nature chemical biology

REVIEW ARTICLE

https://doi.org/10.1038/s41589-021-00805-7

RNA structure probing uncovers RNA structure-dependent biological functions

Xi-Wen Wang[1,2,6], Chu-Xiao Liu[3,6], Ling-Ling Chen[3,4,5] ✉ and Qiangfeng Cliff Zhang[1,2] ✉

RNA molecules fold into complex structures that enable their diverse functions in cells. Recent revolutionary innovations in transcriptome-wide RNA structural probing of living cells have ushered in a new era in understanding RNA functions. Here, we summarize the latest technological advances for probing RNA secondary structures and discuss striking discoveries that have linked RNA regulation and biological processes through interrogation of RNA structures. In particular, we highlight how different long noncoding RNAs form into distinct secondary structures that determine their modes of interactions with protein partners to realize their unique functions. These dynamic structures mediate RNA regulatory functions through altering interactions with proteins and other RNAs. We also outline current methodological hurdles and speculate about future directions for development of the next generation of RNA structure-probing technologies of higher sensitivity and resolution, which could then be applied in increasingly physiologically relevant studies.

The versatile roles for RNA as a genetic material carrying heritable information and as a ribozyme catalyzing essential biological processes are now widely appreicated[1]. While the genetic information is encoded in the RNA sequence, ribozyme activity depends on RNA folding into specific secondary or tertiary structures[2]. One of the best-known examples is ribosomal RNA, which packs into complex tertiary structures and catalyzes protein synthesis during translation[3]. Also, strikingly, the Human Genome Project revealed that the fraction of RNAs responsible for protein coding is actually very small, with the majority of other RNAs falling into the noncoding class[4]. The past two decades have witnessed a burst of studies into the function and regulatory mechanisms of the so-called long noncoding RNAs (lncRNAs)[5]. These lncRNAs have been shown to have a variety of biological functions, including regulating gene expression by acting as signals, decoys, guides and scaffolds[6]. Notably, this research area has largely focused on the functional roles of lncRNAs in various systems, with less attention paid to the mechanistic connections between lncRNA structure and function, mainly due to the lack of methodologies to access and test these attributes.

Thanks to rapid advances in RNA structure-probing methods, structure–function relationships for lncRNAs, as well as other types of RNAs, have recently started to yield important biological insights. In the past 10 years, techniques have been developed to capture transcriptome-wide RNA structures (that is, RNA structuromes) in many species and across conditions[7–36] (Fig. 1). Systems biology analyses have well established how the overall RNA structures can affect RNA regulation and functions, including RNA processing, localization and translation (see review[37]). Accumulating evidence is also starting to highlight the diverse mechanisms through which lncRNAs can form different structural motifs to participate in gene expression regulation[38–44]. For example, different structure-probing methods combined with crosslinking and immunoprecipitation data have revealed that *Xist* contains multiple well-defined structure domains, which associate with various functional protein complexes[23,39]. These studies support existing domain-based models of X chromosome inactivation (XCI).

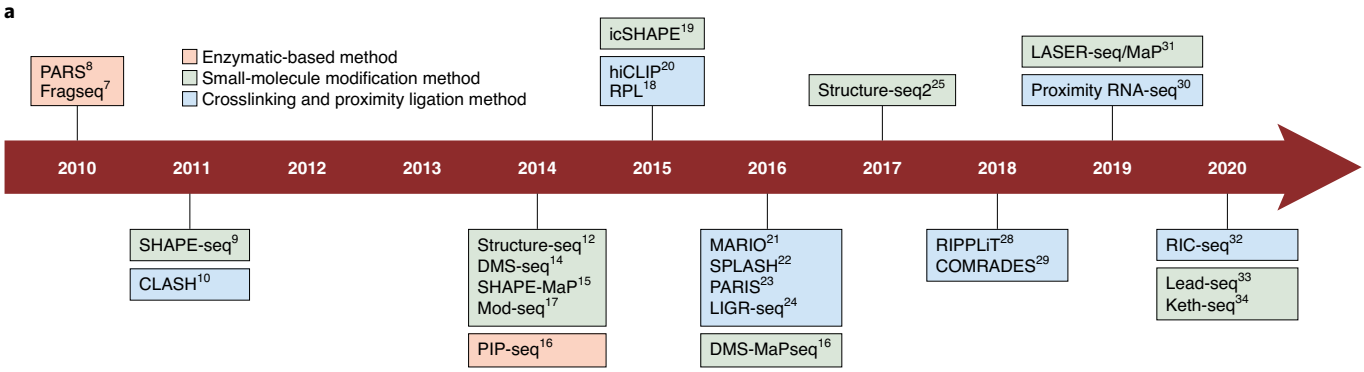
In this Review, we summarize the most recent technological advances and discoveries that link RNA regulation and function through interrogation of RNA structures. First, we provide an overview of the recent technological innovations of RNA structure probing and discuss technical challenges in the development of RNA structure-probing methods with higher resolution and in increasingly informative biological contexts. Second, we focus on the important scientific frontier of elucidating the roles of RNA structure in diverse biological processes and systems. In the final section, we speculate on future technological developments and directions that promise to enrich our understanding of RNA structural elements that control fundamental aspects of biology.

Recent technological innovation of RNA structure probing

Present technologies for RNA structure probing can be categorized into two main classes. One is based on small-molecule modification, while the other depends on crosslinking and proximity ligation. To achieve different experimental aims, individual methods or combinations of methods are used to uncover the biological roles of RNA structures.

Small-molecule modification-based methods. RNase cleavage-based methods, such as fragmentation sequencing (FragSeq)[7] and parallel analysis of RNA structures (PARS)[8], are among the first technologies that achieve transcriptome-wide RNA structure probing with deep sequencing. However, usually, these methods can only be applied in vitro, since RNases are too big to penetrate

[1]MOE Key Laboratory of Bioinformatics, Beijing Advanced Innovation Center for Structural Biology and Frontier Research Center for Biological Structure, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing, China. [2]Tsinghua-Peking Center for Life Sciences, Beijing, China. [3]State Key Laboratory of Molecular Biology, Shanghai Key Laboratory of Molecular Andrology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, University of the Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [4]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [5]School of Life Sciences, Hangzhou Institute for Advanced Study, University of the Chinese Academy of Sciences, Hangzhou, China. [6]These authors contributed equally: Xi-Wen Wang, Chu-Xiao Liu. ✉e-mail: linglingchen@sibcb.ac.cn; qczhang@tsinghua.edu.cn



b

Highlights of recent RNA secondary structure technology in vivo

Method	Samples	Highlights
Structure-seq	Etiolated seedlings of *A. thaliana*	Global structure pattern of mRNA in vivo[12]
DMS-seq	Seedlings of *A. thaliana*, yeast, K562, human foreskin fibroblast	Less structure in vivo than in vitro[14]
SHAPE-MaP	Mouse TSCs	Discover conserved structured elements of RNA as protein binding motifs[39,86]
Mod-seq	Yeast	Identify L26 binding site on the ribosome[17]
icSHAPE	Mouse ESCs, HEK293T	In vivo RNA structure for four bases in genome wide[19]
DMS-MaPseq	Yeast, HEK293T, *D. melanogaster* ovaries	Discover a functional structure in the low-abundance human FXR2 mRNA[27]
Structure-seq2	Seedlings of rice	Identify a new N^1-methyladenosine (m^1A) in a rRNA[25]
LASER-seq/MaP	K562	Rapidly survey for ligand-binding sites[31]
Lead-seq	*Y. pseudotuberculosis* YPIII	Monitor rRNA-binding ligands[33]
Keth-seq	Mouse ESCs and HeLa	Detect rG4 region in vivo[34]
CLASH	HEK293T	miRNA–mRNA interaction[10,50]
PLA	Yeast, GM12878	Generate contact probability maps for ribosomal and other abundant RNAs[18]
hiCLIP	HEK293T	Staufen1-mediated XBP1 mRNA duplex to regulate its splicing[20]
SPLASH	HeLa, GM12892, human ESCs, yeast and *E. coli*	RNA genome structure of dengue and Zika viruses; influenza virions[22,48]
PARIS	HeLa, HEK293T, mouse ESCs	Structure of repeat A of XIST; RNA genome structure of Zika virus[23,49]
MARIO	Mouse ESCs and mouse brain	Malat1 lincRNA–slc2a3 mRNA interaction[21]
LIGR-seq	HEK293T	snoRNAs interact with mRNAs to regulate their gene expression[24]
RIPPLiT	HEK293	Pretranslated mRNPs are tightly packed into linearly organized flexible rod-like structures[28]
COMRADES	JEG-3 cells infected by Zika virus	Zika virus interacts with host mir-21[29]
Proximity RNA-seq	SH-SY5Y cells	Study of the spatial organization of transcripts in the nucleus[30]
RIC-seq	HeLa	Long-range RNA–RNA interaction regulate gene expression by modulate chromatin looping[32]

Fig. 1 | Summary of RNA secondary structure-probing methods. a, Timeline of technology development for RNA secondary structure probing. Enzymatic-based methods, small-molecule modification methods and crosslinking and proximity ligation methods are labeled in orange, green and blue, respectively. hiCLIP, hybrid and individual-nucleotide resolution ultraviolet crosslinking and immunoprecipitation; keth-seq, N^3-kethoxal probing with deep sequencing; LASER-seq/MaP, light-activated structural examination of RNA by high-throughput sequencing/mutational profiling; lead-seq, combined lead(II) acetate-mediated cleavage of single-stranded RNA regions with high-throughput sequencing; MARIO, mapping RNA interactome in vivo; PIP-seq, protein interaction profile sequencing; RIPPLiT, RNA immunoprecipitation and proximity ligation in tandem; RPL, RNA proximity ligation; RNA-seq, RNA sequencing. **b**, Exciting biological discoveries revealed by different RNA structure-probing methods in vivo. The labels are coloured as in **a**. lincRNA, long intergenic non-coding RNA; snoRNA, small nucleolar RNA. XIST, X-inactive specific transcript.

cells. An alternative method is to replace RNases with small molecules, which also modify RNA targets with structure specificity (Fig. 2a). As small molecules are much easier to pass through cell membranes, these approaches are now widely used to probe RNA structure in vivo.

Two general classes of small molecules have been used for RNA structure probing. One is base specific and the other is ribose specific. Base-specific probes directly sense base-pairing interactions or solvent accessibility. For example, dimethyl sulfate (DMS) reacts with N1 of adenine and N3 of cytosine on the Watson–Crick face[12,14]; N_3-kethoxal is used to probe the Watson–Crick face of guanine[34]. In contrast, ribose-specific molecules probe the RNA structure by reacting with the RNA backbone to monitor all four nucleotides. For example, 2′-hydroxyl acylation analyzed by primer extension (SHAPE) reagents comprise a large family of RNA structure-probing molecules that acylate the 2′-OH of the sugar ring[45]. Lead(II) is also an informative probe and it is known to cleave single-stranded RNA by directly hydrolyzing the phosphodiester backbone[33]. Particular probes are suitable for addressing different biological questions. Guidelines for choosing appropriate probes have been provided in some comprehensive reviews[46,47]. For example, probes that react with RNA with slow kinetics are more suitable to probing RNA structure in vivo, while probes that react with RNA rapidly are suited for analyzing RNA folding dynamics or biological processes that occur on fast time scales.

The procedures underlying high-throughput, small-molecule-modification-based RNA structure-probing methods are quite similar. Probes are first incubated with cells. Then, after RNA extraction, modified nucleotides are recorded by reverse transcription with a truncation (RT stop) in the resulting complementary DNA libraries. To improve the structure-probing accuracy, different optimal treatments are used in different strategies (Fig. 2b). A common way of enhancing coverage is through RNA fragmentation, although false signals can be introduced into the library. Methods to minimize RT stop signals derived from unmodified reads have been developed. For example, in vivo click selective 2′-hydroxyl acylation and

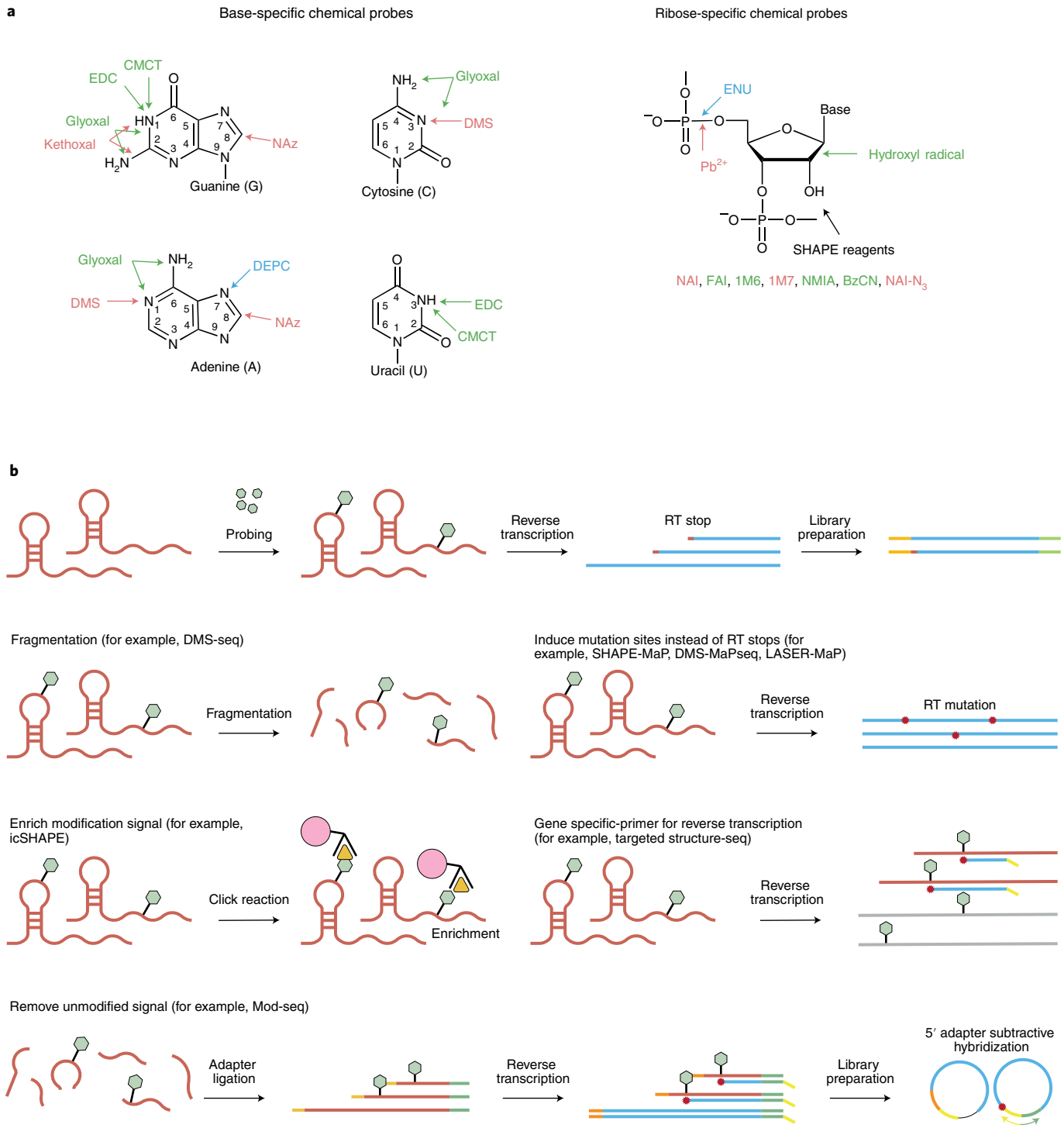


Fig. 2 | Recent technological innovations in RNA secondary structure probing. a, Small-molecule-modification-based methods. Different probing reagents target diverse positions of RNA molecules. The reagents in red indicate those that have been used for in vivo high-throughput probing. The reagents in green indicate those that have been used for in vivo low-throughput probing. The reagents in blue indicate those that have only been used in vitro. 1M6, 1-methyl-6-nitroisatoic anhydride; 1M7, 1-methyl-7-nitroisatoic anhydride; BzCN, Benzoyl cyanide; CMCT, *N*-cyclohexy-*N*′(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate; DEPC, diethyl pyrocarbonate; EDC, 1-ethyl-3-(3-dimethylam inopropyl)-carbodiimide; ENU, ethylnitrosourea FAI, 2-methyl-3-furoic acid imidazolide; NAI, 2-methylnicotinic acid imidazolide; NAz, nicotinoyl azide. **b**, General procedures of RNA structure probing (top) and alternative strategies used to improve the sensitivity and accuracy of structure probing (bottom). The processing steps can be optimized by RNA fragmentation, modification signal enrichment and unmodified signal depletion; introducing an RT mutation increases the RNA structure information; and a gene-specific primer can be used to distinguish between different RNA isoform species and to characterize the structures of low-abundance transcripts.

profiling experiment (icSHAPE) enriches modified RNAs to improve the signal-to-noise ratio via a refined SHAPE reagent (for example, NAI-N_3), which allows the introduction of a biotin molecule to the modified RNA through a click reaction[19]. Mod-seq removes the unmodified reads using 5′ adapter subtractive hybridization[17]. Both strategies decrease unwanted RT stop signals in probing data. Since only one RT stop can be generated from a sequencing read, methods have been developed to increase RNA structure information content in sequencing data, by reading RT mutations instead of RT stops caused by modified nucleotides[15,27]. However, mutation-based RNA structure-probing methods usually require a greater sequencing depth in order to obtain accurate transcriptome-wide mutational profiles. Another improvement for detecting less abundant transcripts is to use gene-specific primers during reverse transcription. This approach increases the sensitivity of the method and can even detect isoform-specific RNA structure[27].

Crosslinking and proximity ligation-based methods. Small-molecule-modification-based probing methods generate one-dimensional RNA structure information for each nucleotide base but do not directly identify specific base-pairing partners. In contrast, crosslinking-based methodologies can directly capture intra- and intermolecular RNA–RNA interactions by proximity ligation. After crosslinking in cells, RNA is extracted and fragmented and pairs of interacting fragments are ligated together. RNA interaction information is then preserved in chimeras of ligation products and read out by sequencing (Fig. 3a). A number of crosslinking and proximity ligation-based technologies have been developed to capture the RNA–RNA interactome across organisms, including yeast[22], humans[22–24,29,32], mice[21,23] and viruses[48,49]. Crosslinking ligation and sequencing of hybrids (CLASH) was among the first[10]. This focuses on target protein-associated RNA–RNA interactions that are crosslinked upon ultraviolet light exposure[50]. Methods used to map transcriptome-wide RNA–RNA interactions have been reported using psoralen-based crosslinking[22,23]. Of note, psoralen can only capture direct RNA–RNA interactions. Indirect RNA–RNA interactions mediated by proteins can be captured through formaldehyde crosslinking[32]. These recently developed crosslinking methods provide a set of high-throughput toolkits that are allowing the development of comprehensive maps of RNA–RNA interaction in living cells.

A substantial improvement in accurately capturing the RNA interactome in libraries has been achieved by target RNA enrichment (Fig. 3b). The psoralen analysis of RNA interactions and structures (PARIS) technique, for example, incorporates two-dimensional polyacrylamide gel electrophoresis selection to enrich crosslinked RNAs[23]. Sequencing of psoralen crosslinked, ligated, and selected hybrids (SPLASH) mediates crosslinking by a biotin-psoralen, allowing purification of the crosslinked RNA using streptavidin beads[22]. LIGR-seq (ligation of interacting RNA followed by high-throughput sequencing) uses RNase R, a 3′–5′ exoribonuclease, to digest uncrosslinked RNAs to enrich crosslinked fragments[24]. COMRADES (crosslinking of matched RNAs and deep sequencing) uses biotin-labeled gene-specific primers to enrich target RNA duplexes, followed by a click reaction in the second purification step to enhance the accuracy of structure probing[29]. After purification, crosslinked RNAs are fragmentated and end-repaired before proximity ligation to produce chimeric RNAs before deep sequencing.

In contrast with RNA secondary structures that are dependent on inter- or intrastrand base-paring interactions, higher-order RNA structures are formed by long-range inter- or intramolecular interactions within the RNA molecules independent of base pairing. How are these higher-order RNA structures packed? How do they relate to the regulatory role of RNA? Some methods have been developed to answer these questions on a transcriptome-wide scale. RIPPLiT (RNA immunoprecipitation and proximity ligation in tandem) is a method that captures higher-order RNA structures by first enriching ribonucleoprotein (RNP) complexes through stringent double immunoprecipitation and then using RNA proximity ligation to investigate three-dimensional organization of the stable RNP core. It has been used to study the exon junction complex before translation and revealed that pre-translational messenger RNAs (mRNAs) compact with associated proteins to form rod-like structures[28]. Another method, proximity RNA sequencing, uses barcoded beads to capture RNAs in close proximity, reporting on how the nuclear transcriptome is partitioned[30]. Since this method does not contain a ligation step, it is not restricted to pairwise interactions, but it can detect multiple spatially neighboring RNAs. Recently, RNA in situ conformation sequencing (RIC-seq) has been developed to profile RNA–RNA spatial interactions in situ globally, by performing RNA proximity ligation after formaldehyde crosslinking and incorporating a biotin-labeled cytidine (pCp-biotin) into the 3′ end of RNA to enrich the chimeric fragments. Application of RIC-seq confirms that enhancer and promoter RNA–RNA interactions can regulate gene expression through modulating chromatin looping[32,51].

Integrative methods. As with all technologies, each of the individual probing methods detailed above has its own limitations. For example, DMS can only probe unpaired adenine and cytosine. It is therefore necessary to combine multiple probing regents if structural information for each nucleotide is desired. Chemical inference of RNA structures followed by massive parallel sequencing (CIRS-seq) combines *N*-cyclohexy-*N*′(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate and DMS to probe all four nucleotides in single-stranded transcript conformations[13]. Applying CIRS-seq to determine the RNA structurome in mouse embryonic stem cells (ESCs) in vitro revealed that Lin28a preferentially binds to RNA motifs in a single-stranded conformation[13]. SHAPE reagents can react with the 2′-hydroxyl group of the backbone of four nucleotides. However, stacking interactions can make unpaired bases poorly reactive towards those reagents[46]. So, additional DMS and terbium probing was performed to validate the *HOTAIR* (*HOX* transcript antisense intergenic RNA) secondary structure derived from the SHAPE data. The study confirmed good agreement between the distinct forms of probing data and proposed four independent architectural modules for *HOTAIR*, of which two precisely correspond to the predicted protein-binding domains[40].

SHAPE and SHAPE-like methods have also been integrated with crosslinking and ligation-based RNA-probing strategies, especially for detecting virus genome structures. Beyond local RNA base pairing, long-range base pairings have been discovered within RNA viruses to mediate fundamental viral processes, such as replication and translation[52]. Combining icSHAPE and PARIS, or selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) and SPLASH, has revealed how these secondary structures are integrated and regulate viral infectivity in Zika and dengue viruses[48,49]. These studies discovered numerous new RNA structural elements, many of which are potentially functional ones supported by coevolution evidence. In particular, a strain-specific long-range interaction between the 5′ untranslated region (UTR) and the coding region of the E protein was found to contribute to the infectivity of Zika viruses[49].

Overall, recent innovations in RNA structure probing can be grouped into three thematic areas. First, different probes are developed to obtain a greater signal-to-noise ratio and higher specificity. Second, by improving proximity ligation-based methods, larger RNA structures and higher-order structures in the RNP complex are able to be detected. Additionally, multiple methods are combined to integrate and take advantage of complementary information provided by different probes.

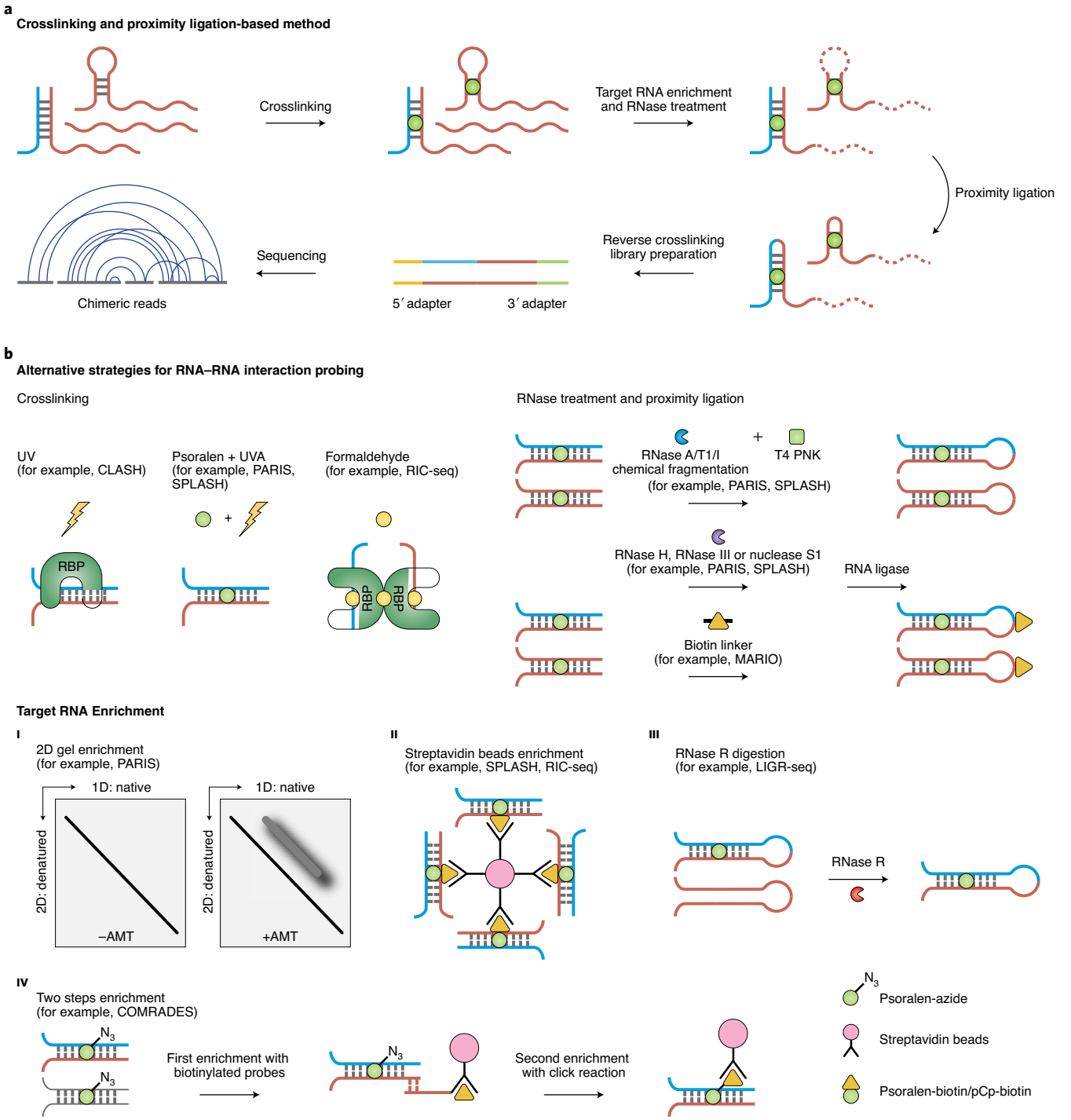


Fig. 3 | Recent technological innovations of crosslinking and proximity ligation-based methods. a, General procedures of crosslinking and proximity ligation-based methods. RNA duplexes are first stabilized by crosslinking, then target RNAs are enriched to enhance the signal. This is followed by RNase treatment and proximity ligation, producing chimeric RNA fragments. Finally, RNAs are reverse crosslinked and converted into a complementary DNA sequencing library. **b**, Alternative strategies for RNA–RNA interaction detection in each step. Crosslinking steps can be mediated by ultraviolet (UV) light, psoralen or formaldehyde. The step of RNase treatment and proximity ligation can be achieved using different strategies. One way is to treat RNA with RNase A/T1 or chemical fragmentation to produce RNA fragments with 5′-hydroxyl and 3′-phosphate; such fragments are incompatible with RNA ligation. The dual activities of T4 polynucleotide kinase can then be used to add a phosphate at the 5′ end of the RNA and to dephosphorylate the 3′ end. An alternative strategy is using RNases (such as S1 nuclease) to generate ligation-compatible ends. Additionally, some methods, such as MARIO, introduce a selectable biotinylated RNA linker to mediate RNA–RNA ligation to minimize false positive signals for RNA–RNA interactions. The methods used to enrich target RNA include two-dimensional (2D) gel enrichment, streptavidin bead enrichment of modified RNA duplexes, RNase R digestion to remove uncrosslinked RNA, and so-called two-step enrichment. Image for two-dimensional gel enrichment in **b** adapted with permission from ref. [23], Elsevier.

Linking RNA structures to RNA functions and regulations

Recent innovations in RNA structure probing are particularly exciting as they allow characterization of the secondary structural features in large RNAs. This is particularly useful for lncRNAs, which are increasingly being recognized as important modulators in diverse biological processes (see reviews[5,53,54]) and potentially possess greater structural complexity than mRNAs to carry out their regulatory functions[12,55,56]. The recent advance of lncRNA secondary structure studies reveals different modes of interaction with proteins (Fig. 4).

Discrete structure motifs for binding multiple proteins. Perhaps the most well-studied lncRNA of this group is *Xist*, which is essential for XCI during female development in mammals. *Xist* is ~18,000 nucleotides in length and, once expressed, binds to critical nucleation sites and spreads to establish chromosome-wide silencing through step-wise interaction with a large number of proteins[57–59]. Earlier studies revealed the importance of the secondary structure in specific regions of *Xist* for XCI[60], but a functional structural map of *Xist* has remained poorly defined until recently.

SHAPE-MaP analysis of the full-length *Xist*, both in mouse trophoblast stem cells and under protein-free conditions (ex vivo)[39], revealed that *Xist* can form 33 well-defined secondary structure domains, which is comparable to the number of functional elements within ribosomal RNAs. These findings support a model in which different domains enable *Xist* to interact with distinct proteins during XCI. The analysis of long-range and alternative RNA structures by icSHAPE and PARIS revealed that the A-repeat region at the 5′ end of *Xist*, which is critical for epigenetic silencing, tends to form duplexes that facilitate interaction with the key silencing factor SPEN[23]. Upon localization of *Xist* to specific nucleation sites, it directly interacts with SPEN to form the *Xist*–SPEN–SMRT complex and recruits histone deacetylase 3, leading to transcriptional silencing and chromatin remodeling. *Xist* then recruits Polycomb repressive complex 2 (PRC2) in a histone deacetylase 3-dependent manner to maintain the inactive state across local chromatin (Fig. 4a)[57]. Given the large size of *Xist*, this recent progress in structural mapping unveils the RNA structure landscape that guides protein binding linked to XCI, and also provides important directions for higher-order structural studies of *Xist* in the future.

It is worthwhile noting that RNA-on-X 1 and 2 (*roX1* and *roX2*) can mediate dosage compensation in *Drosophila*[61]. Analysis of the *roX* RNA secondary structure by SHAPE and iCLIP revealed that both *roX1* and *roX2* contain common and conserved structured tandem stem loops, allowing these lncRNAs to bind the male lethal (MLE) RNA helicase and other components in the male sex lethal (MSL) dosage compensation complex. This observation further highlights the view that *roX1* and *roX2* differ greatly in size and sequences but function similarly using comparable structural motifs[62].

Similar structural analysis has now been applied to many lncRNAs, revealing their secondary structures and modes of interaction with proteins. *HOX* transcript antisense intergenic RNA (*HOTAIR*) is a 2,148-nuclotide-long polyadenylated lncRNA encoded by the *HOXC* locus. *HOTAIR* is proposed to mediate chromatin remodeling of the *HOXD* locus by interacting with PRC2 components and is required for H3K27 trimethylation and transcriptional silencing of the *HOXD* locus[63]. Comparative structure analysis of human and mouse *HOTAIR* revealed a number of conserved elements, including the PRC2-binding region[40]. Notably, it has become clear that PRC2 can bind RNA[64] but does so promiscuously[65]. It remains an open question how RNA structure influences PRC2–RNA interaction-mediated gene inactivation.

The mouse *Braveheart* (*Bvht*) RNA acts in *trans* to regulate cardiovascular lineage commitment[66]. SHAPE and DMS probing on the in vitro-transcribed full-length (~590-nucleotide) *Bvht* found that this lncRNA contains three major domains, as well as a 5′ asymmetric G-rich internal loop, which interacts with the negative regulator CNBP (zinc-finger transcription factor cellular nucleic acid-binding protein) in the cardiac developmental program[41].

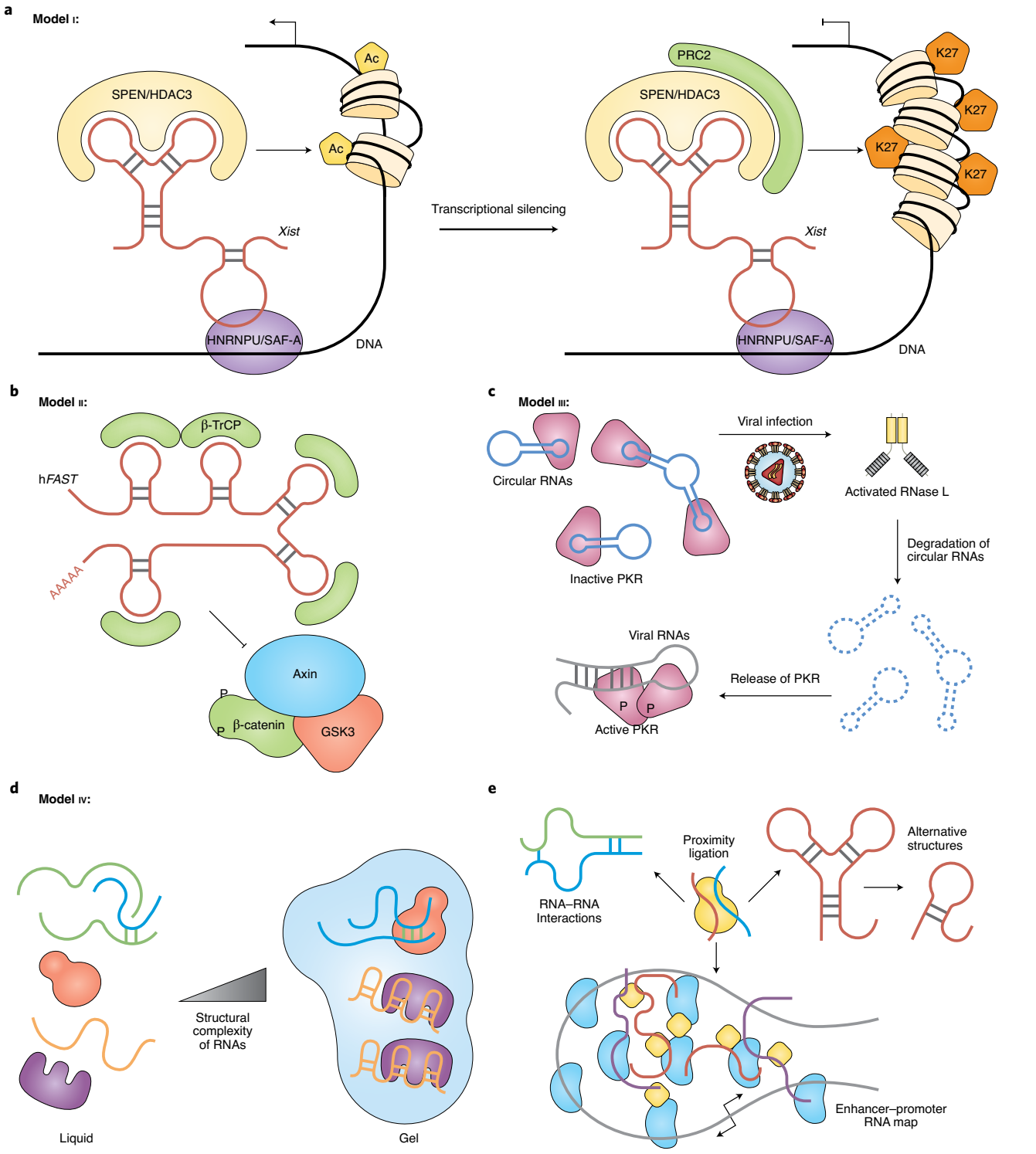
More examples include the steroid receptor RNA activator (*SRA*), an 870-nucleotide lncRNA that can activate several human sex hormone receptors in breast cancer[67]. Structural probing of SRA identified four independent domain structures containing 25 helices (domains I–IV). Importantly, the overall *SRA* secondary architecture is highly conserved across 45 species, and such a highly structured *SRA* interacts with a variety of proteins, including its own translation product SRA protein (or SRAP)[42]. In *Arabidopsis thaliana*, *COOLAIR* is an lncRNA expressed from the *FLOWERING LOCUS C* (*FLC*) locus and is important for vernalization[68]. SHAPE analysis revealed that *COOLAIR* forms numerous secondary structures with two unusual asymmetric 5′ right-hand-turn motifs showing evolutionary conservation[69]. Overall, these studies on functionally important lncRNAs exemplify how an lncRNA can form different structure motifs to interact with multiple proteins for different functions.

Binding of the same proteins with a multivalent platform. The second category comprises lncRNAs that interact with the same effector protein through multiple structural modules. Human FOXD3 antisense transcript 1 (h*FAST*) is a 547-nucleotide lncRNA that maintains Wnt signaling in human ESCs. h*FAST* specifically binds the E3 ubiquitin ligase β-transducin repeat-containing protein (β-TrCP), blocking its interaction with phosphorylated β-catenin. This interaction prevents β-catenin degradation in the deconstruction complex, leading to active WNT signaling, which promotes human ESC pluripotency[38]. In-cell SHAPE-MaP revealed that h*FAST* tends to form five similar but independent stem loops, making each h*FAST* molecule a multivalent β-TrCP-binding platform (Fig. 4b). Stoichiometry analyses revealed that >20% of β-TrCP can be sequestered by h*FAST* in the cytoplasm, suggesting that forming a multivalent protein-binding platform in a specific lncRNA may represent highly effective decoys. β-TrCP belongs to the Fbw (F-box/WD40 repeat-containing) protein family and contains seven WD40 repeats at the carboxy terminus. Truncation and in vitro binding assays revealed that h*FAST* interacts with β-TrCP

Fig. 4 | Understanding RNA secondary structures facilitates RNA functional studies. a, Model I. One lncRNA forms different structure motifs to interact with multiple proteins to support biological functions. *Xist* forms well-defined secondary structural domains that interact with multiple proteins in a domain-based, step-wise sequence of interactions during XCI. HDAC3, histone deacetylase 3; HNRNPU, heterogeneous nuclear ribonucleoprotein U; SAF-A, scaffold attachment factor A. **b**, Model II. One lncRNA forms a multivalent platform for binding a single effector protein. h*FAST* adopts five similar but independent stem loops for β-TrCP binding and acts as a protein decoy of β-TrCP to modulate Wnt signaling in human ESCs. GSK3, glycogen synthase kinase 3. **c**, Model III. Multiple circular RNAs form modules that interact with a group of proteins. Most examined endogenous circular RNAs tend to form compact intramolecularly double-stranded duplexes that allow circular RNAs act as a group to interact with the same proteins and regulate their functions. **d**, Model IV. Secondary structure-based RNA–RNA interactions determine LLPS. RNA–RNA interactions regulate LLPS or initiate the formation of condensates (which can regulate mRNA sorting) with known impacts related to diseases. e, Model V. RNA functions and regulations are deciphered from RNA secondary structure-based RNA interactomes. Crosslinking and proximity ligation-based approaches uncover RNA interactomes in cells, which greatly facilitate the identification of long-range RNA duplexes and even intermolecular RNA interactions.

via the WD40 repeat domain, which is also required for recognition of phosphorylated β-catenin. Importantly, h*FAST* is highly specific, barely interacting with other WD40 domain-containing proteins. It remains to be tested whether RNA secondary structures underlie this exquisite specificity, and whether other lncRNAs fall into this multivalent binding class.

Multiple circular RNAs form a binding module. Circular RNAs are covalently closed RNAs produced from precursor mRNA back-splicing of exons. They are widely expressed in eukaryotes with tissue- and cell-specific expression patterns (for reviews, see Wilusz[70] and Chen[71]). Despite generally low expression, an optimized SHAPE-MaP assay that can distinguish the structural conformation of the circular from the linear cognate RNAs revealed that many circular RNAs, but not their cognate linear RNAs, tend to form 16- to 26-base pair imperfect intramolecular RNA duplexes (Fig. 4c). Such structures allow different circular RNAs as a group to bind and regulate the innate immune double-stranded RNA



(dsRNA) receptor protein kinase R (PKR)[43], probably keeping PKR from unnecessary activation in normal cells. Upon poly(I:C) stimulation or encephalomyocarditis infection, which induces innate immune responses, circular RNAs are globally and rapidly degraded by the activated endonuclease RNase L—a process required for PKR activation in early cellular innate immune responses[43]. Whether this mode of RNA–protein interaction is prevalent remains unclear. However, many dsRNA-binding proteins bind to short or long dsRNA independent of primary sequences[72,73], and circular RNAs appear to interact with other nucleic acid-binding proteins, including OAS[43] and NF90/NF110 (ref. [74]). It is thus possible that circular RNAs also associate with these additional proteins for functions in immune response. Nevertheless, the secondary structural information about individual circular RNAs is incomplete, and additional work is needed to reveal the precise molecular basis of interactions between circular RNAs and PKR.

RNA structure underlies phase separation. Recent work has revealed that RNAs and proteins can organize into phase-separated membraneless organelles or granules that regulate gene expression by compartmentalizing and concentrating specific molecules[75,76]. The secondary structure of an mRNA may determine whether it will be recruited to or excluded from liquid organelles[77]. For example, the *CLN3* and *BNI1* mRNAs can interact with each other independent of their secondary structures, but they both interact with Whi3, a polyQ protein, via a secondary structure-dependent manner. As a result, association of *CLN3* and *BNI1* mRNAs with disrupted secondary structure occurred in the presence and absence of Whi3, indicating that mRNA self-association can be initiated in a protein-independent manner. SHAPE-MaP analysis of *CLN3* and *BNI1* mRNAs revealed that their structures can regulate mRNA sorting into distinct droplets via intermolecular interactions. In the presence of Whi3, most Whi3-binding sites revealed by SHAPE-MaP are exposed on stem loops in *CLN3* and *BNI1* mRNAs (Fig. 4d). Such secondary structure-dependent Whi3 binding can alter the secondary structures of target RNAs, thus promoting assembly of distinct droplets. How mRNA secondary structure influences selective uptake of cellular molecules and impacts liquid–liquid phase separation of messenger RNPs remains to be explored. In another study, multivalent base pairing of RNAs was found to be essential during gel formation without proteins. The length and number of RNA repeats often determine intermolecular base pairing. Examples include CAG repeats in Huntington's disease and spinocerebellar ataxias, CTG in myotonic dystrophy and the hexanucleotide GGGGCC associated with familial amyotrophic lateral sclerosis and frontotemporal dementia[78]. Interactions between repeat regions can result in the formation of microscopically observable RNA foci and influence subcellular RNA localization[79] (Fig. 4d).

Long noncoding RNAs can act as structural scaffolds within membraneless nuclear condensates[80]. Nuclear paraspeckle assembly transcript 1 (*NEAT1*) is a key structural component of paraspeckles and is essential for paraspeckle formation[81]. Although SHAPE probing of the *NEAT1* secondary structure revealed putative long-range RNA–RNA base-pairing interactions between its 5′ and 3′ ends[82], how this structure contributes to the liquid–liquid phase separation (LLPS)-based paraspeckle assembly remains unclear. In addition, interactions with key paraspeckle proteins, such as NONO, which are probably synergistically required for de novo paraspeckle formation, require further exploration. Future studies of the key structural modules of *NEAT1* (that is, 8–16.6 kilobases of *NEAT1*[83]) with and without the addition of NONO will provide new insights into the *NEAT1*-organized condensates.

Although many more RNA structural analyses are needed to decipher commonalities and differences between mechanisms of RNA function and regulation, these examples nevertheless highlight an important view that understanding the structural details of RNAs and their interacting partners is the key to understanding their modes of actions in gene expression regulation.

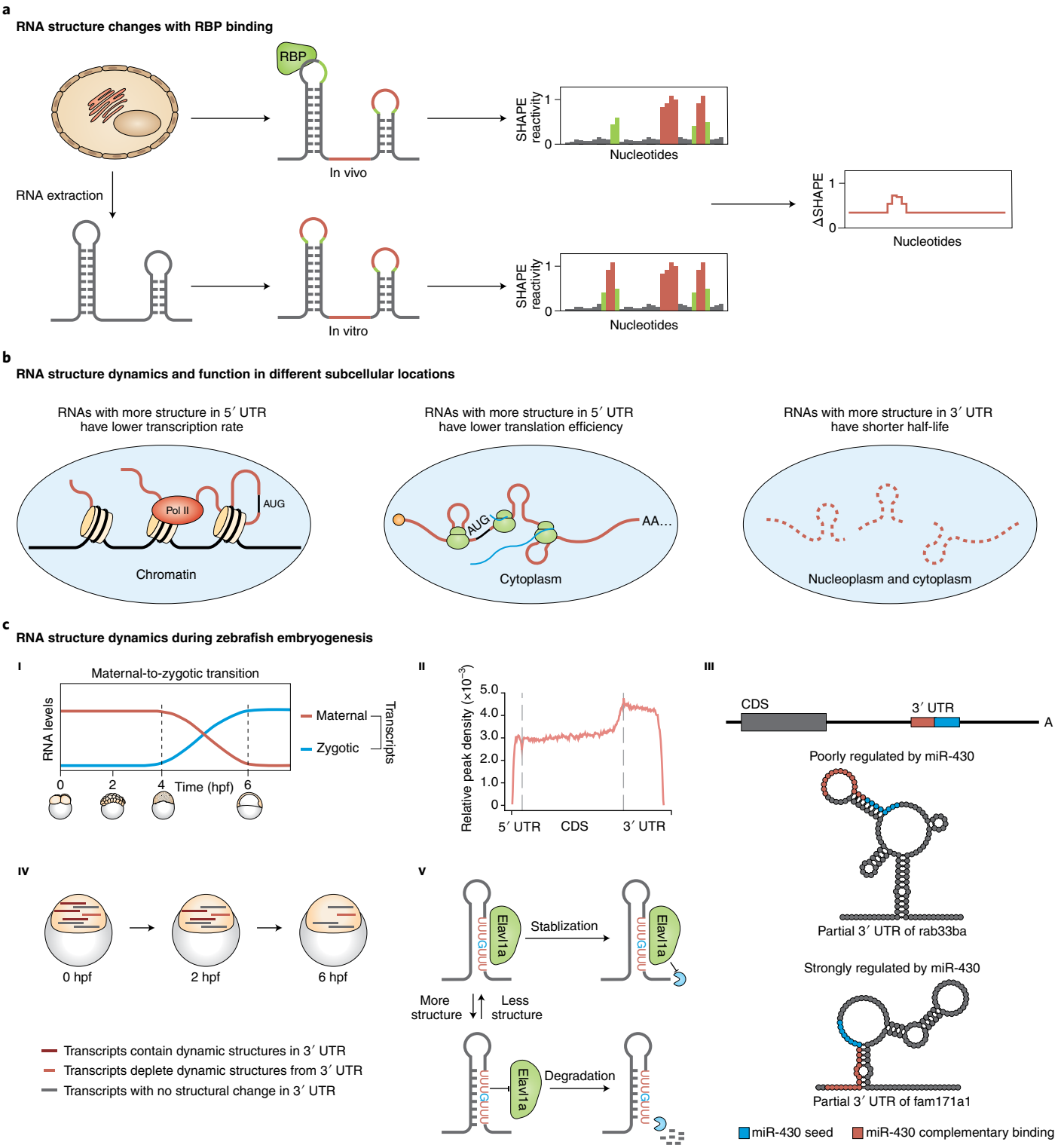
RNA secondary structure-based RNA interactomes. Many crosslinking and proximity ligation-based technologies, including CLASH, PARIS, LIGR-seq and SPLASH (Fig. 3), have been developed to dissect global interaction maps of RNA in cells, but several limitations and challenges exist due to false positive proximity ligation or biotinylation of soluble proteins in cell lysates. The recently developed RIC-seq method carries out RNA proximity ligation in situ, thus efficiently capturing global RNA duplexes and long-range loop–loop interactions efficiently[32]. Application of RIC-seq constructed an intra- and intermolecular RNA connectivity map in human cells and identified 642 RNA interaction hubs in HeLa cells. The conditions under which these distinct nuclear subdomain-localized lncRNAs form *trans* interactions remain to be explored; future global RNA structural mapping at a single-cell level will probably provide some clues (Fig. 4e).

Crosslinking and proximity ligation-based technologies have greatly facilitated the identification of RNA duplexes formed across long distances, and even intermolecular interactions. An important notion is that alternative structures frequently occur when one sequence base pairs with different intra- or interpartners under different conditions[23,32], suggesting the dynamic and flexible nature of RNA secondary structures. In contrast, many long-range alternative RNA structures are evolutionarily conserved[23], indicating that sophisticated regulatory mechanisms can be adopted to modulate structure-based RNA interactomes in cells.

RNA structural changes in gene regulation. In addition to the aforementioned individual cases that have uncovered specific RNA secondary structural mechanisms and functions, studies have also uncovered some general rules and principles of RNA secondary structure underlying key steps in RNA processing, localization and translation.

Fig. 5 | RNA regulation revealed by differential RNA structure analysis. a, In vivo and in vitro RNA structure differences. RNA is probed in living cells (in vivo) or following extraction from cells (in vitro). When comparing in vivo-probed and in vitro-probed RNA structures, a higher SHAPE reactivity of RNA elements in vitro indicates their interaction with other biomolecules, such as RBPs. Image adapted with permission from ref. [85], American Chemical Society. **b**, RNA structure differences across subcellular locations. RNA structure connects transcription, translation and RNA degradation. Image adapted with permission from ref. [84], Springer Nature. **c**, RNA structure changes and their functional roles during zebrafish embryogenesis. i. During the maternal-to-zygotic transition in zebrafish, maternal mRNAs decay rapidly, while zygotic genes are simultaneously activated. hpf, hours post-fertilization. ii. Metagene profile represents the dynamic structure regions are enriched in the 3′ UTR. CDS, coding sequence. iii. In vivo predicted secondary structure of miR-430 target sites identified in the 3′ UTRs of *rab33ba* and *fam171a1*. The 3′ UTR of *fam171a1*, which contains an miR-430 target site located in a single-stranded region, is strongly regulated by miR-430. For *rab33ba*, whose mir-430 target site is located in a double-stranded region, the miR-430-mediated regulation is less pronounced. iv. Transcripts with dynamic structures in their 3′ UTRs decay during the maternal-to-zygotic transition; depletion of portions of these dynamic regions extends their cellular half-life. v. Schematic showing that Elavl1a regulates maternal RNA in a structure-dependent manner. Image in (ii) reproduced with permission from ref. [87], Springer Nature; images in (iii) and (iv) adapted with permission from ref. [88], Springer Nature; image in (v) adapted with permission from ref. [87], Springer Nature.

Several genome-wide probing studies have revealed that RNA adopts very different structures in vivo and in vitro[14,19,84]. Interactions with proteins, metal ions and ligands are known to be a major cause of in vivo and in vitro RNA structural differences. SHAPE-MaP was developed to identify RNA binding protein (RBP) binding sites using RNA structure-probing data[85] (Fig. 5a). Differences in SHAPE reactivities obtained from SHAPE-MaP are calculated by subtracting in vivo SHAPE reactivities from in vitro reactivities. A positive difference in the SHAPE signal reports sites protected from modification in the cellular environment, while a negative difference in the SHAPE signal indicates enhanced reactivity in cells. By comparing mRNA structure between cell and cell-free environments in *Escherichia coli*, Mustoe et al. found that translation efficiency is correlated with ribosomal-binding-site (RBS) structure. Conserved structural elements were also discovered in 35% of UTRs, some of which were validated as novel functional protein-binding motifs[86].



The subcellular localization of RNA is intimately related to its function, so it is not surprising that RNA structures differ across subcellular locations. Probing RNA structure in chromatin, nucleoplasm and cytoplasm reveals dynamic structure information related to RNA regulation[84] (Fig. 5b). RNAs with a more double-stranded structure in the 5′ UTR have a lower transcription rate and a lower translation efficiency, presumably due to decreased ribosome binding. In addition, RNAs with a more double-stranded structure in the 3′ UTR tend to have a shorter half-life, probably as they are more accessible to RNA degradation machines.

Previous studies have focused on the global analysis of RNA structures in vivo at steady state. However, like other macromolecules, RNAs dynamically alternate conformational states to realize their biological function. Recently, the structural dynamics of mRNA was characterized during embryonic development of the zebrafish using icSHAPE[87] and DMS-seq[88] (Fig. 5c). After fertilization, maternal transcripts quickly begin to degrade[89]. It was found that specific 3′ UTR structures regulate maternal RNA degradation by affecting mir-430 activity[90]. The 3′ UTRs of these transcripts show more structural changes compared with the 5′ UTRs and coding regions (affecting the process of maternal RNA degradation[87,88]) and probably more time-controlled RNA events. The structurally variable regions within the 3′ UTRs of maternal transcripts contain known motifs for the binding of many RBPs, including those important for RNA degradation. The study shows that one RBP called Elavl1a binds to the UUUGUUU motif in the 3′ UTR of some transcripts and protects them from exonuclease digestion[87]. During development, these motifs become more structured, which corresponds to the release of Elavl1a from the RNA and triggers transcript degradation.

Future technology development for RNA structure probing

Sequencing-based RNA structure-probing technologies have developed rapidly and represent substantial progress over previous techniques. Datasets generated with these technologies have started to provide a global overview of how RNA structure is organized in vitro and/or in vivo. However, these strategies still have limitations, with probably the biggest challenge being the reliable association of structure data with specific biological functions. That is, although small-molecule-modification-based probing methods coupled with computational analysis can provide base pair-level structural models, experimental data that directly link biological functions to those structural models at scale are lacking.

To date, the most commonly practiced validation approach is the use of nucleotide mutations to disrupt RNA structures associated with a particular function or phenotype, and to rescue the phenotype using compensatory mutations predicted to rescue the structural change[27]. However, with the notable exception of some viruses, such methods have mainly been performed using synthesized RNA molecules. Another potentially insightful validation approach is the use of super-resolution imaging (such as atomic force microscopy (AFM); also called scanning force microscopy) to directly monitor conformational changes in multidomain structured RNAs. AFM is an emerging platform for studies of RNA structural dynamics[91]. It allows visualization, probing and manipulation of RNA molecules under physiological conditions without the need for labeling or staining[92]. For example, AFM analysis has revealed long-range interactions in the H11–H27 pseudoknot that are essential for the execution of maternally expressed gene 3 (*MEG3*) in stimulating the p53 tumor suppressor pathway[44].

Another area ripe for development is fine-scale structural probing. Current analyses have focused solely on signals representing the average behavior of the structuromes. However, given that RNA localization is closely tied to its function[93], the capacity to experimentally resolve differences between RNA structures in different subcellular compartments should deepen our understanding of how RNA functions are dynamically regulated in different locations of cells. Since one RNA can interact with different RBPs to form heterogeneous complexes in cells, subcellular RNA structure probing should help decouple how different RNA conformations are involved in distinct regulatory pathways or carry out different functions. Studies have also shown that mRNAs can form different secondary structures, which bind the same protein but function disparately[77]. It has also been found that the flexibility of RNA molecules can change and affect the formation of long-range interactions and higher-order RNA structures[94]. Combining fine-scale subcellular RNA enrichment and RNA structure probing may shed light on the dynamic changes in the spatial RNA structure within cells.

It must be emphasized that current methods for probing genome-wide RNA structures require large amounts of starting materials; thus, more sensitive techniques are sorely needed to support fine-scale structure-probing efforts. Such methods will enable the capture of structure information from rare samples, such as tissues from early development or clinical samples, thereby potentially uncovering new regulatory roles from RNA structures in development and diseases.

Finally, better bioinformatics pipelines are also required to enforce data quality control and support accurate quantifications of RNA structural data. RNA structural data obtained using different profiling methods are currently difficult to compare directly. There is an unmet need for bioinformatics methods that can perform normalization, comparison and/or integrative modeling of RNA structures from different datasets, including datasets obtained with different profiling methods. Last but not least, there is fast-growing interest in using artificial intelligence algorithms for RNA structure research. The most obvious application of artificial intelligence is the development of accurate and robust methods for predicting RNA secondary structure, with reported methods including SPOT-RNA[95] and MXfold2 (ref. [96]). Artificial intelligence has also been used to discover and model structural patterns related to RNA function. PrismNet, a deep neural network, has recently been developed to characterize structural patterns for protein–RNA binding. It can accurately predict protein–RNA interactions and their changes in a given cellular environment based on information including in vivo RNA structures and protein–RNA binding data for matched cell lines[97]. The capacity for accurate prediction of protein interaction partners for an lncRNA of interest will almost certainly help to uncover the biological role of lncRNAs.

Concluding remarks

Rapidly developing sequencing-based RNA secondary structure-probing technologies have supported exciting new insights at the levels of both individual- and large-scale RNA function. Overcoming the current challenges can usher in RNA structural insights at the subcellular, single-cell and systems levels. Development of the next generation of technologies and analysis algorithms should enable investigations of the interplay between RNA structural changes, RNA modifications and RBP binding, which are likely to be complex and sophisticated. For example, post-transcriptional m6A modification has been shown to regulate HNRNPC binding through mRNA and lncRNA structural remodeling[98]. An array of methods are needed to dissect the functional impacts of chemically identified RNA secondary structural motifs. Efforts are also warranted to obtain native-shaped, full-length structures of known modules of RNA/RNA–protein complexes. Isolating such complexes will also allow methods such as AFM, small-angle X-ray scattering, crystallography, NMR and cryogenic electron microscopy to be applied for deriving tertiary RNA structures. Ultimately, tertiary RNA structures of lncRNA can provide precise base-pairing information at single-base resolution, which can deepen understanding of their functional mechanisms[44,53,99,100].

Received: 5 October 2020; Accepted: 23 April 2021;
Published online: 25 June 2021

References

1. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
2. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* **8**, 776–790 (2007).
3. Thomas, C. & Gluick, D. E. D. Tertiary structure of ribosomal RNA. *Curr. Opin. Struct. Biol.* **2**, 338–344 (1992).
4. Mattick, J. S. RNA regulation: a new genetics? *Nat. Rev. Genet.* **5**, 316–323 (2004).
5. Yao, R. W., Wang, Y. & Chen, L. L. Cellular functions of long noncoding RNAs. *Nat. Cell Biol.* **21**, 542–551 (2019).
6. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904–914 (2011).
7. Underwood, J. G. et al. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7**, 995–1001 (2010).
8. Kertesz, M. et al. Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103–107 (2010).
9. Lucks, J. B. et al. Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl Acad. Sci. USA* **108**, 11063–11068 (2011).
10. Kudla, G., Granneman, S., Hahn, D., Beggs, J. D. & Tollervey, D. Cross-linking, ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast. *Proc. Natl Acad. Sci. USA* **108**, 10010–10015 (2011).
11. Li, F. et al. Global analysis of RNA secondary structure in two metazoans. *Cell Rep.* **1**, 69–82 (2012).
12. Ding, Y. et al. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).
13. 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.* **15**, 491 (2014).
14. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2014).
15. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–965 (2014).
16. Silverman, I. M. et al. RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. *Genome Biol.* **15**, R3 (2014).
17. Talkish, J., May, G., Lin, Y., Woolford, J. L. Jr. & McManus, C. J. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* **20**, 713–720 (2014).
18. Ramani, V., Qiu, R. & Shendure, J. High-throughput determination of RNA structure by proximity ligation. *Nat. Biotechnol.* **33**, 980–984 (2015).
19. Spitale, R. C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486–490 (2015).
20. Sugimoto, Y. et al. hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature* **519**, 491–494 (2015).
21. Nguyen, T. C. et al. Mapping RNA–RNA interactome and RNA structure in vivo by MARIO. *Nat. Commun.* **7**, 12023 (2016).
22. Aw, J. G. et al. In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* **62**, 603–617 (2016).
23. Lu, Z. et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267–1279 (2016).
24. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B. J. Global mapping of human RNA–RNA interactions. *Mol. Cell* **62**, 618–626 (2016).
25. Ritchey, L. E. et al. Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure in vivo. *Nucleic Acids Res.* **45**, e135 (2017).
26. Foley, S. W. et al. A global view of RNA–protein interactions identifies post-transcriptional regulators of root hair cell fate. *Dev. Cell* **41**, 204–220.e5 (2017).
27. Zubradt, M. et al. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **14**, 75–82 (2017).
28. Metkar, M. et al. Higher-order organization principles of pre-translational mRNPs. *Mol. Cell* **72**, 715–726.e3 (2018).
29. Ziv, O. et al. COMRADES determines in vivo RNA structures and interactions. *Nat. Methods* **15**, 785–788 (2018).
30. Morf, J. et al. RNA proximity sequencing reveals the spatial organization of the transcriptome in the nucleus. *Nat. Biotechnol.* **37**, 793–802 (2019).
31. Zinshteyn, B. et al. Assaying RNA structure with LASER-seq. *Nucleic Acids Res.* **47**, 43–55 (2019).
32. Cai, Z. et al. RIC-seq for global in situ profiling of RNA–RNA spatial interactions. *Nature* **582**, 432–437 (2020).
This paper presents RIC-seq, a technology to globally capture higher-order transcriptome structure via proximity ligation. The study also uncovered a widespread role of RNA in gene regulation by remodeling chromatin structure.
33. Twittenhoff, C. et al. Lead-seq: transcriptome-wide structure probing in vivo using lead(II) ions. *Nucleic Acids Res.* **48**, e71 (2020).
34. Weng, X. et al. Keth-seq for transcriptome-wide RNA structure mapping. *Nat. Chem. Biol.* **16**, 489–492 (2020).
35. Li, F. et al. Regulatory impact of RNA secondary structure across the *Arabidopsis* transcriptome. *Plant Cell* **24**, 4346–4359 (2012).
36. Gosai, S. J. et al. Global analysis of the RNA–protein interaction and RNA secondary structure landscapes of the *Arabidopsis* nucleus. *Mol. Cell* **57**, 376–388 (2015).
37. Bevilacqua, P. C., Ritchey, L. E., Su, Z. & Assmann, S. M. Genome-wide analysis of RNA secondary structure. *Annu. Rev. Genet.* **50**, 235–266 (2016).
38. Guo, C. J. et al. Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. *Cell* **181**, 621–636.e22 (2020).
39. Smola, M. J. et al. SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the *Xist* lncRNA in living cells. *Proc. Natl Acad. Sci. USA* **113**, 10322–10327 (2016).
This paper developed SHAPE-MaP by mutational profiling and used the technology to obtain a detailed structural architecture of Xist.
40. Somarowthu, S. et al. HOTAIR forms an intricate and modular secondary structure. *Mol. Cell* **58**, 353–361 (2015).
This study illustrates a tour-de-force example for advancing understanding of lncRNA function (for example, protein–RNA interaction) through RNA secondary structure probing.
41. Xue, Z. et al. 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 (2016).
42. Coleman, K. M., Lam, V., Jaber, B. M., Lanz, R. B. & Smith, C. L. SRA coactivation of estrogen receptor-α is phosphorylation-independent, and enhances 4-hydroxytamoxifen agonist activity. *Biochem. Biophys. Res. Commun.* **323**, 332–338 (2004).
43. Liu, C. X. et al. Structure and degradation of circular RNAs regulate PKR activation in innate immunity. *Cell* **177**, 865–880.e21 (2019).
44. Uroda, T. et al. Conserved pseudoknots in lncRNA *MEG3* are essential for stimulation of the p53 pathway. *Mol. Cell* **75**, 982–995.e9 (2019).
This study characterizes the secondary and tertiary structures of MEG3 in vivo and in vitro, and identifies the conserved long-range tertiary interactions between motifs, which is essential for its tumor suppressive function.
45. Spitale, R. C. et al. RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9**, 18–20 (2013).
46. Strobel, E. J., Yu, A. M. & Lucks, J. B. High-throughput determination of RNA structures. *Nat. Rev. Genet.* **19**, 615–634 (2018).
47. Bevilacqua, P. C. & Assmann, S. M. Technique development for probing RNA structure in vivo and genome-wide. *Cold Spring Harb. Perspect. Biol.* https://doi.org/10.1101/cshperspect.a032250 (2018).
48. Huber, R. G. et al. Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat. Commun.* **10**, 1408 (2019).
49. Li, P. et al. Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity. *Cell Host Microbe* **24**, 875–886.e5 (2018).
50. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665 (2013).
51. Xiang, J. F. et al. Human colorectal cancer-specific *CCAT1-L* lncRNA regulates long-range chromatin interactions at the *MYC* locus. *Cell Res.* **24**, 513–531 (2014).
52. Nicholson, B. L. & White, K. A. Functional long-range RNA–RNA interactions in positive-strand RNA viruses. *Nat. Rev. Microbiol.* **12**, 493–504 (2014).
53. Chillon, I. & Marcia, M. The molecular structure of long non-coding RNAs: emerging patterns and functional implications. *Crit. Rev. Biochem. Mol. Biol.* **55**, 662–690 (2020).
54. Kopp, F. & Mendell, J. T. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**, 393–407 (2018).
55. Wan, Y. et al. Genome-wide measurement of RNA folding energies. *Mol. Cell* **48**, 169–181 (2012).
56. Clark, M. B. et al. Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–898 (2012).
57. McHugh, C. A. et al. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
58. Chu, C. et al. Systematic discovery of *Xist* RNA binding proteins. *Cell* **161**, 404–416 (2015).
59. Minajigi, A. et al. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* https://doi.org/10.1126/science.aab2276 (2015).

60. Maenner, S. et al. 2-D structure of the A region of *Xist* RNA and its implication for PRC2 association. *PLoS Biol.* **8**, e1000276 (2010).
61. Ilik, I. & Akhtar, A. *roX* RNAs: non-coding regulators of the male X chromosome in flies. *RNA Biol.* **6**, 113–121 (2009).
62. Ilik, I. A. et al. Tandem stem-loops in *roX* RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol. Cell* **51**, 156–173 (2013).
63. Rinn, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
64. Tsai, M. C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
65. Davidovich, C. et al. Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol. Cell* **57**, 552–558 (2015).
66. Klattenhoff, C. A. et al. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* **152**, 570–583 (2013).
67. Lanz, R. B. et al. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* **97**, 17–27 (1999).
68. Csorba, T., Questa, J. I., Sun, Q. & Dean, C. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc. Natl Acad. Sci. USA* **111**, 16160–16165 (2014).
69. Hawkes, E. J. et al. COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures. *Cell Rep.* **16**, 3087–3096 (2016).
70. Wilusz, J. E. A 360 degrees view of circular RNAs: from biogenesis to functions. *Wiley Interdiscip. Rev. RNA* **9**, e1478 (2018).
71. Chen, L. L. The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat. Rev. Mol. Cell Biol.* **21**, 475–490 (2020).
72. Chen, T. et al. ADAR1 is required for differentiation and neural induction by regulating microRNA processing in a catalytically independent manner. *Cell Res.* **25**, 459–476 (2015).
73. Patino, C., Haenni, A. L. & Urcuqui-Inchima, S. NF90 isoforms, a new family of cellular proteins involved in viral replication? *Biochimie* **108**, 20–24 (2015).
74. Li, X. et al. Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. *Mol. Cell* **67**, 214–227.e7 (2017).
75. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
76. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* https://doi.org/10.1126/science.aaf4382 (2017).
77. Langdon, E. M. et al. mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **360**, 922–927 (2018).
This paper is one of the first reports about experimentally investigating how mRNA secondary structure can drive phase separation.
78. La Spada, A. R. & Taylor, J. P. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* **11**, 247–258 (2010).
79. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
80. Chujo, T., Yamazaki, T. & Hirose, T. Architectural RNAs (arcRNAs): a class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochim. Biophys. Acta* **1859**, 139–146 (2016).
81. West, J. A. et al. Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J. Cell Biol.* **214**, 817–830 (2016).
82. Yu, X., Li, Z., Zheng, H., Chan, M. T. & Wu, W. K. NEAT1: a novel cancer-related long non-coding RNA. *Cell Prolif.* https://doi.org/10.1111/cpr.12329 (2017).
83. Yamazaki, T. et al. Functional domains of *NEAT1* architectural lncRNA induce paraspeckle assembly through phase separation. *Mol. Cell* **70**, 1038–1053.e7 (2018).
84. Sun, L. et al. RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.* **26**, 322–330 (2019).
This study generated a comprehensive resource on the secondary structures of RNAs isolated from different subcelluar locations, compared the structural changes for the same RNA in different locations and studied how protein binding and RNA modification correlated with these structural changes.
85. Smola, M. J., Calabrese, J. M. & Weeks, K. M. Detection of RNA–protein interactions in living cells with SHAPE. *Biochemistry* **54**, 6867–6875 (2015).
86. Mustoe, A. M. et al. Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* **173**, 181–195.e18 (2018).
This paper applied SHAPE-MaP to derive high-resolution structure models for endogenous transcripts in *E. coli*, and provides new insights for understanding the pervasive and fundamental roles of RNA structure in gene expression.
87. Shi, B. et al. RNA structural dynamics regulate early embryogenesis through controlling transcriptome fate and function. *Genome Biol.* **21**, 120 (2020).
88. Beaudoin, J. D. et al. Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* **25**, 677–686 (2018).
By applying DMS-seq and icSHAPE—two RNA secondary structure-probing methods—this paper and ref. [87], respectively, characterized mRNA structuromes during early zebrafish embryogenesis and revealed regulatory roles of RNA structures in mRNA translation and maternal RNA degradation.
89. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–3042 (2009).
90. Giraldez, A. J. et al. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
91. Schon, P. Imaging and force probing RNA by atomic force microscopy. *Methods* **103**, 25–33 (2016).
92. Uroda, T. et al. Visualizing the functional 3D shape and topography of long noncoding RNAs by single-particle atomic force microscopy and in-solution hydrodynamic techniques. *Nat. Protoc.* **15**, 2107–2139 (2020).
93. Buxbaum, A. R., Haimovich, G. & Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **16**, 95–109 (2015).
94. Van Treeck, B. & Parker, R. Emerging roles for intermolecular RNA–RNA interactions in RNP assemblies. *Cell* **174**, 791–802 (2018).
95. Singh, J., Hanson, J., Paliwal, K. & Zhou, Y. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat. Commun.* **10**, 5407 (2019).
96. Sato, K., Akiyama, M. & Sakakibara, Y. RNA secondary structure prediction using deep learning with thermodynamic integration. *Nat. Commun.* **12**, 941 (2021).
97. Sun, L. et al. Predicting dynamic cellular protein–RNA interactions by deep learning using in vivo RNA structures. *Cell Res.* https://doi.org/10.1038/s41422-021-00476-y (2021).
This study describes the development of PrismNet, a deep neural network that accurately predicts in vivo protein–RNA interactions based on models generated from RNA structure data alongside RNA–RBP profiling data obtained from the same cell types of interest.
98. Liu, N. et al. N^6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* **518**, 560–564 (2015).
99. Kim, D. N. et al. Zinc-finger protein CNBP alters the 3-D structure of lncRNA *Braveheart* in solution. *Nat. Commun.* **11**, 148 (2020).
100. Liu, F., Somarowthu, S. & Pyle, A. M. Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat. Chem. Biol.* **13**, 282–289 (2017).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) (91940303, 31725009, 31830108, 31821004) and the HHMI International Program (55008728) to L.-L.C., and the National Natural Science Foundation of China (Grants No. 91740204, 91940306 and 31761163007) and the State Key Research Development Program of China (Grant No. 2018YFA0107603) to Q.C.Z. L.-L.C. acknowledges the support from the XPLORER PRIZE.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to L.-L.C. or Q.C.Z.

Peer review information *Nature Chemical Biology* thanks Yiliang Ding, Brian Gregory and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.