




## Review

## Revealing lncRNA Structures and Interactions by Sequencing-Based Approaches

Xingyang Qian,<sup>1,3</sup> Jieyu Zhao ,<sup>2,3</sup> Pui Yan Yeung ,<sup>2,3</sup> Qiangfeng Cliff Zhang,<sup>1,\*</sup> and Chun Kit Kwok ,<sup>2,\*</sup>

Long noncoding RNAs (lncRNAs) have emerged as significant players in almost every level of gene function and regulation. Thus, characterizing the structures and interactions of lncRNAs is essential for understanding their mechanistic roles in cells. Through a combination of (bio)chemical approaches and automated capillary and high-throughput sequencing (HTS), the complexity and diversity of RNA structures and interactions has been revealed in the transcriptomes of multiple species. These methods have uncovered important biological insights into the mechanistic and functional roles of lncRNA in gene expression and RNA metabolism, as well as in development and disease. In this review, we summarize the latest sequencing strategies to reveal RNA structure, RNA–RNA, RNA–DNA, and RNA–protein interactions, and highlight the recent applications of these approaches to map functional lncRNAs. We discuss the advantages and limitations of these strategies, and provide recommendations to further advance methodologies capable of mapping RNA structure and interactions in order to discover new biology of lncRNAs and decipher their molecular mechanisms and implication in diseases.

### Biological Significance of lncRNAs and RNA Structure

In the human genome, approximately 93% of DNA can be transcribed as RNA, only 2% of which is protein-encoding mRNAs, while the remaining 98% is known as noncoding RNAs [1,2]. Among these noncoding RNA ‘dark matters’, RNAs longer than 200 bases are classified as lncRNAs. Since the advent of the genomic era in the 2000s, significant progress has been made toward the understanding of the prevalence, abundance, biogenesis, and functions of lncRNAs across different cell types and species [3,4]. In particular, lncRNAs have been demonstrated to play important roles in epigenetic control and the regulation of transcription, translation, RNA metabolism (Table 1), as well as stem cell maintenance and differentiation, cell autophagy and apoptosis, and embryonic development [5–7]. In addition, lncRNAs have been implicated in major diseases including different types of cancer, and neurological and cardiovascular diseases [8–10]. With the accumulating knowledge of genomic variations and expanding lncRNA repository, many disease-associated single nucleotide polymorphisms (SNPs) have been mapped to lncRNA genes [11–14]. Databases such as LincSNP and LincSNP 2.0 have been created to facilitate the exploration of the potential functions of lncRNA-associated SNPs [15,16].

The discovery of the catalytic and regulatory functions of RNA have refined the central dogma of molecular biology and highlighted the multifaceted biological roles of RNA [17,18]. A significant body of research has shown that the higher-order structures as well as interactions of RNA serve

### Highlights

Higher-order structures and interactions of lncRNA are critical for its diverse roles in gene function and regulation.

Novel chemical and sequencing tools are being developed to decipher RNA structures and interactions *in vitro* and *in vivo*.

Application of these innovative methods to lncRNA has revealed new and important structural motifs and interaction groups.

The methods and results reviewed here can help to better understand and further investigate the lncRNA structure–function relationship.

<sup>1</sup>MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China  
<sup>2</sup>Department of Chemistry, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China  
<sup>3</sup>These authors contributed equally to this work

\*Correspondence: qc Zhang at [qc Zhang@tsinghua.edu.cn](mailto:qc Zhang@tsinghua.edu.cn) (Q.Q. Zhang) and ckkwok42@cityu.edu.hk (C.K. Kwok).

Table 1. Roles of Representative lncRNAs in Gene Expression and RNA Metabolism

Biological process	lncRNA example	Role in gene expression and RNA metabolism	Refs
Transcription	<i>NRON</i>	<i>NRON</i> interacts with importin- $\beta$ proteins and inhibits the trafficking of NFAT transcription factor from the cytoplasm to the nucleus, which can lead to inactivation of target genes.	[154]
	<i>HSR1</i>	<i>HSR1</i> can interact with eEF1A, forming a <i>HSR1</i> -eEF1A complex, which can capture and activate the transcription factor HSF1, resulting in the transcription of <i>Hsp</i> and expression of HSPs in response to heat and other stress stimuli.	[155]
Splicing	<i>MALAT1</i>	<i>MALAT1</i> regulates alternative splicing by controlling the phosphorylation and distribution of serine/arginine splicing factors in nuclear speckle domains.	[156]
	<i>ASCO-lncRNA</i>	The <i>ASCO</i> -lncRNA is a nuclear alternative splicing regulator and influences the splicing patterns through binding with nuclear speckle RBP during development in <i>Arabidopsis</i> .	[157]
Translation	Antisense <i>Uchl1</i>	The <i>Uchl1</i> mRNA is complemented by an antisense lncRNA <i>Uchl1</i> , which is shuttled from nucleus to the cytoplasm under stress condition, to increase UCHL1 protein synthesis.	[158]
	<i>HULC</i>	<i>HULC</i> is upregulated in hepatocellular carcinoma, which can bind to miR-372 and downregulates its activity, leading to reduced translational suppression of its target transcript PRKACB.	[159]
RNA localization	<i>Xist</i>	A-repeat within the lnc <i>Xist</i> contains two long stem loop structures, which can recruit PRC2, while C-repeat binds YY1 transcription factor assisting <i>Xist</i> -PRC2 complex in targeting the specific sites on X-inactivation center, then lead to X-linked gene silencing.	[90,160]
	<i>ENOD40</i>	A novel nuclear speckle RBP, MTRBP1 ( <i>Medicago truncatula</i> RNA binding protein 1) can be transported into cytoplasmic granules during nodule organogenesis by interacting with <i>ENOD40</i> in the leguminous plants.	[161,162]
RNA decay	<i>1/2-sbsRNAs</i>	Alu elements within cytoplasmic lncRNA (1/2-sbsRNAs) can form imperfect complementary RNA duplexes with another Alu elements in the 3' untranslated regions (UTRs) of mRNAs, then STAU1 protein subsequently recognizes and binds the resultant dsRNA elements and initiates target mRNA degradation.	[107]
	<i>gadd7</i>	<i>gadd7</i> can regulate the cell cycle G1/S checkpoint in response to UV irradiation. UV-induced <i>gadd7</i> can directly bind to TAR DNA-binding protein (TDP-43) and interfere with the interaction between TDP43 and cyclin-dependent kinase 6 ( <i>Cdk6</i> ) mRNA, resulting in <i>Cdk6</i> mRNA degradation.	[163]
RNA editing	<i>CTN-RNA</i>	The 3' UTR of <i>CTN-RNA</i> contains inverted repeat sequences that can form stem loop recognized by ADAR enzyme for adenosine-to-inosine editing, then the edited RNA interacts with p54nrb, promoting its nuclear retention. This nuclear retention can be involved in the regulation of <i>mCAT2</i> gene expression.	[164]
	<i>sas-10</i>	The <i>sas-10</i> transcripts pair with <i>4f-mp</i> mRNA to form double-stranded molecules as target for A-to-G editing by	[165]

## Glossary

**Bivalent RNA-DNA linker:** a linker that can ligate RNA to proximal DNA.

**Click reaction:** a chemical reaction that is selective, high yielding, and simple to perform.

**Crosslinking and immunoprecipitation (CLIP):** a method that couples UV crosslinking with immunoprecipitation to identify transcripts that interacted with a specific protein.

**Cross-species control:** a control assay performed using samples from two species. For example, MARIO was used in *Drosophila* S2 cells and mouse ES cells to test the extent of random ligation of RNA molecules.

**Fragmentation sequencing (Frag-seq):** a method that couples RNase P-mediated cleavage with HTS.

**G quadruplex:** a nucleic acid secondary structure formed by a G-rich sequence that can self-assemble into two or more G-quartet planes, which then stack on top of each other.

**Parallel analysis of RNA structure (PARS):** a method that couples RNase V1- or RNase S1-mediated cleavage with HTS.

**Proximity ligation:** a ligation of two physically proximal nucleic acid termini.

**Ribonucleoprotein particle:** a biomolecular complex that consists of RNA and RBPs.

**Riboswitch:** an RNA molecule that can sense small ligands, such as metabolites or ions, and induce RNA conformational changes to affect gene expression.

**Ribozyme:** an RNA molecule that can act as an enzyme and catalyze reactions, such as RNA ligation or cleavage.

**RNA interactome:** a term to describe the interactions of all transcripts in the transcriptome, such as RNA-RNA, RNA-DNA, RNA-protein, and others.

**RNA immunoprecipitation sequencing (RIP-seq):** a method that couples native RNA immunoprecipitations with HTS to identify transcripts that interacted with a specific protein.

**RNA structurome:** a term to describe the structures of all transcripts in the transcriptome.

**Selective 2'hydroxyl acylation analyzed by primer extension**

Table 1. (continued)

Biological process	lncRNA example	Role in gene expression and RNA metabolism	Refs
		dADAR editase in the 3' UTR, leading to downregulation of <i>4f-mp</i> mRNA levels.	
Epigenetic remodeling	<i>pRNA</i>	<i>pRNA</i> is a lncRNA that is complementary to the rDNA promoter, which can interact with the target site of the transcription factor TTF-I, forming a DNA: RNA triplex that is specifically recognized by the DNA methyltransferase DNMT3b, then mediate <i>de novo</i> CpG of rRNA gene to repress its expression.	[166]
	<i>HOTAIR</i>	<i>HOTAIR</i> forms multiple double stem-loop structures that bind to PRC2 histone-modification complexes and lysine-specific demethylase 1, mediating different pattern of histone modifications on target genes related to cancer diseases.	[167]

versatile functions, as exemplified by **ribozymes**, **riboswitches**, and **ribonucleoprotein complexes** (see [Glossary](#) [19,20]). RNA adopts diverse structural motifs, such as stem-loop, pseudoknot, triplex, **G-quadruplex**, and is capable of long-range interactions, contributing to its basic biological functions [21]. These structural elements can form through *cis* (intramolecular) interactions within the same RNA molecule, or through *trans* (intermolecular) interactions with other biomolecules such as RNA, DNA, and proteins, to regulate fundamental cellular processes. Identifying RNA structures and interactions that are involved in gene regulation and function is thus critical for the elucidation of the underlying biochemical mechanisms. In addition, major efforts have been dedicated to predicting the impact of SNPs on lncRNA secondary structures and lncRNA–miRNA interactions, especially with respect to their recently uncovered mechanistic roles in various diseases [22–24]. To facilitate these efforts, there is a need to experimentally obtain lncRNA structures and interactions *in vivo* across diverse disease and cancer models. Combining these experimental data with a robust computational pipeline will likely generate more accurate candidates of functional, disease-related lncRNA SNPs for further mechanistic characterization and potential therapeutic intervention.

One of the main mysteries of lncRNAs is the discrepancy between their low sequence conservation and functionally important roles. Thus, many studies have been dedicated to the search for conserved structural elements [25,26]. For example, a large number of correlated positions in lncRNA were revealed by multiple alignments, suggesting evolutionary conservation of lncRNA secondary structures [27]. Additionally, many conserved structure elements were found to be enriched in lncRNAs by screening for functional RNA structures conserved between mice and 59 other vertebrates [26]. However, all this evidence is from computational predictions. It remains to be seen how many of these predicted conserved structures are real and functionally important *in vivo*.

Another mystery of lncRNAs is their association with the ribosome and potential for encoding peptides/proteins. Studies have shown that many lncRNAs in the cytosol are bound by ribosomes [28–30]. While a number of studies have suggested that these ribosome-bound lncRNAs do not yield peptide/protein products [31,32], implying that the function of these lncRNAs is at the RNA level, others have suggested that some lncRNAs are likely to generate peptide/proteins [33,34]. It is still not entirely clear how and why ribosomes and translation regulators recognize and interact with lncRNAs, and possibly lead to productive peptide/protein synthesis.

**(SHAPE):** a technique that uses an acylating agent, such as 1M7 and NAI, to react with flexible 2'OH groups of RNA, followed by primer extension reaction for readout.

**Shotgun secondary structure (3S) approach:** a technique that breaks down long RNA into smaller fragments for structure probing, and allows the identification of RNA structural domains.

**Single-stranded/double-stranded RNA sequencing (ss/dsRNA-seq):** a method that couples cleavage by ss/dsRNA ribonucleases with HTS.

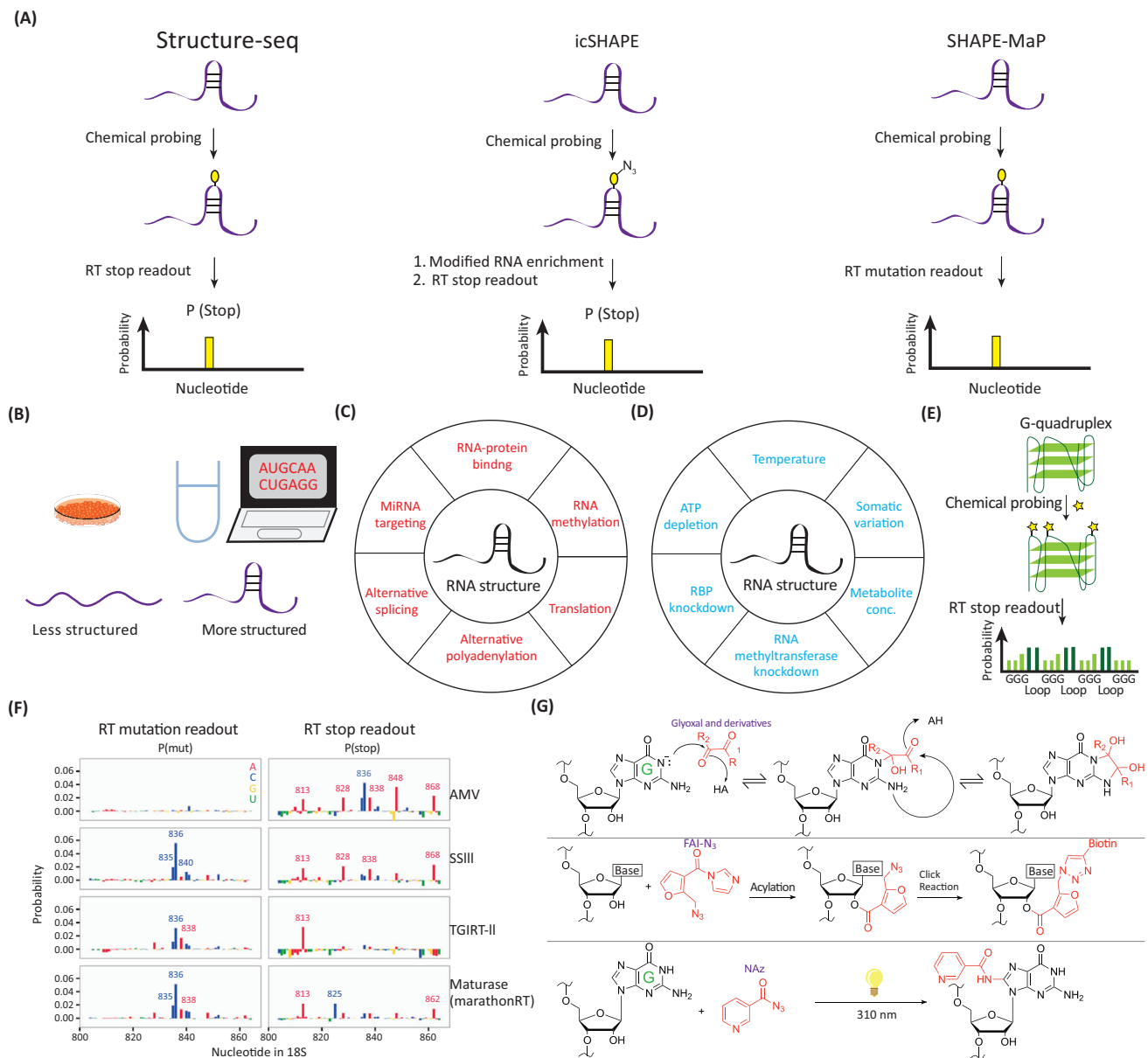
**Structural probing of elongating transcripts sequencing (SPET-seq):** a method that couples treatment with fast-reacting DMS probe and HTS to determine RNA secondary structures of transcription intermediates.

Despite the importance of structural information for the understanding of lncRNA functions, our knowledge of lncRNA structures is limited. According to the PDB database [35], the only items that contain tertiary RNA structure information are the classical ncRNAs such as rRNAs, tRNAs, and small nuclear RNAs. The flexibility and the relatively large size of lncRNAs have made their structures difficult to be resolved by traditional 3D structural determination methods. The computational approach is a useful alternative to predict RNA secondary structure and interactions between two RNAs [36–39]. Classically, the default module of most computational methods predicts the most thermodynamically stable structure of an isolated RNA molecule using minimum free energy approach, with accuracy of prediction decreasing for longer and more complex RNAs [37]. Recent approaches have suggested that the real biological structure is more likely to be found when considering a Boltzmann ensemble of suboptimal folding states [40]. Parallel to computation approaches, experimental enzymatic and chemical RNA probing methods were developed to analyze the structure of individual RNA transcripts [41]. Nevertheless, in contrast to the classical RNAs such as rRNAs and tRNAs, the structure and interaction of lncRNAs and other RNA types remained elusive until recently. In the last few years, HTS methodology development has allowed us to discover and appreciate the elaborate structure and interaction landscape on a transcriptome-wide scale, and this progress is discussed below.

### Recent Advances in RNA Structure Probing

Methods for probing **RNA structuromes**, which couple HTS with ribonuclease cleavage or chemical probing, have facilitated the transcriptome-wide mapping of RNA structure [21,42]. The initial approaches such as **parallel analysis of RNA structure (PARS)** [43], **fragmentation sequencing(Frag-seq)** [44], **single-stranded/double-stranded RNA sequencing (ss/dsRNA-seq)** [45,46] and **selective 2'hydroxyl acylation analyzed by primer extension(SHAPE)-seq** [47] were only able to determine transcriptome-wide RNA secondary structures *in vitro*. Subsequent development of a new generation of methods, including Structure-seq [48], DMS-seq [49], Mod-seq [50], SHAPE-MaP [51], and icSHAPE [52], enabled *in vivo* transcriptome-wide RNA structure probing, thus providing a better understanding and support of RNA functions in cellular environment. The *in vivo* methods can be broadly classified into three categories: (i) chemical probing – reverse transcriptase (RT) stop readout, such as Structure-seq; (ii) chemical probing – modified RNA enrichment – RT stop readout, such as icSHAPE; and (iii) chemical probing – RT mutation readout, such as SHAPE-MaP (Figure 1A). Detailed experimental and bioinformatic pipelines of these methods have been summarized elsewhere [42,53–55].

To date, studies of RNA structuromes have revealed several important findings that warrant further in-depth mechanistic and functional investigation. First, comparison of *in vivo*- versus *in vitro/in silico*-obtained RNA structures has revealed that RNA, including lncRNAs, is less structured *in vivo* than is observed *in vitro*/predicted by *in silico* analyses, across multiple species [48,49,52,56,57] (Figure 1B). The finding highlights that cellular factors, such as RNA-binding proteins (RBPs), RNA helicases, and ribosomes, are important contributors in regulating the RNA structures in cells [56–59]. Second, distinct RNA structural signatures have been identified and associated with translation processes, alternative splicing and polyadenylation, RNA–protein binding, miRNA targeting, RNA modifications, and other events [46,48,52,60–63], suggesting the significance of RNA structure in gene regulation (Figure 1C). Notably, RNA structural reactivity has also been correlated with protein domain folding [64,65], linking RNA structure with protein domain structure and opening up the possibility of coevolution of protein and RNA structures. Third, comparative studies have revealed that RNA structure changes under different reaction conditions, such as temperature, somatic variation, ATP depletion,



Trends in Biochemical Sciences

**Figure 1. Recent Developments in RNA Structure Probing by the Application of Chemical Probes and Sequencing Methods.** (A) Three categories of RNA structure sequencing methods, with representative examples. Briefly, for Structure-seq, chemical probing was performed, followed by detection using RT stop readout approach to detect reactivity at each RNA nucleotide. For icSHAPE, chemical probing was performed, followed by modified RNA enrichment and RT stop readout approach. For SHAPE-MaP, chemical probing was performed, followed by detection using RT mutation readout approach. (B) Structurome studies suggest that RNAs are less structured in cells compared to those observed *in vitro* or predicted *in silico*. (C) Distinct RNA structural profiles were identified and associated with RNA-protein binding, RNA modifications, translation process, alternative polyadenylation, alternative splicing, and miRNA targeting. (D) Genetic, epigenetic, and environmental factors can impact the structure of RNA. (E) *In vitro* SHALiPE analysis on canonical rG4s reveals a distinct modification pattern. (F) The effect of different readout approaches and RTs on a region of 18S rRNA. Adapted from [71]. (G) Chemical structure and mechanism of recently developed RNA structure probes, for example, glyoxal and its derivatives, 2-methyl-3-furoic acid imidazole azide, and nicotinoyl azide. The structures and nucleotide specificities of classical probes, such as dimethyl sulfate, methylnicotinic acid imidazole, 1-methyl-7-nitrosatoic anhydride, and others, can be found in earlier reviews [42,54]. Abbreviation: RT, reverse transcriptase.

RBP knockdown, RNA methyltransferase knockdown, or metabolite concentration [49,52,61,66–69]. These interesting results suggest the impact of genetic/epigenetic variation and environment on RNA structures and functions (Figure 1D). These studies were mostly centered on mRNAs, and thus future research with the focus on lncRNAs will likely expand the multifaceted biological roles of RNA structure.

The deeper exploration of RNA-structure-probing methods and analyses has encouraged remarkable technological advances. First, new strategies have been used to study higher-order RNA structures such as pseudoknot and G-quadruplex. It has been shown that SHAPE reactivity generated from SHAPE-MaP could be used to verify known and identify new pseudoknots [51]. Using SHALiPE, a specific chemical-induced modification pattern was obtained for RNA G quadruplexes *in vitro* in several transcripts [70] (Figure 1E). Second, analyses of 18S rRNA data using the RT-stop versus RT-mutation approaches have indicated distinct structural features on the same RNA, whereby the results were also dependent on the choice of RT enzymes, indicating the complementary nature of these two methods and the need to further examine the properties of RT enzymes [71] (Figure 1F). Third, new types of RNA structure-probing chemicals were developed. For example, glyoxal and its derivatives were used to target the guanine nucleotides at the Watson–Crick face *in vivo* [72]. Another bifunctional intracellular SHAPE reagent, 2-methyl-3-furoic acid imidazolide azide (FAI-N3 or FAz), was shown to form a more stable adduct and longer reactive lifetime than 2-methylnicotinic acid imidazolide azide (NAI-N3), which allows better control of the experimental reaction by dithiothreitol quenching [73]. Moreover, nicotinoyl azide (NAz) was designed as a light-activated, fast-reacting reagent that can measure solvent accessibility of purine nucleobases upon light irradiation, and supports the detection of RNA–protein interactions and intracellular RNA structures [74] (Figure 1G). Together with the commonly used dimethyl sulfate (DMS), and classical SHAPE reagents such as 1-methyl-7-nitroisatoic anhydride (1M7) and 2-methylnicotinic acid imidazolide (NAI) [42,54], these chemicals and probing strategies often generate complementary structural information, and can be directly applied to study lncRNA structure both *in vitro* and *in vivo*.

The sequencing-based RNA structure profiling technologies rely heavily on the development of robust computational analysis method. However, several key challenges remain to be addressed in this field [54,55,75], and development in this area is ongoing. Recently, a statistical machine learning pipeline used a beta-uniform mixture hidden Markov model to analyze structure probing data, and was able to obtain structural information with high sensitivity even at low sequencing coverage [76]. Another study, PROBer, was able to identify isoform-specific chemical modification profiles [77]. In addition to methods that generate full-length structure profiles or models, algorithms that mine short RNA structure elements directly from the sequencing data are also of interest in the field. PATTERNNA used a pattern recognition machine-learning algorithm to detect RNA structure motifs, and was shown to achieve a high accuracy comparable to that of thermodynamic models [78,79].

### RNA Structure Probing Applied to lncRNAs

With the development of the RNA-structure-probing techniques, secondary structures of a number of important lncRNAs were recently uncovered (Table 2). For instance, several independent studies used a similar approach to investigate lncRNA structure *in vitro* [80–85]. In these studies, RNA transcripts were first generated *in vitro*, followed by either denaturation or native purification of the desired RNA fragments. Typically, native purification should allow physiologically relevant lncRNA structures to be preserved. Next, structure probing by means of chemical modification, enzymatic cleavage or UV crosslinking was carried

Table 2. Representative lncRNAs Studied by Structure Probing and/or Crosslinking Strategy

lncRNA studied (length)	Method used	Species and system	Structural and biological features	Refs
<i>SRA</i> (874 nt)	–SHAPE, in-line, DMS, RNase V1 probing –3S approach –Capillary electrophoresis	Human; <i>in vitro</i> (RNA heated and renatured being probing)	–Four structural domains, consisting of 25 helices, 16 terminal loops, 15 internal loops, and 5 junction regions –Domains I–III are more evolutionarily conserved than Domain IV3	[80]
<i>HOTAIR</i> (2148 nt)	–SHAPE, DMS, and terbium probing –3S approach –Capillary electrophoresis	Human; <i>in vitro</i> (RNA natively folded before probing)	–Four structural domains, consisting of 56 helices, 38 terminal loops, 34 internal loops, and 19 junction regions –Many of the structural elements are evolutionarily conserved	[81]
	–ChIRP	Human MDA-MB-231 breast cancer cells	–GA-rich polypurine motif of <i>HOTAIR</i> –Binding sites of <i>HOTAIR</i> are related to more broad domains of PRC2 occupancy and H3K27me3	[114]
<i>COOLAIR</i> (~750 nt)	–SHAPE, CMCT probing –3S approach –Capillary electrophoresis	<i>Arabidopsis thaliana</i> ; <i>in vitro</i> (RNA heated and renatured being probing)	–Three major structural domains were observed: 5' domain, 3' major domain and 3' minor domain –Distal COOLAIR lncRNA consists of 12 helices, 7 stem loops, a 3-way junction, a 5-way junction, and 2 rare r-turns –The structures show conservation and covariation across several Brassicaceae species	[82]
	–ChIRP	<i>Arabidopsis thaliana</i>	–COOLAIR enriched in the nucleation region and the 3' region of the gene FLC	[168]
<i>Braveheart</i> (~590 nt)	–SHAPE and DMS probing –3S approach –Capillary electrophoresis	Mouse; <i>in vitro</i> (RNA natively folded before probing)	–Three structural domains involved: 5' domain, central domain, and 3' domain, consisting 12 helices, 8 terminal loops, 5 internal loops, a 5-way junction. –An internal G-rich RNA motif (AGIL) interacts with CNBP	[83]
<i>lncRNA-p21</i> (2882 nt, short isoform; 3898 nt, long isoform)	–SHAPE probing –Capillary electrophoresis	Human; <i>in vitro</i> (RNA natively folded before probing)	–Contains inverted repeat <i>Alu</i> elements –Left and right arm of <i>Alu</i> are linked by single-stranded RNA region –Each arm consist of a central 3-way junction, followed by a long stem loop – <i>Alu</i> elements are conserved among primates and embedded in lncRNA-p21	[169]
<i>REPA</i> (1630 nt)	–SHAPE, DMS, UV crosslinking probing –3S approach –Capillary electrophoresis	Mouse; <i>in vitro</i> (RNA natively folded before probing)	–Contains 3 structural domains linked by a central junction –Domain I exhibits the highest dynamics among the three domains, while Domains II and III are more stable in their structures –Domains I and II are generally highly conserved across species in their general structures, but the sequence of Domain III is poorly conserved in mammals.	[84]

Table 2. (continued)

lncRNA studied (length)	Method used	Species and system	Structural and biological features	Refs
<i>NEAT1</i> (shorter isoform: <i>hNEAT1_S</i> , 3735 nt; <i>mNEAT1_S</i> , 3176 nt)	–SHAPE probing –3S approach –Mod-seq sequencing	Human and mouse; <i>in vitro</i> (RNA natively folded before probing)	–Formed with four structural domains –RNA–RNA interactions between the 5′ end and 3′ ends of <i>NEAT1_L</i> , which are close to TARDBP binding sites	[85]
	–CHART	Human MCF7 cells; <i>in situ</i> (crosslinked nuclear extract)	–Interacts with hundreds of genomic sites –Colocalize with <i>MALAT1</i> lncRNA –Interacts with many overlapping proteins with <i>MALAT1</i> , but each lncRNA also has its unique protein sets	[170]
<i>MALAT1</i> (~8300 nt)	–SHAPE and DMS probing –Capillary electrophoresis	Humans, lizards, zebrafish; <i>in vitro</i> (RNA heated and renatured being probing)	–Triple-helix and t-RNA-like structure at the 3′ end of <i>MALAT1</i> –The triplex helix contains base triplets (U:A:U or C:G:C), and stem loops with purine-rich loops	[171]
	–RAP	Mouse ESs; <i>in situ</i> (crosslinked nuclear extract)	–Interacts with thousands of genomic sites –Indirectly interacts with nascent pre-mRNAs through proteins	[109]
	–CHART	Human MCF7 cells; <i>in situ</i> (crosslinked nuclear extract)	–Interacts with hundreds of genomic sites –Colocalizes with <i>NEAT1</i> lncRNA –Interacts with many overlapping proteins with <i>NEAT1</i> , but each lncRNA also has its unique protein sets	[170]
<i>roX</i> ( <i>roX1</i> : ~3700 nt; <i>roX2</i> : ~600 nt)	–SHAPE probing, followed by denaturing PAGE gel –PARS sequencing	<i>Drosophila melanogaster</i> ; <i>in vitro</i> (RNA heated and renatured being probing)	–3′-terminal domain of <i>roX1</i> consists of 3 helices linked by a flexible linker –Global structure of <i>roX2</i> consists of 8 stem loops. –For both <i>roX1</i> and <i>roX2</i> , the stem regions (but not the loops or linkers) are highly conserved	[172]
	–dChIRP	<i>Drosophila melanogaster</i> ; <i>in situ</i> (crosslinked nuclear extract)	–The 3 U domains of <i>roX1</i> are associated in 3D space, whereas the 3D domains of <i>roX1</i> are spatially distant from each other – <i>roX1</i> interacts with hundreds of genomic sites, with majority in X chromosome – <i>roX1</i> D domains interacts with MSL complex	[127]



out, followed by capillary electrophoresis or HTS to infer the secondary structure and tertiary base pair interactions. The experimental constraints derived from the structure probing were combined with comparative sequence analysis to generate a consensus structure model. In some studies, a strategy termed **Shotgun Secondary Structure (3S)** determination was used, in which the full-length lncRNA was divided into shorter fragments that were independently subjected to structural probing to identify major structural domains. The results of these studies suggested that lncRNAs are generally folded into multiple well-defined structural domains that are important for their functions (e.g., protein binding) (Table 2). It is particularly interesting to note that while it is generally speculated that lncRNAs may contain abundant conserved structural elements and modules with covarying base pairs, studies also suggest that it needs more significant covariation support [86–88].

A lncRNA of exceptional interest, *Xist*, is an ~18-kb nuclear lncRNA that is essential for X-chromosome inactivation in mammalian cells [89], and has been extensively characterized by multiple structural studies (Table 2) [90–93]. The structure of the conserved A-repeat region (~400 nt, 8.5 and 7.5 repeats in human and mouse, respectively) at the 5' end of *Xist* RNA has been given special attention due to its importance in recruiting the PRC2 Polycomb group complexes to the inactive X chromosome [94]. While prior structural models suggested that each repeat base pairs within itself (intra-repeat interaction) [95] or other repeats (inter-repeat interaction) exclusively [90], *in vivo* targeted Structure-seq analysis revealed a combination of both types of structures [91]. In combination with PARIS (discussed in later sections), icSHAPE probing of human *Xist* RNA *in vivo* showed that inter-repeat structures are more prevalent *in vivo* and showed that the A repeat is mostly folded as an isolated domain and does not form base pairs with distant regions [93]. However, SHAPE-MaP of mouse *Xist* indicated that the A-repeat region exhibits large structural variability, and likely interacts with other segments of the *Xist* RNA [92]. It is possible that the discrepancies of these results are due to the structural dynamics of the *Xist* RNA, the variation in cellular environment or reaction conditions, as well as different experimental and bioinformatic methods used. Nevertheless, these studies revealed *in vivo* lncRNA structure for the first time and provided invaluable insights. Future efforts may focus on dissecting the *in vivo* lncRNA structures across the transcriptome by optimizing the existing methods for large scale analysis.

As discussed below, *Xist* has also been one of the key lncRNA targets in the studies of RNA–RNA, RNA–chromatin, and RNA–protein interactions.

### Strategies for Probing RNA–RNA Interactions and Their Application to lncRNAs

RNA functions are often governed by interacting with other molecules, including RNA, DNA, and proteins. Recently, we have witnessed breakthroughs in all the three aspects, allowing us to reveal the full spectrum of **RNA interactomes**. In the following sections, strategies and methods for transcript-specific and transcriptome-wide detection of RNA–RNA, RNA–DNA, and RNA–protein interactions are described and summarized in Table 3.

RNA–RNA interactions (RRIs) can be generally classified into two groups: interactions mediated by proteins, or those effected by direct RNA base pairing. Accordingly, methods to decode RRIs can also be classified into two corresponding categories. The first category includes protein pull-down-dependent methods such as CLASH [96], hiCLIP [97], and MARIO [98], and the second category includes direct RRI detection methods such as PARIS [93], SPLASH [99], and LiGR-Seq [100] (Table 3).

Table 3. *In vivo* Approaches to Detect RNA Structure and Interaction

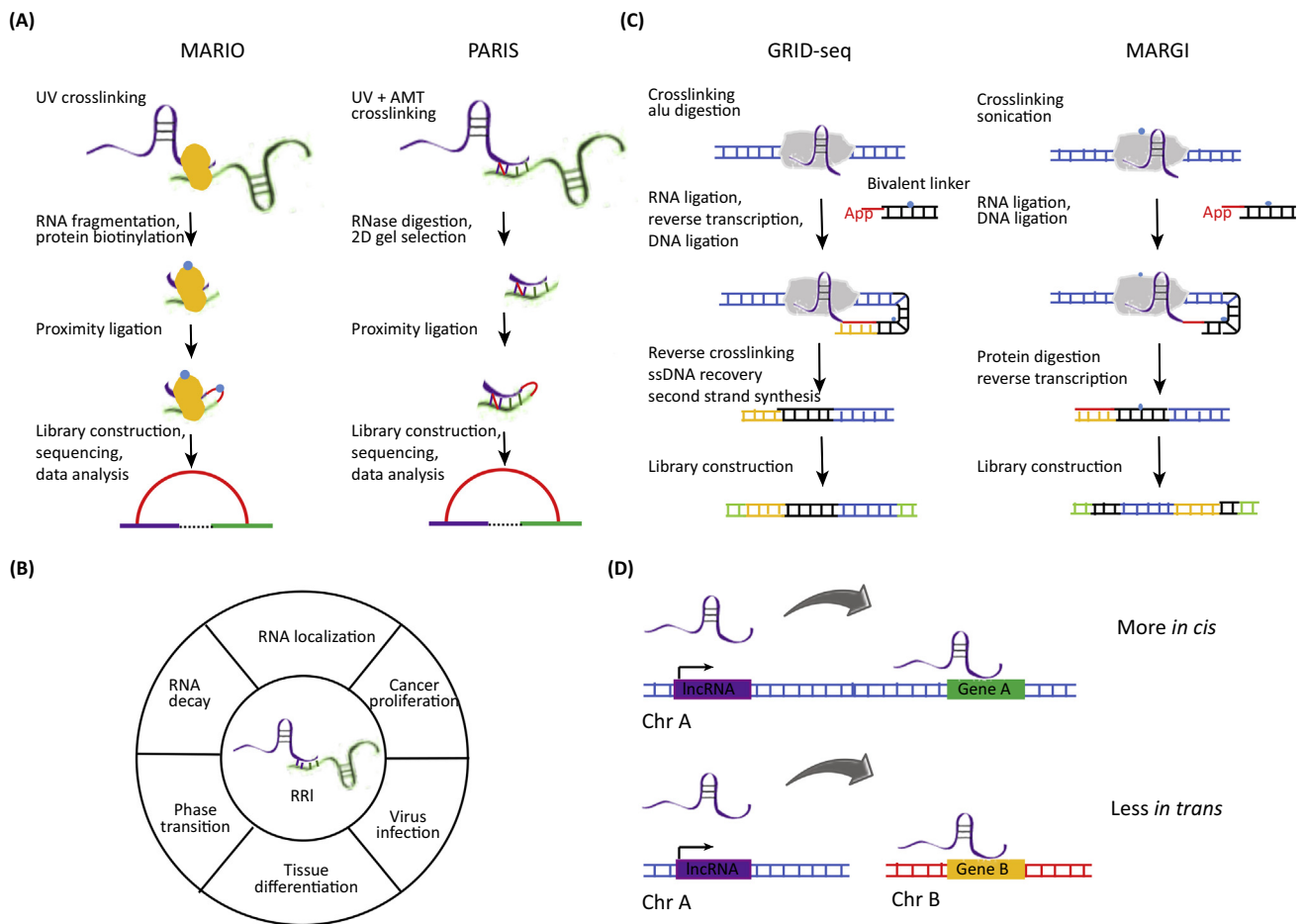
Methods	<i>In vivo</i> systems applied to date	Features	Limitations	Refs
<i>Chemical probing – RT stop readout</i>				
Structure-seq DMS-seq Mod-seq In cell SHAPE-seq	<i>Arabidopsis thaliana</i> , yeast, human, mouse, <i>E. coli</i> , cucumber mosaic virus, rice, zebrafish	–Applied to diverse species –Simple library preparation –Ensemble measurement	–No direct base pairing information	[48–50,56,58,59,173–177]
<i>Chemical probing – modified RNA enrichment – RT stop readout</i>				
icSHAPE	Mouse, human	–Lower background due to modified RNA enrichment –Ensemble measurement	–More steps involved in library preparation –No direct base pairing information	[52,93]
<i>Chemical probing – RT mutation readout</i>				
SHAPE-MaP DMS-MaPseq	<i>E. coli</i> , HIV, <i>Drosophila</i> , yeast, human, influenza A	–Direct base pairing information –Single molecule measurement	–High background due to low rate of mutation –The RT mutation readout site is often distinct from RT stop site	[51,57,62,92,178,179]
<i>RNA–RNA interaction – protein mediated</i>				
CLASH hiCLIP MARIO	Mouse, human, yeast, <i>Drosophila</i>	–Inter-molecules and intra- molecules RRIs information	–Protein-mediated RRIs –Limited crosslinking and ligation efficiency	[96,97]
<i>RNA–RNA interaction – direct base pairing</i>				
PARIS SPLASH LIGR-seq	Mouse, human	–Near base solution –Direct base-paired RRIs –Long-range RRIs	–Psoralen or AMT crosslinking –Low mapping rate of duplex reads	[93,99,100]
<i>RNA–DNA interaction – specific target – probe hybridization capture</i>				
CHART ChIRP RAP dChIRP	Mouse, human, <i>Drosophila</i>	–Effectively enrich target lncRNA	–Biotinylated probes needed –Specific lncRNA	[113,114,127]
<i>RNA–DNA interaction – Specific target – RNA–DNA adenine methylase identification</i>				
RNA-DamID	<i>Drosophila</i>	–High accuracy and sensitivity –Cell-type-specific interactions	–Methylate adenine residues in the sequence GATC –Plasmid construction and transfection needed	[116]
<i>RNA–DNA interaction – transcriptome – proximity ligation</i>				
MARGI GRID-seq ChAR-seq	Mouse, human, <i>Drosophila</i>	–Proximity ligation –Cross-species control	–Limited ligation efficiency	[117–119]
<i>RNA–DNA interaction – transcriptome – split pool</i>				
SPRITE	Human, mouse	–Proximity ligation independent –Cross-species control –DNA–DNA interaction included	–Interactions that occur across larger spatial distances	[120]
<i>RNA–protein interaction – specific target – probe hybridization capture</i>				
CHART-MS ChIRP-MS RAP-MS dChIRP-MS	Mouse, human, <i>Drosophila</i>	–Effectively enrich target lncRNA	–Biotinylated probes needed –Specific lncRNA	[125–127]

Table 3. (continued)

Methods	<i>In vivo</i> systems applied to date	Features	Limitations	Refs
<i>RNA–protein interaction – specific target – RBP biotinylation</i>				
RaPID	Mouse, human, Zika virus	–RBPs binding to target motif or region –RBPs binding to mutated motif or region	–Plasmid construction and transfection needed –Dependent on specificity of RBP biotinylation	[128]
<i>RNA–protein interaction – transcriptome</i>				
RICK RBR-ID CARIC	Mouse, human	–Proteins binding to newly transcribed RNA –For RICK and CARIC, click reaction is needed –For RBR-ID, RNA-binding peptide information can be detected	–4SU or EU label –No direct RNA–protein pair information	[133–135]

The CLASH method represents one of the first efforts to detect RRIs by crosslinking interacting RNAs with an RBP of interest. Early development of CLASH focused on describing small nucleolar (sno)RNA–rRNA interactions by small nucleolar ribonucleoprotein copurification. Subsequently, the technique was applied to characterize different types of RRIs, such as miRNA–mRNA interactions via Argonaute protein copurification [96,101]. The basis of hiCLIP is similar, but this method achieves higher specificity by adding a ligation linker to allow a more efficient ligation reaction and thus unambiguous hybrid mapping. Both CLASH and hiCLIP rely on protein pull-down, and therefore in principle identify a class of RRIs involving a certain RBP of interest. Caution should be taken as immunoprecipitation of some RNA duplexes with the target RBP may result from promiscuous ligations, even with application of high salt washing or nitrocellulose membrane transfer in CLASH or hiCLIP, respectively. Considering the low expression level of lncRNA, the limited number of high-affinity RNA duplex binding proteins, and the interaction instability without direct RNA–RNA crosslinking, both methods can reveal only limited information on RRIs of lncRNAs. Recently, MARIO was developed to enrich RRIs mediated by the whole proteome through protein biotinylation and pull-down (Figure 2A). In this approach, three control experiments are performed to reduce background noise and random ligations, including a non-cross-linking control, a nonbiotinylation control and a **cross-species control**. The use of strict control assays greatly expands the type and number of discoverable RRIs, which include RRIs between lncRNAs and, for example, mRNAs and snoRNAs.

The second category of RRI discovery methods identifies interactions by exploiting the ability of certain small molecules, for example, psoralen or its analogs, to directly crosslink nucleotide base pairing. The main distinguishing features of these methods (Table 3: PARIS, SPLASH, and LIGR-seq) reside in the separation and purification of RNA duplexes. In PARIS (Figure 2A), 2D gel separation is used to select duplex regions before **proximity ligation**. In SPLASH, biotinylated psoralen can be used to allow capturing crosslinked duplexes with streptavidin beads. And in LIGR-seq, RNA ligase CircLigase performs proximity ligation and converts crosslinked RNAs into circular form, followed by RNase-mediated digestion to remove unpaired regions. The details of these pipelines are described elsewhere [54,55]. All the three methods can potentially generate RNA–RNA interactome maps including lncRNA interactions. However, given that psoralen can only crosslink uridines with limited efficiency, current methods only revealed a limited number of lncRNA interactions *in vivo*.



Trends in Biochemical Sciences

**Figure 2. Recent Developments in the Detection of RNA–RNA and RNA–chromatin interactions.** (A) Two categories of methods for transcriptome-wide RRI detection, with representative examples. MARIO detects protein-mediated RRIs, and the procedures include RNA crosslinking by UV, RNA fragmentation, RNA-binding protein biotinylation and capture, proximity ligation, library construction, HTS and bioinformatic analysis. PARIS detects RRIs using direct base pairing, and the procedures include RNA crosslinking by UV and AMT, digestion by RNase, RNA duplexes enrichment by 2D gel selection, proximity ligation, library construction, HTS, and bioinformatic analysis. (B) Functions of lncRNA-based RRIs, including RNA localization, RNA decay, RNA phase transition, cancer immunity, virus infection, and tissue differentiation. (C) Two examples of methods for transcriptome-wide RNA–DNA interaction detection. Both in GRID-seq and MARGI, biotin-labeled bivalent linker is used to ligate RNA to proximal DNA. The procedures of GRID-seq include UV crosslinking, fragmentation, RNA ligation, reverse transcription, DNA ligation, reverse crosslinking, library construction, and sequencing. The procedures of MARGI include crosslinking, fragmentation, RNA ligation, DNA ligation, reverse transcription, library construction, and sequencing. (D) Transcriptome-wide RNA–chromatin interaction studies show that *cis* lncRNA–chromatin interactions are more prevalent than *trans*. Abbreviations: HTS, high-throughput screening; lncRNA, long noncoding RNA; RRI, RNA–RNA interaction.

Notably, for all these technologies, bioinformatic pipelines play critical roles in revealing RRIs from the sequencing data. The key challenges are detecting base-pairing duplexes effectively and estimating the significance of each [102]. Recently, a new algorithm tool named Cross-Linked reads Analysis (CLAN) was developed to find and merge two nonoverlapping mappings with the largest read length and then select the merged mappings by a dynamic programming-based chaining algorithm [103]. The results showed that CLAN can achieve high computational efficiency and high sensitivity and accuracy of duplex detection at the same time. Meanwhile, to facilitate the exploration of RRIs of interest, a database called RISE has been created to gather rapidly accumulating RRI data in humans, mice, and yeast [104].

Previous studies have revealed diverse functions of lncRNA-based RRI in RNA localization, RNA decay, RNA phase transition, cancer immunity, virus infection, and tissue differentiation (Figure 2B) [105–110]. A particularly interesting type of RRI is endogenous sense–antisense interactions, as one-third of human protein-coding genes are overlapped by antisense lncRNAs in the same locus [27]. Natural antisense lncRNAs are transcribed from the opposite strand of protein-coding genes and perform diverse functional roles [111]. For example, lncRNAs termed half-sbsRNAs interact with their sense genes through imperfect base pairing of Alu element to mediate their decay [107]. Furthermore, lncRNA termed BACE-AS1 interacts with its sense gene *BACE1* to stabilize the target gene [112]. Overall, the newly developed RRI detection methods have expanded the available RRI information, and are expected to identify important lncRNA targets and to generate more insights into lncRNA functions.

### Strategies for Probing RNA–DNA Interactions and Their Application to lncRNAs

Nuclear lncRNAs can bind chromatin and regulate chromatin state and gene expression. Accordingly, methods have been developed to characterize RNA–chromatin interactions. Initially, such methods were focused on a particular RNA of interest, and included CHART [113], ChIRP [114], and RAP [115] (Table 3). Through these methods, much information has been gathered on chromatin binding sites of several well-known lncRNAs including *Xist*, *NEAT1*, and *MALAT1*. For example, ChIRP revealed that *HOTAIR* binding sites are enriched for genes of pattern specification processes, consistent with the fact that *HOTAIR* enforces the epigenomic state of distal and posterior positional identity. RAP showed that *Xist* binds broadly across the X chromosome during the maintenance of X-chromosome inactivation. Using CHART, *roX2* was shown to be able to bind MSL, a critical protein complex for dosage compensation in *Drosophila*. Different to antisense oligohybridization strategies, RNA–DamID combined the UAS–GAL4 control system and transgenic expression of fusions of DNA adenine methyltransferase to lncRNAs to map cell-type-specific lncRNA–chromatin interactions *in vivo* with high sensitivity and accuracy, and showed that the binding sites of *roX1* differ in neural stem cells from those in salivary glands [116].

All the above-mentioned RNA–chromatin interaction methods can only provide information of interacting DNA location for a target lncRNA. Recently, three new proximity ligation dependent methods, GRID-seq [117], MARGI [118], and ChAR-seq [119], were developed to profile the map of all RNA–DNA interactions in cells (Figure 2C). The key innovation of these methods is the design of a **bivalent RNA–DNA linker** that ligates RNA to proximal DNA *in situ* in fixed nuclei. In order to remove random ligation noise, both methods used a mixture of *Drosophila* cells and human cells as input to construct a background of cross-species control for nonspecific RNA–chromatin interactions. The results showed that few RNA–DNA reads are mapped to the human and *Drosophila* genomes at the same time, indicating that ligations among different species are limited. GRID-seq further developed an endogenous background using trans-chromosomal mRNA–chromatin interactions for further normalization. Except for the proximity ligation strategy, SPRITE used several rounds of split-pool tagging to label crosslinked complexes with specific barcodes, which provide the information of genome wide RNA–DNA interaction, even for long-range interactions [120]. Global analysis of RNA–chromatin interactions showed that the majority of lncRNAs exhibit predominant local or *cis* interactions, while some specific ones, including *MALAT1*, *NEAT1*, and *roX2*, interact with the chromatin in a *trans* fashion (Figure 2D). Indeed, an increasing number of functional studies revealed that many lncRNAs function locally [121–124]. With more lncRNA–chromatin interactions being reported and validated, the elaborate mechanism of diverse functional formats of lncRNAs may be elucidated.

## Strategies for Probing RNA–Protein Interactions and Their Application to lncRNAs

Mass spectrometry (MS) can be used to identify the interacting proteins of an RNA, often in combination with target-specific antisense oligonucleotide pull-down. The relevant methods, such as CHART-MS, ChIRP-MS, and RAP-MS, have a similar underlying strategy consisting of crosslinking, biotinylated probe hybridization, beads-assisted purification, and detection of the binding proteins by MS (Table 3). This strategy provides a systematic way to discover the binding proteins of a target RNA. For example, ChIRP-MS analysis of *Xist* identified 81 interacting proteins, including hnRNPK, which is involved in *Xist*-mediated gene silencing [125]. RAP-MS analysis of *Xist* also revealed ten significant enriched binding proteins, including SHARP, which is involved in *Xist*-mediated recruitment of PRC2 across the X chromosome [126]. As a further development of ChIRP, the dChIRP approach [127] can identify binding proteins at specific regions by designing the corresponding targeting probes.

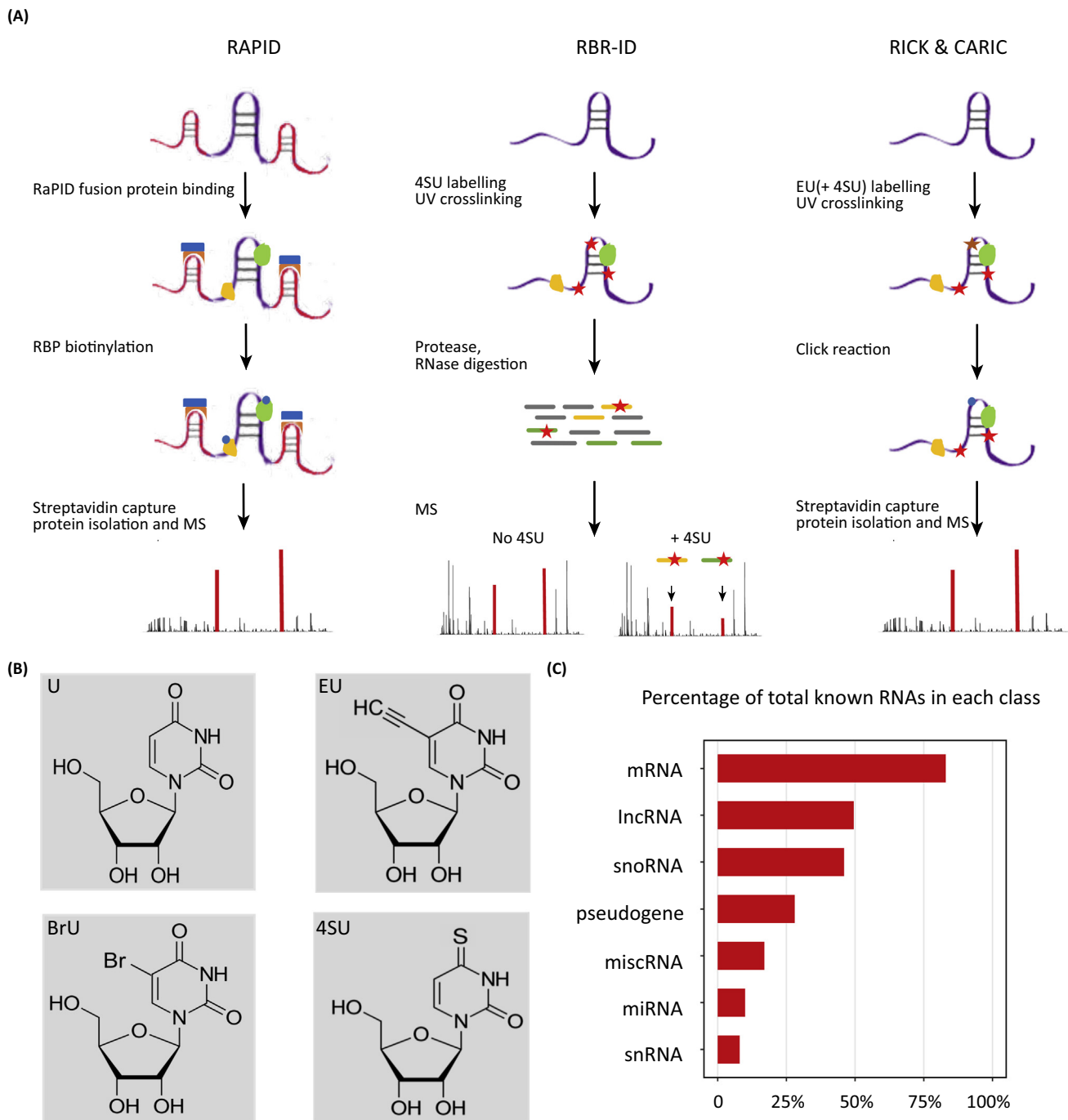
Recently, a new strategy termed RaPID [128] was designed to detect the binding proteins of target RNA motifs based on biotin ligase BirA (Figure 3A). The application of RaPID identified known and novel RBPs, including, for example, essential host proteins that interact with Zika virus RNA. It is particularly useful to reveal the binding proteins for mutated RNA motifs, which then can be used for studies in cancer or other human genetic disorders.

The techniques describe above studied the binding of proteome to certain individual RNAs, however, it should be note that several methods to capture the full proteome bound to a specific class of RNAs in cells have been developed. Among them, oligod(T) hybridization and capture was well established to study the mRNA-bound proteome [129,130]. The efficiency and specificity were later improved by the use of nucleoside analogs such as 4-thiouridine (4SU), a photoactivatable uridine analog, and ethynyluridine (EU), which can be labeled with biotin by a **Click reaction** [131,132] (Figure 3B). RBR-ID [133] and RICK [134] use 4SU and EU, respectively, to label RNAs and then identify the interacting proteins, while CARIC [135] uses both 4SU and EU. This shows that RICK can identify proteins interacting with newly transcribed RNAs, including many ncRNAs (Figure 3C). Upon application of RBR-ID, peptides crosslinked with 4SU-labeled RNA demonstrate a decreased intensity in MS analysis compared to those without crosslinking, indicating the detection of true RNA-binding peptides. However, while the methods described above have revealed many lncRNA-binding proteins, the exact associations of lncRNAs and RBPs remain unresolved.

Finally, in addition to the above-mentioned methods that capture RNA–protein interactions in an RNA-centric way, **RNA immunoprecipitation sequencing (RIP)** and **crosslinking and immunoprecipitation (CLIP)** are protein-centric methods that detect RNA targets of a certain protein of interest [136,137]. Combined with HTS, different CLIP-seq protocols including HITS-CLIP [138], iCLIP [139], PAR-CLIP [140], irCLIP [141], and eCLIP [142] have mapped diverse RNA–protein interactions in multiple species and cell lines. In addition to RNA–protein interactions, RIP and CLIP can also reveal RRIs. For example, a recent study mapped the AGO-miRNA and AGO-mRNA binding sites from the same HITS-CLIP experiments, and then identified miRNA–mRNA interactions based on a linear regression model [143]. Computational methods have been developed to detect miRNA–mRNA, miRNA–lncRNA, and other types of interactions from the CLIP data sets [144–147].

## Future Perspectives and Challenges

The number of lncRNAs being identified and functionally validated far exceeds the ones being structurally characterized, hindering the investigation and establishment of lncRNA structure–



Trends in Biochemical Sciences

**Figure 3. Recent Developments in the Detection of RNA–Protein Interactions.** (A) Three representative examples of RNA–protein interaction detection methods, all of which are based on MS. RaPID designs two constructs, one containing the target motif and two flanking BoxB stem loops, which binds the  $\lambda$ N peptide with high affinity, and the other one containing biotin ligase BirA fused with  $\lambda$ N peptide, which biotinylates proteins bound to the target motif. RBR-ID uses 4SU to label RNAs and then identifies RBPs with decreased intensity compared to that without the 4SU label. RICK and CARIC use EU and 4SU to label RNAs, capture the labeled RNA after the click reaction, and then identify RBPs. (B) Nucleoside analogs of uridine widely used in RNA–protein detection methods: EU, BrU, and 4SU. (C) Transcriptome-wide RNA–protein interaction detection methods reveal RBPs of diverse kinds of RNAs, including large amounts of lncRNAs. Adapted from [134]. Abbreviations: 4SU, 4-thiouridine; BrU, bromouridine; EU, ethynyluridine; miscRNA, miscellaneous RNA; MS, mass spectroscopy; RBP, RNA-binding protein; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA.

function relationship. Several major challenges remain and need to be addressed to gain full understanding of the biological importance of lncRNA structure and interactions (see Outstanding Questions).

First, although some lncRNAs such as MALAT1 and NEAT1 are highly expressed, many (or most) lncRNAs are expressed only weakly, and thus the information on lncRNAs obtained from transcriptome-wide studies is usually too limited to drive functional investigation of lncRNAs. One way to resolve this is to enrich target lncRNAs by antisense oligonucleotides, as exemplified by the CHART, CHIRP, and RAP protocols. Parameters such as probe length, numbers, and density are important to ensure high specificity and efficiency in lncRNA capture.

Second, RNA is dynamic and often adopts multiple structural conformations. Most existing methods measure the averaged structure of many copies of the same RNA molecule over the course of reaction. It is a major challenge to dissect individual conformations from a structure or an interactome ensemble. A few experimental approaches have now started to generate structural information at the single RNA molecule level [93,99,100,148]. However, the information obtained from these studies remains complex and entangled with conflicting alternative interpretations. In parallel to experimental approaches, computational methods are also in development to resolve this structural ensemble deconvolution issue [149–151]. Nevertheless, without prior knowledge, the *de novo* identification of individual, and sometimes subpopulated, RNA structural conformation or interactions, especially for lncRNAs *in vivo*, will require significant development in both experimental and bioinformatic directions.

Third, the biological roles of RNA, including lncRNAs, are likely to be linked to their functional RNA folds, either structured RNA motif and/or unstructured RNA sequence under particular settings such as their physical location (e.g., nucleus, cytoplasm, and mitochondria), timing of action (e.g., during transcription, cotranscription, and post-transcription), and specific cellular conditions (e.g., normal vs. stressed), or cell/tissue types (e.g., healthy vs. cancerous). Studies have started to characterize the dynamic patterns of RNA structure in different compartments during their life cycle [152] and across different developmental stages [59]. Several recent methods have been developed to investigate nascent RNA structure and nascent RNA–protein interactions, for example, **structural probing of elongating transcripts sequencing (SPET-seq)** [153] and RICK [134], and their findings are promising and should be applicable to lncRNA studies.

Finally, as mentioned in the Introduction, there is a huge knowledge gap in our understanding of whether lncRNA structures and interactions have significant roles in driving lncRNA evolution, controlling ribosome association, and regulating their own decay and translation. In addition, it will be interesting to examine if the generated transcriptome-wide lncRNA structure and interaction data can help reveal functional disease-related SNPs, and facilitate the determination of high-resolution 3D lncRNA structures. Addressing these long-standing questions in the lncRNA field will move us one step closer to establishing the much sought after lncRNA structure–function relationship.

### Concluding Remarks

The discovery of lncRNAs has revolutionized the way we understand the role of RNA in biology. With the current methods (Table 3), we are beginning to explore and appreciate the complexity and diversity of the RNA structure and interactions at an unprecedented pace and with increasingly high sensitivity and resolution. With optimism, we think that further improvement of the described methods or invention of new sequencing strategies will allow us to address the



experimental and bioinformatic noise, artifacts or biases of the existing approaches. Future applications of these strategies to lncRNAs will enable us to investigate the crucial regulatory roles of lncRNAs *in vivo*, provide a mechanistic understanding of their functions, and establish their relationship with diseases. With extensive and collective efforts from both experimentalists and bioinformaticians around the globe, we anticipate that these innovative methods and groundbreaking discoveries will be unveiled in the near future.

### Acknowledgments

The Kwok Laboratory is supported by Hong Kong RGC Project No. N\_CityU110/17, CityU 21302317, Croucher Foundation Project No. 9500030, and City University of Hong Kong Projects No. 9610363, 7200520. The Zhang Laboratory is supported by the National Natural Science Foundation of China (Grant No. 31671355, 91740204, and 31761163007) and the National Thousand Young Talents Program of China. We apologize to colleagues whose works are not cited due to space limitation.

### Disclaimer Statement

The authors declare no competing financial interests.

### References

- Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* 2, 986–991
- Clark, M.B. *et al.* (2011) The reality of pervasive transcription. *PLoS Biol.* 9, e1000625
- Yang, L. *et al.* (2014) Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem. Sci.* 39, 35–43
- Wu, H. *et al.* (2017) The diversity of long noncoding RNAs and their generation. *Trends Genet.* 33, 540–552
- Mercer, T.R. *et al.* (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159
- Rinn, J.L. and Chang, H.Y. (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166
- Chen, L.L. (2016) Linking long noncoding RNA localization and function. *Trends Biochem. Sci.* 41, 761–772
- Fenoglio, C. *et al.* (2013) An emerging role for long non-coding RNA dysregulation in neurological disorders. *Int. J. Mol. Sci.* 14, 20427–20442
- Uchida, S. and Dimmeler, S. (2015) Long noncoding RNAs in cardiovascular diseases. *Circ. Res.* 116, 737–750
- Schmitt, A.M. and Chang, H.Y. (2016) Long noncoding RNAs in cancer pathways. *Cancer Cell* 29, 452–463
- Pasmant, E. *et al.* (2011) ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J.* 25, 444–448
- Jendrzewski, J. *et al.* (2012) The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc. Natl. Acad. Sci. U. S. A.* 109, 8646–8651
- Zhang, X. *et al.* (2014) The identification of an ESCC susceptibility SNP rs920778 that regulates the expression of lncRNA HOTAIR via a novel intronic enhancer. *Carcinogenesis* 35, 2062–2067
- Huarte, M. (2015) The emerging role of lncRNAs in cancer. *Nat. Med.* 21, 1253–1261
- Ning, S. *et al.* (2014) LincSNP: a database of linking disease-associated SNPs to human large intergenic non-coding RNAs. *BMC Bioinformatics* 15, 152
- Ning, S. *et al.* (2017) LincSNP 2.0: an updated database for linking disease-associated SNPs to human long non-coding RNAs and their TFBSs. *Nucleic Acids Res.* 45, D74–D78
- Sharp (2009) The centrality of RNA. *Cell* 136, 577–580
- Cech, T.R. and Steitz, J.A. (2014) The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* 157, 77–94
- Dreyfuss, G. (1986) Structure and function of nuclear and cytoplasmic ribonucleoprotein particles. *Annu. Rev. Cell Biol.* 2, 459–498
- Serganov, A. and Patel, D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* 8, 776–790
- Wan, Y. *et al.* (2011) Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12, 641–655
- Sabarathanan, R. *et al.* (2013) RNAsnp: efficient detection of local RNA secondary structure changes induced by SNPs. *Hum. Mutat.* 34, 546–556
- Gong, J. *et al.* (2015) lncRNAsnp: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res.* 43, D181–D186
- Ren, C. *et al.* (2018) Lnc2Catlas: an atlas of long noncoding RNAs associated with risk of cancers. *Sci. Rep.* 8, 1909
- Seemann, S.E. *et al.* (2017) The identification and functional annotation of RNA structures conserved in vertebrates. *Genome Res.* 27, 1371–1383
- Thiel, B.C. *et al.* (2018) RNA structure elements conserved between mouse and 59 other vertebrates. *Genes* 9, 392
- Derrien, T. *et al.* (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789
- Ingolia, N.T. *et al.* (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802
- van Heesch, S. *et al.* (2014) Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* 15, R6
- Carlevaro-Fita, J. *et al.* (2016) Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA* 22, 867–882
- Banfai, B. *et al.* (2012) Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* 22, 1646–1657
- Guttman, M. *et al.* (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240–251
- Ji, Z. *et al.* (2015) Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* 4, e08890
- Nelson, B.R. *et al.* (2016) A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* 351, 271–275

### Outstanding Questions

RNA expression issue: can the structures and interactions of poorly expressed lncRNAs be revealed on a large scale? Can new strategies be developed to enrich rare lncRNAs in a cost-effective fashion and in a higher-throughput manner?

RNA structural ensemble issue: can subpopulated RNA structural conformations be detected unambiguously? Can new experimental approaches and bioinformatic pipelines be implemented to deconvolute structural ensembles into individual structural conformations?

Spatial-temporal RNA folding issue: how does RNA fold in different cell compartments and during biochemical processes? Can we devise innovative methods to map RNA structures/interactions in certain cell compartments/biochemical processes? Can specific chemical probes or crosslinking agents be designed to control the RNA structure modification/crosslinking in a spatial-temporal manner?

Cellular environment and cell/tissue-specific RNA folding issue: how does RNA fold in different cellular conditions and across different cell lines/tissues? How does the structural variation in RNA folds affect its function and biology? Can the structural variation be used as a biomarker for diseases, or as a target for drug development?

lncRNA evolution issue: where did these genes come from? How did they acquire functions? Why are nearly all primate-specific genes lncRNAs? Do lncRNAs consist of functional structural elements connected by nonstructural linkers that can be identified across species? Do intermolecular RNA–RNA, RNA–DNA, and RNA–protein interactions from transacting partners coevolve with the lncRNA structure? How does evolution select and shape lncRNA functional structural elements?

lncRNA SNP issue: can the generated transcriptome-wide lncRNA structure and interaction data help distil functional lncRNA-exonic, disease-related SNPs for further mechanistic characterization?

35. Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
36. Turner, D.H. *et al.* (1988) RNA structure prediction. *Annu. Rev. Biophys. Biophys. Chem.* 17, 167–192
37. Mathews, D.H. *et al.* (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288, 911–940
38. Bernhart, S.H. *et al.* (2006) Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.* 1, 3
39. DiChiacchio, L. *et al.* (2016) AccessFold: predicting RNA–RNA interactions with consideration for competing self-structure. *Bioinformatics* 32, 1033–1039
40. Leamy, K.A. *et al.* (2016) Bridging the gap between *in vitro* and *in vivo* RNA folding. *Q. Rev. Biophys.* 49, e10
41. Ehresmann, C. *et al.* (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res.* 15, 9109–9128
42. Kwok, C.K. *et al.* (2015) The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. *Trends Biochem. Sci.* 40, 221–232
43. Kertesz, M. *et al.* (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107
44. Underwood, J.G. *et al.* (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7, 995–1001
45. Zheng, Q. *et al.* (2010) Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in *Arabidopsis*. *PLoS Genet.* 6, e1001141
46. Li, F. *et al.* (2012) Global analysis of RNA secondary structure in two metazoans. *Cell Rep.* 1, 69–82
47. Lucks, J.B. *et al.* (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U. S. A.* 108, 11063–11068
48. Ding, Y. *et al.* (2014) *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505, 696–700
49. Rouskin, S. *et al.* (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* 505, 701–705
50. Talkish, J. *et al.* (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* 20, 713–720
51. Siegfried, N.A. *et al.* (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* 11, 959–965
52. Spitale, R.C. *et al.* (2015) Structural imprints *in vivo* decode RNA regulatory mechanisms. *Nature* 519, 486–490
53. Aviran, S. and Pachter, L. (2014) Rational experiment design for sequencing-based RNA structure mapping. *RNA* 20, 1864–1877
54. Kwok, C.K. (2016) Dawn of the *in vivo* RNA structurome and interactome. *Biochem. Soc. Trans.* 44, 1395–1410
55. Piao, M. *et al.* (2017) RNA regulations and functions decoded by transcriptome-wide RNA structure probing. *Genomics Proteomics Bioinformatics* 15, 267–278
56. Burkhardt, D.H. *et al.* (2017) Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *eLife* 6, e22037
57. Mustoe, A.M. *et al.* (2018) Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* 173, 181–195 e118
58. Zhang, Y. *et al.* (2018) A stress response that monitors and regulates mRNA structure is central to cold shock adaptation. *Mol. Cell* 70, 274–286 e277
59. Beauclain, J.D. *et al.* (2018) Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* 25, 677–686
60. Li, F. *et al.* (2012) Regulatory impact of RNA secondary structure across the *Arabidopsis* transcriptome. *Plant Cell* 24, 4346–4359
61. Wan, Y. *et al.* (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709
62. Smola, M.J. *et al.* (2015) Detection of RNA-protein interactions in living cells with SHAPE. *Biochemistry* 54, 6867–6875
63. Saha, K. *et al.* (2018) Exonic unpaired elements modulate pre-mRNA structure for splice site recognition. *bioRxiv* <http://dx.doi.org/10.1101/292003>
64. Watts, J.M. *et al.* (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460, 711–716
65. Tang, Y. *et al.* (2016) Protein structure is related to RNA structural reactivity *in vivo*. *J. Mol. Biol.* 428, 758–766
66. Wan, Y. *et al.* (2012) Genome-wide measurement of RNA folding energies. *Mol. Cell* 48, 169–181
67. Righetti, F. *et al.* (2016) Temperature-responsive *in vitro* RNA structurome of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* 113, 7237–7242
68. Lackey, L. *et al.* (2018) Allele-specific SHAPE-MaP assessment of the effects of somatic variation and protein binding on mRNA structure. *RNA* 24, 513–528
69. Tapsin, S. *et al.* (2018) Genome-wide identification of natural RNA aptamers in prokaryotes and eukaryotes. *Nat. Commun.* 9, 1289
70. Kwok, C.K. *et al.* (2016) Structural analysis using SHALiPE to reveal RNA G-quadruplex formation in human precursor micro-RNA. *Angew. Chem. Int. Ed.* 55, 8958–8961
71. Sexton, A.N. *et al.* (2017) Interpreting reverse transcriptase termination and mutation events for greater insight into the chemical probing of RNA. *Biochemistry* 56, 4713–4721
72. Mitchell, D. *et al.* (2018) Glyoxals as *in vivo* RNA structural probes of aguanine base-pairing. *RNA* 24, 114–124
73. Chan, D. *et al.* (2018) Facile synthesis and evaluation of a dual-functioning furoyl probe for *in-cell* SHAPE. *Bioorg. Med. Chem. Lett.* 28, 601–605
74. Feng, C. *et al.* (2018) Light-activated chemical probing of nucleobase solvent accessibility inside cells. *Nat. Chem. Biol.* 14, 276–283
75. Choudhary, K. *et al.* (2017) Comparative and integrative analysis of RNA structural profiling data: current practices and emerging questions. *Quant. Biol.* 5, 3–24
76. Selega, A. *et al.* (2017) Robust statistical modeling improves sensitivity of high-throughput RNA structure probing experiments. *Nat. Methods* 14, 83–89
77. Li, B. *et al.* (2017) PROBER provides a general toolkit for analyzing sequencing-based toeprinting assays. *Cell Syst.* 4, 568–574 e567
78. Ledda, M. and Aviran, S. (2018) PATTERN: transcriptome-wide search for functional RNA elements via structural data signatures. *Genome Biol.* 19, 28
79. Radecki, P. *et al.* (2018) Automated recognition of RNA structure motifs by their SHAPE data signatures. *Genes (Basel)* 9, E300
80. Novikova, I.V. *et al.* (2012) Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.* 40, 5034–5051
81. Somarowthu, S. *et al.* (2015) HOTAIR forms an intricate and modular secondary structure. *Mol. Cell* 58, 353–361
82. Hawkes, E.J. *et al.* (2016) COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures. *Cell Rep.* 16, 3087–3096
83. Xue, Z. *et al.* (2016) A G-rich motif in the lncRNA Braveheart interacts with a zinc-finger transcription factor to specify the cardiovascular lineage. *Mol. Cell* 64, 37–50
84. Liu, F. *et al.* (2017) Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat. Chem. Biol.* 13, 282–289
85. Lin, Y. *et al.* (2018) Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. *Nucleic Acids Res.* 46, 3742–3752

lncRNA involvement in ribosome association and protein translation issue: do lncRNA structures and interactions have roles in controlling ribosome association, and regulating their own decay and translation?

lncRNA tertiary structure and function issue: can the generated transcriptome-wide lncRNA structure and interaction data facilitate the determination of high-resolution 3D lncRNA structures, for example, as additional constraints to cryo-electron microscopy determined structures? How specifically can we use lncRNA structural elements and/or empirically documented interactions with other RNAs and with proteins to understand the function of nonconserved sequences in the lncRNAs?

86. Rivas, E. *et al.* (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* 14, 45–48
87. Tavares, R.C.A. *et al.* (2018) Covariation analysis with improved parameters reveals conservation in lncRNA structures. *bioRxiv* <http://dx.doi.org/10.1101/364109>
88. Rivas, E., Eddy, S.R. Response to Tavares *et al.*, "Covariation analysis with improved parameters reveals conservation in lncRNA structures". *bioRxiv* (in press)
89. Plath, K. *et al.* (2002) Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* 36, 233–278
90. Maenner, S. *et al.* (2010) 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol.* 8, e1000276
91. Fang, R. *et al.* (2015) Probing Xist RNA structure in cells using targeted Structure-seq. *PLoS Genet.* 11, e1005668
92. Smola, M.J. *et al.* (2016) SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl. Acad. Sci. U. S. A.* 113, 10322–10327
93. Lu, Z. *et al.* (2016) RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165, 1267–1279
94. Zhao, J. *et al.* (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756
95. Wutz, A. *et al.* (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* 30, 167–174
96. Helwak, A. *et al.* (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153, 654–665
97. Sugimoto, Y. *et al.* (2015) hiCLIP reveals the *in vivo* atlas of mRNA secondary structures recognized by Staufen 1. *Nature* 519, 491–494
98. Nguyen, T.C. *et al.* (2016) Mapping RNA-RNA interactome and RNA structure *in vivo* by MARIO. *Nat. Commun.* 7, 12023
99. Aw, J.G. *et al.* (2016) *In vivo* mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* 62, 603–617
100. Sharma, E. *et al.* (2016) Global mapping of human RNA-RNA interactions. *Mol. Cell* 62, 618–626
101. Kudla, G. *et al.* (2011) Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10010–10015
102. Weidmann, C.A. *et al.* (2016) Direct duplex detection: an emerging tool in the RNA structure analysis toolbox. *Trends Biochem. Sci.* 41, 734–736
103. Zhong, C. and Zhang, S. (2017) CLAN: the CrossLinked reads ANalysis tool. *bioRxiv* <http://dx.doi.org/10.1101/233841>
104. Gong, J. *et al.* (2018) RISE: a database of RNA interactome from sequencing experiments. *Nucleic Acids Res.* 46, D194–D201
105. Lerner, M.R. *et al.* (1981) Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus. *Proc. Natl. Acad. Sci. U. S. A.* 78, 805–809
106. Karreth, F.A. *et al.* (2011) *In vivo* identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell* 147, 382–395
107. Gong, C. and Maquat, L.E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* 470, 284–288
108. Kretz, M. *et al.* (2013) Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 493, 231–235
109. Engreitz, J.M. *et al.* (2014) RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* 159, 188–199
110. Fox, A.H. *et al.* (2018) Paraspeckles: where long noncoding RNA meets phase separation. *Trends Biochem. Sci.* 43, 124–135
111. Chen, J. *et al.* (2004) Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res.* 32, 4812–4820
112. Faghihi, M.A. *et al.* (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat. Med.* 14, 723–730
113. Simon, M.D. *et al.* (2011) The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20497–20502
114. Chu, C. *et al.* (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* 44, 667–678
115. Engreitz, J.M. *et al.* (2013) The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341, 1237973
116. Cheetham, S.W. and Brand, A.H. (2018) RNA-DamID reveals cell-type-specific binding of roX RNAs at chromatin-entry sites. *Nat. Struct. Mol. Biol.* 25, 109–114
117. Li, X. *et al.* (2017) GRID-seq reveals the global RNA-chromatin interactome. *Nat. Biotechnol.* 35, 940–950
118. Sridhar, B. *et al.* (2017) Systematic mapping of RNA-chromatin interactions *in vivo*. *Curr. Biol.* 27, 602–609
119. Bell, J.C. *et al.* (2018) Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *eLife* 7, e27024
120. Quinodoz, S.A. *et al.* (2018) Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* 174, 744–757 e724
121. Paralkar, V.R. *et al.* (2016) Unlinking an lncRNA from its associated cis element. *Mol. Cell* 62, 104–110
122. Engreitz, J.M. *et al.* (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539, 452–455
123. Anderson, K.M. *et al.* (2016) Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature* 539, 433–436
124. Joung, J. *et al.* (2017) Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* 548, 343–346
125. Chu, C. *et al.* (2015) Systematic discovery of Xist RNA binding proteins. *Cell* 161, 404–416
126. McHugh, C.A. *et al.* (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521, 232–236
127. Quinn, J.J. *et al.* (2014) Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat. Biotechnol.* 32, 933–940
128. Ramanathan, M. *et al.* (2018) RNA-protein interaction detection in living cells. *Nat. Methods* 15, 207–212
129. Baltz, A.G. *et al.* (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* 46, 674–690
130. Castello, A. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393–1406
131. Melvin, W.T. *et al.* (1978) Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. *Eur. J. Biochem.* 92, 373–379
132. Jao, C.Y. and Salic, A. (2008) Exploring RNA transcription and turnover *in vivo* by using click chemistry. *Proc. Natl. Acad. Sci. U. S. A.* 105, 15779–15784
133. He, C. *et al.* (2016) High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Mol. Cell* 64, 416–430
134. Bao, X. *et al.* (2018) Capturing the interactome of newly transcribed RNA. *Nat. Methods* 15, 213–220
135. Huang, R. *et al.* (2018) Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* 115, E3879–E3887

136. Ule, J. *et al.* (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302, 1212–1215
137. Zhao, J. *et al.* (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* 40, 939–953
138. Licatalosi, D.D. *et al.* (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464–469
139. Konig, J. *et al.* (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* 17, 909–915
140. Hafner, M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141
141. Zarnegar, B.J. *et al.* (2016) irCLIP platform for efficient characterization of protein-RNA interactions. *Nat. Methods* 13, 489–492
142. Van Nostrand, E.L. *et al.* (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* 13, 508–514
143. Chi, S.W. *et al.* (2009) Argonaute HITS-CLIP decodes micro-RNA-mRNA interaction maps. *Nature* 460, 479–486
144. Khorshid, M. *et al.* (2011) CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res.* 39, D245–D252
145. Anders, G. *et al.* (2012) doRINA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.* 40, D180–D186
146. Jalali, S. *et al.* (2013) Systematic transcriptome wide analysis of lncRNA-miRNA interactions. *PLoS One* 8, e53823
147. Li, J.H. *et al.* (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42, D92–D97
148. Homan, P.J. *et al.* (2014) Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U. S. A.* 111, 13858–13863
149. Woods, C.T. *et al.* (2017) Comparative visualization of the RNA suboptimal conformational ensemble *in vivo*. *Biophys. J.* 113, 290–301
150. Spasic, A. *et al.* (2018) Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic Acids Res.* 46, 314–323
151. Li, H. and Aviran, S. (2018) Statistical modeling of RNA structure profiling experiments enables parsimonious reconstruction of structure landscapes. *Nat. Commun.* 9, 606
152. Sun, I. *et al.* (2018) Dynamic regulation of RNA structure in mammalian cells. *bioRxiv* <http://dx.doi.org/10.1101/399386>
153. Incamato, D. *et al.* (2017) *In vivo* probing of nascent RNA structures reveals principles of cotranscriptional folding. *Nucleic Acids Res.* 45, 9716–9725
154. Willingham, A.T. *et al.* (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* 309, 1570–1573
155. Shamovsky, I. *et al.* (2006) RNA-mediated response to heat shock in mammalian cells. *Nature* 440, 556–560
156. Tripathi, V. *et al.* (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39, 925–938
157. Bardou, F. *et al.* (2014) Long noncoding RNA modulates alternative splicing regulators in *Arabidopsis*. *Dev. Cell* 30, 166–176
158. Carrieri, C. *et al.* (2012) Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491, 454–457
159. Wang, J.Y. *et al.* (2010) CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res.* 38, 5366–5383
160. Jeon, Y. and Lee, J.T. (2011) YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146, 119–133
161. Li, R. *et al.* (2016) Understanding the functions of long non-coding RNAs through their higher-order structures. *Int. J. Mol. Sci.* 17, E702
162. Campalans, A. *et al.* (2017) Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell* 29, 912
163. Liu, X.F. *et al.* (2012) Long non-coding RNA gadd7 interacts with TDP-43 and regulates Cdk6 mRNA decay. *EMBO J.* 31, 4415–4427
164. Prasanth, K.V. *et al.* (2005) Regulating gene expression through RNA nuclear retention. *Cell* 123, 249–263
165. Peters, N.T. *et al.* (2003) RNA editing and regulation of *Drosophila* 4f-mp expression by sas-10 antisense readthrough mRNA transcripts. *RNA* 9, 698–710
166. Schmitz, K.M. *et al.* (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* 24, 2264–2269
167. Tsai, M.C. *et al.* (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693
168. Csorba, T. *et al.* (2014) Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc. Natl. Acad. Sci. U. S. A.* 111, 16160–16165
169. Chillon, I. and Pyle, A.M. (2016) Inverted repeat Alu elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function. *Nucleic Acids Res.* 44, 9462–9471
170. West, J.A. *et al.* (2014) The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* 55, 791–802
171. Zhang, B. *et al.* (2017) Identification and characterization of a class of MALAT1-like genomic loci. *Cell Rep.* 19, 1723–1738
172. Ilik, I.A. *et al.* (2013) Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol. Cell* 51, 156–173
173. Incamato, D. *et al.* (2014) Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* 15, 491
174. Takahashi, M.K. *et al.* (2016) Using in-cell SHAPE-Seq and simulations to probe structure-function design principles of RNA transcriptional regulators. *RNA* 22, 920–933
175. Ritchey, L.E. *et al.* (2017) Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure *in vivo*. *Nucleic Acids Res.* 45, e135
176. Deng, H. *et al.* (2018) Rice *in vivo* RNA structure reveals RNA secondary structure conservation and divergence in plants. *Mol. Plant* 11, 607–622
177. Watters, K.E. *et al.* (2018) Probing of RNA structures in a positive sense RNA virus reveals selection pressures for structural elements. *Nucleic Acids Res.* 46, 2573–2584
178. Zubradt, M. *et al.* (2017) DMS-MaPseq for genome-wide or targeted RNA structure *in vivo*. *Nat. Methods* 14, 75–82
179. Dadonaite, B. *et al.* (2017) The structure of the influenza A virus genome. *bioRxiv* <http://dx.doi.org/10.1101/236620>