HHS Public Access

Author manuscript

Curr Opin Biotechnol. Author manuscript; available in PMC 2017 June 01.

Published in final edited form as:

Curr Opin Biotechnol. 2016 June ; 39: 182–191. doi:10.1016/j.copbio.2016.03.019.

RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs

Eric J. Strobel[1], **Kyle E. Watters**[1], **David Loughrey**[1], and **Julius B. Lucks**[1,*]

[1]School of Chemical and Biomolecular Engineering, Cornell University, Ithaca NY

Abstract

RNAs assume sophisticated structures that are active in myriad cellular processes. In this review, we highlight newly identified ribozymes, riboswitches and small RNAs, some of which control the function of cellular metabolic and gene expression networks. We then examine recent developments in genome-wide RNA structure probing technologies that are yielding new insights into the structural landscape of the transcriptome. Finally, we discuss how these RNA 'structomic' methods can address emerging questions in RNA systems biology, from the mechanisms behind long non-coding RNAs to new bases for human diseases.

Graphical abstract

Introduction

The ability of RNA to encode both genetic and structural information is paramount to its biological centrality. Its predominantly single-stranded nature allows RNA to serve as both the physical template of protein synthesis and adopt intricate structures that influence genetic processes. For example, catalytic RNAs (ribozymes) perform essential cellular functions including translation, tRNA maturation, and splicing. Even more diverse are the roles of non-coding RNAs in regulating gene expression. These roles are frequently

Correspondence: Prof. Julius Lucks, 120 Olin Hall, Ithaca, NY 14853, jblucks@cornell.edu.

mediated by *cis*- and *trans*-acting RNA structures that block or expose regulatory elements within mRNAs that control transcription, translation, or RNA degradation [1]. Further regulatory roles for RNAs include protein recruitment, molecular scaffolding, and RNA interference, with many others being discovered at an accelerating rate [2]. These advances frame an emerging picture of diverse RNAs acting together in a networked, systems-level fashion to regulate the fundamental processes of the cell (Figure 1).

Alongside exciting discoveries about the breadth of RNA *function* is the development of tools to uncover 'omics'-level views of RNA *structure*. As RNA function is intimately tied to RNA structure, these technologies provide powerful strategies for elucidating RNA structure-function relationships on a systems-level scale by accessing structural information for entire transcriptomes in their native cellular context.

In this review, we unite exciting developments in the growing knowledge of systems-level RNA functions and new capabilities used to uncover the RNA structures that give rise to those functions. We start by highlighting new discoveries that expose the prevalent and varied nature of RNA functions in biological systems. Next, we discuss recent experimental developments in high-throughput RNA structure analysis at the transcriptome level, the bioinformatic advances necessary to analyze the generated data sets, and the insights these studies have provided. Finally, we highlight questions that can be asked with a systems-level knowledge of RNA structure-function relationships and consider the new role that their answers will play in the future of RNA biology.

I. Unearthing New Global Roles for RNAs in Regulating Cellular Processes

Recent efforts to identify and characterize RNA-mediated regulatory pathways have led to an appreciation for the role of RNA in governing global cellular processes such as metabolic and gene expression networks. The identification of new RNA mechanisms and functional roles suggests that others remain hidden within the transcriptome.

Twister, Twister Sister, Pistol, and Hatchet – New Ribozymes Hiding in Plain Sight

In spite of their involvement in major cellular functions such as translation and tRNA processing, as of 2013 only 10 classes of natural ribozymes had been identified [3]. Recently, Roth *et al.* discovered a new 'twister' class of self-cleaving ribozyme using a comparative genomics approach that incorporated RNA structure prediction through sequence covariation analysis to identify over 2,700 sequences that match the twister motif across diverse organisms from bacteria to eukarya [3]. This sequence and structure-based search methodology also suggested that the twister ribozyme forms a compact, double-pseudoknotted structure that is further supported by biochemical evidence and three crystallographic studies [3–6] (Figure 2).

Despite this exciting discovery, the functional roles of twister ribozymes remain a mystery. Furthermore, by searching near genetic elements frequently associated with twister or hammerhead ribozymes, Weinberg *et al.* recently added the 'twister sister', 'pistol', and 'hatchet' classes of self-cleaving ribozymes to the list of new ribozyme classes with unknown function [7–9] (Figure 2). While the enrichment of ribozymes in specific genetic

contexts could yield clues to their function, their apparent ubiquity across broad organism classes suggests new roles of ribozymes in controlling core cellular processes that have yet to be uncovered.

Riboswitches – A Network of Small-Molecule Regulators

Like ribozymes, riboswitches have functions that are closely linked to their structures. Whereas ribozyme structures enable catalysis, riboswitch structures switch between distinct conformations in the presence or absence of a ligand to modulate gene expression (Figure 2). These ligand-mediated structural changes provide a natural sensory feedback mechanism to regulate genes involved in controlling ligand concentration at the transcriptional, translational, and splicing levels.

The recent identification and characterization of new riboswitches has expanded our understanding of their role in modulating cellular state through specific regulation of key transporter genes and even small RNAs. Recent studies have identified roles for riboswitches in prokaryotic metal ion homeostasis. Specifically, an Mn^{2+}-responsive riboswitch (*yybyko Y)* was shown to control the expression of an Mn^{2+} exporter [10,11] and a Ni^{2+}/Co^{2+} riboswitch was shown to control expression of Co^{2+} transporters [12]. Since Mn^{2+} and Co^{2+} are cofactors of protein enzymes, but are toxic at elevated levels, these studies highlight how riboswitches provide feedback mechanisms that affect cell physiology.

In another example of expanded roles of riboswitches, Kim *et al.* recently identified and characterized a class of riboswitch that responds to ZMP, a purine biosynthetic intermediate, and its 5′-triphosphorylated derivative, ZTP [13]. The widespread existence of the ZTP riboswitch provides a molecular basis for a previous proposal that elevated levels of ZTP function as an alarmone to signal low levels of 10f-tetrahydrofolate, a formyl group donor in purine biosynthesis. Finally, Kellenberger *et al.* and *Nelson et al.* have identified the subclass GEMM-1b riboswitch in the bacterium *Geobacter metallireducens* that responds selectively to the cyclic dinucleotide cAG [14,15]. Interestingly, many genes that are regulated by the GEMM-1b riboswitch are associated with extracellular electron transfer [14,15].

Finally, two studies reported a riboswitch that controls the expression of small RNAs (sRNAs) that regulate the *eut* mRNAs involved in ethanolamine catabolism [16,17]. In this system, the protein EutV interacts with the 5′ untranslated region of the *eut* mRNAs to regulate their expression by transcription antitermination. However, an sRNA (Rli55 in *Listeria monocytogenes* and EutX in *Enterococcus faecalis*) can sequester EutV, preventing it from antiterminating the *eut* mRNAs. The sRNA is regulated by an adenosyl cobalamine (AdoCbl) riboswitch that terminates sRNA synthesis before the EutV binding site in the presence of the ethanolamine catabolism cofactor AdoCbl (Figure 2).

Big Roles For Small RNAs in Bacteria

Bacterial sRNAs have long been known to be important regulators of cellular state via regulating specific target genes. Recently, several groups have reported advances in understanding how the RNA-binding protein Hfq mediates these processes by presenting sRNAs for mRNA target recognition [18–21]. These structural studies are revealing key design principles for sRNA structure and function and are supporting newly discovered big

roles of sRNAs across the cell, including as elements in sophisticated regulatory networks that facilitate cellular information processing. For example, Papenfort *et al.* showed that the sRNA RprA controls a coherent feed-forward loop with AND gate logic that regulates *Salmonella* plasmid conjugation by controlling expression of the *ricI* gene [22].

Several recent studies have reported more global roles for sRNAs in the regulation of cellular state. Duss *et al.* recently uncovered the molecular basis behind this capability of the *Pseudomonas fluorescens* sRNA RsmZ by showing that it can sequentially bind five RsmE dimers, as well as release RsmE following RNaseE cleavage of RsmZ [23]. In another example, Miyakoshi *et al.* showed that the sRNA SroC functions as a sponge for another sRNA, GcvB, a global regulator in *Salmonella* [24]. Interestingly, SroC is generated as an mRNA decay product of one of GcvB's targets, creating an sRNA feedback loop within this regulation (Figure 2). Additionally, Chao and Vogel showed that RNaseE cleaves the 3′ untranslated region of the stress chaperone CpxP mRNA to produce the sRNA CpxQ, which represses mRNAs that encode inner membrane proteins [25]. Finally, Guo *et al.* reported the discovery of a new sRNA, MicL, that downregulates the most abundant protein in *E. coli*, the major lipoprotein Lpp, in times of membrane stress [26]. This is particularly interesting because MicL is expressed from a newly identified σ^E-dependent promoter within the coding sequence of *cutC* (Figure 2). This suggests a potential abundance of RNAs playing systems-level regulatory roles that remain hidden throughout the transcriptome waiting to be found.

II. RNA Structomics – A Burgeoning New Field Enabled by New Technologies

The accelerating discovery rate of new RNA functions demands methods that can provide structure-function insights at the same pace. While phylogenetic analysis of RNA structure has been immensely successful in identifying bacterial functional RNAs, the application of such methods to eukaryotes is complicated by increased genomic complexity and reduced sequence divergence [27]. New techniques that marry RNA enzymatic or chemical probing with next-generation sequencing (NGS) provide an experimental framework for the identification of functional RNAs at a transcriptome-wide level. This new 'RNA structomics' field [28] is already uncovering new insights into the global roles of RNA structures across cellular processes.

A Suite of New High-Throughput Methods Characterize RNA Structures Across the Entire Transcriptome

Early approaches to transcriptome-level RNA structure probing include the FragSeq and PARS techniques, which used NGS to sequence and map cleavage positions generated by ssRNA and dsRNA nucleases [29,30] (Table 1, Figure 3). In a similar spirit, PIP-seq combined nuclease-based RNA structure probing with crosslinking methods to access a transcriptome-wide profile of RNA-protein interactions [31]. Transcriptome-wide RNA structure probing techniques underwent another breakthrough with the use of small molecule chemical probes that can diffuse across the cell membrane and thereby probe RNA structures in their native environment [32,33]. Chemical probes also allow the interrogation

of RNA structures at higher resolution because of their small size compared to the more bulky enzymes. Following the development of methods for coupling chemical probing with NGS [34], techniques such as DMS-Seq [35], structure-seq [36], and Mod-Seq [37] were developed to probe the structure of the transcriptome inside the cell (Table 1, Figure 3). These chemical probing-NGS methods consist of a core set of steps outlined in Figure 3. While powerful, the first versions of these techniques were limited by dimethyl sulfate (DMS), which has a strong preference for A and C positions. Incarnato *et al.* partially addressed this by probing with both DMS and N-cyclohexyl-N'-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMCT), which reacts primarily with G and U, although their method, CIRS-seq, did not modify the RNA until after cell lysis [38] (Table 1, Figure 3). Complete nucleotide coverage was achieved with icSHAPE, which uses a clickable version of the SHAPE reagent 2-methylnicotinic acid imidazolide (NAI) [32], NAI-N_3. In icSHAPE, after modification of RNA by NAI-N_3, a biotin moiety is added via click chemistry to enable selective purification of probed RNAs [39]. SHAPE-MaP, which uses mutational profiling to locate modification positions, has also been applied on a transcriptome-wide scale using the well-characterized SHAPE reagent 1-methyl-7nitroisatoic anhydride (1M7) [40]. Finally, a complementary technique called RNA proximity ligation (RPL) was recently developed to characterize the proximity of nucleotides in three-dimensional space using a combination of RNase cleavage and localized ligation [41].

The complexities of the datasets generated by NGS methods have required substantial new developments in bioinformatics pipelines that can ultimately convert NGS reads into RNA structure models. To do this, NGS reads are converted into 'reactivity' values, broadly defined as a measure of the flexibility of a given nucleotide position [34,42,43]. Reactivities can then be used to generate RNA structural models that account for the tendency of more reactive nucleotides to be unpaired [44]. SeqFold presents one particularly interesting modeling approach to select RNA structures that are most consistent with the experimental reactivity data [45]. Seqfold's approach is powerful because it uses reactivity information to pick from clusters of sub-optimal structures rather than relying solely on the minimum free energy structure model. The various reactivity calculation and RNA structural modeling approaches are still in their infancy and represent a challenging new frontier for computational biology to fully utilize the vast datasets generated by the new NGS structure probing techniques.

Global Insights Into the Roles of RNA Structures Across Cellular Processes— Characterization of RNA structures at the transcriptome level is revealing features of RNA structures on a genome-wide scale. Meta-analyses that average reactivities of many different RNAs have revealed structure-function trends across multiple species. Notably, a three-nucleotide periodicity of reactivity was observed within mRNA coding regions [36,38,39,46]. As another example, the Kozak sequence appeared to be highly reactive in *H. sapiens* [46], *A. thaliana* [36], and *M. musculus* [38,39], suggesting that it is generally unstructured to facilitate translation initiation. In addition, Wan *et al.* observed that nucleotides preceding sequences known to interact with miRNAs tend to be unstructured [46].

There are also interesting conclusions gained from comparing reactivities of RNAs refolded and probed *in vitro* to RNAs probed *in vivo*. For example, *S. cerevisiae* RNAs appear more unstructured *in vivo* than *in vitro* [35]. The data collected from icSHAPE in *M. musculus* also support this argument, although the degree of *in vivo* unfolding observed was different across different classes of RNA elements [39]. In *A. thaliana*, Ding *et al*. reported a correlation between less structured mRNAs *in vivo* and mRNAs annotated for stress response and suggested that reduced structure may facilitate stress-mediated RNA structural changes [36]. Spitale *et al*. found that Kozak sequence accessibility observed *in vivo* was preserved *in vitro*, suggesting that this and other structural features of translation regulatory regions are programmed by the mRNA sequence and not through interactions with cellular factors [39]. Smola *et al*. developed the ΔSHAPE analysis framework for characterizing RNA-protein interactions through comparison of RNAs probed *in cellulo* and *ex vivo* [40]. Spitale *et al*. also show that comparison between *in vivo* and *in vitro* reactivity data can uncover specific RNA structural changes due to protein binding [39], which has also been shown in another recent study in *E. coli* with in-cell SHAPE-Seq [47].

III. New Technologies Enable New Questions

Rapid advances in RNA structure characterization technologies promise to change the way we investigate the relationship between RNA structure and function at a systems-level. While many new questions can be addressed, two of the most interesting are the structural basis of long non-coding RNA (lncRNA) function and the role of RNA misfolding in human diseases.

What are the structure-function principles of long non-coding RNAs?

lncRNAs are loosely defined as RNA molecules more than 200 nucleotides long with little-to-no protein-coding capacity [48]. Despite their abundance [49], lncRNAs are one of the least understood RNA classes. While we know lncRNA structure is important [50], we have little detailed knowledge of how specific lncRNA structures mediate their broad arrays of function, although this has begun to change. Recently, Somarowthu *et al*. used several chemical probing techniques to determine the secondary structure of the 2,148 nt long lncRNA HOTAIR, giving structural insights into how this RNA performs the twin functions of regulating epidermal tissue development and repressing tumor and metastasis suppressor genes [51]. It will be exciting to gauge how RNA structures influence the function of newly discovered lncRNAs, such as Firre, which has been shown to act as a platform for organizing trans-chromosomal association [52] (Figure 1). Another interesting new example is the *extra-coding* CEBPA, which controls DNA methylation state at the CEBPA locus using RNA structures that are targeted by a DNA methyltransferase, DNMT-1 [53] (Figure 1). We anticipate this to be the tip of the iceberg, as high-throughput structural studies enable a wealth of new insight into the structure-function principles of these important global RNA regulatory molecules.

How does RNA misfolding contribute to human disease?

The growing appreciation for the role of RNA structure in cellular activities has lead to intriguing questions about the role of RNA structure in human disease. A recent focus of

these studies is the “riboSNitch”, an RNA-encoded regulatory element in which a single nucleotide variant (SNV) significantly alters its structural ensemble, sometimes leading to a disease state such as β-Thalassemia or Chronic Obstructive Pulmonary Disease [54]. Following the initial computational prediction of riboSNitches and their disease associations [54], Wan *et al.* used PARS (described above) to structurally probe the transcriptomes of a mother, father, and child on a genome-wide scale and found that over 1,907 (15%) of identified SNVs altered RNA structures between these relatives [46]. The dataset acquired in this study was then used by Corley *et al.* to benchmark RNA folding algorithms to predict the effect of SNVs on RNA structure and thus accurately predict the locations of riboSNitches from primary sequence information [55]. While there is still fascinating work to be done to improve computational prediction, the link between riboSNitches and disease is one of the most exciting areas of future RNA research, both in terms of understanding the global RNA structure-function relationship and as a new frontier in human disease research.

Conclusion

Far from being a passive carrier of genetic information and an intriguing catalyst of select chemical processes of life, RNAs play diverse roles as regulators of central cellular processes. Our knowledge of these roles is expanding at an accelerated rate, with recent discoveries uncovering RNAs in unexpected places and with unexpected function. These studies suggest that we may need to rethink our view of RNA yet again and may warrant investment in a new study of ‘RNA systems biology’ that can more thoroughly uncover the roles and mechanisms of RNAs in modern biology.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program [Grant No. DGE-1144153 to K.E.W.], and a New Innovator Award through the National Institute of General Medical Sciences of the National Institutes of Health [grant number 1DP2GM110838 to J. B. L.]. K.E.W. is a Fleming Scholar in the School of Chemical and Biomolecular Engineering at Cornell University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Chappell J, Takahashi MK, Meyer S, Loughrey D, Watters KE, Lucks J. The centrality of RNA for engineering gene expression. Biotechnol J. 2013; 8:1379–1395. [PubMed: 24124015]
2. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 2014; 157:77–94. [PubMed: 24679528]
3**. Roth A, Weinberg Z, Chen AG, Kim PB, Ames TD, Breaker RR. A widespread self-cleaving ribozyme class is revealed by bioinformatics. Nat Chem Biol. 2014; 10:56–60. This work reports the identification of a new class of self-cleaving ribozyme, called ‘twister’, present in all domains of life, but with an unknown biological role. This work prompted the discovery of three additional self-cleaving ribozyme classes [7]. Together, these works suggest broad biological roles for self-cleaving ribozymes that have yet to be identified. [PubMed: 24240507]
4. Liu Y, Wilson TJ, McPhee SA, Lilley DM. Crystal structure and mechanistic investigation of the twister ribozyme. Nat Chem Biol. 2014; 10:739–744. [PubMed: 25038788]
5. Eiler D, Wang J, Steitz TA. Structural basis for the fast self-cleavage reaction catalyzed by the twister ribozyme. Proc Natl Acad Sci U S A. 2014; 111:13028–13033. [PubMed: 25157168]

6. Ren A, Kosutic M, Rajashankar KR, Frener M, Santner T, Westhof E, Micura R, Patel DJ. In-line alignment and Mg(2)(+) coordination at the cleavage site of the env22 twister ribozyme. Nat Commun. 2014; 5:5534. [PubMed: 25410397]
7. Weinberg Z, Kim PB, Chen TH, Li S, Harris KA, Lunse CE, Breaker RR. New classes of self-cleaving ribozymes revealed by comparative genomics analysis. Nat Chem Biol. 2015; 11:606–610. [PubMed: 26167874]
8. Li S, Lunse CE, Harris KA, Breaker RR. Biochemical analysis of hatchet self-cleaving ribozymes. RNA. 2015; 21:1845–1851. [PubMed: 26385510]
9. Harris KA, Lunse CE, Li S, Brewer KI, Breaker RR. Biochemical analysis of pistol self-cleaving ribozymes. RNA. 2015; 21:1852–1858. [PubMed: 26385507]
10. Price IR, Gaballa A, Ding F, Helmann JD, Ke A. Mn(2+)-sensing mechanisms of yybP-ykoY orphan riboswitches. Mol Cell. 2015; 57:1110–1123. [PubMed: 25794619]
11. Dambach M, Sandoval M, Updegrove TB, Anantharaman V, Aravind L, Waters LS, Storz G. The ubiquitous yybP-ykoY riboswitch is a manganese-responsive regulatory element. Mol Cell. 2015; 57:1099–1109. [PubMed: 25794618]
12. Furukawa K, Ramesh A, Zhou Z, Weinberg Z, Vallery T, Winkler WC, Breaker RR. Bacterial riboswitches cooperatively bind Ni(2+) or Co(2+) ions and control expression of heavy metal transporters. Mol Cell. 2015; 57:1088–1098. [PubMed: 25794617]
13. Kim PB, Nelson JW, Breaker RR. An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. Mol Cell. 2015; 57:317–328. [PubMed: 25616067]
14. Kellenberger CA, Wilson SC, Hickey SF, Gonzalez TL, Su Y, Hallberg ZF, Brewer TF, Iavarone AT, Carlson HK, Hsieh YF, et al. GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. Proc Natl Acad Sci U S A. 2015; 112:5383–5388. [PubMed: 25848022]
15. Nelson JW, Sudarsan N, Phillips GE, Stav S, Lunse CE, McCown PJ, Breaker RR. Control of bacterial exoelectrogenesis by c-AMP-GMP. Proc Natl Acad Sci U S A. 2015; 112:5389–5394. [PubMed: 25848023]
16**. DebRoy S, Gebbie M, Ramesh A, Goodson JR, Cruz MR, van Hoof A, Winkler WC, Garsin DA. Riboswitches. A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. Science. 2014; 345:937–940. This paper reports the discovery of a riboswitch that controls the expression of an sRNA in *Enterococcus faecalis*. This is a new mechanistic role for riboswitches that expands their functions as systems level regulators of cell physiology. [PubMed: 25146291]
17**. Mellin JR, Koutero M, Dar D, Nahori MA, Sorek R, Cossart P. Riboswitches. Sequestration of a two-component response regulator by a riboswitch-regulated noncoding RNA. Science. 2014; 345:940–943. Occurring back to back with DeBroy *et al.*, this paper reports a similar discovery of a riboswitch-regulated sRNA in *Listeria monocytogenes*. [PubMed: 25146292]
18. Malecka EM, Strozecka J, Sobanska D, Olejniczak M. Structure of bacterial regulatory RNAs determines their performance in competition for the chaperone protein Hfq. Biochemistry. 2015; 54:1157–1170. [PubMed: 25582129]
19. Wang L, Wang W, Li F, Zhang J, Wu J, Gong Q, Shi Y. Structural insights into the recognition of the internal A-rich linker from OxyS sRNA by Escherichia coli Hfq. Nucleic Acids Res. 2015; 43:2400–2411. [PubMed: 25670676]
20. Dimastrogiovanni D, Frohlich KS, Bandyra KJ, Bruce HA, Hohensee S, Vogel J, Luisi BF. Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. Elife. 2014; 3
21. Peng Y, Curtis JE, Fang X, Woodson SA. Structural model of an mRNA in complex with the bacterial chaperone Hfq. Proc Natl Acad Sci U S A. 2014; 111:17134–17139. [PubMed: 25404287]
22. Papenfort K, Espinosa E, Casadesus J, Vogel J. Small RNA-based feedforward loop with AND-gate logic regulates extrachromosomal DNA transfer in Salmonella. Proc Natl Acad Sci U S A. 2015; 112:E4772–4781. [PubMed: 26307765]
23. Duss O, Michel E, Yulikov M, Schubert M, Jeschke G, Allain FH. Structural basis of the non-coding RNA RsmZ acting as a protein sponge. Nature. 2014; 509:588–592. [PubMed: 24828038]

24*. Miyakoshi M, Chao Y, Vogel J. Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. EMBO J. 2015; 34:1478–1492. This paper shows that the sRNA SroC acts as a sponge to downregulate the global regulatory sRNA GcvB in *Salmonella enterica*. Since SroC is generated from the decay of one of GcvB's targets, this shows that sRNAs can play sophisticated roles in regulatory feedback loops. [PubMed: 25630703]

25*. Chao Y, Vogel J. A 3′ UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. Mol Cell. 2016; 61:352–363. This work identifies the sRNA CpxQ as a decay product of the *cpxP* mRNA that represses synthesis of inner membrane proteins in response to inner membrane stress in *Salmonella*. CpxQ is the first 3′ UTR-derived sRNA that has been identified as a multi-target regulator. [PubMed: 26805574]

26**. Guo MS, Updegrove TB, Gogol EB, Shabalina SA, Gross CA, Storz G. MicL, a new sigmaE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. Genes Dev. 2014; 28:1620–1634. This paper presents the discovery of the sRNA MicL, which controls the expression of one of the most abundant proteins in *E. coli*, Lpp. Intriguingly MicL was shown to be encoded in the CutC coding region and all phenotypes associated with CutC were shown to be attributed to MicL. This paper challenges our understanding of protein-mediated phenotypes by finding that sRNAs can be embedded in coding sequences and calls for a deeper systems level investigation of sRNA biogenesis. [PubMed: 25030700]

27. Breaker RR. Prospects for riboswitch discovery and analysis. Mol Cell. 2011; 43:867–879. [PubMed: 21925376]

28. Kwok CK, Tang Y, Assmann SM, Bevilacqua PC. The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. Trends Biochem Sci. 2015; 40:221–232. [PubMed: 25797096]

29. Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, Lowe TM, Salama SR, Haussler D. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. Nat Methods. 2010; 7:995–1001. [PubMed: 21057495]

30. Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E. Genome-wide measurement of RNA secondary structure in yeast. Nature. 2010; 467:103–107. [PubMed: 20811459]

31. Silverman IM, Li F, Alexander A, Goff L, Trapnell C, Rinn JL, Gregory BD. RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. Genome Biol. 2014; 15:R3. [PubMed: 24393486]

32*. Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. RNA SHAPE analysis in living cells. Nat Chem Biol. 2013; 9:18–20. This paper reports the development of SHAPE reagents that can be used to probe cellular RNA structures, thus opening the door to cellular RNA structomics. This paper also shows that the influence of protein binding on RNA structure can be uncovered by the comparison of *in vivo* and *in vitro* structure measurements. [PubMed: 23178934]

33*. Tyrrell J, McGinnis JL, Weeks KM, Pielak GJ. The cellular environment stabilizes adenine riboswitch RNA structure. Biochemistry. 2013; 52:8777–8785. This paper demonstrates that the SHAPE reagent 1M7 can be used to study RNA structure inside cells. This paper also provides early glimpses of the comparison between *in vivo* and *in vitro* RNA folding. [PubMed: 24215455]

34. Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA, Arkin AP. 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]

35. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature. 2014; 505:701–705. [PubMed: 24336214]

36. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature. 2014; 505:696–700. [PubMed: 24270811]

37. Talkish J, May G, Lin Y, Woolford JL Jr, McManus CJ. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. RNA. 2014; 20:713–720. [PubMed: 24664469]

38. Incarnato D, Neri F, Anselmi F, Oliviero S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. Genome Biol. 2014; 15:491. [PubMed: 25323333]

39**. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, et al. Structural imprints in vivo decode RNA regulatory mechanisms. Nature. 2015; 519:486–490. This work presents in-vivo click SHAPE (icSHAPE), a technique that can be used to study cellular RNA structures transcriptome-wide. Using icSHAPE, the authors uncovered specific global roles of RNA structure in the regulation of broad cellular processes, such as translation. icSHAPE and related techniques have created a new field of RNA structomics, which studies global RNA structure in cellular regulation. [PubMed: 25799993]

40**. Smola MJ, Calabrese JM, Weeks KM. Detection of RNA-Protein Interactions in Living Cells with SHAPE. Biochemistry. 2015; 54:6867–6875. This work presents the ΔSHAPE analysis framework for characterization of RNA-protein interactions by comparing SHAPE reactivities of RNAs probed *in cellulo* and *ex vivo*. [PubMed: 26544910]

41. Ramani V, Qiu R, Shendure J. High-throughput determination of RNA structure by proximity ligation. Nat Biotechnol. 2015; 33:980–984. [PubMed: 26237516]

42. Aviran S, Trapnell C, Lucks JB, Mortimer SA, Luo S, Schroth GP, Doudna JA, Arkin AP, Pachter L. Modeling and automation of sequencing-based characterization of RNA structure. Proc Natl Acad Sci U S A. 2011; 108:11069–11074. [PubMed: 21642536]

43. Aviran S, Lucks JB, Pachter L. RNA structure characterization from chemical mapping experiments. Proc 49th Allerton Conf on Communication, Control, and Computing. 2011a:1743–1750.

44. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proc Natl Acad Sci U S A. 2013; 110:5498–5503. [PubMed: 23503844]

45. Ouyang Z, Snyder MP, Chang HY. SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. Genome Res. 2013; 23:377–387. [PubMed: 23064747]

46. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, et al. Landscape and variation of RNA secondary structure across the human transcriptome. Nature. 2014; 505:706–709. [PubMed: 24476892]

47. Watters KE, Abbott TR, Lucks JB. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. Nucleic Acids Res. 2016; 44:e12. [PubMed: 26350218]

48. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 2007; 316:1484–1488. [PubMed: 17510325]

49. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015; 47:199–208. [PubMed: 25599403]

50. Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol. 2013; 20:300–307. [PubMed: 23463315]

51. Somarowthu S, Legiewicz M, Chillon I, Marcia M, Liu F, Pyle AM. HOTAIR forms an intricate and modular secondary structure. Mol Cell. 2015; 58:353–361. [PubMed: 25866246]

52. Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR, et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat Struct Mol Biol. 2014; 21:198–206. [PubMed: 24463464]

53. Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, Figueroa ME, De Figueiredo Pontes LL, Alberich-Jorda M, Zhang P, et al. DNMT1-interacting RNAs block gene-specific DNA methylation. Nature. 2013; 503:371–376. [PubMed: 24107992]

54. Halvorsen M, Martin JS, Broadaway S, Laederach A. Disease-associated mutations that alter the RNA structural ensemble. PLoS Genet. 2010; 6:e1001074. [PubMed: 20808897]

55*. Corley M, Solem A, Qu K, Chang HY, Laederach A. Detecting riboSNitches with RNA folding algorithms: a genome-wide benchmark. Nucleic Acids Res. 2015; 43:1859–1868. This paper

describes the state of the art in computational riboSNitch prediction. By benchmarking the performance of a group of RNA structure prediction tools to predict experimentally validated riboSNitches, the authors were able to establish several algorithmic properties that are important for accurate prediction. Interestingly they find that some riboSNitches were not predicted by any algorithm, suggesting the existence of a class of 'environmental riboSNitches' that are influenced more by environmental factors than the thermodynamics of RNA folding. [PubMed: 25618847]

56. Kuo MY, Sharmeen L, Dinter-Gottlieb G, Taylor J. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. J Virol. 1988; 62:4439–4444. [PubMed: 3184270]
57. Michel F, Jacquier A, Dujon B. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie. 1982; 64:867–881. [PubMed: 6817818]
58. Mauger DM, Golden M, Yamane D, Williford S, Lemon SM, Martin DP, Weeks KM. Functionally conserved architecture of hepatitis C virus RNA genomes. Proc Natl Acad Sci U S A. 2015; 112:3692–3697. [PubMed: 25775547]
59. Lavender CA, Gorelick RJ, Weeks KM. Structure-Based Alignment and Consensus Secondary Structures for Three HIV-Related RNA Genomes. PLoS Comput Biol. 2015; 11:e1004230. [PubMed: 25992893]

Highlights

- We review recent progress in uncovering the global roles of RNAs across the cell.
- Regulatory RNAs are more ubiquitous and play more global, networked roles than previously thought.
- ‘RNA Structomics’ allows cellular RNA structure to be characterized across transcriptomes.
- New ways in which RNA structures globally regulate cellular processes have been uncovered.
- RNA Structomics is enabling progress in understanding the role of RNA misfolding in human disease.

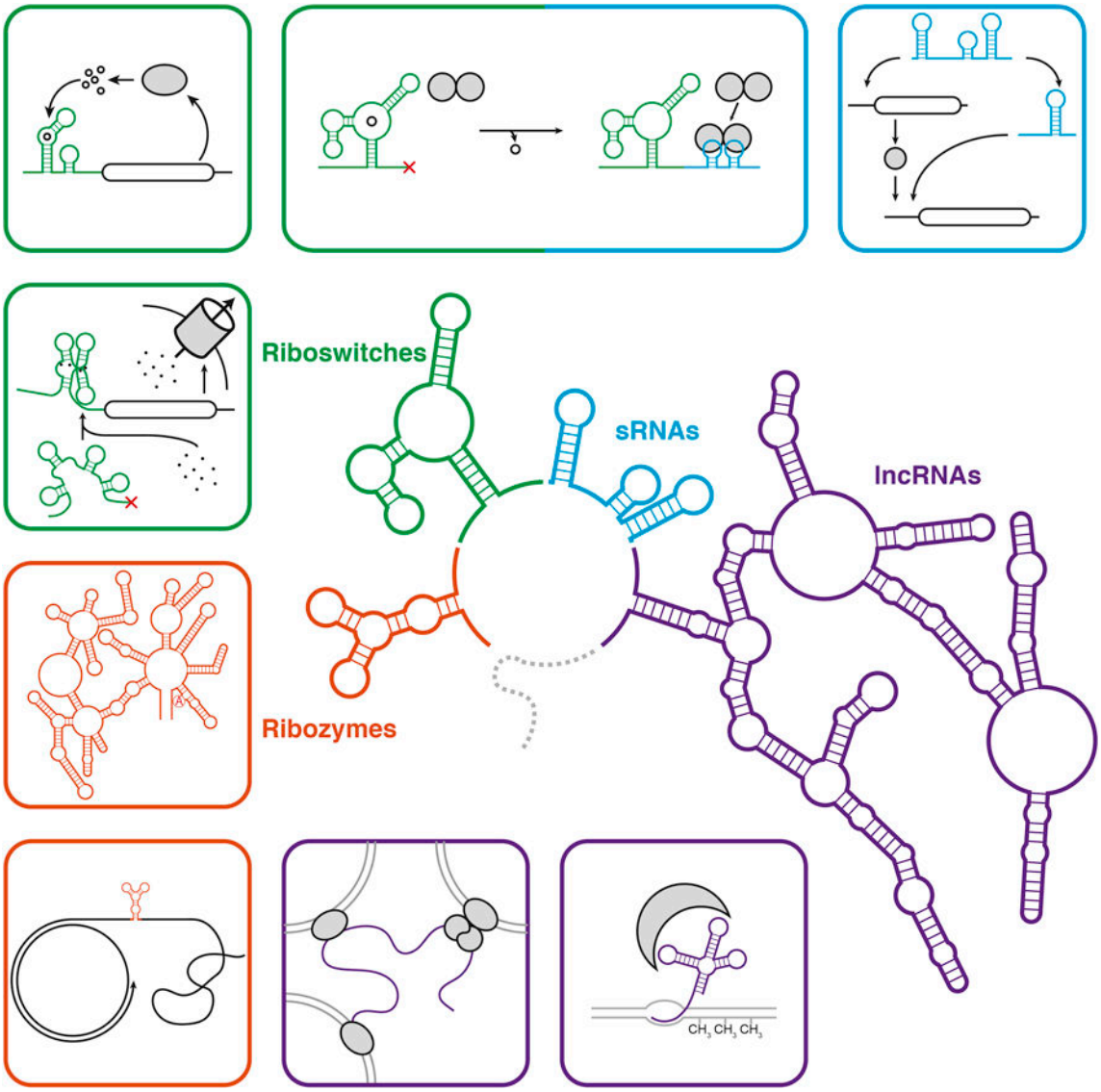

Figure 1. The centrality of RNA structures in regulating cellular processes

Non-coding RNAs (ncRNAs) play widespread and diverse roles in the regulation of cellular processes. (Center) A schematic of representative classes of the structures formed by ncRNAs including ribozymes (orange), riboswitches (green), small RNAs (sRNAs, blue), and long non-coding RNAs (lncRNAs, purple). The surrounding panels depict representative functions of each of these classes including (clockwise) concatemer cleavage in rolling circle replication [56] and group II intron splicing [57] (ribozymes); metal-ion sensing [10–12], regulation of biosynthetic operons [13], and regulation of sRNA expression [16,17] (riboswitches); sequestration of regulatory factors [16,17] and information processing in regulatory networks [22] (sRNAs); and controlling DNA methylation [53] and scaffolding for inter-chromosomal structures [52] (lncRNAs). Functional RNA motifs are highlighted in colors corresponding to the center schematic. Protein components are shaded grey.

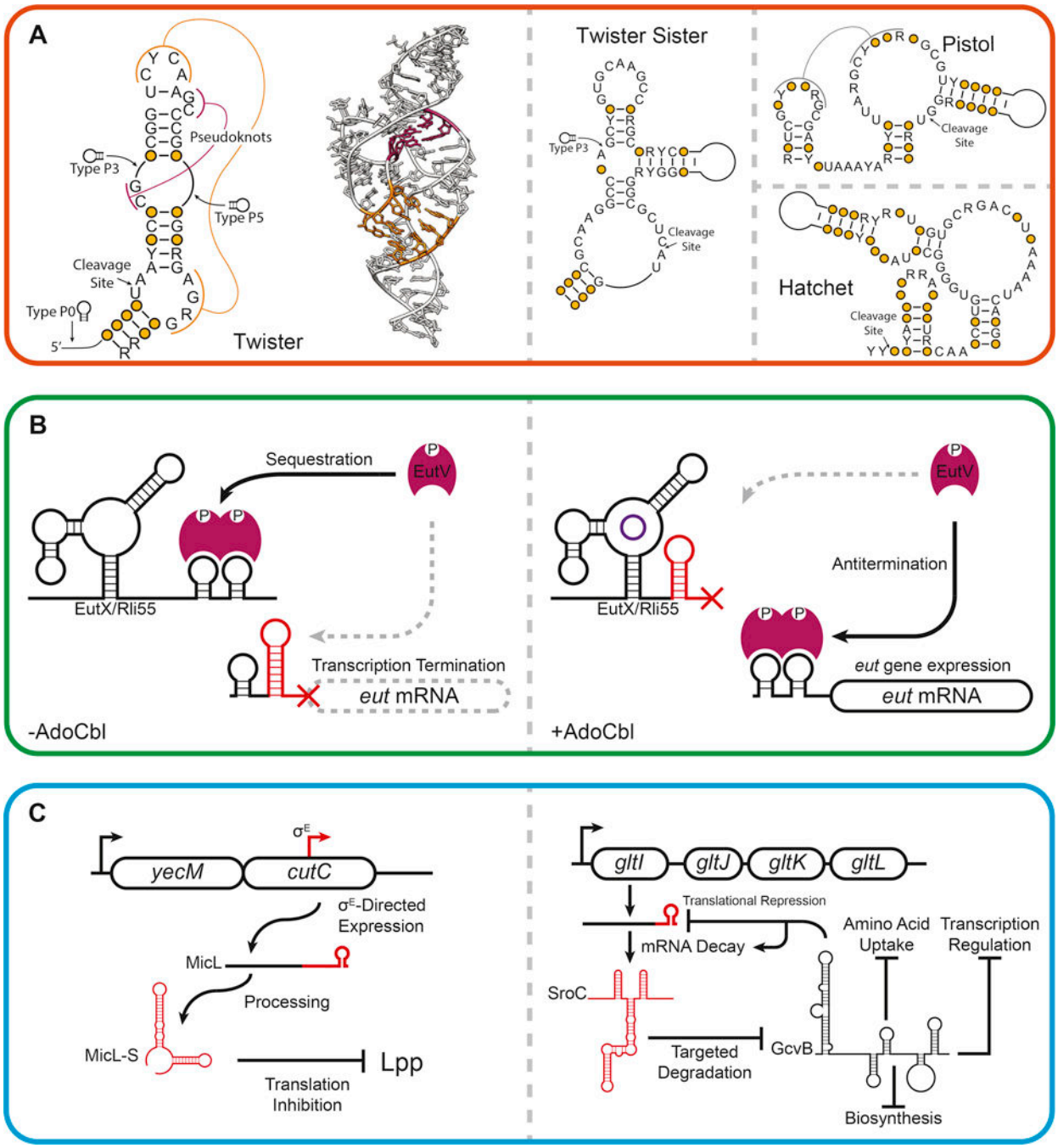

Figure 2. New roles for non-coding RNAs

(A) Widespread identification of novel self-cleaving ribozymes. Secondary structures for the twister, twister sister, pistol, and hatchet ribozymes are shown [3,7]. A crystal structure of a twister ribozyme is shown (PDBID: 4QJH) [5]. Pseudoknot interactions are shown in magenta and orange in the twister ribozyme secondary structure and correspond to magenta and orange nucleotides in the twister ribozyme crystal structure. (B) An AdoCbl riboswitch regulates the expression of an sRNA sponge [16,17]. In the absence of AdoCbl, the full length EutX (or Rli55) sRNA is synthesized and sequesters the transcription antiterminator

EutV. In the presence of AdoCbl, premature termination of EutX/Rli55 permits EutV to antiterminate transcription of the *eut* operon, leading to the expression of proteins involved in ethanolamine catabolism. (C) New roles for sRNAs in the global regulation of cellular processes. (Left) In response to outer membrane stress in γ-proteobacteria, the sRNA MicL is expressed from a σ^E-dependent promoter embedded in the *cutC* coding sequence and is processed into MicL-S, which inhibits translation of the outer membrane protein Lpp [26]. (Right) Decay of the gltI mRNA yields the sRNA SroC, which promotes degradation of the global regulatory sRNA GcvB, which regulates translation of the *gltI* mRNA [24].

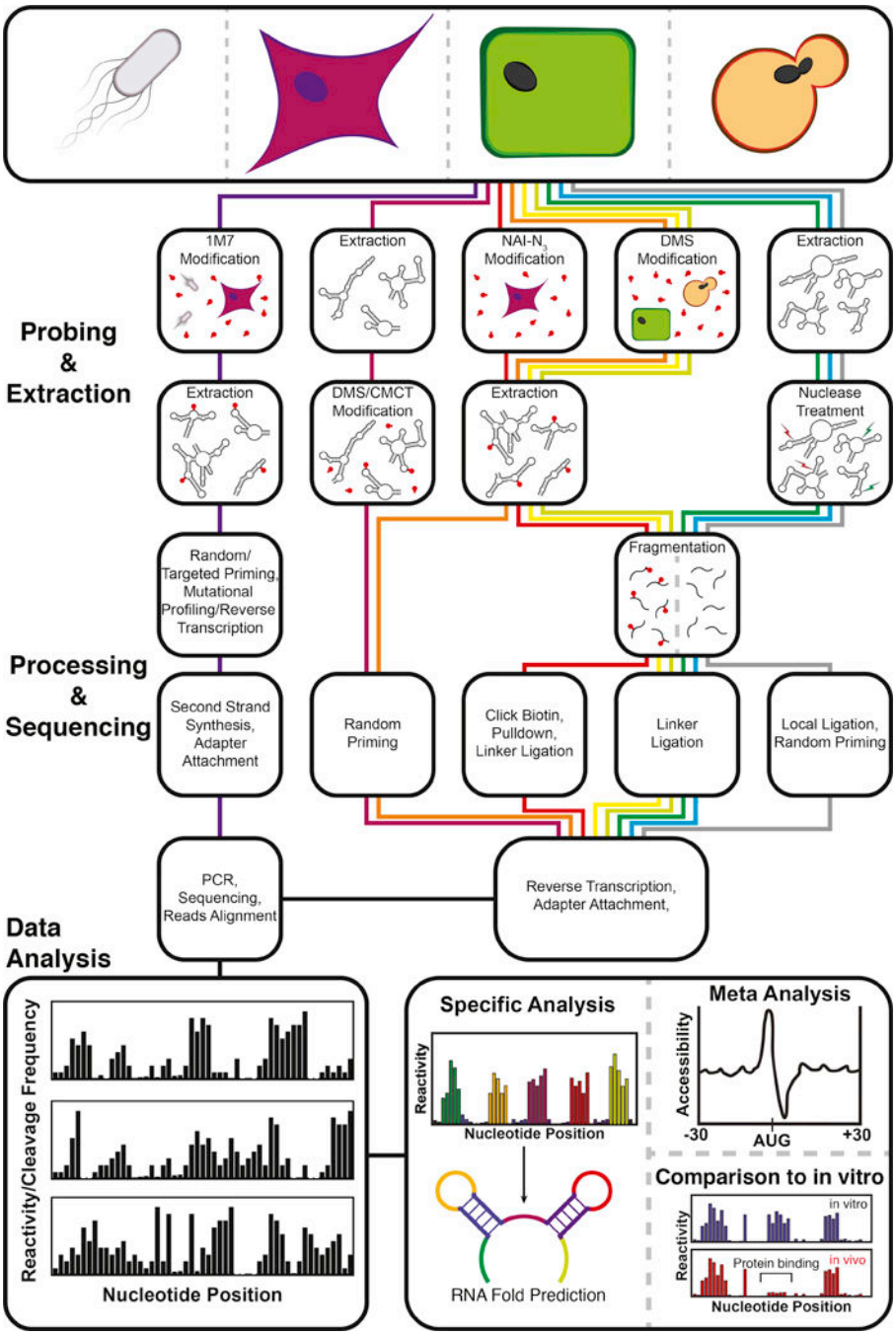

Figure 3. Transcriptome-wide RNA structure probing technologies

Transcriptome-wide RNA structure probing uses chemical probing or enzymatic cleavage to introduce covalent modifications or directly cleave RNA in a structure dependent fashion, respectively. Modification or cleavage positions are detected through processing steps followed by next-generation sequencing. Bioinformatic processing of the resulting sequencing reads yields a measure of chemical modification ‘reactivity’ or enzymatic cleavage frequency at each nucleotide. High reactivities correspond to flexible nucleotide positions that are not participating in RNA structures or bound by cellular factors. High

enzymatic cleavage frequencies give information on structure depending on the characteristics of the nucleases used. These values can be used for several types of specific analyses, such as constrained RNA folding, averaging meta-analysis of reactivities across the entire transcriptome, and comparisons with *in vitro* probing data. An outline of the steps for SHAPE-Map [40] and in-cell SHAPE-Seq [47] (purple) CIRS-seq [38] (maroon), icSHAPE [39] (red), structure-seq [36] (orange), DMS-seq [35] (yellow), Mod-seq [37] (light green), PARS [30] (green), FragSeq [29] (blue), and RPL [41] (grey), is shown. Further technique details can be found in Table 1.

Table 1

Characteristics of high throughput RNA structure probing methods

These techniques use a series of structure probing, extraction, processing, next generation sequencing library preparation, sequencing and bioinformatic data analysis steps (Figure 3) to characterize RNA structures in high throughput and in some cases transcriptome-wide. Detailed differences are included below.

Name	Modifying reagent or enzyme	Probing, Extraction and Processing before Sequencing Library Preparation	Organism(s) studied	Current Analyzing Software	Reference(s)
PARS (parallel analysis of RNA structure)	V1 (dsRNase) & S1 (ssRNase)	RNA extraction, equilibrium refolding, enzyme treatment, RNA fragmentation & linker ligation	*Saccharomyces cerevisiae*	Bowtie2, custom scripts	Kertesz, 2010, Nature [30]
FragSeq	P1 (ssRNase)	RNA extraction, equilibrium refolding, enzyme treatment, fragmentation & linker ligation	*Mus musculus*	FragSeq algorithm	Underwood, 2010, Nat. Meth. [29]
PIP-Seq	RNase ONE (ssRNase) & V1 (dsRNase)	crosslinking, cell lysis, enzyme treatment, crosslink reversal, RNA extraction	*Homo sapiens, Arabid opsis thaliana*	Tophat, CSAR	Silverman, 2014, Genome Biol. [31]
DMS-Seq	DMS	in-cell DMS modification, RNA extraction, fragmentation & linker ligation	*Saccharomyces cerevisiae*	SOAP	Rouskin, 2014, Nature [35]
structure-seq	DMS	in-cell DMS modification, RNA extraction, random priming	*Arabidopsis thaliana*	StructureF old (as part of Galaxy suite)	Ding, Nature, 2014 [36]
Mod-Seq	DMS	in-cell DMS modification, RNA extraction, fragmentation & linker ligation	*Saccharomyces cerevisiae*	Mod-seeker	Talkish, RNA, 2014 [37]
CIRS-Seq	DMS, CMCT	RNA extraction, DMS or CMCT modification, random priming	*Mus musculus*	custom scripts	Incarnato, Genome Biol, 2014 [38]
icSHAPE (*in vivo* click selective 2′-hydroxyl acylation and profiling experiment)	NAI-N_3 (2-methylnico tinic acid imidazolid e)	in-cell NAI-N_3 modification, RNA extraction, fragmentation, biotin click & purification	*Mus musculus*	Bowtie2, custom scripts	Spitale, Nature, 2015 [39]
RPL (RNA proximity ligation)	Endogeno us RNases	spheroplast,endogenous RNase cleavage, RNA cross-strand ligation, fragmentation & ligation	*Saccharomyces cerevisiae*	STAR aligner, custom scripts	Ramani, Nat. Biotech., 2015 [41]
SHAPE-MaP (selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling)	1M7 (1-methyl-nitroisatoic anhydride)	*in vitro* synthesis or viral purification, equilibrium folding, in-cell 1M7 modification, RNA extraction and fragmentation, random/targeted priming & ligation	Hepatitis C & HIV, *Mus musculus*	ShapeMa pper	Mauger, PNAS, 2015 & Lavender, Plos Comput. Biol., 2015 [40,58,59]

Name	Modifying reagent or enzyme	Probing, Extraction and Processing before Sequencing Library Preparation	Organism(s) studied	Current Analyzing Software	Reference(s)
In-cell SHAPE-Seq	1M7	In-cell 1M7 modification, RNA extraction, specific priming	*Eschericia coli* (natural and synthetic sRNAs, riboswitches and RNase P)	Spats	Watters, NAR 2016 [47]