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 previously known or that have undergone post-transcriptional modifications, such as alternative splicing or RNA editing. By contrast, for capillary sequencing (or SHAPE-seq, as is currently practised), the nucleotide sequence, as well as how the RNA is spliced, needs to be known to enable primer design along the length of the RNA. This process is tedious and also restricts capillary sequencing to structure probing of transcripts that are well-annotated in the transcriptome.

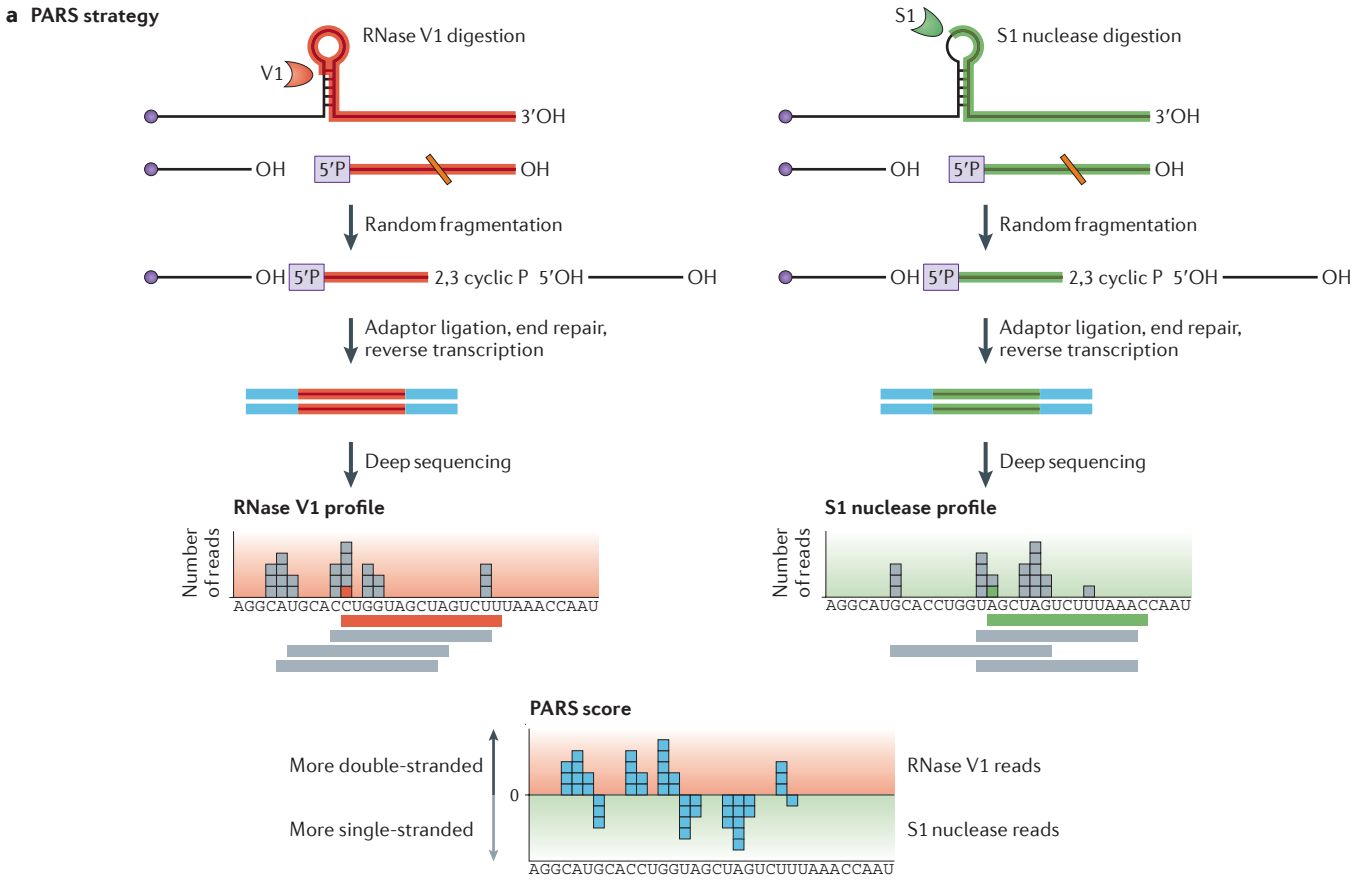
Despite these potential advantages, care and thoughtful controls are necessary to design and interpret genome-scale RNA structure maps, as has been done with RNA footprinting by capillary sequencing[99]. Key points to consider in experimental design include replicates to examine reproducibility, titration of structural probes to maintain single-hit kinetics and controls to assess various biases that may arise from library preparation, deep sequencing or mapping[100]. The addition of positive control RNAs with well-known structures to genome-scale reactions is a useful measure of assessing the quality of structural information generated by deep sequencing.

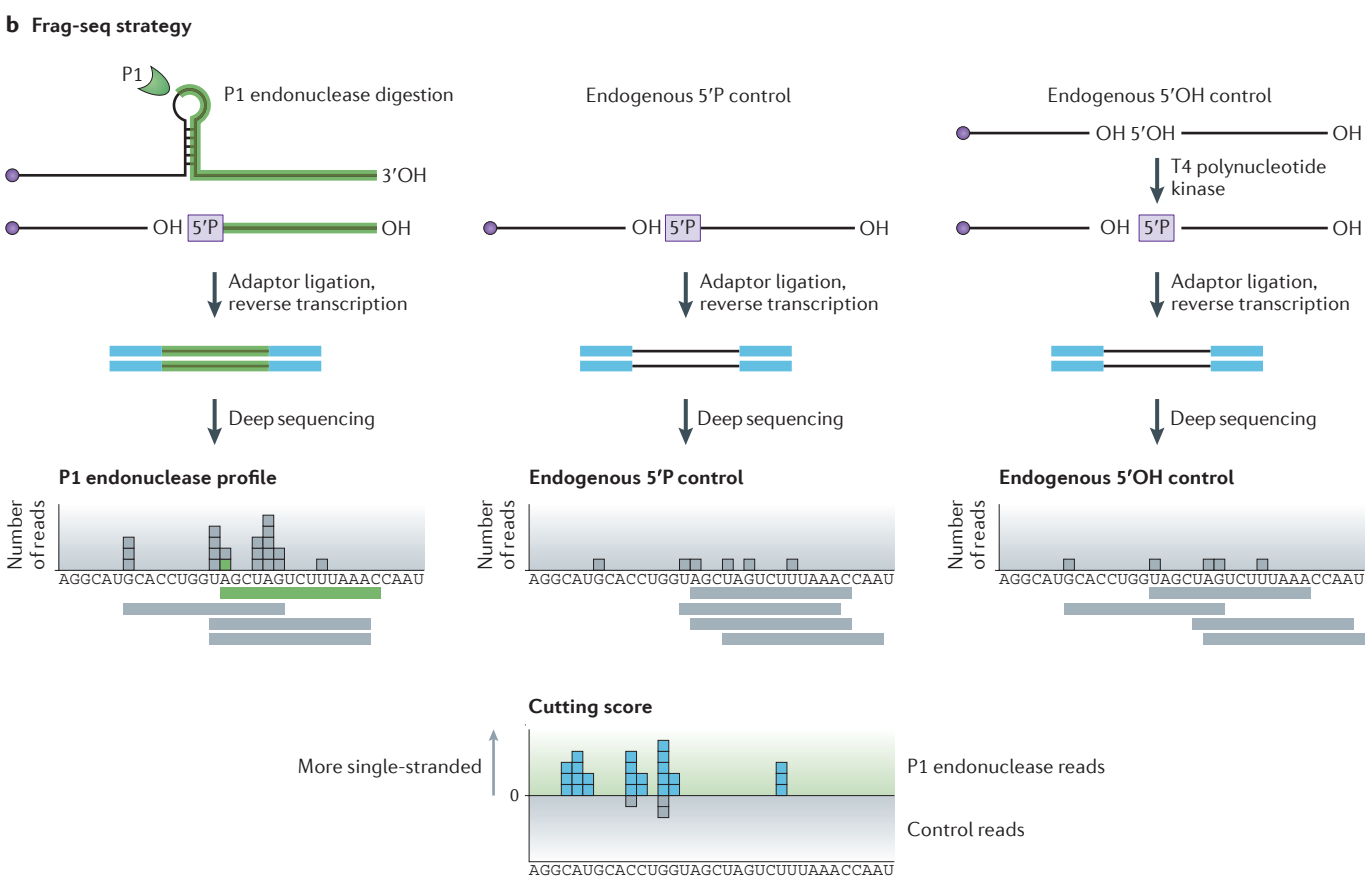
Figure 4 | **PARS and Frag-seq methods. a** | Parallel analysis of RNA structure (PARS) strategy. In PARS, poly(A) selected RNA is folded *in vitro* and incubated with either RNase V1 or S1 nuclease to probe for double- and single-stranded regions, respectively. RNase V1 and S1 nuclease cleave, resulting 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 adaptors and reverse transcribed. The cDNA library is sequenced using high-throughput sequencing and the resulting reads are mapped to the genome to identify double- or 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** | Fragmentation sequencing (Frag-seq) strategy. 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 adaptor 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. Frag-seq also contains controls that 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 that incorporates reads from P1 nuclease and reads from endogenous degradation or fragmentation products. A positive cutting score indicates that the base is single-stranded. Part **a** is modified, with permission, from REF. 21

Future directions

Much remains to be done and learned from genome-wide maps of RNA structure. First, it is likely that many 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 several of these reagents, including DMS, Pb^{2+} 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 effects of alternative splicing of exons that are located hundreds or thousands of bases apart from each other can be more simply evaluated.

Second, *in vivo* and dynamic RNA structure maps will yield a crucial understanding of how RNA structures may change and will help to regulate different





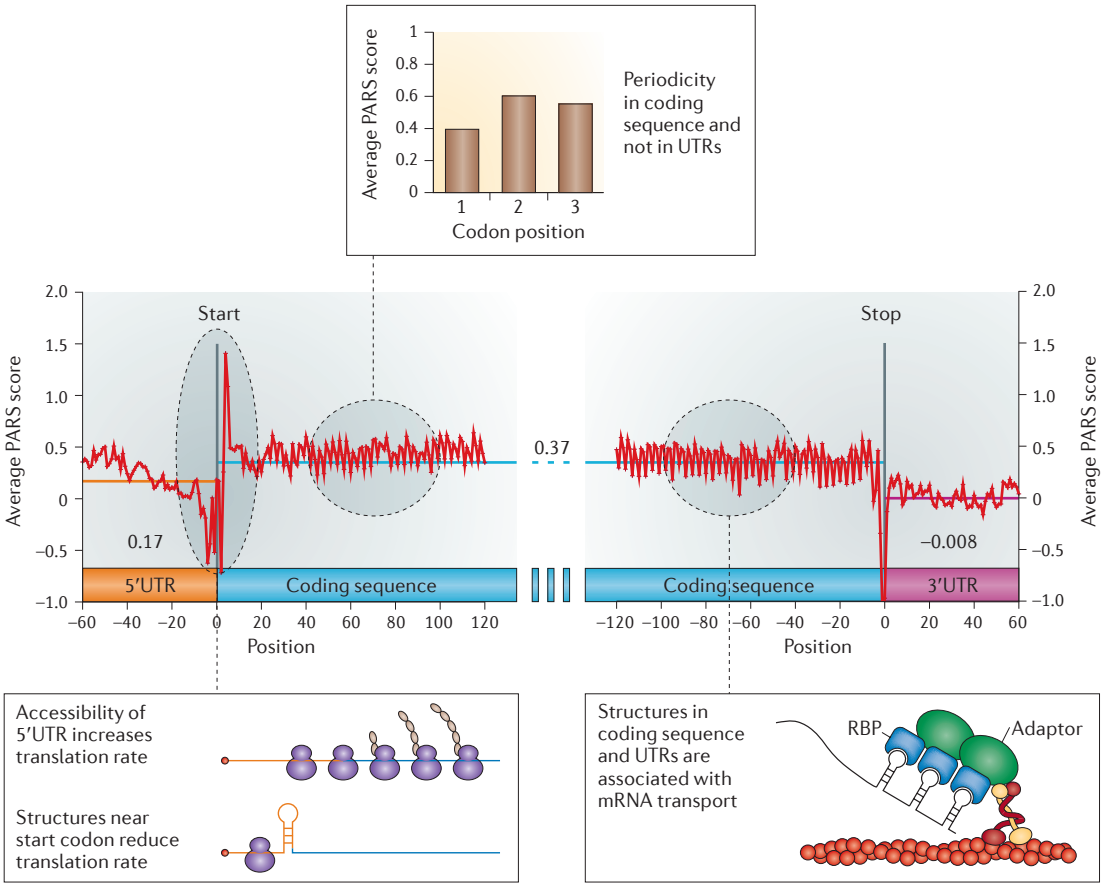


Figure 5 | **Structural organization of the mRNA transcriptome.** Thousands of yeast mRNAs are structurally probed in parallel analysis of RNA structure (PARS) and aligned according to their start and stop codons. The average PARS score of the coding sequence (CDS) is shown in blue, the 5′UTR is shown in yellow and the 3′UTR is shown in red. The organization of secondary structures within the transcriptome revealed an increased accessibility of RNA structure near the start codon that is 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 purple lines). Some of these structures are important for cellular processes such as mRNA transport. A three-nucleotide periodicity (top inset box) in RNA is also seen in the coding region and is absent from the UTRs. RBP, RNA-binding protein. This figure is modified, with permission, from REF. 21

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 could be applied to native RNA that has been isolated without denaturation. Several chemical probes — such as Pb^{2+} ions, DMS, NMIA and hydroxyl radicals — have been successfully used to probe RNA structures *in vivo* by penetrating cellular membranes[82,85,93,95]. RNA footprinting can also be performed under different conditions, such as alterations in temperature or the presence of specific proteins or small-molecule ligands, to probe the impact of these perturbations on RNA structure[10,22,101,102].

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 to further predict impacts of changes in the RNA structure (caused by a SNP, changes in biological state or a drug) on the 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].

1. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nature Rev. Mol. Cell Biol.* **8**, 113–126 (2007).
2. Warf, M. B. & Berglund, J. A. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* **35**, 169–178 (2010).
3. Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–730 (2009).
4. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13–37 (2005).
5. Breaker, R. R. Riboswitches and the RNA World. *Cold Spring Harb. Perspect. Biol.* 24 Nov 2010 (doi:10.1101/cshperspect.a003566).
6. Park, P. J. ChIP–seq: advantages and challenges of a maturing technology. *Nature Rev. Genet.* **10**, 669–680 (2009).

7. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
8. Blencowe, B. J., Ahmad, S. & Lee, L. J. Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. *Genes Dev.* **23**, 1379–1386 (2009).
9. Henkin, T. M. Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev.* **22**, 3383–3390 (2008).
10. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).
11. Weinberg, Z. *et al.* Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.* **11**, R31 (2010).
12. Barrick, J. E. & Breaker, R. R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* **8**, R239 (2007).
13. Dann, C. E. *et al.* Structure and mechanism of a metal-sensing regulatory RNA. *Cell* **130**, 878–892 (2007).
14. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**, 577–586 (2003).
15. Mandal, M. *et al.* A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**, 275–279 (2004).
16. Nahvi, A., Barrick, J. E. & Breaker, R. R. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.* **32**, 143–150 (2004).
17. Croft, M. T., Moulin, M., Webb, M. E. & Smith, A. G. Thiamine biosynthesis in algae is regulated by riboswitches. *Proc. Natl Acad. Sci. USA* **104**, 20770–20775 (2007).
18. Sudarsan, N., Barrick, J. E. & Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* **9**, 644–647 (2003).
19. Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Struct. Mol. Biol.* **11**, 29–35 (2004).
This paper discusses the discovery that a single base in the adenine riboswitch determines the affinity of the aptamer for adenine versus guanine. This demonstrates the specificity of RNA structures in binding to their substrates.
20. Ray, P. S. *et al.* A stress-responsive RNA switch regulates VEGFA expression. *Nature* **457**, 915–919 (2009).
21. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103–107 (2010).
This was the first description of genome-wide RNA structure probing, which studied double- and single-stranded regions in yeast RNAs *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. *Nature Cell Biol.* **12**, 1014–1020 (2010).
This paper showed that Pumilio 1 protein binding results in a conformational change that allows miRNA binding sites in *p27* mRNA to be accessible for regulation. This demonstrates that RNA structure dynamics contributes to the complexity of gene regulation.
23. Zhao, J. *et al.* Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40**, 939–953 (2010).
24. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
25. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
26. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
27. Hofacker, I. L., Fekete, M. & Stadler, P. F. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* **319**, 1059–1066 (2002).
28. Do, C. B., Woods, D. A. & Batzoglou, S. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* **22**, e90–e98 (2006).
29. Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911–940 (1999).
This paper reports the measurement of thermodynamic parameters for the stability of secondary-structure motifs. These parameters have been extensively used in prediction algorithms.
30. Rabani, M., Kertesz, M. & Segal, E. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. *Proc. Natl Acad. Sci. USA* **105**, 14885–14890 (2008).
31. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nature Genet.* **39**, 1278–1284 (2007).
32. Schattner, P., Brooks, A. N. & Lowe, T. M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**, W686–W689 (2005).
33. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93–D97 (2009).
34. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2008).
35. Rivas, E. & Eddy, S. R. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **2**, 8 (2001).
36. Lu, Z. J. *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.* **21**, 276–285 (2011).
37. Pedersen, J. S. *et al.* Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput. Biol.* **2**, e33 (2006).
38. Gautheret, D. & Lambert, A. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol.* **313**, 1003–1011 (2001).
39. Hochsmann, M., Toller, T., Giegerich, R. & Kurtz, S. Local similarity in RNA secondary structures. *Proc. IEEE Comput. Soc. Bioinform. Conf.* **2**, 159–168 (2003).
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.* **32**, 3258–3269 (2004).
41. Gorodkin, J., Heyer, L. J. & Stormo, G. D. Finding the most significant common sequence and structure motifs in a set of RNA sequences. *Nucleic Acids Res.* **25**, 3724–3732 (1997).
42. Eddy, S. R. & Durbin, R. RNA sequence analysis using covariance models. *Nucleic Acids Res.* **22**, 2079–2088 (1994).
43. Sun, F. J. & Caetano-Anolles, G. The origin and evolution of tRNA inferred from phylogenetic analysis of structure. *J. Mol. Evol.* **66**, 21–35 (2008).
44. Washietl, S., Hofacker, I. L. & Stadler, P. F. Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA* **102**, 2454–2459 (2005).
45. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. Rfam: an RNA family database. *Nucleic Acids Res.* **31**, 439–441 (2003).
46. Yao, Z., Weinberg, Z. & Ruzzo, W. L. CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics* **22**, 445–452 (2006).
47. Knudsen, B. & Hein, J. Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.* **31**, 3423–3428 (2003).
48. Seemann, S. E., Gorodkin, J. & Backofen, R. Unifying evolutionary and thermodynamic information for RNA folding of multiple alignments. *Nucleic Acids Res.* **36**, 6355–6362 (2008).
49. Mathews, D. H. & Turner, D. H. Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.* **16**, 270–278 (2006).
50. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–W74 (2008).
This paper describes the Vienna RNA package, which is one of the most commonly used software suites for folding single and aligned sequences and predicting RNA–RNA interactions.
51. Dowell, R. D. & Eddy, S. R. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinformatics* **5**, 71 (2004).
52. Doshi, K. J., Cannone, J. J., Cobaugh, C. W. & Gutell, R. R. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* **5**, 105 (2004).
53. Hamada, M., Kiryu, H., Sato, K., Mituyama, T. & Asai, K. Prediction of RNA secondary structure using generalized centroid estimators. *Bioinformatics* **25**, 465–473 (2009).
54. Lu, Z. J., Gloor, J. W. & Mathews, D. H. Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA* **15**, 1805–1813 (2009).
55. Halvorsen, M., Martin, J. S., Broadaway, S. & Laederach, A. Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet.* **6**, e1001074 (2010).
This paper describes SNPs that are present in different disease states that 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.* **9**, 133–148 (1981).
57. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA* **101**, 7287–7292 (2004).
58. Quarrier, S., Martin, J. S., Davis-Neulander, L., Beauregard, A. & Laederach, A. Evaluation of the information content of RNA structure mapping data for secondary structure prediction. *RNA* **16**, 1108–1117 (2010).
59. 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).
60. Xia, T. *et al.* Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry* **37**, 14719–14735 (1998).
61. Liu, B., Diamond, J. M., Mathews, D. H. & Turner, D. H. Fluorescence competition and optical melting measurements of RNA three-way multibranch loops provide a revised model for thermodynamic parameters. *Biochemistry* **50**, 640–653 (2011).
62. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51–55 (2008).
63. Sharma, S., Ding, F. & Dokholyan, N. V. iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics* **24**, 1951–1952 (2008).
64. Das, R. & Baker, D. Automated *de novo* prediction of native-like RNA tertiary structures. *Proc. Natl Acad. Sci. USA* **104**, 14664–14669 (2007).
65. van Batenburg, F. H., Gultyaev, A. P. & Pleij, C. W. PseudoBase: structural information on RNA pseudoknots. *Nucleic Acids Res.* **29**, 194–195 (2001).
66. Brierley, I., Pennell, S. & Gilbert, R. J. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nature Rev. Microbiol.* **5**, 598–610 (2007).
67. Chen, H. L., Condon, A. & Jabbari, H. An O(n(5)) algorithm for MFE prediction of kissing hairpins and 4-chains in nucleic acids. *J. Comput. Biol.* **16**, 803–815 (2009).
68. Lyngso, R. B. & Pedersen, C. N. RNA pseudoknot prediction in energy-based models. *J. Comput. Biol.* **7**, 409–427 (2000).
69. Rivas, E. & Eddy, S. R. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.* **285**, 2053–2068 (1999).
70. Dirks, R. M. & Pierce, N. A. An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *J. Comput. Chem.* **25**, 1295–1304 (2004).
71. Reeder, J. & Giegerich, R. Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC Bioinformatics* **5**, 104 (2004).
72. Ruan, J., Stormo, G. D. & Zhang, W. An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. *Bioinformatics* **20**, 58–66 (2004).
73. Ren, J., Rastegari, B., Condon, A. & Hoos, H. H. HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *RNA* **11**, 1494–1504 (2005).

74. Chen, X. *et al.* FlexStem: improving predictions of RNA secondary structures with pseudoknots by reducing the search space. *Bioinformatics* **24**, 1994–2001 (2008).
75. Bellaousov, S. & Mathews, D. H. ProbKnot: fast prediction of RNA secondary structure including pseudoknots. *RNA* **16**, 1870–1880 (2010).
76. 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* **27**, i85–i93 (2011).
77. Weeks, K. M. Advances in RNA structure analysis by chemical probing. *Curr. Opin. Struct. Biol.* **20**, 295–304 (2010).
78. Ehresmann, C. *et al.* Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 9109–9128 (1987).
79. Romby, P. *et al.* Ribosomal 5S RNA from *Xenopus laevis* oocytes: conformation and interaction with transcription factor IIIA. *Biochimie* **72**, 437–452 (1990).
80. Wurst, R. M., Vournakis, J. N. & Maxam, A. M. Structure mapping of 5′-32P-labeled RNA with S1 nuclease. *Biochemistry* **17**, 4493–4499 (1978).
81. 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).
82. Wells, S. E., Hughes, J. M., Igel, A. H. & Ares, M. Jr. Use of dimethyl sulfate to probe RNA structure *in vivo*. *Methods Enzymol.* **318**, 479–493 (2000).
83. 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).
This paper describes SHAPE reagents that react with 2′OH of all four RNA bases to probe for flexible structural regions in an RNA. They have been found to be excellent chemical probes.
84. Lowman, H. B. & Draper, D. E. On the recognition of helical RNA by cobra venom V1 nuclease. *J. Biol. Chem.* **261**, 5396–5403 (1986).
85. Adilakshmi, T., Lease, R. A. & Woodson, S. A. Hydroxyl radical footprinting *in vivo*: mapping macromolecular structures with synchrotron radiation. *Nucleic Acids Res.* **34**, e64 (2006).
86. Shcherbakova, I. & Mitra, S. Hydroxyl-radical footprinting to probe equilibrium changes in RNA tertiary structure. *Methods Enzymol.* **468**, 31–46 (2009).
87. Das, R., Laederach, A., Pearlman, S. M., Herschlag, D. & Altman, R. B. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* **11**, 344–354 (2005).
88. Mitra, S., Shcherbakova, I. V., Altman, R. B., Brenowitz, M. & Laederach, A. High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res.* **36**, e63 (2008).
89. Vasa, S. M., Guex, N., Wilkinson, K. A., Weeks, K. M. & Giddings, M. C. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14**, 1979–1990 (2008).
90. Zemora, G. & Waldsich, C. RNA folding in living cells. *RNA Biol.* **7**, 634–641 (2010).
91. Liebeg, A. & Waldsich, C. Probing RNA structure within living cells. *Methods Enzymol.* **468**, 219–238 (2009).
92. Russell, R. RNA misfolding and the action of chaperones. *Front. Biosci.* **13**, 1–20 (2008).
93. Lindell, M., Romby, P. & Wagner, E. G. Lead(II) as a probe for investigating RNA structure *in vivo*. *RNA* **8**, 534–541 (2002).
94. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2009).
The RNA structure probing of the entire 9 kb HIV RNA genome that is described in this paper provided many insights into differentially structured regions and their biological functions.
95. 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).
96. Underwood, J. G. *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nature Methods* **7**, 995–1001 (2010).
This paper reports a method for genome-wide RNA structure probing in using P1 nuclease to probe for single-stranded regions in mouse cells *in vitro*.
97. 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).
In this study, chemical structure probing was coupled to high-throughput sequencing by using a SHAPE reagent to probe flexible structural regions of seven RNAs *in vitro*.
98. Aviran, S. *et al.* Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Natl Acad. Sci. USA* **108**, 11069–11074 (2011).
99. Weeks, K. M. RNA structure probing dash seq. *Proc. Natl Acad. Sci. USA* **108**, 10933–10934 (2011).
100. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. & Pachter, L. Improving RNA-seq expression estimates by correcting for fragment bias. *Genome Biol.* **12**, R22 (2011).
101. Leipply, D. & Draper, D. E. Dependence of RNA tertiary structural stability on Mg^{2+} concentration: interpretation of the Hill equation and coefficient. *Biochemistry* **49**, 1843–1853 (2010).
102. Chowdhury, S., Maris, C., Allain, F. H. & Narberhaus, F. Molecular basis for temperature sensing by an RNA thermometer. *EMBO J.* **25**, 2487–2497 (2006).
103. Elemento, O., Slonim, N. & Tavazoie, S. A universal framework for regulatory element discovery across all genomes and data types. *Mol. Cell* **28**, 337–350 (2007).
104. Riordan, D. P., Herschlag, D. & Brown, P. O. Identification of RNA recognition elements in the *Saccharomyces cerevisiae* transcriptome. *Nucleic Acids Res.* **39**, 1501–1509 (2011).
105. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
106. Ray, D. *et al.* Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nature Biotech.* **27**, 667–670 (2009).
107. Li, X., Quon, G., Lipshitz, H. D. & Morris, Q. Predicting *in vivo* binding sites of RNA-binding proteins using mRNA secondary structure. *RNA* **16**, 1096–1107 (2010).
108. Montange, R. K. & Batey, R. T. Structure of the *S*-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441**, 1172–1175 (2006).
109. Gilbert, S. D., Rambo, R. P., Van Tyne, D. & Batey, R. T. Structure of the SAM-II riboswitch bound to *S*-adenosylmethionine. *Nature Struct. Mol. Biol.* **15**, 177–182 (2008).
110. Lu, C. *et al.* Crystal structures of the SAM-III/S_{MK} riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nature Struct. Mol. Biol.* **15**, 1076–1083 (2008).
111. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
112. Kanhere, A. *et al.* Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* **38**, 675–688 (2010).
113. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
114. Kaneko, S. *et al.* Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev.* **24**, 2615–2620 (2010).
115. 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* **30**, 1956–1962 (2011).
116. Wanrooij, P. H., Uhler, J. P., Simonsson, T., Falkenberg, M. & Gustafsson, C. M. G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proc. Natl Acad. Sci. USA* **107**, 16072–16077 (2010).
117. Henkin, T. M. & Grundy, F. J. Sensing metabolic signals with nascent RNA transcripts: the T box and S box riboswitches as paradigms. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 231–237 (2006).
118. Wang, J. & Nikonowicz, E. P. Solution structure of the K-turn and Specifier Loop domains from the *Bacillus subtilis* tyrS T-box leader RNA. *J. Mol. Biol.* **408**, 99–117 (2011).
119. Lu, C. *et al.* SAM recognition and conformational switching mechanism in the *Bacillus subtilis yitJ* S box/SAM-I riboswitch. *J. Mol. Biol.* **404**, 803–818 (2010).
120. Deikus, G. & Bechhofer, D. H. *Bacillus subtilis* trp Leader RNA: RNase J1 endonuclease cleavage specificity and PNPase processing. *J. Biol. Chem.* **284**, 26394–26401 (2009).
121. Butler, E. B., Xiong, Y., Wang, J. & Strobel, S. A. Structural basis of cooperative ligand binding by the glycine riboswitch. *Chem. Biol.* **18**, 293–298 (2011).
122. Zhang, Q., Kang, M., Peterson, R. D. & Feigon, J. Comparison of solution and crystal structures of preQ1 riboswitch reveals calcium-induced changes in conformation and dynamics. *J. Am. Chem. Soc.* **133**, 5190–5193 (2011).
123. 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.* **31**, 1812–1821 (2011).
124. Warf, M. B., Diegel, J. V., von Hippel, P. H. & Berglund, J. A. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc. Natl Acad. Sci. USA* **106**, 9203–9208 (2009).
125. Oikawa, D., Tokuda, M., Hosoda, A. & Iwawaki, T. Identification of a consensus element recognized and cleaved by IRE1α. *Nucleic Acids Res.* **38**, 6265–6273 (2010).
126. Yang, Y. *et al.* RNA secondary structure in mutually exclusive splicing. *Nature Struct. Mol. Biol.* **18**, 159–168 (2011).
127. Cheah, M. T., Wachter, A., Sudarsan, N. & Breaker, R. R. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**, 497–500 (2007).
128. Lee, E. R., Baker, J. L., Weinberg, Z., Sudarsan, N. & Breaker, R. R. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* **329**, 845–848 (2010).
129. Aragon, T. *et al.* Messenger RNA targeting to endoplasmic reticulum stress signalling sites. *Nature* **457**, 736–740 (2009).
130. Gonsalvez, G. B., Urbinati, C. R. & Long, R. M. RNA localization in yeast: moving towards a mechanism. *Biol. Cell* **97**, 75–86 (2005).
131. Bullock, S. L., Ringel, I., Ish-Horowicz, D. & Lukavsky, P. J. A′-form RNA helices are required for cytoplasmic mRNA transport in *Drosophila*. *Nature Struct. Mol. Biol.* **17**, 703–709 (2010).
132. Subramanian, M. *et al.* G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep.* **12**, 697–704 (2011).
133. Chao, J. A. *et al.* ZBP1 recognition of β-actin zipcode induces RNA looping. *Genes Dev.* **24**, 148–158 (2010).
134. Van Etten, R. A. *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.* **124**, 325–340 (1994).
135. 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.* **9**, 774–780 (2008).
136. Parsons, C. J. *et al.* Mutation of the 5′-untranslated region stem-loop structure inhibits α1(I) collagen expression *in vivo*. *J. Biol. Chem.* **286**, 8609–8619 (2011).
137. Cho, H. H. *et al.* Selective translational control of the Alzheimer amyloid precursor protein transcript by iron regulatory protein-1. *J. Biol. Chem.* **285**, 31217–31232 (2010).
138. Goforth, J. B., Anderson, S. A., Nizzi, C. P. & Eisenstein, R. S. Multiple determinants within iron-responsive elements dictate iron regulatory protein binding and regulatory hierarchy. *RNA* **16**, 154–169 (2010).
139. Shahid, R., Bugaut, A. & Balasubramanian, S. The BCL-2 5′ untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* **49**, 8300–8306 (2010).

140. Derecka, K. *et al.* Occurrence of a quadruplex motif in a unique insert within exon C of the bovine estrogen receptor α gene (*ESR1*). *Biochemistry* **49**, 7625–7633 (2010).
141. Gomez, D. *et al.* A G-quadruplex structure within the 5′-UTR of *TRF2* mRNA represses translation in human cells. *Nucleic Acids Res.* **38**, 7187–7198 (2010).
142. Reineke, L. C., Komar, A. A., Caprara, M. G. & Merrick, W. C. A small stem loop element directs internal initiation of the URE2 internal ribosome entry site in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **283**, 19011–19025 (2008).
143. Feng, S. *et al.* Alternate rRNA secondary structures as regulators of translation. *Nature Struct. Mol. Biol.* **18**, 169–176 (2011).
144. Waldminghaus, T., Heidrich, N., Brantl, S. & Narberhaus, F. FourU: a novel type of RNA thermometer in *Salmonella*. *Mol. Microbiol.* **65**, 413–424 (2007).
145. Giuliodori, A. M. *et al.* The *cspA* mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Mol. Cell* **37**, 21–33 (2010).
146. Kortmann, J., Sczodrok, S., Rinnenthal, J., Schwalbe, H. & Narberhaus, F. Translation on demand by a simple RNA-based thermosensor. *Nucleic Acids Res.* **39**, 2855–2868 (2011).
147. Badis, G., Saveanu, C., Fromont-Racine, M. & Jacquier, A. Targeted mRNA degradation by deadenylation-independent decapping. *Mol. Cell* **15**, 5–15 (2004).
148. Prouteau, M., Daugeron, M. C. & Seraphin, B. Regulation of ARE transcript 3′ end processing by the yeast Cth2 mRNA decay factor. *EMBO J.* **27**, 2966–2976 (2008).
149. Fukuchi, M. & Tsuda, M. Involvement of the 3′-untranslated region of the brain-derived neurotrophic factor gene in activity-dependent mRNA stabilization. *J. Neurochem.* **115**, 1222–1233 (2010).
150. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281–286 (2004).
151. McCown, P. J., Roth, A. & Breaker, R. R. An expanded collection and refined consensus model of glmS ribozymes. *RNA* **17**, 728–736 (2011).

Acknowledgements
We gratefully acknowledge the support of the US National Institutes of Health (NIH) (R01-HG004361), the Agency of Science, Technology and Research of Singapore (Y.W.) and the 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.

Competing interests statement
The authors declare competing financial interests: see Web version for details.

FURTHER INFORMATION
Eran Segal's homepage: http://genie.weizmann.ac.il
Howard Chang's homepage: http://changlab.stanford.edu/
CAFA and SNPfold: http://ribosnitch.bio.unc.edu/The_Laederach_Lab/Software.html
iPARS application for iPhone and iPad: http://itunes.apple.com/us/app/ipars/id394343332?mt=8
SAFA: https://simtk.org/home/safa

ALL LINKS ARE ACTIVE IN THE ONLINE PDF