Advanced Biotechnology

REVIEW

RNA structure: implications in viral infections and neurodegenerative diseases

Suiru Lu[2,3†], Yongkang Tang[2†], Shaozhen Yin[2] and Lei Sun[1,2,3,4*]

Abstract

RNA is an intermediary between DNA and protein, a catalyzer of biochemical reactions, and a regulator of genes and transcripts. RNA structures are essential for complicated functions. Recent years have witnessed rapid advancements in RNA secondary structure probing techniques. These technological strides provided comprehensive insights into RNA structures, which significantly contributed to our understanding of diverse cellular regulatory processes, including gene regulation, epigenetic regulation, and post-transactional regulation. Meanwhile, they have facilitated the creation of therapeutic tools for tackling human diseases. Despite their therapeutic applications, RNA structure probing methods also offer a promising avenue for exploring the mechanisms of human diseases, potentially providing the key to overcoming existing research constraints and obtaining the in-depth information necessary for a deeper understanding of disease mechanisms.

Keywords RNA structure probing, RNA secondary structure, Post-transcription regulation, Human diseases

1 Introduction

RNA was discovered over a century ago, igniting intensive investigations into its biochemical properties. Over time, scientists unveiled the intricate connections between RNA and DNA, shedding light on RNA's pivotal role in the complicated process of protein synthesis (Brenner 1961; Crick 1970). Despite its primary role in polypeptide chain encoding, the improvements in molecular biology tools and the availability of sequencing technology led to the discovery of a wide variety of non-coding RNAs and associated phenomena, such as X Inactive Specific Transcript (Xist) and X chromosome

inactivation (Lyon 1972). Concurrently, another wave is emerging with the revelation that RNAs exhibit catalytic functions (Cech 1989; Kruger et al. 1982). Moreover, the discovery of RNA viruses in the mid-twentieth century and the identification of human immunodeficiency virus (HIV)(Yilmaz 2001) led to significant advancements in investigating RNA viruses, demonstrating that RNA can act as a viable genetic material. The revelation that RNA, capable of carrying genetic information and possessing catalytic functions, could potentially be the genesis of life, led to the formulation of the "RNA world" hypothesis (Bartel and Szostak 1993; Noller et al. 1992).

In the last 20 years, RNA research has entered a new era, mainly focusing on its functions, disease-related mechanisms, and potential clinical applications. For example, scientists are interested in exploring RNA-interacting molecules, developing messenger RNA (mRNA) vaccines, and uncovering mechanisms that connect RNA with human diseases. One of the fundamental topics that underlie the functional study, mechanism discovery, and application development is RNA structure. Just like the structure–function relationship in proteins, the understanding of RNA structures would provide a more

---

†Suiru Lu and Yongkang Tang are co-first authors.

*Correspondence:
Lei Sun
sunlei0227@sdu.edu.cn
1 Pingyuan Laboratory, Xinxiang, Henan 453007, China
2 Shandong Provincial Key Laboratory of Animal Cell and Developmental Biology, School of Life Sciences, Shandong University, Qingdao 266237, China
3 Taishan College, Shandong University, Qingdao 266237, China
4 State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, China

Lu *et al. Advanced Biotechnology*   (2024) 2:3

Page 2 of 16

rational and precise approach to studying or engineering methods for RNA drug/vaccine development and human disease mechanism research.

Nowadays, many researchers in this field are trying to construct the structure–function maps for RNA molecules. Importantly, RNA molecules exhibit a notably high degree of structural variability, particularly in the absence of associated proteins. This diversity encompasses a broad spectrum of sequence-to-secondary structure relationships, making RNA research particularly challenging. Conventional techniques, including X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM) frequently produce structures that deviate significantly from the actual biological configurations(Schroeder et al. 2004). Computational methodologies have sought to address this issue by employing thermodynamic calculations to derive the structure of RNA in its minimum free energy state (Hofacker 2003; Mathews et al. 2007). This approach, however, remains imperfect as the structure of RNA is regulated by multiple factors within the cellular environment.

Fortunately, there are now both more advanced experimental and computational methods available for RNA structure determination. People are now utilizing machine learning (Zhao et al. 2021) and deep learning (Fu et al. 2022; Sato et al. 2021; Singh et al. 2019; Wang et al. 2023; Townshend et al. 2021; Szikszai 2022) technologies for RNA structure prediction, which have been proven to be effective. By leveraging these technologies, researchers can analyze large amounts of RNA data and predict the structures of RNA molecules with greater accuracy. For the experimental methods, novel approaches that investigate RNA structures using enzymes and small molecule reactions were developed (Silverman et al. 2016). These methods enable researchers to extract information from the overall or specific RNA structures in vivo after serval technology iterations. These new methods, which have paved the way for studying the RNA structurome, not only facilitate the construction of structure–function connections in mechanism research but also provide solid information for developing and refining clinically applicable strategies.

In this review, we will delve into how RNA structures intricately participate in regulating intracellular functions, and, more importantly, how these functional insights can be applied to understand and manage human diseases. We will commence by briefly introducing pertinent structure probing techniques. Meanwhile, we will provide an overview of the significance of RNA structures concerning transcriptional and post-transcriptional regulations. Furthermore, we will discuss the state-of-the-art tools presently in development for potential

applications in clinical therapeutics. Lastly, we will present two practical case studies illustrating the application of structure probing methods in exploring disease mechanisms.

2 RNA structurome probing technologies and their applications in SARS-CoV-2

2.1 Methods for probing RNA structurome

Over the last decade, substantial progress has been made in RNA chemical probing methods, enabling the comprehensive analysis of RNA structures in living cells at a genome-wide scale (Table 1). These RNA structure probing technologies can be broadly classified into two main categories (Piao et al. 2017).

The first class is based on small-molecule modifications, which can provide insights into base-pairing possibilities (Wang et al. 2021). For instance, Dimethyl sulfide (DMS) and 2-(Azidomethyl) nicotinic acid imidazolide (NAI-N_3) are two commonly used chemicals to probe single-stranded RNA (ssRNA) regions. The DMS can modify unpaired adenines (A) and cytosines (C) (Lightfoot and Hall 2014), while NAI-N_3 can unbiasedly react with the unpaired ribose of unstructured nucleotides (Spitale et al. 2015). Consequently, this allows the identification of modified sites based on reverse transcription (RT), detecting stop sites of reverse transcriptase extension or chemically induced mutation positions.

The second class relies on crosslinking and proximity ligation (Kudla et al. 2020). Distinguishing itself from base modification techniques, crosslinking-based methodologies offer the advantage of directly capturing long-range RNA-RNA interactions. These processes involve crosslinking RNA within cells, followed by the extraction and fragmentation of RNA. Pairs of interacting fragments are then joined through ligation. The crucial information about RNA interactions is retained within these chimeric ligation products, which can be subsequently identified through sequencing and bioanalysis (Lu et al. 2016).

These aforementioned methods have enabled scientists to analyze the RNA secondary structure of various transcripts, including those in human, mouse cell lines, and certain viruses (Sun et al. 2019; Sun et al. 2021a; Lan et al. 2022). However, small chemical molecules, such as DMS, are unable to modify certain bases, and they also exhibit low efficiency in modifying bases with low solvent accessibility. Hence, certain studies are dedicated to refining the chemical molecules employed in structural probing techniques, for example, the development of 2-aminopyridine-3-carboxylic acid imidazolide (2A3) (Marinus et al. 2021), 2-methyl-3-furoic acid imidazolide (FAI) (Spitale et al. 2013), and NAI-N_3 for in vivo RNA studies (Spitale et al. 2015). The RNA structural information obtained can subsequently be used for

Table 1 Summary of high-throughput approaches to probing RNA structure

Methods	Reagent	Features	Limitations	Refs
Chemical probing – RT stop readout				
SHAPE-seq 1.0/2.0	1M7, NMIA, BzCN	Unbiased probing four bases	SHAPE reagents cannot penetrate cell membranes	Lucks et al. 2011; Loughrey, et al. 2014)
DMS-seq	DMS	DMS can penetrate cell membranes which makes it suitable for probing in vivo RNA bases	Nucleotide bias—Only probing adenines and cytosines	Rouskin et al. 2014)
Structure-seq	DMS	Similar to DMS-Seq with additional DMSO controls	Nucleotide bias	Fang et al. 2015; Ding et al. 2014)
Structure-seq2	DMS	Additional biotinylated nucleotide eliminating purification steps to improve library building efficiency	Nucleotide bias	Ritchey et al. 2017)
Mod-seq	DMS	Similar to DMS-Seq with a particular focus on rRNA	Nucleotide bias	Talkish et al. 2014)
icSHAPE	NAI-N_3	Probing genome-wide in vivo RNA structures without nucleotide bias Modified RNA molecules are enriched by biotin isolation to reduce background noises	Cannot generate direct base pairing information at modification sites Complicated library construction process	Spitale et al. 2015)
smartSHAPE	NAI-N_3	Similar with icSHAPE but eliminated endogenous modification by RNaseI and biotin enrichment Control groups are not required Ultra-low input	Adopting the RT-stop strategy, cannot generate direct base pairing information at modification sites	Piao et al. 2022)
LASER-Seq	NAz	Light-generated probing Reactivity detected correlates with solvent accessibility	Nucleotide bias	Zinshteyn et al. 2019)
Lead-seq	Lead(II) irons	Probing single base, but modification patterns for probing sites are different from other probing signals	Lead(II) irons are toxic	Twittenhoff et al. 2020)
Keth-seq	N3-kethoxal	Probing G-quadruplexes	Only probing guanosine	Weng et al. 2020)
Chemical probing – RT mutation readout				
SHAPE-MaP	NAI,NAI-N_3, 1M7, 1M6, NMIA	Single sequencing reads include more structure information than stop strategy	Higher sequencing depth Complicated calculation methods	Siegfried et al. 2014; Luo et al. 2021)
DMS-MaPseq	DMS	Similar to SHAPE-MaP	Nucleotide bias	Zubradt et al. 2017)
RNA–RNA interaction –proximity ligation				
PARIS	AMT + UVA	Direct capture of long-range in vivo RNA-RNA interactions RNA enrichment is conducted by 2D gel filtration UV light exposure	Efficiency of proximity ligation is low AMT crosslinking	Lu et al. 2016)
SPLASH	Biotinylated psoralen	Direct capture of in vivo RNA-RNA interactions	Efficiency of proximity ligation is low Psoralen crosslinking	Aw et al. 2016)
COMRADES	Psoralen	Direct capture of long-range in vivo RNA-RNA interactions Involved two enrichment steps	Efficiency of proximity ligation is low Psoralen crosslinking	Ziv et al. 2018)
RIC-seq	Formaldehyde	Capturing protein-mediated RNA-RNA interactions	Efficiency of proximity ligation is low Formaldehyde crosslinking	Cai et al. 2020)
CLASH	UVA	Direct capture of long-range RNA-RNA interactions	Efficiency of proximity ligation is low	Helwak et al. 2013)

Lu *et al. Advanced Biotechnology*   (2024) 2:3

Page 4 of 16

Table 1 (continued)

Methods	Reagent	Features	Limitations	Refs
MARIO	UVA	Direct capture of RNA-RNA interactions	RNA-RNA interactions detection has to be protein-mediated Efficiency of proximity ligation is low	Nguyen et al. 2016)
LIGR-seq	AMT	Direct capture of long-range RNA-RNA interactions	Efficiency of proximity ligation is low AMT crosslinking	Sharma et al. 2016)

1M7 1- methyl-7-nitroisatoic anhydride (1M7), *Amt* psoralen derivative 4'-aminomethyltrioxalen, *COMRADES* Crosslinking of matched RNAs and deep sequencing, *DMS* Dimethyl sulfide, *icSHAPE* in vivo SHAPE; LIGR-seq, Ligation of interacting RNA and high-throughput sequencing, *MARIO* Mapping RNA interactome *in vivo*, *SHAPE* Selective 2'-Hydroxyl Acylation analyzed by Primer Extension), *Kethoxal* 1,1-dihydroxy-3-ethoxy-2-butanone, *LASER-Seq* Light Activated Structural Examination of RNA by high throughput sequencing, *NAI-N$_3$* Azide-modified 2-methylnicotinic acid imidazolide, *NAz* Nicotinoyl azide, *smartSHAPE* Small amount random RT icSHAPE

various applications, including but not limited to designing targeted drugs, developing vaccines, and searching for RNA-binding proteins. For detailed information on probing technologies, one could kindly refer to the following reviews (Wang et al. 2021; Qian et al. 2019; Spitale and Incarnato 2023).

2.2 Application of RNA structurome probing methods in SARS-CoV-2 research

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for the 2020 COVID-19 pandemic, and it belongs to the Betacoronavirus genus (Lu et al. 2020a). The genome of SARS-CoV-2 consists of single-stranded, positive-sense RNA and is known to be among the largest RNA viruses with approximately 30 kilobases (kb) in size (Jackson et al. 2022; Masters 2006). RNA genome itself, especially the RNA complex structures are integral to their functions and life cycle. Therefore, RNA structural elements and their potential for drug development against SARS-CoV-2 are actively researched to tackle virus-induced pathologies (Sun et al. 2021a). To achieve this, it is key to understand the RNA secondary structures of SARS-CoV-2.

Chemical probing is a critical strategy that has long been used for characterizing complex structures of RNA molecules. In recent studies, various models for the RNA secondary structure of the entire SARS-CoV-2 genome in human or monkey cells have been determined using RNA chemical probing techniques (Sun et al. 2021a; Huston et al. 2021). Research involving the chemical probing of the SARS-CoV-2 RNA genome can be broadly categorized into four approaches, which belong to two strategies mentioned in the previous section (Fig. 1). As demonstrated by the small-molecule modifications strategies shown in Fig. 1a, first, *in vivo* click selective 2-hydroxyl acylation and profiling experiment (icSHAPE) study conducted by Sun et al. used random N RT-primers to acquire structural information for both SARS-CoV-2 and the human transcriptome (Sun et al. 2021a). Similarly, the library construction could also be performed by adding an adaptor to the RNA before reverse transcription and using the adaptor sequence as the RT primer target. However, an additional genome fragmentation step needs to be included before adapter ligation (Spitale et al. 2015). Meanwhile, SHAPE with mutational profiling (SHAPE-MaP) and DMS mutational profiling with sequencing (DMS-MaPseq) obtained structural information specifically for the virus with specific primers designed for SARS-CoV-2 (Lan et al. 2022; Huston et al. 2021; Manfredonia et al. 2020).

One major concern for DMS or NAI-N$_3$ probing methods is that they could not get long-range RNA-RNA interaction information. Fortunately, this can be tackled by utilizing the proximity ligation method to uncover interactions between viral RNA and other viral RNA molecules, as well as interactions between viral RNA and host RNA. With crosslinking and proximity ligation strategies (Fig. 1b), sequencing of psoralen cross-linked, ligated and selected hybrids (SPLASH) technique was used to acquire virus-virus, host-host, and virus-host RNA interactions (Zhang et al. 2021; Yang et al. 2021). Other methods involving an initial specific enrichment of SARS-CoV-2 RNA or virion are COMRADES (Ziv et al. 2020) and RNA in situ conformation sequencing (RIC-seq) (Cao et al. 2021), which highly improved the interested RNA probing depth. The development of these approaches has been instrumental in revealing the structure of hairpin-type pseudoknots, such as the frameshifting element (FSE) (Zhang et al. 2021; Ziv et al. 2020).

Thanks to the aforementioned RNA-structure probing technologies, numerous RNA structural elements have been identified within the SARS-CoV-2 genome (Sun et al. 2021a; Huston et al. 2021; Manfredonia et al. 2020). For example, multiple conserved RNA structural regions have been discovered, particularly those that are well-suited for designing antisense oligonucleotide
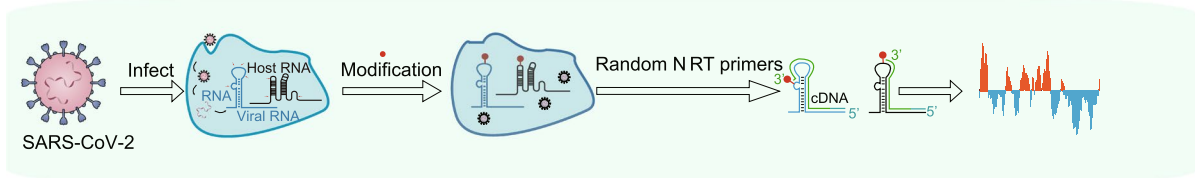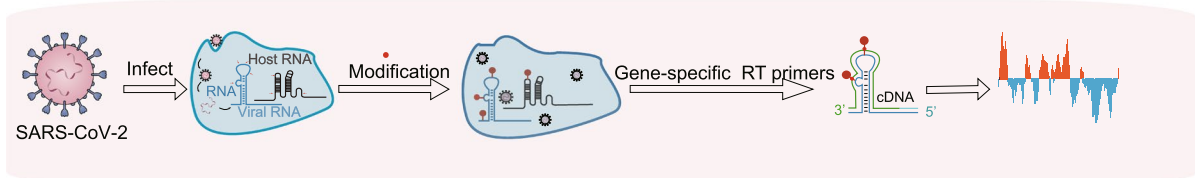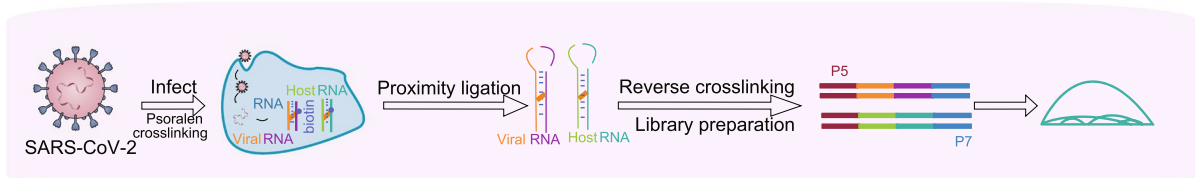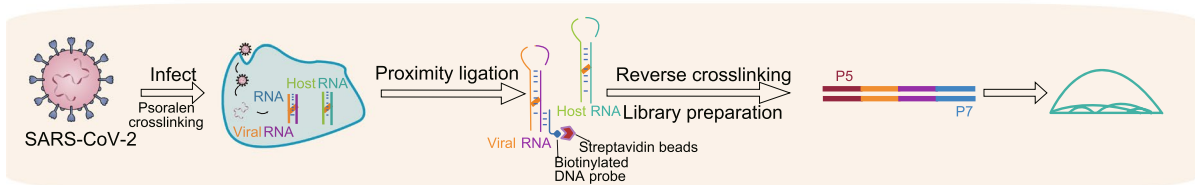
Fig. 1 Illustration of the application of two classes of RNA structure probing methods in the context of SARS-CoV-2. **a,** Foot Printing-Based Probing Methods Workflow. Following SARS-CoV-2 infection, DMS or NAI-N$_3$ is employed to modify the single-stranded RNA bases, including A and C, or the ribose of all four bases (A, U, C, G), respectively. These modifications result in distinct markers (red dots in the picture), which are subsequently converted into RT truncations or RT-induced mutations during reverse transcription. This process is facilitated by the use of either random N primers or gene-specific primers. The random N primer strategy provides information about both host and virus RNA structures, while the gene-specific primers exclusively yield information about the virus's RNA structure. The sites of RT truncation or RT mutations are then identified through sequencing and further analyzed using bioinformatics tools. **b,** Proximity Ligation-Based Probing Methods Workflow. After SARS-CoV-2 infection, RNAs are crosslinked by psoralen at regions where base pairing or interactions occur. The crosslinked RNAs are then fragmented. Proximity ligation is performed to connect viral RNA with other viral RNA molecules, as well as interactions between viral RNA and host RNA. Proximity ligation products can be enriched using gene-specific DNA probes. However, it's important to note that when enriching for virus RNA using this approach, it may lead to the specific acquisition of virus RNA interaction information while potentially losing information about host RNA interactions. Upon reversing the crosslinking, chimeric RNAs are generated, containing information about base pairing or interactions. This information can be retrieved from chimeric reads in sequencing libraries. The P5 and P7 shown in the figure represent two sequencing adapters ligated for sequencing during library preparations

(ASO) therapeutics (Manfredonia et al. 2020). ASOs can bind to conserved viral RNA regions and disrupt RNA structural functions to attenuate viral infection according to several published studies (Sun et al. 2021a; Manfredonia et al. 2020). For instance, specific ASOs were employed to disrupt SARS-CoV-2 structural elements within the open reading frame (ORF) ab (ORF1ab_6449 and ORF1ab_9456) and structural protein-coding regions

(N_29502). These functional study results demonstrated the inhibited replication of SARS-CoV-2. To validate the structural significance in SARS-CoV-2 replication, mutations were specifically introduced to perturb base-pairing propensities in the ORF1ab_6449 structural element. As expected, viral RNA levels were drastically reduced, and this massive reduction could be partly rescued by tunning the displaced RNA structure back (Sun et al.

2021a). Apart from conserved regions, a wide range of techniques were adopted to unveil other dynamic RNA structures and interactions. Functional analysis, such as locked nucleic acids (LNAs), was employed to disrupt RNA structures to reveal the functional roles of certain RNA motifs within well-folded regions during the SARS-CoV-2 life cycle (Huston et al. 2021). Additionally, information about the secondary structure of RNA can be utilized to predict interactions between host proteins and the SARS-CoV-2 RNA genome. To give an example, the deep-learning tool Protein-RNA Interaction by Structure-informed Modeling using a deep neural network (PrismNet) was utilized to predict RNA-binding protein (RBP) and RNA interactions using icSHAPE scores (Sun et al. 2021b). Essentially, some of the interactions identified by PrismNet can serve as potential drug targets, thereby facilitating the repurposing of existing drugs to inhibit SARS-CoV-2 (Sun et al. 2021a).

Together, these developed tools and methods have made substantial contributions to our comprehension of the RNA secondary structure of SARS-CoV-2 and its interactions with host proteins and RNA. These discoveries offer hope for the development of effective therapeutics and our capacity to combat viral infections.

3 RNA secondary structure-dependent regulation and applications

The remarkable flexibility and diversity of RNA structures have established a fundamental basis for RNA's pivotal role as a key regulatory mediator. Through the application of advanced RNA structure probing technologies, we have gained unprecedented insights into the secondary structural characteristics of a myriad of RNA molecules at a transcriptome-wide scale. Consequently, these revelations have contributed to our growing understanding of the integral roles that RNA structures play in both transcriptional and post-transcriptional regulation. As our comprehension of RNA secondary structure advances, there is a growing expectation that this knowledge will open new avenues for the treatment of human diseases. In this review, we explore crucial insights into the regulatory roles of RNA, particularly concerning its secondary structure, and its wide-ranging applications in the context of disease treatment (Fig. 2).

3.1 RNA secondary structures are involved in transcriptional regulation

RNA assumes a critical role in the regulation of transcription via two primary mechanisms. First, it serves as an intermediary, facilitating interactions with various proteins. Second, it exerts control over the transcriptional state by modulating its structure. RNA molecules, characterized by intricate secondary structures, provide

an ideal platform for accommodating multiple protein bindings, thereby bolstering the efficiency of these vital interactions (Sanchez de Groot, et al. 2019). Meanwhile, RNA has the capability to react to alterations in external stimuli and undergo conformational changes correspondingly.

3.1.1 LncRNA Xist plays a role in epigenetic regulation

Long non-coding RNA (lncRNA) represents a category of RNA molecules with a length exceeding 200 nucleotides. These molecules have been implicated in various pathological conditions, including prostate cancer (Takayama et al. 2020), hereditary diseases (Ramos et al. 2023; Cannavicci et al. 2020; Yang, et al. 2022), and neurodegenerative disease (Tan et al. 2023; Li and Wang 2023), as evidenced by numerous experiments. Nevertheless, the majority of lncRNAs remain unexplored in terms of their specific functions. Notably, Xist stands out as one of the most extensively studied among these lncRNAs and has been found to play a crucial role in the epigenetic inactivation of the X chromosome (Pacini et al. 2021).

The secondary structure of Xist has been successfully elucidated (Lu et al. 2016; Lu et al. 2020b; Fang et al. 2015; Smola et al. 2016). In 2015, Rui Fang et al*.* introduced an innovative method named "Targeted Structure-Seq" for the comprehensive investigation of RNA secondary structures, and was applied to probe the complete secondary structure of Xist (Fang et al. 2015). Their research led to an updated comprehension of the repeat A region and the discovery of a previously unidentified conserved element within the repeat C region. The conserved structure in the repeat A region is essential for gene repression and the disruption of the conserved elements in the C region can displace the Xist from chromatin. Researchers used SHAPE-Map, a method that can determine RNA secondary structure with single-nucleotide resolution, to investigate the complex structure of Xist and its interactions within the cellular environment. Also, they found that the E region's repeat structure can act as a binding platform for several protein cofactors, including CELF1, PTBP1, and HuR (Smola et al. 2016). These bonded proteins can regulate processes such as Xist splicing and RNA decay, which further regulate X chromosome deactivation.

Another technique that also provides near single-nucleotide resolution, known as psoralen analysis of RNA interactions and structures (PARIS), relies on reversible psoralen crosslinking. This innovative approach can identify long-range RNA-RNA interactions, thereby revealing higher-order RNA structures (Lu et al. 2016). They found that the inter-repeat duplex structure of A region can recruit SPEN. SPEN protein is a key epigenetic silencing
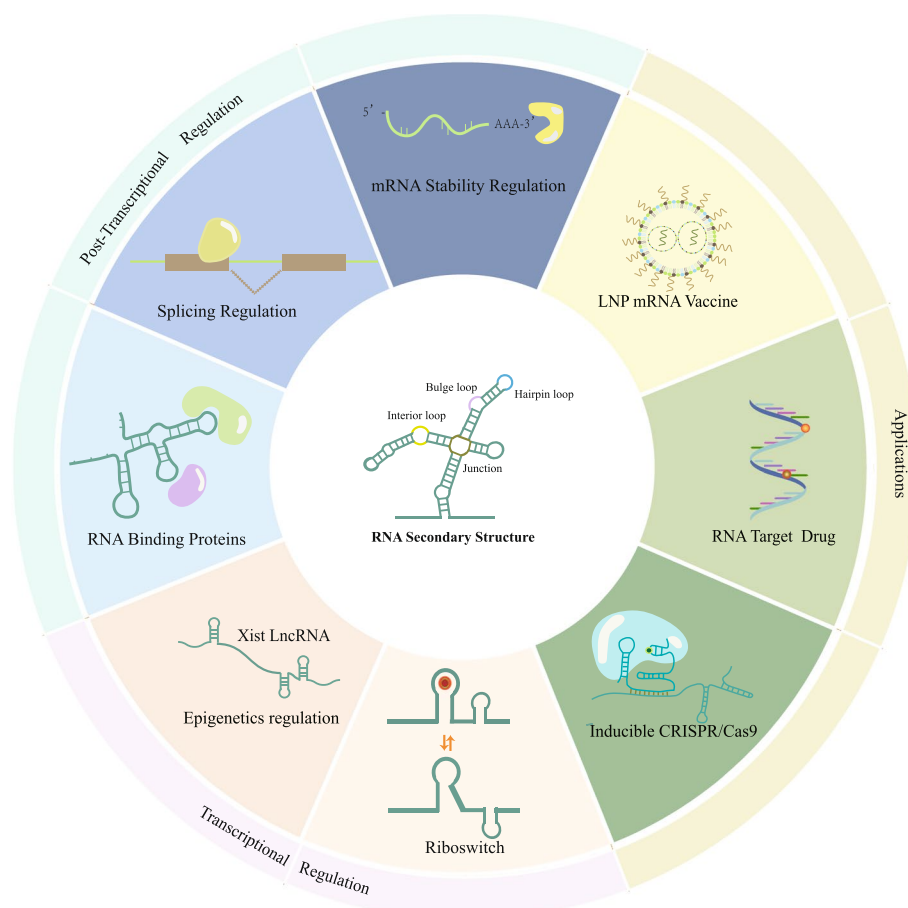
Fig. 2 RNA secondary structure-dependent regulations and applications. The secondary structure of RNA serves a crucial function in both transcriptional and post-transcriptional regulation. These regulatory roles have a broad spectrum of practical applications. (i) Transcriptional regulation: epigenetic regulation can be modulated by the secondary structure of LncRNA. When riboswitches function as trans-acting elements, they have the capacity to influence transcription in various ways. (ii) Post-transcriptional regulation: post-transcriptional regulation activities, for example, RNA stability, RNA splicing, and RNA binding proteins, are closely related to RNA structure. (iii) Applications: RNA secondary structure can facilitate the research on therapeutic tools such as LNP mRNA vaccines, inducible CRISPR/Cas9 systems, and RNA-target drugs

protein that also carries an indispensable role in initiating X-chromosome inactivation.

In a separate study, a structure–function map of the Xist RNA−protein complex was created through the integration of fRIP-seq, eCLIP, and a combination of PARIS, icSHAPE, and conservation analysis (Lu et al. 2020b). During the research, they altered locations of the repeat A region in Xist and found that related proteins can nucleate correspondingly, which means that the A region is the center for ribonucleoprotein (RNP) assembly. They also identified that the modular structure of Xist can facilitate N[6]-methyladenosine (m[6]A) modification, which is proven to have various roles in gene expression regulations (Jiang et al. 2021).

These studies collectively indicate that Xist is a highly dynamic lncRNA characterized by several conserved structural domains, serving as a platform for interaction with numerous proteins (e.g. SPEN). These conserved structural elements underlie Xist's crucial roles in epigenetic inactivation of the X chromosome (Patil et al. 2016).

3.1.2 Riboswitch—RNA as a regulator to control transcription

Riboswitch represents a class of structured non-coding RNA elements that has the capacity to bind to small molecules. Upon binding, riboswitches undergo secondary structural changes that exert a direct impact on gene expressions (Bédard et al. 2020). Riboswitches typically comprise two distinct domains: a relatively conserved aptamer domain and an expression platform. Riboswitches are characterized by their acute responses, high sensitivity, and a broad array of available ligands or corresponding components. These features make them attractive candidates for potential therapeutic applications. As

our understanding of structural and functional aspects of riboswitches grows, computational approaches, including rational or semi-rational design, can be used to discover natural and synthetic riboswitches with specificity, allowing researchers to target and regulate gene expression in normal and disease settings (Wu et al. 2023; Gong, et al. 2017). Nonetheless, due to the dynamic nature of RNA structure within living cells, designed riboswitches that perform effectively in vitro may not function as expected in vivo. Essentially, the regulation of riboswitches relies on the transitions between two relatively stable structural conformations (Bédard et al. 2020). This transition can be triggered by various factors such as small molecules, including metabolites, metal ions, and even other RNA molecules, offering substantial opportunities for exploitation and fine-tuning.

In nature, riboswitches generally serve as cis-acting regulatory elements, influencing the expression of downstream genes via RNA structural changes. A prominent instance of such regulatory elements is the Theophylline-dependent riboswitch (Nakahira et al. 2013). In this particular case, the gene subjected to regulation is positioned downstream of the riboswitch. As the concentration of theophylline attains a certain threshold, the riboswitch can activate the gene as required. This mode of action predominantly occurs in post-transcriptional regulation, encompassing processes such as translation regulation, splicing control, and mRNA degradation. In contrast, in the realm of transcriptional regulation, riboswitches frequently serve as trans-acting elements. For example, they may collaborate with the clustered regularly interspaced short palindromic repeats (CRISPR) system to facilitate gene editing when the respective effectors present (Cengic, et al. 2022), which we will discuss in the third part.

3.2 Post-transcriptional regulation is highly dependent on RNA structure

Post-transcriptional regulation encompasses a diverse array of processes occurring after RNA transcription. These regulatory mechanisms include RNA localization, stabilization, splicing, and translation, among others. RNA structure plays a pivotal role in shaping specific processes within this intricate regulatory network, forming a fundamental aspect that directly influences various functions. In the following sections, we will elucidate several pivotal processes and delineate the role of RNA structure in each.

3.2.1 *RNA stability regulation mainly associates with RNA structure in the 3′ end*

Numerous intracellular processes demand precise coordination with the RNA life cycle, necessitating meticulous regulation of RNA abundance. This involves the precise control of RNA stabilization and degradation for the seamless operation of the cell. RNA decay encompasses multiple intricate pathways. The mRNA, for instance, can undergo either $3′ \rightarrow 5′$ decay (e.g., through the exosome pathway) or $5′ \rightarrow 3′$ decay (e.g., via the XRN1 pathway) (Garneau et al. 2007). The stability of RNA is contingent on various factors, including RNA modifications, sequence characteristics, RNA-binding proteins, and cellular states, among others. Multiple studies also showed that the balance between RNA decay and RNA stability can be influenced by RNA structure. Structured 3′ untranslated regions (UTRs) tend to undergo decay, while more flexible 3′ UTRs tend to remain stable (Fischer et al. 2020; Rasekhian et al. 2021). A potential RNA decay mechanism has been proposed involving two key proteins, UPF1 (an RNA-binding protein) and G3BP1 (associated with UPF1). The depletion of either protein leads to the accumulation of mRNAs with highly structured 3′ UTRs, underlying a structure-dependent mechanism of RNA regulation. Additionally, research revealed that the predominant factor influencing RNA stability is the base-pairing density in the 3′ UTR (Fischer et al. 2020). The high density of base-pairing indicates proximity and readiness to initiate structure-mediated RNA decay (SRD).

The 3′ UTR of OX40 mRNA, as one of the examples of structural regulation of RNA stability, was initially characterized by Tants et al. They identified four distinct structural elements, namely the bulge, the ADE, the CDE, and the kinked-helical terminus (Tants et al. 2022). Interestingly, rather than the sequence features of the ADE and CDE elements, the key mRNA stability regulator Roquin tends to recognize the stem-loop structure (Braun et al. 2018). Moreover, Goodarzi et al. developed TEISER (Tool for Eliciting Informative Structural Elements in RNA) and have identified 8 structural elements that are crucial in the stability regulation of mammalian mRNAs (Goodarzi et al. 2012). Among these eight structural elements, sRSM1 plays a pivotal role in global mRNA regulation by interacting with the binding protein HNRPA2B1. Essentially, the stability of RNA is a crucial factor to be addressed in the context of RNA vaccine development, which we will explore further in the next part.

3.2.2 *Splicing regulation*

Pre-mRNA contains both exons and introns, and splicing is the pivotal process involving the removal of introns and the ligation of exons. Multiple arrangements of exons on the same mRNA give rise to alternative splicing (Marasco and Kornblihtt 2023). In this process, the RNA, acting as a ribozyme, assumes a significant role (Wilkinson

et al. 2020), which underscores the influence of RNA structure on this catalytic process.

A substantial body of evidence underscores the critical role of RNA structure in the regulation of splicing. For instance, one study identified distinct asymmetric RNA secondary structures at exon-exon junctions in both denatured and naturally deproteinized transcripts. Specifically, the 5' exon ends tend to be more accessible, while the 3' exon ends are more structured. This implies that certain structures may be involved in the splicing process (Wan et al. 2014). Another investigation established a robust connection between splice site GC content and splice site usage, suggesting that compact secondary structures formed by high GC content may facilitate alternative splicing (Zhang et al. 2011). Furthermore, competing RNA secondary structures represent a significant mechanism in mutually exclusive splicing. As a widely applicable mechanism proposed in 2011 suggests, inter-intronic RNA pairing ensures the selection of only one exon in mutually exclusive splicing (Yang et al. 2011). Additionally, Hou et al. discovered that competing RNA pairs facilitate both cis-splicing and trans-splicing in Dscam1, leading to the generation of diverse isoforms (Hou, et al. 2022). The subsequent section will delve into the discussion of how unexpected splicing events can give rise to human diseases.

3.2.3 RNA Binding Protein
RBPs are the primary agents through which RNA regulates cellular states. It is also noteworthy that RBPs represent important drug-targeted sites (Prasad et al. 2022). Traditionally, the characterization of protein binding has often relied on sequence information. However, the importance of structural information yielding more precise insights for research has gradually gained traction. On the one hand, these RBPs interact with RNA, regulating functions such as RNA splicing, RNA stability and decay, and RNA virus activities (see related part). On the other hand, highly structured RNAs provide platforms for the interplay between proteins that work collectively to carry out certain functions, such as epigenetic regulation.

There is increasing evidence to suggest that interactions between RBPs and RNAs are structural-dependent. For example, Roquin recognizes the particular RNA stem-loop structure to regulate the mRNA stability (Tants et al. 2022). Additionally, researchers have discovered that the accuracy of deep learning models for RBP prediction can be significantly improved by incorporating RNA secondary structure information during training. This improvement has been demonstrated by PrismNet (Xu et al. 2023) and high-throughput dynamic cellular RNA-binding event identification using deep neural network (HDRNet) (Zhu et al. 2023). In another study, transfer learning was employed by integrating RNA secondary structure information to predict potential protein target sites, resulting in increased accuracy of the prediction results (Vaculík, et al. 2023).

Most importantly, RNA secondary structure can regulate virus activities through RBP. To give an example, SND1 (Staphylococcal Nuclease And Tudor Domain Containing 1) is proven to have higher affinity with RNA hairpin structure compared with single strand RNA (Xu et al. 2023). In another study, it was reported that the host SND1 protein can recognize the antisense RNA of SARS-CoV-2 and can regulate its replication in the host cell (Schmidt et al. 2023). The lower expression of SND1 is corresponded to reduced viral replication.

There are several experimental methods available for studying protein-RNA interactions in vivo, including RNA immunoprecipitation (RIP) (Gilbert and Svejstrup 2006) and a technique that combines crosslinking and immunoprecipitation (CLIP) and sequencing (Nostrand et al. 2016; Hafner et al. 2010; Licatalosi et al. 2008). Furthermore, many researchers also employ machine learning or deep learning methods to integrate experimental data for predicting RBP binding (Sun et al. 2021b; Xu et al. 2023; Zhu et al. 2023; Vaculík, et al. 2023; Ghanbari and Ohler 2020).

3.3 From regulation to application: Insights of RNA secondary structure facilitates development of therapeutic tools for human diseases
RNA plays a pivotal role in a wide spectrum of biological processes and has been implicated in the development of a broad array of diseases. RNA exhibits a vast array of structures and engages in interactions with numerous proteins. When these interactions go awry, diseases are initiated. As our comprehension of RNA, particularly its structural intricacies continue to advance, researchers start to explore the potential applications of RNA in clinical disease management.

The following section provides an overview of how our understanding of RNA secondary structure has contributed to the development of clinical therapeutic tools for human diseases. We will focus on three prominent examples: mRNA vaccines, the CRISPR/Cas9 system, and RNA-targeted drugs.

3.3.1 Lipid nanoparticles (LNP)-mRNA vaccines
Traditional vaccines often entail a lengthy and time-consuming development process, whereas mRNA vaccines can be developed with greater expediency (Gote, et al. 2023). Notably, mRNA vaccines offer enhanced safety and cost-effectiveness with superior efficacy in comparison to conventional vaccines (Zhang et al. 2023a).

Lu *et al. Advanced Biotechnology*   (2024) 2:3

Page 10 of 16

Research on mRNA vaccines can be categorized into two primary domains: vaccine development and the creation of effective vaccine delivery systems. The former is primarily concerned with optimizing the desired effects of mRNA, while the latter focuses on enabling RNA to target specific areas within the body. Collectively, mRNA vaccine stability and the development of efficient delivery systems are the current hindrances to the advancement of mRNA vaccines. A comprehensive overview of the general progress in mRNA vaccine development is available elsewhere (Gote, et al. 2023). Expanding upon our previous assertion that the structural component located at the UTR ends of the mRNA enhances RNA stability, we will now delve into how this knowledge is applied in the actual production of vaccines.

One study has enhanced the immunogenicity and protective efficacy of the SARS-CoV-2 vaccine in non-human primates by optimizing the unmodified RNA's 5' UTR (Gebre et al. 2022). They discovered that the structural modifications in the 5' UTR of CV2CoV, a second-generation mRNA vaccine with an optimized non-coding region, significantly contribute to robust protection. Thus, optimizing the 5' UTR becomes imperative, given its association with protein translation efficiency (Ryczek et al. 2023). Zhang et al. have proposed an algorithm, LinearDesign, designed to optimize RNA secondary structure, stability, and codon usage for mRNA vaccines in less than 16 min. This innovation not only reduces labor costs but also enhances the effectiveness of mRNA vaccines (Zhang et al. 2023b).

3.3.2 Inducible CRISPR/Cas9 system based on riboswitch
The CRISPR/Cas gene-editing technology, which initially originated in the immune systems of prokaryotes, has emerged as a groundbreaking tool (Barrangou et al. 2007; Brouns et al. 2008; Liu et al. 2022). Among the best-known Cas proteins consisting of Cas9 (type II), Cas12 (type V), and Cas13 (type VI), CRISPR/Cas9 is the most extensively utilized system. An overview of the advances and challenges in the use of CRISPR/Cas9 is available elsewhere (Li et al. 2023). In this context, we will discuss the inducible CRISPR/Cas9 system, which relies on riboswitches.

The inducible CRISPR/Cas9 system represents a precision-enhanced approach with reduced off-target effects, activated by external or internal signals. There are three primary methods to achieve inducibility in CRISPR systems. The first approach involves controlling the timing of guide RNA (gRNA) or Cas nuclease expression, ensuring their release at specific times to initiate CRISPR. The second method entails engineering nucleases, including the regulation of protein conformation. The last method involves manipulating the gRNA, which can be further categorized into two similar strategies reminiscent of riboswitches. One approach is to incorporate a non-functional ribozyme at the 5' end of gRNA, while the other method involves adding a sequence that prevents gRNA targeting. When a signaling molecule, such as a small molecule drug, binds to the added sequence, it induces a change in secondary structure, consequently, releasing the CRISPR/Cas9 system or exposing the gRNA sequence.

In 2016, Liu et al. introduced a CRISPR/Cas9-based signal conductor by incorporating a signal-responsive aptamer. They utilized this conductor as a 'switch' to redirect nucleophosmin (NPM)-mediated signaling from promoting proliferation to inducing quiescence in bladder cancer (Liu et al. 2016). Another study employed internal signals, specifically microRNAs (miRNAs), to trigger the CRISPR/Cas9 system when they are released and bind to the binding site in the pre-single guide RNA (pre-sgRNA) (Wang et al. 2019). The pre-sgRNA is non-functional, and upon miRNA binding to the pre-sgRNA's binding site, the pre-sgRNA undergoes cleavage, releasing mature and functional sgRNA molecules. The design of these aforementioned riboswitches relies heavily on RNA secondary structures.

3.3.3 RNA-targeted drug
As mentioned earlier, RNA's involvement in numerous diseases makes it a vital target for therapeutic interventions. Drug delivery systems often face challenges in traversing the blood–brain barrier, while the practical application of gene editing tools is hindered by collateral activities and off-target effects. In light of these challenges, small-molecule drugs, known for their ability to readily penetrate the blood–brain barrier, emerge as a compelling choice.

Accurate RNA structure models are crucial in the development of RNA-targeted drugs. Presently, most drug development efforts are grounded in the secondary structure of RNA to identify functional elements suitable for drug design. For instance, Sarah et al. employed SHAPE-Map to explore interactions between small molecules and RNAs (Martin et al. 2019). Essentially, the utilization of SHAPE allows us to acquire in vivo RNA secondary structure models better suited for the design of small molecule drugs (Martin et al. 2019; Smola et al. 2015).

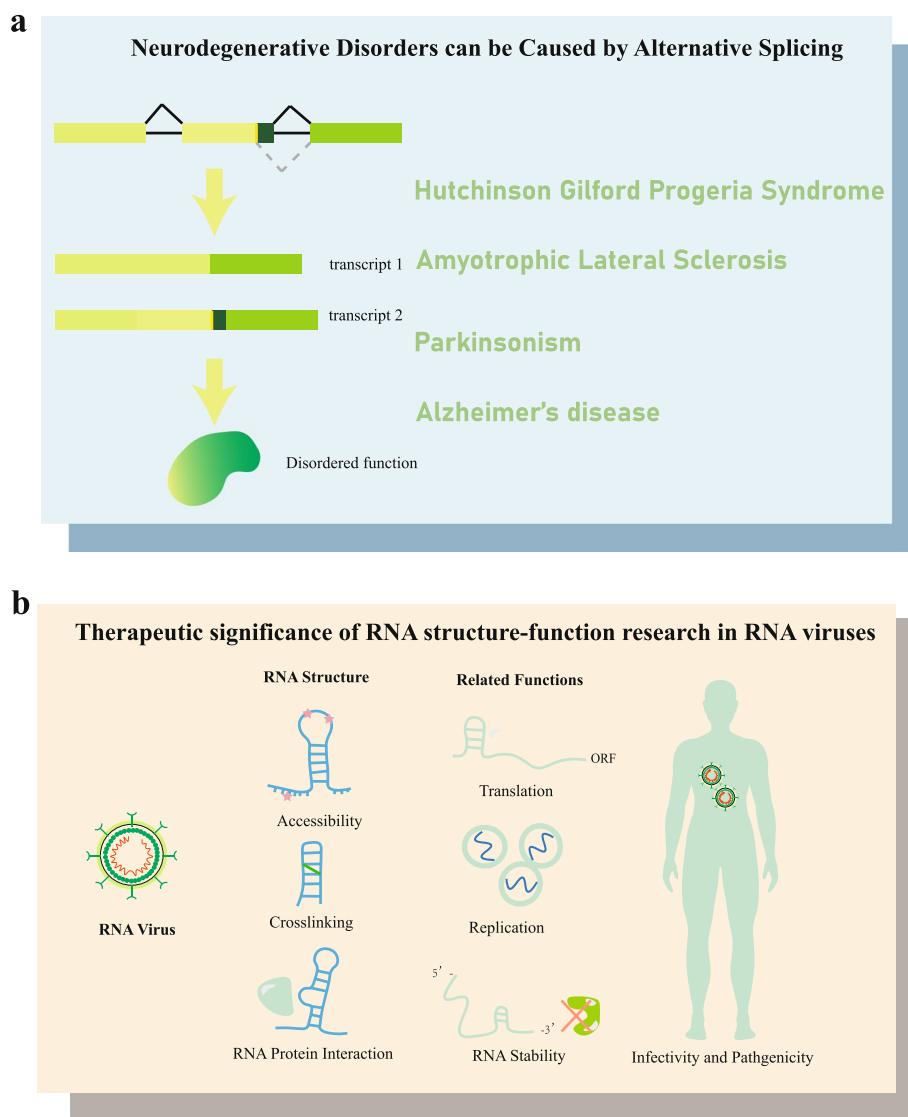## 4 RNA Structure Probing Methods Aids in the Discovery of Human Disease Mechanisms
Numerous disease mechanisms remain the subject of ongoing exploration, limited by the current extent of our knowledge. As our understanding of RNA continues to expand, its role in various diseases becomes

increasingly evident. As highlighted earlier, RNA structure plays a pivotal role in multiple regulatory processes. It is now clear that a deeper comprehension of disease mechanisms often arises when RNA structure is taken into account. RNA structure exerts regulatory control over alternative splicing (as discussed in Part III), and it is noteworthy that at least 10% of diseases can be attributed to erroneous alternative splicing (Krawczak et al. 2007). Additionally, RNA structure has a substantial impact on the behavior of disease-causing viruses. In the sections that follow, we will elucidate how RNA structure probing technology facilitates the exploration of these two critical mechanisms (Fig. 3).

4.1 Neurodegenerative disorders can be caused by alternative splicing

Changes in the gene, such as point mutations, insertions, or deletions, can result in altered RNA secondary structures. As illustrated in the third part, changes in RNA secondary structure affect cellular functions via transcriptional and post-transcriptional regulations, and potentially lead to human diseases. In most cases, human diseases are triggered by abnormal RNA secondary

Fig. 3 RNA structure probing methods aid the discovery of human disease mechanisms. **a** Neurodegenerative disorders can be caused by alternative splicing, and RNA structural information can provide insights into the research of the mechanisms. **b** Therapeutic significance of RNA structure–function research in RNA viruses. Various RNA structure probing methods are now accessible, facilitating a deeper comprehension of RNA virus infectivity and pathogenicity

structures that cause unusual alternative splicing or RBP interactions. Here, we will focus on how RNA probing methods can help explore the mechanisms of diseases that are caused by alternative splicing.

Diseases such as Hutchinson-Gilford Progeria Syndrome (HGPS) (Rahman et al. 2021), Amyotrophic Lateral Sclerosis (ALS) (Feldman et al. 2022; Masrori and Damme 2020), Parkinsonism (Jakubauskienė and Kanopka 2021), Alzheimer's disease (AD) (Waheed et al. 2023), and various types of cancers, such as bladder, brain, lung, gastric, and prostate cancers (Xie et al. 2019; Babic et al. 2013; Sheng et al. 2018; Ailiken et al. 2020; Bradley and Anczuków 2023), are closely linked to alternative splicing of RNA (Fig. 3a). Alternative splicing generates new transcripts that encode proteins with altered structures, prohibiting them from performing normal functions (Rahman et al. 2021; Waheed et al. 2023).

HGPS is a rare genetic disorder characterized by accelerated aging in affected patients (Rahman et al. 2021). This condition arises from aberrant alternative splicing of exon 11 of the *LMNA* gene. The presence of a mutation in the *LMNA* gene leads to the replacement of the original 5' splice site with an internal site. This, in turn, results in the production of a structurally disordered protein (Rahman et al. 2021; Lopez-Mejia et al. 2011). To investigate this internal site selection mechanism, Asaf et al. employed SHAPE-Map to examine the structural features of both the wild-type and C1824U mutant transcripts (Shilo et al. 2019). The results revealed striking similarities in the overall structures of the two variants. However, upon the disruption of the surrounding structural elements, there was an increased utilization of the alternative 5' splice site, suggesting that structural information holds more significance than sequence information in the pathogenesis of HGPS. This example underscores the pivotal role of RNA structural information, particularly at the 5' splice site, in determining the selection of alternative splicing sites. Various supporting factors, including regulatory proteins, also contribute to this process. For a comprehensive history of understanding the 5' splice site, a detailed review is available elsewhere (Roca et al. 2013).

Another example is Alzheimer's disease (AD), which is associated with the dysregulation of the tau protein variants ratio (4R:3R ratio) (encoded by the *MART* gene). The role of RNA secondary structure in the splicing regulation of the *MART* gene has been investigated (Strang et al. 2019; Kumar, et al. 2022). In one of the studies, DMS-Map was used to examine the structures of pre-mRNA and mature *MART-encoded* mRNA, and data were collected for predicting alternative splicing outcomes (Kumar, et al. 2022). They found that the mutations can change the structural ensembles in the exon10-intron10 junction, differentiating alternative splicing events. Within normal human brain tissue, the 4R:3R ratio is approximately 1. When the structure of the 5' splicing site of exon10-intron10 is rather stabilized (stem-loop formation), exon10 tends to be skipped during alternative splicing and producing 3R isoform. When the stabilized structure of this junction region is shifted, the 5' splicing site becomes more accessible and gives rise to the 4R isoform. This result can be verified in previous research observations (Donahue et al. 2006), which found that mutations modifying the stem-loop structure at the exon–intron interface of exon10 ultimately led to the increased 3R mRNA isoform production. This study shows the integration of RNA structure probing method with computational methods to delve into the mechanisms of diseases.

These two specific diseases caused by unusual alternative splicing provide insights into how RNA secondary structure probing methods can be used to facilitate disease mechanism explorations. Moreover, these probing methods can also be combined with computational methods to predict disease outcomes.

4.2 Therapeutic significance of RNA structure–function research in RNA viruses

RNA viruses belong to a category of viruses that utilize RNA as their genetic material (V'Kovski et al. 2021). Many RNA viruses can cause diseases in human bodies, and new disease-causing RNA viruses continue to be discovered. For instance, SARS-CoV-2 and Ebola viruses can cause severe acute respiratory syndrome (Hu et al. 2021) and Ebola Virus Disease (EVD) (Liu, et al. 2022) respectively. Also, Zika virus (ZIKV) is associated with infectious diseases, birth defects, and neurological disorders (Marbán-Castro et al. 2021). There are many other RNA viruses, such as HIV, hepacivirus, and influenza A virus, that can initiate human diseases and have caused global concern.

Despite the comparatively simple genomes of RNA viruses, their genetic material is remarkably compact and organized by intricate structural elements. As research has been well-demonstrated, RNA structure is closely related to its function. Numerous viral processes, including virus translation, RNA decay, and viral replication are intricately intertwined with RNA structure (Fig. 3b). These processes significantly impact viral infectivity and pathogenicity which warrant in-depth investigation.

Virus translation is vital for infectivity and pathogenicity, serving as the initial step to fulfill its role in host cells for positive-sense RNA viruses. According to research, translation efficiency is related to the frequency of single-stranded regions in SARS-CoV-2, highlighting the importance of RNA secondary structures in viral

translation (Sun et al. 2021a). The internal ribosome entry site (IRES) is a class of structural elements located in the 5' UTR of viral RNA (Pelletier and Sonenberg 1988; Jang et al. 1988), exhibiting complex structures and capable of recruiting translation machinery within host cells. There is a more detailed review that discusses how viral RNA structures manipulate RNA translation (Jaafar and Kieft 2019). Different RNA virus families are characterized by distinct IRES elements which have been reviewed by other reviews (Peng et al. 2023; Niepmann and Gerresheim 2020; Nakashima and Uchiumi 2009; Lozano and Martínez-Salas 2015). One study uses RNase V1 probing to determine the 5' UTR in SARS-CoV-2 and found that there is a highly stable four-way junction in this region, which may be involved in translation initiation (Miao et al. 2021). Additionally, structural elements situated in the 3' UTR influence translation processes as well, which is a topic thoroughly reviewed elsewhere (Rasekhian et al. 2021).

Virus replication is also regulated by RNA secondary structure. Liu et al. found that disruption of the pseudoknots structures, achieved through IRES point mutations in the of Senecavirus A (SVA), inhibits virus replication rather than its translation (Liu et al. 2021a; Liu et al. 2021b). They identified other essential structural elements (stem II and stem Ib) in SVA that are crucial for virus replications. Enterovirus has a structural element in the 5' UTR called the Cloverleaf (CL) structure, which provides platforms for 3CD and PCBP2 recruiting to promote viral replication (Das et al. 2023). Some reviews can provide overall perspectives on how RNA structure can regulate viral replications (Malone et al. 2022; Szczesniak, et al. 2023).

RNA structure probing methods, such as SHAPE-Map(Huston et al. 2021), MarathonRT (Guo et al. 2022), DMS-MaPseq (Lan et al. 2022), icSHAPE (Sun et al. 2021a; Li et al. 2018), etc., have been instrumental in determining the genome-wide in vivo structures of viruses like SARS-CoV-2 and ZIKV. These methods have paved the way for understanding the virus RNA structurome, allowing the discovery of more drug-targetable sites. Despite the comprehensive structural-function map of SARS-CoV-2 generated by Sun et al. (Sun et al. 2021a), Li et al. also employed icSHAPE and PARIS techniques to analyze the viral genome structure of ZIKV (Li et al. 2018). They identified key structural elements that are crucial for viral infestation of cells and constructed structure–function maps. SHAPE-Map was also used to detect interactions between RNA and small molecules, which has provided us with a new way for antiviral drug screening (Martin et al. 2019). RNA viruses exemplify the close interplay between the structure and function of RNA. Utilizing RNA secondary structure probing methods, researchers can uncover the in vivo secondary structure of RNA and RNA-RNA interactions (Lu et al. 2016). This approach facilitates the identification of crucial structural elements, enables the analysis of structure–function relationships, and reveals potential RNA drug targets. Therefore, when combined with an in-depth analysis of RNA–protein interactions, it not only establishes the groundwork for a comprehensive understanding of these relationships but also extends the pool of potential targets for drug screening.

5 Conclusion

Precise investigation of RNA structure is vital in advancing our understanding of life processes. Recent years have witnessed rapid advancements in RNA secondary structure probing techniques, with many of them achieving single-nucleotide resolution. These technological strides have led to the development of RNA Structurome research, also providing comprehensive insights into RNA secondary structures. These insights have significantly contributed to our understanding of diverse cellular regulatory processes, including epigenetic regulation, riboswitch mechanisms, RNA splicing, and RNA stability. Meanwhile, they have facilitated the creation of therapeutic tools for tackling human diseases. Despite their therapeutic applications, RNA structure probing methods also offer a promising avenue for exploring the mechanisms of human diseases, potentially providing the key to overcoming existing research constraints and obtaining the in-depth information necessary for a deeper understanding of disease mechanisms.

Remarkably, RNA structural research has entered a new era, in which advanced computational and experimental methods are used to study RNA structurome. This enables the development of various clinical tools and opens new perspectives for understanding disease mechanisms and developing more efficient drugs. However, there still exists hurdles in both determining the interested RNA structures and the applications in human diseases. First, the existing RNA secondary probing methods inadequately cover all positions in the transcriptome or genome, resulting in certain positions within a specific transcript lacking corresponding icSHAPE scores. This may lead to an unrealistic RNA secondary structure in these uncertain positions, which can hinder scientists from studying the related RNA and disease mechanisms. In the future, more sensitive and precise probing methods need to be developed by changing the experimental conditions or improving the efficiency of enzymes. Second, existing tools developed for clinical treatment are rarely applied due to potential side effects, such as off-target effects caused by inducible CRISPR/Cas9 systems based on riboswitches. Scientists should

develop more effective ways to evaluate these tools and improve their safety. Third, deep learning models have certain limitations, notably issues like overfitting and inaccurate predictions. These challenges can arise due to several factors, including overcomplicated parameterization during model establishment and the presence of high background noise in the training datasets. Additionally, the accuracy of prediction models is often closely tied to the quantity of training data, making precise secondary structure predictions for low-abundant RNAs particularly challenging. Overall, addressing these limitations requires researchers to optimize network architectures and enhance the quality of experimental datasets.

Availability of data and materials
No datasets were generated or analyzed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors agreed with the content and that all gave explicit consent to submit.

Competing interests
None declared.

References

Ailiken G, et al. Post-transcriptional regulation of BRG1 by FIRΔexon2 in gastric cancer. Oncogenesis. 2020;9(2):26.

Aw JG, et al. In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. Mol Cell. 2016;62(4):603–17.

Babic I, et al. EGFR mutation-induced alternative splicing of Max contributes to growth of glycolytic tumors in brain cancer. Cell Metab. 2013;17(6):1000–8.

Barrangou R, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315(5819):1709–12.

Bartel DP, Szostak JW. Isolation of new ribozymes from a large pool of random sequences [see comment]. Science. 1993;261(5127):1411–8.

Bédard AV, Hien EDM, Lafontaine DA. Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. Biochim Biophys Acta Gene Regul Mech. 2020;1863(3): 194501.

Bradley RK, Anczuków O. RNA splicing dysregulation and the hallmarks of cancer. Nat Rev Cancer. 2023;23(3):135–55.

Braun J, et al. Identification of new high affinity targets for Roquin based on structural conservation. Nucleic Acids Res. 2018;46(22):12109–25.

Brenner S. RNA, ribosomes, and protein synthesis. Cold Spring Harb Symp Quant Biol. 1961;26:101–10.

Brouns SJ, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008;321(5891):960–4.

Cai Z, et al. RIC-seq for global in situ profiling of RNA-RNA spatial interactions. Nature. 2020;582(7812):432–7.

Cannavicci A, Zhang Q, Kutryk MJB. Non-Coding RNAs and Hereditary Hemorrhagic Telangiectasia. J Clin Med. 2020;9(10):3333.

Cao C, et al. The architecture of the SARS-CoV-2 RNA genome inside virion. Nat Commun. 2021;12(1):3917.

Cech TR. RNA chemistry. Ribozyme self-replication? Nature. 1989;339(6225):507–8.

Cengic I, et al. Inducible CRISPR/Cas9 Allows for Multiplexed and Rapidly Segregated Single-Target Genome Editing in Synechocystis Sp. PCC 6803. ACS Synth Biol. 2022;11(9):3100–13.

Crick F. Central dogma of molecular biology. Nature. 1970;227(5258):561–3.

Das NK, et al. Crystal structure of a highly conserved enteroviral 5' cloverleaf RNA replication element. Nat Commun. 2023;14(1):1955.

Ding Y, et al. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature. 2014;505(7485):696–700.

Donahue CP, et al. Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing. J Biol Chem. 2006;281(33):23302–6.

Fang R, et al. Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. PLoS Genet. 2015;11(12): e1005668.

Feldman EL, et al. Amyotrophic lateral sclerosis. Lancet. 2022;400(10360):1363–80.

Fischer JW, et al. Structure-Mediated RNA Decay by UPF1 and G3BP1. Mol Cell. 2020;78(1):70-84.e6.

Fu L, et al. UFold: fast and accurate RNA secondary structure prediction with deep learning. Nucleic Acids Res. 2022;50(3): e14.

Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. Nat Rev Mol Cell Biol. 2007;8(2):113–26.

Gebre MS, et al. Optimization of non-coding regions for a non-modified mRNA COVID-19 vaccine. Nature. 2022;601(7893):410–4.

Gilbert C, Svejstrup JQ. RNA immunoprecipitation for determining RNA-protein associations in vivo. Curr Protoc Mol Biol. 2006:Chapter 27:Unit 27.4.

Ghanbari M, Ohler U. Deep neural networks for interpreting RNA-binding protein target preferences. Genome Res. 2020;30(2):214–26.

Goodarzi H, et al. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. Nature. 2012;485(7397):264–8.

Gong S, et al. Computational Methods for Modeling Aptamers and Designing Riboswitches. Int J Mol Sci. 2017;18(11):2442.

Gote V, et al. A Comprehensive Review of mRNA Vaccines. Int J Mol Sci. 2023;24(3):2700.

Guo LT, et al. Direct tracking of reverse-transcriptase speed and template sensitivity: implications for sequencing and analysis of long RNA molecules. Nucleic Acids Res. 2022;50(12):6980–9.

Hafner M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010;141(1):129–41.

Helwak A, et al. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell. 2013;153(3):654–65.

Hofacker IL. Vienna RNA secondary structure server. Nucleic Acids Res. 2003;31(13):3429–31.

Hou S, et al. Trans-splicing facilitated by RNA pairing greatly expands sDscam isoform diversity but not homophilic binding specificity. Sci Adv. 2022;8(27):eabn9458.

Huston NC, et al. Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. Mol Cell. 2021;81(3):584-598.e5.

Hu B, et al. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. 2021;19(3):141–54.

Jaafar ZA, Kieft JS. Viral RNA structure-based strategies to manipulate translation. Nat Rev Microbiol. 2019;17(2):110–23.

Jackson CB, et al. Mechanisms of SARS-CoV-2 entry into cells. Nat Rev Mol Cell Biol. 2022;23(1):3–20.

Jakubauskienė E, Kanopka A. Alternative Splicing and Hypoxia Puzzle in Alzheimer's and Parkinson's Diseases. Genes (Basel). 2021;12(8):1272.

Jang SK, et al. A segment of the 5′nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. J Virol. 1988;62(8):2636–43.

Jiang X, et al. The role of m6A modification in the biological functions and diseases. Signal Transduct Target Ther. 2021;6(1):74.

Krawczak M, et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. Hum Mutat. 2007;28(2):150–8.

Kruger K, et al. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell. 1982;31(1):147–57.

Kudla G, Wan Y, Helwak A. RNA Conformation Capture by Proximity Ligation. Annu Rev Genomics Hum Genet. 2020;21:81–100.

Kumar J, et al. Quantitative prediction of variant effects on alternative splicing in MAPT using endogenous pre-messenger RNA structure probing. Elife. 2022;11:e73888.

Lan TCT, et al. Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. Nat Commun. 2022;13(1):1128.

Li K, Wang Z. lncRNA NEAT1: Key player in neurodegenerative diseases. Ageing Res Rev. 2023;86: 101878.

Li P, et al. Integrative Analysis of Zika Virus Genome RNA Structure Reveals Critical Determinants of Viral Infectivity. Cell Host Microbe. 2018;24(6):875-886.e5.

Li T, et al. CRISPR/Cas9 therapeutics: progress and prospects. Signal Transduct Target Ther. 2023;8(1):36.

Licatalosi DD, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature. 2008;456(7221):464–9.

Lightfoot HL, Hall J. Endogenous polyamine function–the RNA perspective. Nucleic Acids Res. 2014;42(18):11275–90.

Liu CH, et al. Therapeutic Strategies against Ebola Virus Infection. Viruses. 2022;14(3):579.

Liu F, et al. Motif mutations in pseudoknot stem I upstream of start codon in Senecavirus A genome: Impacts on activity of viral IRES and on rescue of recombinant virus. Vet Microbiol. 2021b;262: 109223.

Liu F, et al. Stem II-disrupting pseudoknot does not abolish ability of Senecavirus A IRES to initiate protein expression, but inhibits recovery of virus from cDNA clone. Vet Microbiol. 2021a;255: 109024.

Liu G, et al. The CRISPR-Cas toolbox and gene editing technologies. Mol Cell. 2022;82(2):333–47.

Liu Y, et al. Directing cellular information flow via CRISPR signal conductors. Nat Methods. 2016;13(11):938–44.

Lopez-Mejia IC, et al. A conserved splicing mechanism of the LMNA gene controls premature aging. Hum Mol Genet. 2011;20(23):4540–55.

Loughrey D, et al. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. Nucleic Acids Res. 2014;42(21).

Lozano G, Martínez-Salas E. Structural insights into viral IRES-dependent translation mechanisms. Curr Opin Virol. 2015;12:113–20.

Lu R, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020a;395(10224):565–74.

Lu Z, et al. RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell. 2016;165(5):1267–79.

Lu Z, et al. Structural modularity of the XIST ribonucleoprotein complex. Nat Commun. 2020b;11(1):6163.

Lucks JB, et al. Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc Natl Acad Sci U S A. 2011;108(27):11063–8.

Luo QJ, et al. RNA structure probing reveals the structural basis of Dicer binding and cleavage. Nat Commun. 2021;12(1):3397.

Lyon MF. X-chromosome inactivation and developmental patterns in mammals. Biol Rev Camb Philos Soc. 1972;47(1):1–35.

Malone B, et al. Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design. Nat Rev Mol Cell Biol. 2022;23(1):21–39.

Manfredonia I, et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. Nucleic Acids Res. 2020;48(22):12436–52.

Marbán-Castro E, et al. Zika virus infection in pregnant women and their children: A review. Eur J Obstet Gynecol Reprod Biol. 2021;265:162–8.

Marasco LE, Kornblihtt AR. The physiology of alternative splicing. Nat Rev Mol Cell Biol. 2023;24(4):242–54.

Marinus T, et al. A novel SHAPE reagent enables the analysis of RNA structure in living cells with unprecedented accuracy. Nucleic Acids Res. 2021;49(6): e34.

Martin S, et al. Using SHAPE-MaP to probe small molecule-RNA interactions. Methods. 2019;167:105–16.

Masrori P, Van Damme P. Amyotrophic lateral sclerosis: a clinical review. Eur J Neurol. 2020;27(10):1918–29.

Masters PS. The molecular biology of coronaviruses. Adv Virus Res. 2006;66:193–292.

Mathews DH, Turner DH, Zuker M. RNA secondary structure prediction. Curr Protoc Nucleic Acid Chem, 2007:Chapter 11:Unit 11.2.

Miao Z, et al. Secondary structure of the SARS-CoV-2 5′-UTR. RNA Biol. 2021;18(4):447–56.

Nakahira Y, et al. Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in Cyanobacterium Synechococcus elongatus PCC 7942. Plant Cell Physiol. 2013;54(10):1724–35.

Nakashima N, Uchiumi T. Functional analysis of structural motifs in dicistroviruses. Virus Res. 2009;139(2):137–47.

Niepmann M, Gerresheim GK. Hepatitis C Virus Translation Regulation. Int J Mol Sci. 2020;21(7):2328.

Nguyen TC, et al. Mapping RNA-RNA interactome and RNA structure in vivo by MARIO. Nat Commun. 2016;7:12023.

Noller HF, Hoffarth V, Zimniak L. Unusual resistance of peptidyl transferase to protein extraction procedures. Science. 1992;256(5062):1416–9.

Pacini G, et al. Integrated analysis of Xist upregulation and X-chromosome inactivation with single-cell and single-allele resolution. Nat Commun. 2021;12(1):3638.

Patil DP, et al. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature. 2016;537(7620):369–73.

Pelletier J, Sonenberg N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature. 1988;334(6180):320–5.

Peng T, et al. Structural diversity and biological role of the 5′ untranslated regions of picornavirus. RNA Biol. 2023;20(1):548–62.

Piao M, et al. An ultra low-input method for global RNA structure probing uncovers Regnase-1-mediated regulation in macrophages. Fundamental Research. 2022;2(1):2–13.

Piao M, Sun L, Zhang QC. RNA Regulations and Functions Decoded by Transcriptome-wide RNA Structure Probing. Genomics Proteomics Bioinformatics. 2017;15(5):267–78.

Prasad K, et al. The SARS-CoV-2 targeted human RNA binding proteins network biology to investigate COVID-19 associated manifestations. Int J Biol Macromol. 2022;217:853–63.

Qian X, et al. Revealing lncRNA Structures and Interactions by Sequencing-Based Approaches. Trends Biochem Sci. 2019;44(1):33–52.

Rahman MM, et al. Hutchinson-Gilford Progeria Syndrome: An Overview of the Molecular Mechanism, Pathophysiology and Therapeutic Approach. Curr Gene Ther. 2021;21(3):216–29.

Rasekhian M, et al. The Role of 3′UTR of RNA Viruses on mRNA Stability and Translation Enhancement. Mini Rev Med Chem. 2021;21(16):2389–98.

Ramos TAR, et al. Single-cell transcriptional landscape of long non-coding RNAs orchestrating mouse heart development. Cell Death Dis. 2023;14(12):841.

Ritchey LE, et al. Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure in vivo. Nucleic Acids Res. 2017;45(14): e135.

Roca X, Krainer AR, Eperon IC. Pick one, but be quick: 5′ splice sites and the problems of too many choices. Genes Dev. 2013;27(2):129–44.

Rouskin S, et al. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature. 2014;505(7485):701–5.

Ryczek N, Łyś A, Makałowska I. The Functional Meaning of 5′UTR in Protein-Coding Genes. Int J Mol Sci. 2023;24(3):2976.

Lu *et al. Advanced Biotechnology*   (2024) 2:3

Page 16 of 16

Sanchez de Groot N, et al. RNA structure drives interaction with proteins. Nat Commun. 2019;10(1):3246.

Sato K, Akiyama M, Sakakibara Y. RNA secondary structure prediction using deep learning with thermodynamic integration. Nat Commun. 2021;12(1):941.

Schmidt N, et al. SND1 binds SARS-CoV-2 negative-sense RNA and promotes viral RNA synthesis through NSP9. Cell. 2023;186(22):4834-4850.e23.

Schroeder R, Barta A, Semrad K. Strategies for RNA folding and assembly. Nat Rev Mol Cell Biol. 2004;5(11):908–19.

Sharma E, et al. Global Mapping of Human RNA-RNA Interactions. Mol Cell. 2016;62(4):618–26.

Sheng J, et al. SRSF1 modulates PTPMT1 alternative splicing to regulate lung cancer cell radioresistance. EBioMedicine. 2018;38:113–26.

Shilo A, et al. Interplay of primary sequence, position and secondary RNA structure determines alternative splicing of LMNA in a pre-mature aging syndrome. Nucleic Acids Res. 2019;47(11):5922–35.

Siegfried NA, et al. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat Methods. 2014;11(9):959–65.

Singh J, et al. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. Nat Commun. 2019;10(1):5407.

Silverman IM, et al. Genome-Wide Approaches for RNA Structure Probing. Adv Exp Med Biol. 2016;907:29–59.

Smola MJ, et al. 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. 2016;113(37):10322–7.

Smola MJ, et al. Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. Nat Protoc. 2015;10(11):1643–69.

Spitale RC, et al. Structural imprints in vivo decode RNA regulatory mechanisms. Nature. 2015;519(7544):486–90.

Spitale RC, et al. RNA SHAPE analysis in living cells. Nat Chem Biol. 2013;9(1):18–20.

Spitale RC, Incarnato D. Probing the dynamic RNA structurome and its functions. Nat Rev Genet. 2023;24(3):178–96.

Strang KH, Golde TE, Giasson BI. MAPT mutations, tauopathy, and mechanisms of neurodegeneration. Lab Invest. 2019;99(7):912–28.

Sun L, et al. RNA structure maps across mammalian cellular compartments. Nat Struct Mol Biol. 2019;26(4):322–30.

Sun L, et al. Predicting dynamic cellular protein-RNA interactions by deep learning using in vivo RNA structures. Cell Res. 2021b;31(5):495–516.

Sun L, et al. In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. Cell. 2021a;184(7):1865-1883.e20.

Szczesniak I, et al. Structural and Functional RNA Motifs of SARS-CoV-2 and Influenza A Virus as a Target of Viral Inhibitors. Int J Mol Sci. 2023;24(2):1232.

Szikszai M, et al. Deep learning models for RNA secondary structure prediction (probably) do not generalise across families. Bioinformatics. 2022;38(16):3892–9.

Takayama KI, et al. Identification of long non-coding RNAs in advanced prostate cancer associated with androgen receptor splicing factors. Commun Biol. 2020;3(1):393.

Talkish J, et al. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. RNA. 2014;20(5):713–20.

Tan Q, et al. Dexmedetomidine Promoted HSPB8 Expression via Inhibiting the lncRNA SNHG14/UPF1 Axis to Inhibit Apoptosis of Nerve Cells in AD : The Role of Dexmedetomidine in AD. Neurotox Res. 2023;41(5):471–80.

Tants JN, et al. NMR-derived secondary structure of the full-length Ox40 mRNA 3′UTR and its multivalent binding to the immunoregulatory RBP Roquin. Nucleic Acids Res. 2022;50(7):4083–99.

Townshend RJL, et al. Geometric deep learning of RNA structure. Science. 2021;373(6558):1047–51.

Twittenhoff C, et al. Lead-seq: transcriptome-wide structure probing in vivo using lead(II) ions. Nucleic Acids Res. 2020;48(12): e71.

Vaculík O, et al. Transfer Learning Allows Accurate RBP Target Site Prediction with Limited Sample Sizes. Biology (Basel). 2023;12(10):1276.

Van Nostrand EL, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods. 2016;13(6):508–14.

V'Kovski P, et al. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol. 2021;19(3):155–70.

Waheed Z, et al. The Role of Tau Proteoforms in Health and Disease. Mol Neurobiol. 2023;60(9):5155–66.

Wan Y, et al. Landscape and variation of RNA secondary structure across the human transcriptome. Nature. 2014;505(7485):706–9.

Wang XW, et al. A microRNA-inducible CRISPR-Cas9 platform serves as a microRNA sensor and cell-type-specific genome regulation tool. Nat Cell Biol. 2019;21(4):522–30.

Wang XW, et al. RNA structure probing uncovers RNA structure-dependent biological functions. Nat Chem Biol. 2021;17(7):755–66.

Wang W, et al. trRosettaRNA: automated prediction of RNA 3D structure with transformer network. Nat Commun. 2023;14(1):7266.

Weng X, et al. Keth-seq for transcriptome-wide RNA structure mapping. Nat Chem Biol. 2020;16(5):489–92.

Wilkinson ME, Charenton C, Nagai K. RNA Splicing by the Spliceosome. Annu Rev Biochem. 2020;89:359–88.

Wu Y, et al. High content design of riboswitch biosensors: All-around rational module-by-module design. Biosens Bioelectron. 2023;220: 114887.

Xie R, et al. Polypyrimidine tract binding protein 1 promotes lymphatic metastasis and proliferation of bladder cancer via alternative splicing of MEIS2 and PKM. Cancer Lett. 2019;449:31–44.

Xu Y, et al. PrismNet: predicting protein-RNA interaction using in vivo RNA structural information. Nucleic Acids Res. 2023;51(W1):W468-w477.

Yang SL, et al. Comprehensive mapping of SARS-CoV-2 interactions in vivo reveals functional virus-host interactions. Nat Commun. 2021;12(1):5113.

Yang W, et al. Long Noncoding RNAs in the Pathogenesis of Insulin Resistance. Int J Mol Sci. 2022;23(24):16054.

Yang Y, et al. RNA secondary structure in mutually exclusive splicing. Nat Struct Mol Biol. 2011;18(2):159–68.

Yilmaz G. Diagnosis of HIV infection and laboratory monitoring of its therapy. J Clin Virol. 2001;21(3):187–96.

Zhao Q, et al. Review of machine learning methods for RNA secondary structure prediction. PLoS Comput Biol. 2021;17(8): e1009291.

Zhang G, et al. mRNA vaccines in disease prevention and treatment. Signal Transduct Target Ther. 2023a;8(1):365.

Zhang H, et al. Algorithm for optimized mRNA design improves stability and immunogenicity. Nature. 2023b;621(7978):396–403.

Zhang J, Kuo CC, Chen L. GC content around splice sites affects splicing through pre-mRNA secondary structures. BMC Genomics. 2011;12:90.

Zhang Y, et al. In vivo structure and dynamics of the SARS-CoV-2 RNA genome. Nat Commun. 2021;12(1):5695.

Zhu H, et al. Dynamic characterization and interpretation for protein-RNA interactions across diverse cellular conditions using HDRNet. Nat Commun. 2023;14(1):6824.

Zinshteyn B, et al. Assaying RNA structure with LASER-Seq. Nucleic Acids Res. 2019;47(1):43–55.

Ziv O, et al. The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. Mol Cell. 2020;80(6):1067-1077.e5.

Ziv O, et al. COMRADES determines in vivo RNA structures and interactions. Nat Methods. 2018;15(10):785–8.

Zubradt M, et al. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. Nat Methods. 2017;14(1):75–82.

Publisher's Note