The ins and outs of lncRNA structure: how, why and what comes next?

Amanda J Blythe^{a,1}, Archa Fox^b, Charles S Bond^{a,1}.

^aSchool of Chemistry and Biochemistry, University of Western Australia, Western Australia, 6009 Australia

^bHarry Perkins Institute of Biomedical Research, Western Australia, 6009 Australia

¹To whom correspondence should be addressed.

Charles Bond,
School of Chemistry and Biochemistry,
University of Western Australia,
35 Stirling Hwy,
Crawley, WA 6009
Australia
Tel: +61 8 6488 4406, Fax: +61 8 6488 1005
email: charles.bond@uwa.edu.au

Amanda Blythe,
School of Chemistry and Biochemistry,
University of Western Australia,
35 Stirling Hwy,
Crawley, WA 6009
Australia
Tel: +61 8 6488 4464, Fax: +61 8 6488 1005
email: amanda.blythe@uwa.edu.au

Abstract

The field of structural biology has the unique advantage of being able to provide a comprehensive picture of biological mechanisms at the molecular and atomic level. Long noncoding RNAs (lncRNAs) represent the new frontier in the molecular biology of complex organisms yet it remains the least characterised of all the classes of RNA. Thousands of new lncRNAs are being reported each year yet very little structural data exists for this rapidly expanding field. The length of lncRNAs range from 200 nt to over 100 kb in length and generally exhibit low cellular abundance. Therefore, obtaining sufficient quantities of lncRNA to use for structural analysis is challenging. However, as technologies develop structures of lncRNAs are starting to emerge providing important information regarding their mechanism of action. Here we review the current methods used to determine the structure of lncRNA and lncRNA:protein complexes and describe the significant contribution structural biology has and will make to the field of lncRNA research.

Keywords

lncRNA; structure; RNA structure; RNA-protein interaction; structural biology; non-coding

1. Introduction

Structural biology has long been at the forefront of being able to provide detailed mechanistic views of important cellular processes. The field of structural biology has made significant contributions to the understanding of protein function, with over 100,000 protein structures in the Protein Data Bank (PDB) to date¹. When this is compared to the

1,078 RNA and 1,522 RNA:protein structures, it is clear that structural biology is lagging behind in the field of RNA research. The ability of structural biology to contribute to RNA research has long been hindered by the difficulties in obtaining sufficient RNA, either by isolating the RNA target or producing it via *in vitro* transcription. These difficulties are greatly exacerbated in the field of long non-coding RNA (lncRNA) research due to the length of the transcripts, which range from 200nt to in excess of 100kb, and their generally low cellular abundance. As such it is no surprise that, with the exception of ribosomal RNA (rRNA), lncRNA is the least characterised in terms of structure and function amongst the various classes of RNA.

lncRNAs represent a new frontier in the molecular biology of complex organisms as it is increasingly evident that they are involved in the regulation of almost every stage of gene expression, as well as being implicated in a variety of disease states. Unlike proteins, whose sequences are often highly conserved across species indicating functional importance, the primary structure of lncRNA generally displays extremely poor evolutionary conservation²⁻⁹. This may be due to nucleotide substitutions in protein-coding sequences being more deleterious compared with non-coding sequence, and that conservation may be through functional or structural mechanisms, rather than being sequence based⁹. To this effect, with the structural characterisation of some lncRNAs we are beginning to see evidence of secondary and tertiary structure conservation¹⁰⁻¹³. The structure-function relationship has long since been proven for proteins, and it is becoming evident that the same can be said for RNA.

Certain functional properties of RNA are determined by their primary structure, with sequence specific processes like the base pairing of mature micro RNA and short interfering RNA molecules to their targets¹⁴. However, RNA also has a unique ability to

adopt a variety of complex secondary and tertiary folds and, because of its long length, lncRNA has the potential for increased structural complexity. Indeed, recent studies have shown that lncRNAs exhibit greater structural complexity than messenger RNA with comparable stability profiles^{15, 16}. The structural versatility of RNA allows it to perform various functions ranging from precise protein recognition (ribonucleoproteins; RNPs), to catalysis (ribozymes¹⁷⁻¹⁹) and metabolite sensing (riboswitches²⁰⁻²²). Current research is demonstrating that many lncRNAs act in conjunction with proteins and other nucleic acids and that their secondary or tertiary structures are important for mediating these interactions.

Currently there are 111, 685 human annotated lncRNA transcripts (LNCipedia 3.1^{23, 24}); however, despite (or perhaps because of) their wide distribution across the genomes of complex organisms only a minute fraction of lncRNAs have been functionally and/or structurally characterised. Indeed, there still exists wide debate over the proportion of these lncRNAs that will eventually be proven to be functional²⁵⁻³². A contributing factor to slow rates of functional determination is certainly the excessive length and low cellular abundance of most lncRNAs, which presents a challenge to most *in vitro* and *in vivo* characterisation techniques. In this review we seek to outline the major advances structural biology has made in the understanding of the molecular mechanism behind lncRNA function, highlight the different methodological approaches the field uses to study lncRNA and to show how these methods can/will be applied in the future to lncRNA research.

2. Methods to determine RNA structure

Determining RNA secondary structure is an important first step towards understanding RNA function. RNA secondary structures comprise of single stranded regions that

connect helices, loops and bulges (Figure 1A). These secondary structures dictate the formation of RNA tertiary structure through the coaxial stacking of adjacent helices at junctions, pseudoknots and kissing loops; and, networks of triple helices, tetraloop-receptor interactions and other structures dictated by sequence-specific or backbone interactions (reviewed in ³³)(Figure 1B).

A variety of methods have been employed to predict and determine structures of RNA and these can be divided into three groups: bioinformatics, biochemical methods and 3-dimensional methods. We have summarised the applications, advantages and disadvantages of the individual methods from each group in Table 1 and briefly describe them here in the context of lncRNA; however, for a more detailed explanation see reviews ³⁴⁻³⁷.

2.1. Bioinformatics

RNA structure prediction programs predict the secondary structure of RNA using the thermodynamics of base pairing, base stacking and hairpins. Whilst these programs have been proven to be accurate for short RNA sequences, the accuracy drops significantly as the length of the transcript increases, thus proving ineffective in predicting the structure of most lncRNAs. Furthermore, these programs are unable to cope with predicting non-Watson-Crick base-pairs and long range tertiary interactions³⁸. Tertiary structure prediction programs have previously accurately predicted the 3D structures of RNA <50 nt in length³⁹; however, by utilizing the experimental information from chemical probing, and as more atomic resolution structures are being determined, these programs are increasingly able to predict tertiary structures of RNAs up to a few hundred nucleotides⁴⁰⁻⁴³. In another novel approach to overcome computational limitations in RNA structure

prediction, “crowdsourcing” games have been developed to capitalize on the collective human intelligence and problem-solving capabilities to answer fundamental questions about RNA folding mechanisms. EteRNA (<http://eterna.cmu.edu/web/>, for 2D structure determination) has led to new RNA design rules⁴⁴ and RNA-Puzzles^{45, 46} (for 3D structure determination) has identified the low accuracy of non-Watson-Crick interactions as the major bottleneck in RNA structure prediction⁴⁶. Through the integration of both 2D and 3D experimental data, computational methods will increasingly be able to become more accurate for longer RNAs.

Phylogenetic analysis looks at the covariance between two nucleotides over a diverse range of species and is currently widely-used for RNA structure prediction^{47, 48}. This technique has been successfully used in the case of ribosomal and riboswitch RNAs⁴⁹⁻⁵²; however, lncRNAs lack sequence diversity and exhibit low sequence conservation which significantly limits the number of available sequences for use in multiple sequence alignments. The development of methods to overcome these challenges and to apply these predictions to cover entire genomes is underway¹³. No matter how precise computational methodologies become, structure prediction techniques will always require experimental validation; however, when used in combination with biochemical structural determination methods, phylogenetic analysis can be a powerful technique in identifying functional regions of structure in lncRNA^{53, 54}.

2.2. Biochemical

There have been many biochemical techniques designed to obtain RNA structural information, most commonly chemical or enzymatic probing. These approaches are able to differentiate between single stranded and helical regions with single nucleotide resolution that result in the generation of 2-D RNA structure maps. RNA cleavage with

nucleases or metal ions serves to map single or double stranded RNA regions depending on the nuclease specificity⁵⁵⁻⁵⁸. Chemical probing relies on chemical base modification at specific single-stranded bases, or binding to mobile regions of the RNA backbone⁵⁹⁻⁶³. These methods are followed by primer extension or end-labelling to detect the cleavage sites or modifications^{36, 64-66}. These techniques have been instrumental in determining the early ribosomal RNA secondary structures, revealing important structural elements in other RNAs and have since been applied to the structural determination of lncRNA as there are generally no length, quantity or heterogeneity restrictions^{49-51, 55, 57, 62, 67-69}. The 2-D information gained from these approaches directs mutagenesis and deletion studies and, in combination with foot-printing studies, can identify interaction motifs and provide important information regarding the mechanism of RNA function⁷⁰.

2.3. *In vivo*

Given the difficulty in obtaining robust and plentiful samples for *in vitro* structure determination, lncRNA structural biology is turning towards the structural interrogations of candidate RNAs *in vivo*, generally from cultured cells. Several chemical probes (Pb²⁺ ions, dimethyl sulfate (DMS), selective 2'-hydroxyl acylation analysed by primer extension (SHAPE) reagents and hydroxyl radicals) have been successful in penetrating cellular membranes, acting on an RNA of interest, and thereby being used to determine RNA secondary structures^{55, 57, 62, 67, 68}. DMS and SHAPE reagents actively modify single-stranded or flexible regions of target RNAs, while Pb²⁺ and hydroxyl radicals cleave the RNA at these regions. Pb²⁺ and hydroxyl radicals have been used to determine the secondary structure of small RNAs, while DMS and SHAPE have been used for the secondary structure determination of longer RNAs such as the HIV RNA genome and the SRA and HOTAIR lncRNAs^{10, 53, 54, 67, 71}. These methods provide us with the ability to assess RNA structure in cells; however, they target single molecules only and still rely on

RNA extraction and primer extension, which is problematic for many lncRNAs that are not present in sufficient abundance for structure probing *in vivo*.

SHAPE is currently the gold standard of RNA secondary structure determination and in theory is applicable to the determination of all lncRNA structures, as it has no length restrictions. However, in practise, obtaining sufficient levels or quantities of RNA for this technique is still a major limitation for its use in lncRNA as it requires either *in vitro* transcription, which assumes the lncRNA will fold autonomously and correctly, or cellular extraction of the lncRNA requiring a high cellular abundance for sufficient detection.

The incorporation of next generation sequencing (NGS) has seen the advancement of secondary structure determination methods such as SHAPE-seq⁷² or DMS-seq⁷³⁻⁷⁵ allowing for lower initial RNA quantities (50 pmol of in-vitro transcribed RNA (SHAPE-seq⁷⁶) compared to 600 pmol (SHAPE⁶⁷) or 500 ng of total cellular input RNA (DMS-Seq⁷⁴) compared to 1-5 µg (DMS⁶²)) with increased resolution. In the future, methods that address the ability to produce significantly longer RNA transcripts *in vitro* or couple *in vivo* RNA chemical modification with NGS to increase the detection sensitivity of low abundant RNA, would mean that SHAPE has the potential to be a highly effective secondary structure determination technique for lncRNA.

2.4. High-throughput / genome wide

The approaches described above are limited to characterising a single target RNA structure; however, with thousands of new lncRNAs being annotated each year, high-throughput structural characterisation techniques are being developed. Parallel analysis of RNA structure (PARS)⁷⁷ and Fragmentation sequencing (FragSeq)⁷⁸ use enzymatic hydrolysis coupled to NGS to analyse the RNA secondary structure content of entire

transcriptomes (Reviewed in ⁷⁹). FragSeq has been successfully used to analyse the structural content of small nucleolar RNAs in mouse cells⁷⁸; however, the technique isolates RNAs up to 100 bases long after nuclease digestion and so many larger RNAs may be under-represented. The use of PARS uncovered a number of general features for yeast mRNAs⁷⁷ and also has the ability to probe structures of long RNAs. An extension of the PARS method incorporating temperature elevation (PARTE) provided a genome wide measurement of RNA folding energies. Diverse RNA structures and functionally important elements were distinguished and identified in the yeast transcriptome¹⁵. PARTE was able to discriminate between classes of RNAs through their melting temperatures and determined that lncRNAs are a stable species which appear to contain a level of structure in between the structural content of the ribosome and a typical mRNA.

2.5. 3-dimensional

3-dimensional (3-D) structural solution is the ultimate goal in macromolecular structure determination as it provides atomic resolution data at the molecular level. 3-D RNA structure determination techniques include small angle x-ray scattering (SAXS), nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography and electron microscopy. For small RNA molecules, 2-D mapping techniques have been almost completely superseded by these atomic resolution approaches; however, they are often labour intensive and typically rely on milligram quantities of robust and homogenous *in vitro* transcribed RNA that is often difficult to obtain for lncRNAs⁸⁰. As such, with the exception of rRNA of which large quantities were readily available, there has been no atomic resolution 3-D structure solution of an entire lncRNA to date.

The various methods used to predict/determine structures of lncRNA all have the same major limitation: obtaining sufficient quantities of robust and homogeneous samples for

very long (over 3 kb) and low abundant lncRNAs. This is particularly problematic for 3-D structural characterisation methods. A limitation of the high-throughput techniques is that the RNAs are extracted from cells and renatured, thus the resulting structures may not be an accurate representation of the *in vivo* structures. Whilst there are now biochemical reagents that can probe RNA structure in living cells, such as SHAPE and DMS, these methods have not yet been adapted to deep sequencing platforms limiting their high-throughput or genome-wide applications. Nevertheless, various combinations of these techniques have been used successfully in characterising various structural elements and probing the entirety of some lncRNAs, which we have described below.

3. Examples of lncRNA structures

Recently, Watts et. al, used SHAPE to interrogate the entire 9173 nt HIV-1 RNA genome identifying new functional elements and long-range RNA interactions⁵⁴. Structures in single-stranded RNA viral genomes are critical to viral replication and host defence evasion. Analysis of the HIV-1 genome structure identified that unstructured protein loops are derived from highly structured RNA elements, suggesting the existence of another level of the genetic code. Additionally, the HIV-1 genome structure revealed that the regions which have high sequence variation that are a critical component for viral host evasion are fundamentally unstructured, while the stable structure regions are evolutionarily conserved. This provides insight into the mechanism of genetic diversity in ssRNA viruses. The HIV-1 genome structure showed the potential for an astonishing amount of information being encoded into the structures of large RNA molecules and makes it apparent that there is much to be gained from the global structural analyses of RNA.

The first experimental characterisation of the secondary structure of an entire human lncRNA, was for the steroid receptor RNA activator (SRA)¹⁰. The authors used a combination of chemical and enzymatic approaches to map the entire secondary structure of the 870 nt lncRNA. The structure determination revealed that the SRA consisted of four major domains organised into a complex secondary structure and, whilst being short enough for mechanistic studies it was long enough to encompass many of the structural characteristics commonly found in rRNA¹⁰. Structure determination of SRA in combination with previous deletion and mutational analysis provided evidence for lncRNA function requiring the complex structural organisation of the entire molecule, rather than being limited to one particular substructure^{10, 81}.

Very recently, the secondary structure of the entire 2.2 kb lncRNA HOTAIR (HOX transcript antisense intergenic RNA) was experimentally determined⁵³. The structural map obtained by using a combination of chemical probes shows that HOTAIR, like SRA, folds into four independent domains. Two of the four sub-domains correspond to predicted protein-binding domains with high conservation displayed in the secondary structure elements surrounding protein-binding motifs. Validation of the map with phylogenetic analysis identified potential key functional regions through highly conserved secondary structure elements.

The secondary structure determination of SRA and HOTAIR provides exciting possibilities for the structural determination of other human lncRNAs of similar length; however, the RNAs for these studies were produced via *in vitro* transcription: a method that is generally not applicable for lncRNAs of lengths over 3 kb. With the many challenges of determining the entire structure of lncRNA meaning that very few lncRNAs have been characterised,

researchers often settle for the high-resolution structure determination of smaller fragments.

For example, the NMR structure of a 32-nucleotide (nt) fragment from the 0.4 kb human telomerase RNA (hTR; also called telomerase RNA component (TERC)) showed a stable hairpin containing an asymmetric internal loop in a domain that was important for telomerase activity⁸² (Figure 2A). The twist in the RNA structure caused by the loop provides a key functional feature by generating a docking site for the enzymatic protein component of the telomerase, hTERT as well as identifying possible sites for drug targeting. Other examples include the NMR structural solution of a 14 nt hairpin from the 17 kb *Xist* lncRNA involved in X-chromosome inactivation, identifying a novel AUCG tetraloop. The tetraloop is unusual in that there is no base-pair between the first and last nucleotide, a turn is introduced between the first two nucleotides rather than the second and third nucleotide, and the final nucleotide of the tetraloop is solvent exposed which is a rare and unusual observation (Figure 2B). Further experiments showed that this novel tetraloop was important for X-chromosome inactivation⁸³. A combined crystallography and NMR approach elucidated the molecular mechanism for the binding of a 20 nt stem loop region within the 600 nt long intergenic non-coding RNA (lincRNA) growth arrest-specific 5 (Gas5) to steroid receptors (SRs)⁸⁴ (Figure 2C). The double stranded stem region of Gas5 competes directly with DNA for the DNA binding domain of the SRs to inhibit their transcriptional activity; and a crystal structure characterising a triple helix from the 8 kb metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) lncRNA that is critical for the prevention of nuclear decay identifies it as a potential target for reducing MALAT1 levels in cancer cells⁸⁵ (Figure 2D).

With the exception of the ribosome, as yet there are no atomic resolution structures of any complete lncRNAs thus the intricacies of tertiary interactions within lncRNA remain a mystery. However, the atomic resolution structures of large RNA molecules illustrate that RNA molecules can display a level of versatility and complexity comparable to proteins. For example, the structure of the 155 nt HIV-1 RNA packaging signal, solved by NMR, adopts a tandem three-way junction. This single RNA conformation is able to inhibit translation and selectively promote the packaging of its unspliced genome during the viral replication of HIV. Formation of the junction results in the sequestration of splice and translation initiation sites and exposes weakly paired or unpaired guanosines enabling their binding to the HIV proteins directing genome packaging (Figure 3A)⁸⁶.

The crystal structure of a bacterial ribonuclease (RNase) P, a 346 nt enzyme that processes the 5' leader sequence of precursor tRNA molecules, is an example of an RNA catalyst which constitutes multiple domains - one for substrate recognition and the other forming the scaffold for the active site that requires the coordination of divalent metal ions for activity⁸⁷ (Figure 3B). The structure identified the location of the active site, the requirement of the protein cofactor in tRNA catalysis and revealed that the enzyme uses shape selective recognition and specific RNA-RNA contacts such as base stacking, A-minor interactions and canonical base-pairing to recognise its tRNA substrate⁸⁷.

Finally the crystal structure of the ribosome, the longest 3-D structure of RNA to be solved to date, illustrates the ability of long RNA molecules to fold into multiple and distinct functional domains. The highly organised RNA of the ribosome is largely an arrangement of helical elements held together by the coaxial stacking between helices, stabilising long-range interactions of single stranded regions, and protein binding (Figure 3C)⁸⁸.

4. lncRNA:protein interactions

Interacting molecules can have an impact on RNA structure and a major recurring theme in lncRNA biology is its ability to function through the recruitment of protein factors. lncRNA:protein complexes function to alter the chromatin state, organize nuclear substructures, and regulate gene expression; however, the mechanisms of these important cellular processes are still largely unknown. In comparison to other classes of RNA, a feature of lncRNA is its propensity to contain a huge number of protein-binding sites. Thus lncRNA can provide a platform for the multimerisation of proteins, such as the *Xist* RNA for Rev proteins⁸³, or can provide a scaffold for the assembly of multiple proteins into larger functional units, such as NEAT1 in the paraspeckle⁸⁹⁻⁹¹. Critical toward understanding the function of lncRNA is the understanding of how they interact with proteins.

The recognition of small RNAs by RNA-binding proteins (RBPs) is through distinct structural elements, RNA-binding domains (RBDs) that are responsible for the specific recruitment of RNA to the target protein⁹²⁻⁹⁵. The RNA recognition motif (RRM) constitutes the most abundant RBD and is hence the most extensively studied. Other common RBDs include the heterogeneous ribonucleoprotein particle homology (KH) domain, zinc-fingers (ZnF), arginine rich motifs (ARMs), and the RGG box. The comprehensive structural analyses of RBDs indicate that they primarily recognise short ssRNA sequences; however, other than occurring in single stranded stretches, these regions can occur in RNA secondary structures such as hairpin loops or in bulges. There are examples of proteins binding to double-stranded (ds) RNA stems and while these proteins seem to recognise the dsRNA shape, they also bind to specific dsRNA targets indicating some element of sequence specificity⁹⁶⁻¹⁰¹. There has been a multitude of RNA:protein structures contributing to the characterisation of how these common RNA binding domains recognise

specific RNA molecules, providing us with a detailed understanding of their mechanism of interaction (Figure 4).

Unlike ribosomal proteins, which interact with the rRNA through electrostatic interactions with the sugar phosphate backbone¹⁰², there is evidence that many lncRNAs bind their proteins partners in a canonical manner. Characterised examples include LARP7 (La-related proteins group 7) that uses its La domain and two RRM domains to bind a stretch of U's at the 3' end and the apical loop of the 3' hairpin of the 331 nt 7SK RNA¹⁰³ (Figure 5A); and, the ARMs of the viral Rev protein homodimer bind with high specificity to the major grooves along the two legs of the 233 nt viral Rev-response element (RRE) lncRNA¹⁰⁴⁻¹⁰⁷ (Figure 5B). Further to these characterised structures, domain analysis shows that many lncRNA-binding proteins (lncRBPs) contain these common RBDs and in multiple copies. For example, the Drosophila behaviour human splicing (DBHS) proteins known to interact with NEAT1 contain two RRMs^{108, 109}, heterogeneous nuclear ribonuclear protein K (HNRNPK) which interacts with both lincRNA-p21 and NEAT1 contains three KH domains and an RGG box^{110, 111}, and the myriad of proteins shown to co-localise to the paraspeckle contain multiple copies and combinations of canonical RBD's¹¹¹.

In contrast to canonical recognition, other lncRBPs such as the polycomb repressor group complex (PRC) proteins do not contain characterised canonical RBD's but are known to interact with multiple lncRNAs¹¹²⁻¹¹⁶. This suggests that they interact with lncRNA through non-canonical backbone interactions, in a similar manner to ribosomal proteins, or they could display unique methods of interactions, such as Gas5 lncRNA binding to the DNA-binding domain of steroid receptors⁸⁴. Further characterization of these lncRNA:protein

assemblies is needed in order to elucidate the molecular mechanisms behind these non-canonical protein:RNA interactions.

5. lncRNA as scaffolds for many proteins.

Many lncRNAs interact with multiple different proteins suggesting that they may act as scaffolds to amalgamate protein complexes into larger functional units. The most well characterised example of this is the ribosome, of which two-thirds is made up of non-coding RNA. The RNA forms two subunits, which in eukaryotes consist of approximately 86 proteins interspersed throughout a 1.9 kb and a 4.7 kb lncRNA and two shorter RNAs (120 nt and 160 nt). Early structural mapping of rRNA using cross-linking and chemical foot-printing methods provided initial physical constraints for building models of the ribosome¹¹⁷⁻¹²⁰ and the initial two subunit morphology of the ribosome was determined by negative-stain electron microscopy. Early cryoEM studies began to provide some details about the mechanism of action^{121, 122} including evidence of the structural dynamics¹²³; however, it was the high resolution crystal structure of the ribosome that provided fundamental insight into its mechanism, paving the way for more complex biochemical and genetic experiments to further evolve our understanding of mechanisms and regulation of ribosome function.

The ribosome represents a highly structured and rigid scaffold that could be used for structural studies due to its high cellular abundance and robustness allowing it to be purified intact from cells. Other lncRNAs are suggested to form flexible scaffolds that are not so abundant or robust, making them significantly harder to characterise. The yeast telomerase RNA, *Xist* and HOTAIR have all been suggested to be lncRNA scaffolds that tether a multitude of proteins together¹²⁴⁻¹²⁷. In the case of HOTAIR it is proposed that the

lncRNA may serve as a scaffold by providing a binding surface to assemble and tether two separate histone modification complexes¹²⁷. For the yeast telomerase, large stretches of the RNA could be deleted without significant loss in function suggesting that the RNA serves primarily to tether the protein binding subunits together¹²⁸, and a similar phenomena, whereby significant segments of the RNA are dispensable for function, has been seen for *Xist*¹²⁹. The mechanism for how these lncRNA:protein complexes are assembled is currently unknown; however, in light of the recent secondary structure determination of HOTAIR, combined with more targeted biochemical and structural experiments, a detailed molecular understanding of its assembly mechanism may soon be revealed.

Flexible lncRNA scaffolds can also play a role in the assembly of nuclear bodies. There are several types of non-coding RNAs that are physiologically enriched within various types of nuclear bodies such as Cajal bodies, nuclear speckles, paraspeckles and nuclear stress bodies (reviewed in ¹³⁰⁻¹³³). In several cases, formation of these bodies occurs through a nucleating RNA, resulting in the recruitment and accumulation of associating proteins. The best-studied lncRNA involved in the formation of nuclear bodies is NEAT1's nucleation of paraspeckles.

Paraspeckles are a dynamic, ribonucleoprotein body found in mammalian cell nuclei consisting of the 23 kb NEAT1 transcript (NEAT1_v2), and a shorter 3.7 kb isoform (NEAT1_v1)^{111, 134}. These two isoforms provide a scaffold for the recruitment of in excess of 40 nuclear RNA binding proteins⁸⁹⁻⁹¹. Unlike the lncRNAs GomaFu or MALAT1, which are both enriched in, but not required for the formation of gomaFu speckles and nuclear speckles respectively^{90, 135}, paraspeckle biogenesis is dependant on the transcription of

NEAT1 and requires seven essential paraspeckle proteins HNRNPK, NONO, RBM14, SFPQ, DAZAP, FUS and HNRNPH3^{90, 91, 111, 136}. The macromolecular structure of the paraspeckle is yet to be elucidated; further, it is unclear as to the stoichiometry of paraspeckle proteins and NEAT1 RNA within a single paraspeckle.

Electron microscopy studies suggest that NEAT1 accumulates in paraspeckles to form a regular network of RNA fibres with NEAT1_v1 and the 5' and 3' ends of NEAT1_v2 localising to the periphery of the paraspeckle, and the central sequence of NEAT1_v2 localising to the inner core¹³⁷. The macromolecular arrangement of NEAT1 and the paraspeckle proteins is currently unknown. The scaffold of the paraspeckle could be formed via RNA-RNA interactions between NEAT1 transcripts; alternatively, multiple NEAT1_v2 transcripts could be connected by means of interaction with the paraspeckle proteins. This is highly likely given that essential DBHS paraspeckle proteins, PSPC1, NONO and SFPQ have the propensity to form dimers and oligomers^{108, 138}, and that many paraspeckle proteins are required for paraspeckle formation without being required to maintain NEAT1 levels¹³⁹. In this latter case, protein-protein interactions may be bridging RNA-protein interactions to form a paraspeckle (Figure 6).

The oligomerisation of the essential DBHS paraspeckle proteins could provide key insight into the mechanism of macromolecular assembly of paraspeckles. Recent structural characterisation of SFPQ revealed a remarkably extended coiled-coil through which extensive interactions mediated infinite polymerization of SFPQ homodimers¹⁰⁸. The extended coiled-coil is a feature of the DBHS proteins and their ability to polymerise through this domain was proven to be critical for paraspeckle formation and localization of SFPQ and PSPC1 to paraspeckles. The oligomerisation of DBHS proteins through the coiled-coil brings together their globular functional domains, such as the RRM domains which are

the proposed sites of NEAT1 interaction. Thus, one model for the assembly of paraspeckle complexes is through the cooperative oligomerisation of DBHS proteins on the NEAT1 scaffold utilizing both protein-protein and protein-RNA interactions¹⁰⁸ (Figure 5). A similar mechanism of protein oligomerisation on lncRNA is seen for the formation of the Rev-RRE complex formation¹⁴⁰. It is unknown what defines the composition and architecture of paraspeckles but interestingly, the polymerisation domains of DBHS homo- and heterodimers were shown to reversibly associate and dissociate. Comparable to the network of coiled-coil interactions Lee et al. observed for the DBHS proteins is the reversible and functional fibrillar-like aggregates of the low complexity domains found in many nucleic acid binding proteins, including several paraspeckle proteins¹⁴¹. These data suggest a mechanism for the dynamic formation and dissociation of the paraspeckle through an equilibrium in and out of the functional oligomerised or aggregated state of paraspeckle proteins.

The paraspeckle represents a unique and intriguing lncRNA:protein assembly of which there are many unanswered questions surrounding the details of its macromolecular structure. Techniques such as CryoEM or cryoelectron tomography are ideal to study the structure of these micron sized nuclear bodies, yet the extreme length and low cellular abundance of NEAT1 makes *in vitro* transcription or cellular extraction difficult and the dynamic nature of paraspeckles renders them challenging to isolate. The difficulties in the structural determination of paraspeckles are representative of the challenges associated with many large lncRNA:protein complexes, and highlight the requirement for technological advancements in current structural biology techniques.

The mechanism for lncRNA:protein complex assembly is unknown, and the task of interrogating these mechanisms requires detailed knowledge of the lncRNA structure and

of the individual RNA:protein interactions. Interacting proteins need to be identified, the RNA binding specificity of the individual proteins characterised, protein binding sites on the RNA identified, biochemical and structural characterisation of the individual protein:RNA complexes, and the global structure of RNA determined. A task such as this is daunting given our current technological limitations, yet necessary if we want to ascertain the molecular mechanisms behind lncRNA function.

6. Discussion

lncRNAs have emerged as essential regulators in almost all aspects of biology and are being found to play a role in a variety of disease states. lncRNAs display greater tissue specificity compared to other classes of RNA making them desirable candidates in the quest for diagnostic markers; therefore, an improved understanding of the role of lncRNA would aid pharmaceutical design and novel therapeutic strategies¹⁴². Structural biology is of critical importance for future studies on lncRNAs through building the necessary groundwork to interpret their functions, study the effect of mutations on lncRNA function and elucidating their direct connection to disease.

The overall architecture of lncRNA remains unknown; however, similar to the domain configuration seen in the ribosome, determining the secondary structure of the SRA and HOTAIR lncRNA revealed the presence of distinct domains within the sequences. With the possibility that all lncRNAs consist of multiple structural or functional domains, it follows that each domain could be studied independently, comparable to the approach used for multidomain proteins or large protein complexes. By focusing on individual domains, lncRNAs could be separated into more manageable lengths for atomic resolution studies. Thus perhaps a more realistic way to interrogate lncRNA structures in the future

would be in first characterizing the secondary structure of lncRNA in order to identify the functional and structural regions to target for tertiary structure determination.

Tertiary nucleic acid structures are publically available through the PDB and the Nucleic Acid Database (NDB)^{143, 144}, and their secondary structure information is able to be extracted for visualization and manipulation through tools such as RnaMLView¹⁴⁵. However, there is no similarly available database for the sharing of RNA secondary structure information, such as the structural annotations of the HOTAIR or SRA lncRNA maps. As more secondary structures of lncRNAs are experimentally determined, the generation of a centralized depository of RNA secondary structure annotations will be important for the dissemination of the data. An obvious choice would be to incorporate RNA secondary structure maps into the NDB, which is already a hub for RNA tertiary structures, as a sequence and secondary structure annotation file. Another candidate system for the distribution of RNA secondary structure data could be the inclusion of experimentally determined RNA secondary structure tracks into the UCSC genome browser¹⁴⁶; analogous perhaps to the work of Smith et al.¹³ whose evolutionary conserved RNA secondary structure predictions can be visualized on the human genome browser via a public track hub (www.martinalexandersmith.com/ecs/). The RNA secondary structures annotated at nucleotide resolution are then easily extracted as input into RNA folding software.

While there are many available 2-D and 3-D techniques to study the structures of proteins and small RNAs, they have not yet been sufficiently adapted for the lncRNA field. Size restrictions are still the major limitation of most available techniques. Out of the 100,000+ annotated lncRNAs, only two have experimentally determined secondary structures. Thus, it is obvious that new technological advances are required for the ability to produce

long RNA transcripts through either *in vitro* or *in vivo* methods in order to obtain high-resolution structural characterisation of lncRNAs and their binding partners. CryoEM is revolutionizing the structural biology of macromolecular protein assemblies. It negates the need for crystals, requires minimal sample and can differentiate biochemical and conformational heterogeneities of the samples. Indeed it has been predicted that cryoEM will largely supersede crystallography¹⁴⁷. The future of lncRNA structural biology will most likely require a combined approach of high-resolution global imaging techniques such as cryoEM in combination with 2D structural determination methods such as SHAPE. Structural maps can then be used to guide detailed biophysical explorations in order to obtain a complete overview of the molecular functions and mechanisms of the new frontier of molecular biology.

Acknowledgements

This work was supported by the National Health and Medical Research Council (NHMRC) grant numbers 1048659 and 1050585.

References

1. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Research* 28, 235-242.
2. Wood, E. J., Chin-Inmanu, K., Jia, H., and Lipovich, L. (2013) Sense-antisense gene pairs: sequence, transcription, and structure are not conserved between human and mouse, *Frontiers in genetics* 4, 183.

3. Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., Samudrala, R., Yu, J., and Wong, G. K. (2004) Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs, *Nature* 431, 1 p following 757; discussion following 757.
4. Pang, K. C., Frith, M. C., and Mattick, J. S. (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function, *Trends in genetics : TIG* 22, 1-5.
5. Kutter, C., Watt, S., Stefflova, K., Wilson, M. D., Goncalves, A., Ponting, C. P., Odom, D. T., and Marques, A. C. (2012) Rapid turnover of long noncoding RNAs and the evolution of gene expression, *PLoS genetics* 8, e1002841.
6. Engstrom, P. G., Suzuki, H., Ninomiya, N., Akalin, A., Sessa, L., Lavorgna, G., Brozzi, A., Luzzi, L., Tan, S. L., Yang, L., Kunarso, G., Ng, E. L., Batalov, S., Wahlestedt, C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Wells, C., Bajic, V. B., Orlando, V., Reid, J. F., Lenhard, B., and Lipovich, L. (2006) Complex Loci in human and mouse genomes, *PLoS genetics* 2, e47.
7. Ponjavic, J., Ponting, C. P., and Lunter, G. (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs, *Genome Res* 17, 556-565.
8. Johnsson, P., Lipovich, L., Grander, D., and Morris, K. V. (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function, *Biochim Biophys Acta* 1840, 1063-1071.
9. Ponting, C. P., Oliver, P. L., and Reik, W. (2009) Evolution and functions of long noncoding RNAs, *Cell* 136, 629-641.
10. Novikova, I. V., Hennelly, S. P., and Sanbonmatsu, K. Y. (2012) Structural architecture of the human long non-coding RNA, steroid receptor RNA activator, *Nucleic Acids Res* 40, 5034-5051.
11. Kino, T., Hurt, D. E., Ichijo, T., Nader, N., and Chrousos, G. P. (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor, *Science signaling* 3, ra8.

12. Theimer, C. A., Blois, C. A., and Feigon, J. (2005) Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function, *Mol Cell* 17, 671-682.
13. Smith, M. A., Gesell, T., Stadler, P. F., and Mattick, J. S. (2013) Widespread purifying selection on RNA structure in mammals, *Nucleic Acids Res* 41, 8220-8236.
14. Carthew, R. W., and Sontheimer, E. J. (2009) Origins and Mechanisms of miRNAs and siRNAs, *Cell* 136, 642-655.
15. Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., Makino, D. L., Nutter, R. C., Segal, E., and Chang, H. Y. (2012) Genome-wide measurement of RNA folding energies, *Mol Cell* 48, 169-181.
16. Clark, M. B., Johnston, R. L., Inostroza-Ponta, M., Fox, A. H., Fortini, E., Moscato, P., Dinger, M. E., and Mattick, J. S. (2012) Genome-wide analysis of long noncoding RNA stability, *Genome Res* 22, 885-898.
17. Wu, Q., Huang, L., and Zhang, Y. (2009) The structure and function of catalytic RNAs, *Science in China. Series C, Life sciences / Chinese Academy of Sciences* 52, 232-244.
18. Lehmann, K., and Schmidt, U. (2003) Group II introns: structure and catalytic versatility of large natural ribozymes, *Critical reviews in biochemistry and molecular biology* 38, 249-303.
19. Cech, T. R. (2002) Ribozymes, the first 20 years, *Biochem Soc Trans* 30, 1162-1166.
20. Serganov, A., and Patel, D. J. (2012) Metabolite recognition principles and molecular mechanisms underlying riboswitch function, *Annual review of biophysics* 41, 343-370.
21. Haller, A., Souliere, M. F., and Micura, R. (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms, *Accounts of chemical research* 44, 1339-1348.
22. Batey, R. T. (2015) Riboswitches: still a lot of undiscovered country, *RNA (New York, N.Y.)* 21, 560-563.

23. Volders, P. J., Helsens, K., Wang, X., Menten, B., Martens, L., Gevaert, K., Vandesompele, J., and Mestdagh, P. (2013) LNCipedia: a database for annotated human lncRNA transcript sequences and structures, *Nucleic Acids Res* 41, D246-251.
24. Volders, P. J., Verheggen, K., Menschaert, G., Vandepoele, K., Martens, L., Vandesompele, J., and Mestdagh, P. (2015) An update on LNCipedia: a database for annotated human lncRNA sequences, *Nucleic Acids Res* 43, D174-180.
25. Clark, M. B., Amaral, P. P., Schlesinger, F. J., Dinger, M. E., Taft, R. J., Rinn, J. L., Ponting, C. P., Stadler, P. F., Morris, K. V., Morillon, A., Rozowsky, J. S., Gerstein, M. B., Wahlestedt, C., Hayashizaki, Y., Carninci, P., Gingeras, T. R., and Mattick, J. S. (2011) The reality of pervasive transcription, *PLoS biology* 9, e1000625; discussion e1001102.
26. Clark, M. B., Choudhary, A., Smith, M. A., Taft, R. J., and Mattick, J. S. (2013) The dark matter rises: the expanding world of regulatory RNAs, *Essays Biochem* 54, 1-16.
27. Dinger, M. E., Amaral, P. P., Mercer, T. R., and Mattick, J. S. (2009) Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications, *Briefings in functional genomics & proteomics* 8, 407-423.
28. Morris, K. V., and Mattick, J. S. (2014) The rise of regulatory RNA, *Nature reviews. Genetics* 15, 423-437.
29. Pheasant, M., and Mattick, J. S. (2007) Raising the estimate of functional human sequences, *Genome Res* 17, 1245-1253.
30. Ponting, C. P., and Belgard, T. G. (2010) Transcribed dark matter: meaning or myth?, *Human molecular genetics* 19, R162-168.
31. Ponting, C. P., and Hardison, R. C. (2011) What fraction of the human genome is functional?, *Genome Res* 21, 1769-1776.
32. van Bakel, H., and Hughes, T. R. (2009) Establishing legitimacy and function in the new transcriptome, *Briefings in functional genomics & proteomics* 8, 424-436.

33. Butcher, S. E., and Pyle, A. M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks, *Accounts of chemical research* 44, 1302-1311.
34. Novikova, I. V., Hennelly, S. P., and Sanbonmatsu, K. Y. (2013) Tackling structures of long noncoding RNAs, *Int J Mol Sci* 14, 23672-23684.
35. Kenyon, J., Prestwood, L., and Lever, A. (2014) Current perspectives on RNA secondary structure probing, *Biochem Soc Trans* 42, 1251-1255.
36. Ziehler, W. A., and Engelke, D. R. (2001) Probing RNA structure with chemical reagents and enzymes, *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.] Chapter 6*, Unit 6 1.
37. Wan, Y., Kertesz, M., Spitale, R. C., Segal, E., and Chang, H. Y. (2011) Understanding the transcriptome through RNA structure, *Nature reviews. Genetics* 12, 641-655.
38. Miao, Z., Adamiak, R. W., Blanchet, M. F., Boniecki, M., Bujnicki, J. M., Chen, S. J., Cheng, C., Chojnowski, G., Chou, F. C., Cordero, P., Cruz, J. A., Ferre-D'amare A, R., Das, R., Ding, F., Dokholyan, N. V., Dunin-Horkawicz, S., Kladwang, W., Krokhotin, A., Lach, G., Magnus, M., Major, F., Mann, T. H., Masquida, B., Matelska, D., Meyer, M., Peselis, A., Popena, M., Purzycka, K. J., Serganov, A., Stasiewicz, J., Szachniuk, M., Tandon, A., Tian, S., Wang, J., Xiao, Y., Xu, X., Zhang, J., Zhao, P., Zok, T., and Westhof, E. (2015) RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures, *RNA (New York, N.Y.)*.
39. Sharma, S., Ding, F., and Dokholyan, N. V. (2008) iFoldRNA: three-dimensional RNA structure prediction and folding, *Bioinformatics (Oxford, England)* 24, 1951-1952.
40. Cheng, C. Y., Chou, F. C., and Das, R. (2015) Modeling complex RNA tertiary folds with Rosetta, *Methods in enzymology* 553, 35-64.

41. Kerpedjiev, P., Honer Zu Siederdissen, C., and Hofacker, I. L. (2015) Predicting RNA 3D structure using a coarse-grain helix-centered model, *RNA (New York, N.Y.)* 21, 1110-1121.
42. Kim, N., Laing, C., Elmetwaly, S., Jung, S., Curuksu, J., and Schlick, T. (2014) Graph-based sampling for approximating global helical topologies of RNA, *Proc Natl Acad Sci U S A* 111, 4079-4084.
43. Krokhotin, A., Houlihan, K., and Dokholyan, N. V. (2015) iFoldRNA v2: folding RNA with constraints, *Bioinformatics (Oxford, England)*.
44. Lee, J., Kladwang, W., Lee, M., Cantu, D., Azizyan, M., Kim, H., Limpaecher, A., Yoon, S., Treuille, A., Das, R., and Ete, R. N. A. P. (2014) RNA design rules from a massive open laboratory, *Proc Natl Acad Sci U S A* 111, 2122-2127.
45. Cruz, J. A., Blanchet, M. F., Boniecki, M., Bujnicki, J. M., Chen, S. J., Cao, S., Das, R., Ding, F., Dokholyan, N. V., Flores, S. C., Huang, L., Lavender, C. A., Lisi, V., Major, F., Mikolajczak, K., Patel, D. J., Philips, A., Puton, T., Santalucia, J., Sijenyi, F., Hermann, T., Rother, K., Rother, M., Serganov, A., Skorupski, M., Soltysinski, T., Sripakdeevong, P., Tuszynska, I., Weeks, K. M., Waldsich, C., Wildauer, M., Leontis, N. B., and Westhof, E. (2012) RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction, *RNA (New York, N.Y.)* 18, 610-625.
46. Miao, Z., Adamiak, R. W., Blanchet, M. F., Boniecki, M., Bujnicki, J. M., Chen, S. J., Cheng, C., Chojnowski, G., Chou, F. C., Cordero, P., Cruz, J. A., Ferre-D'Amare, A. R., Das, R., Ding, F., Dokholyan, N. V., Dunin-Horkawicz, S., Kladwang, W., Krokhotin, A., Lach, G., Magnus, M., Major, F., Mann, T. H., Masquida, B., Matelska, D., Meyer, M., Peselis, A., Popena, M., Purzycka, K. J., Serganov, A., Stasiewicz, J., Szachniuk, M., Tandon, A., Tian, S., Wang, J., Xiao, Y., Xu, X., Zhang, J., Zhao, P., Zok, T., and Westhof, E. (2015) RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures, *RNA (New York, N.Y.)* 21, 1066-1084.

47. Lindgreen, S., Gardner, P. P., and Krogh, A. (2006) Measuring covariation in RNA alignments: physical realism improves information measures, *Bioinformatics (Oxford, England)* 22, 2988-2995.
48. Smit, S., Knight, R., and Heringa, J. (2009) RNA structure prediction from evolutionary patterns of nucleotide composition, *Nucleic Acids Res* 37, 1378-1386.
49. Woese, C. R., Magrum, L. J., Gupta, R., Siegel, R. B., Stahl, D. A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J. J., and Noller, H. F. (1980) Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence, *Nucleic Acids Res* 8, 2275-2293.
50. Noller, H. F., and Woese, C. R. (1981) Secondary structure of 16S ribosomal RNA, *Science* 212, 403-411.
51. Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F., Herr, W., Stahl, D. A., Gupta, R., and Waese, C. R. (1981) Secondary structure model for 23S ribosomal RNA, *Nucleic Acids Res* 9, 6167-6189.
52. Fox, G. E., and Woese, C. R. (1975) 5S RNA secondary structure, *Nature* 256, 505-507.
53. Somarowthu, S., Legiewicz, M., Chillon, I., Marcia, M., Liu, F., and Pyle, A. M. (2015) HOTAIR Forms an Intricate and Modular Secondary Structure, *Mol Cell*.
54. Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr., Swanstrom, R., Burch, C. L., and Weeks, K. M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome, *Nature* 460, 711-716.
55. Adilakshmi, T., Lease, R. A., and Woodson, S. A. (2006) Hydroxyl radical footprinting in vivo: mapping macromolecular structures with synchrotron radiation, *Nucleic Acids Res* 34, e64.
56. Ingle, S., Azad, R. N., Jain, S. S., and Tullius, T. D. (2014) Chemical probing of RNA with the hydroxyl radical at single-atom resolution, *Nucleic Acids Res* 42, 12758-12767.

57. Lindell, M., Romby, P., and Wagner, E. G. (2002) Lead(II) as a probe for investigating RNA structure in vivo, *RNA (New York, N.Y.)* 8, 534-541.
58. Tullius, T. D., and Greenbaum, J. A. (2005) Mapping nucleic acid structure by hydroxyl radical cleavage, *Current opinion in chemical biology* 9, 127-134.
59. Merino, E. J., Wilkinson, K. A., Coughlan, J. L., and Weeks, K. M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE), *Journal of the American Chemical Society* 127, 4223-4231.
60. Mortimer, S. A., and Weeks, K. M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry, *Journal of the American Chemical Society* 129, 4144-4145.
61. Regulski, E. E., and Breaker, R. R. (2008) In-line probing analysis of riboswitches, *Methods in molecular biology (Clifton, N.J.)* 419, 53-67.
62. Wells, S. E., Hughes, J. M., Igel, A. H., and Ares, M., Jr. (2000) Use of dimethyl sulfate to probe RNA structure in vivo, *Methods in enzymology* 318, 479-493.
63. Wilkinson, K. A., Merino, E. J., and Weeks, K. M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution, *Nature protocols* 1, 1610-1616.
64. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P., and Ehresmann, B. (1987) Probing the structure of RNAs in solution, *Nucleic Acids Res* 15, 9109-9128.
65. Peattie, D. A., and Gilbert, W. (1980) Chemical probes for higher-order structure in RNA, *Proc Natl Acad Sci U S A* 77, 4679-4682.
66. Stern, S., Moazed, D., and Noller, H. F. (1988) Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension, *Methods in enzymology* 164, 481-489.
67. Wilkinson, K. A., Gorelick, R. J., Vasa, S. M., Guex, N., Rein, A., Mathews, D. H., Giddings, M. C., and Weeks, K. M. (2008) High-throughput SHAPE analysis reveals

structures in HIV-1 genomic RNA strongly conserved across distinct biological states, *PLoS biology* 6, e96.

68. Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T., and Chang, H. Y. (2013) RNA SHAPE analysis in living cells, *Nature chemical biology* 9, 18-20.

69. Beniaminov, A., Westhof, E., and Krol, A. (2008) Distinctive structures between chimpanzee and human in a brain noncoding RNA, *RNA (New York, N.Y.)* 14, 1270-1275.

70. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine, *Nature structural biology* 10, 701-707.

71. Charpentier, B., Stutz, F., and Rosbash, M. (1997) A dynamic in vivo view of the HIV-I Rev-RRE interaction, *J Mol Biol* 266, 950-962.

72. Lucks, J. B., Mortimer, S. A., Trapnell, C., Luo, S., Aviran, S., Schroth, G. P., Pachter, L., Doudna, J. A., and Arkin, A. P. (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq), *Proc Natl Acad Sci U S A* 108, 11063-11068.

73. Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., and Assmann, S. M. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features, *Nature* 505, 696-700.

74. Talkish, J., May, G., Lin, Y., Woolford, J. L., Jr., and McManus, C. J. (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure, *RNA (New York, N.Y.)* 20, 713-720.

75. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J. S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo, *Nature* 505, 701-705.

76. Loughrey, D., Watters, K. E., Settle, A. H., and Lucks, J. B. (2014) SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing, *Nucleic Acids Res* 42.
77. Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., and Segal, E. (2010) Genome-wide measurement of RNA secondary structure in yeast, *Nature* 467, 103-107.
78. Underwood, J. G., Uzilov, A. V., Katzman, S., Onodera, C. S., Mainzer, J. E., Mathews, D. H., Lowe, T. M., Salama, S. R., and Haussler, D. (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing, *Nature methods* 7, 995-1001.
79. Westhof, E., and Romby, P. (2010) The RNA structurome: high-throughput probing, *Nature methods* 7, 965-967.
80. Flores, J. K., Walshe, J. L., and Ataide, S. F. (2014) RNA and RNA-Protein Complex Crystallography and its Challenges, *Aust J Chem* 67, 1741-1750.
81. Lanz, R. B., Razani, B., Goldberg, A. D., and O'Malley, B. W. (2002) Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA), *Proc Natl Acad Sci U S A* 99, 16081-16086.
82. Leeper, T. C., and Varani, G. (2005) The structure of an enzyme-activating fragment of human telomerase RNA, *RNA (New York, N.Y.)* 11, 394-403.
83. Duszczek, M. M., Wutz, A., Rybin, V., and Sattler, M. (2011) The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization, *RNA (New York, N.Y.)* 17, 1973-1982.
84. Hudson, W. H., Pickard, M. R., de Vera, I. M., Kuiper, E. G., Mourtada-Maarabouni, M., Conn, G. L., Kojetin, D. J., Williams, G. T., and Ortlund, E. A. (2014) Conserved sequence-specific lincRNA-steroid receptor interactions drive transcriptional repression and direct cell fate, *Nature communications* 5, 5395.

85. Brown, J. A., Bulkley, D., Wang, J., Valenstein, M. L., Yario, T. A., Steitz, T. A., and Steitz, J. A. (2014) Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix, *Nat Struct Mol Biol* 21, 633-640.
86. Keane, S. C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Holic, A., Florwick, A., Santos, J., Bolden, N. C., McCowin, S., Case, D. A., Johnson, B. A., Salemi, M., Telesnitsky, A., and Summers, M. F. (2015) RNA structure. Structure of the HIV-1 RNA packaging signal, *Science* 348, 917-921.
87. Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragon, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA, *Nature* 468, 784-789.
88. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonnrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit, *Nature* 407, 327-339.
89. Bond, C. S., and Fox, A. H. (2009) Paraspeckles: nuclear bodies built on long noncoding RNA, *J Cell Biol* 186, 637-644.
90. Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A., and Lawrence, J. B. (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles, *Mol Cell* 33, 717-726.
91. Sasaki, Y. T., and Hirose, T. (2009) How to build a paraspeckle, *Genome biology* 10, 227.
92. Auweter, S. D., Oberstrass, F. C., and Allain, F. H. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition?, *Nucleic Acids Res* 34, 4943-4959.
93. Hudson, W. H., and Ortlund, E. A. (2014) The structure, function and evolution of proteins that bind DNA and RNA, *Nat Rev Mol Cell Biol* 15, 749-760.

94. Messias, A. C., and Sattler, M. (2004) Structural basis of single-stranded RNA recognition, *Accounts of chemical research* 37, 279-287.
95. Varani, G., and Nagai, K. (1998) RNA recognition by RNP proteins during RNA processing, *Annual review of biophysics and biomolecular structure* 27, 407-445.
96. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature* 372, 432-438.
97. Allain, F. H., Bouvet, P., Dieckmann, T., and Feigon, J. (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin, *EMBO J* 19, 6870-6881.
98. Masliah, G., Barraud, P., and Allain, F. H. (2013) RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence, *Cell Mol Life Sci* 70, 1875-1895.
99. Lu, D., Searles, M. A., and Klug, A. (2003) Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition, *Nature* 426, 96-100.
100. Ramos, A., Grunert, S., Adams, J., Micklem, D. R., Proctor, M. R., Freund, S., Bycroft, M., St Johnston, D., and Varani, G. (2000) RNA recognition by a Staufen double-stranded RNA-binding domain, *EMBO J* 19, 997-1009.
101. Ryter, J. M., and Schultz, S. C. (1998) Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA, *EMBO J* 17, 7505-7513.
102. Korobeinikova, A. V., Garber, M. B., and Gongadze, G. M. (2012) Ribosomal proteins: structure, function, and evolution, *Biochemistry. Biokhimiia* 77, 562-574.
103. Uchikawa, E., Natchiar, K. S., Han, X., Proux, F., Roblin, P., Zhang, E., Durand, A., Klaholz, B. P., and Dock-Bregeon, A. C. (2015) Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7, *Nucleic Acids Res.*

104. Daugherty, M. D., D'Orso, I., and Frankel, A. D. (2008) A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA, *Mol Cell* 31, 824-834.
105. Daugherty, M. D., Liu, B., and Frankel, A. D. (2010) Structural basis for cooperative RNA binding and export complex assembly by HIV Rev, *Nat Struct Mol Biol* 17, 1337-1342.
106. DiMattia, M. A., Watts, N. R., Stahl, S. J., Rader, C., Wingfield, P. T., Stuart, D. I., Steven, A. C., and Grimes, J. M. (2010) Implications of the HIV-1 Rev dimer structure at 3.2 Å resolution for multimeric binding to the Rev response element, *Proc Natl Acad Sci U S A* 107, 5810-5814.
107. Fang, X., Wang, J., O'Carroll, I. P., Mitchell, M., Zuo, X., Wang, Y., Yu, P., Liu, Y., Rausch, J. W., Dyba, M. A., Kjems, J., Schwieters, C. D., Seifert, S., Winans, R. E., Watts, N. R., Stahl, S. J., Wingfield, P. T., Byrd, R. A., Le Grice, S. F., Rein, A., and Wang, Y. X. (2013) An unusual topological structure of the HIV-1 Rev response element, *Cell* 155, 594-605.
108. Lee, M., Sadowska, A., Bekere, I., Ho, D., Gully, B. S., Lu, Y., Iyer, K. S., Trewhella, J., Fox, A. H., and Bond, C. S. (2015) The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation, *Nucleic Acids Res.*
109. Passon, D. M., Lee, M., Rackham, O., Stanley, W. A., Sadowska, A., Filipovska, A., Fox, A. H., and Bond, C. S. (2012) Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation, *Proc Natl Acad Sci U S A* 109, 4846-4850.
110. Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M. J., Kenzelmann-Broz, D., Khalil, A. M., Zuk, O., Amit, I., Rabani, M., Attardi, L. D., Regev, A., Lander, E. S., Jacks,

- T., and Rinn, J. L. (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response, *Cell* 142, 409-419.
111. Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y. F., Goshima, N., and Hirose, T. (2012) Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles, *EMBO J* 31, 4020-4034.
112. Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., Song, J. J., Kingston, R. E., Borowsky, M., and Lee, J. T. (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq, *Mol Cell* 40, 939-953.
113. Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs, *Cell* 129, 1311-1323.
114. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome, *Science* 322, 750-756.
115. Yap, K. L., Li, S., Munoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M. J., and Zhou, M. M. (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a, *Mol Cell* 38, 662-674.
116. Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B. R., Protacio, A., Flynn, R. A., Gupta, R. A., Wysocka, J., Lei, M., Dekker, J., Helms, J. A., and Chang, H. Y. (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression, *Nature* 472, 120-124.
117. Doring, T., Mitchell, P., Osswald, M., Bochkariov, D., and Brimacombe, R. (1994) The decoding region of 16S RNA; a cross-linking study of the ribosomal A, P and E sites using tRNA derivatized at position 32 in the anticodon loop, *EMBO J* 13, 2677-2685.

118. Moazed, D., and Noller, H. F. (1986) Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes, *Cell* 47, 985-994.
119. Moazed, D., and Noller, H. F. (1989) Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites, *Cell* 57, 585-597.
120. Moazed, D., Stern, S., and Noller, H. F. (1986) Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension, *J Mol Biol* 187, 399-416.
121. Agrawal, R. K., Penczek, P., Grassucci, R. A., Li, Y., Leith, A., Nierhaus, K. H., and Frank, J. (1996) Direct visualization of A-, P-, and E-site transfer RNAs in the Escherichia coli ribosome, *Science* 271, 1000-1002.
122. Stark, H., Orlova, E. V., Rinke-Appel, J., Junke, N., Mueller, F., Rodnina, M., Wintermeyer, W., Brimacombe, R., and van Heel, M. (1997) Arrangement of tRNAs in pre- and posttranslocational ribosomes revealed by electron cryomicroscopy, *Cell* 88, 19-28.
123. Frank, J., and Agrawal, R. K. (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation, *Nature* 406, 318-322.
124. Zappulla, D. C., and Cech, T. R. (2004) Yeast telomerase RNA: a flexible scaffold for protein subunits, *Proc Natl Acad Sci U S A* 101, 10024-10029.
125. Wutz, A., Rasmussen, T. P., and Jaenisch, R. (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA, *Nature genetics* 30, 167-174.
126. Yoon, J. H., Abdelmohsen, K., Kim, J., Yang, X., Martindale, J. L., Tominaga-Yamanaka, K., White, E. J., Orjalo, A. V., Rinn, J. L., Kreft, S. G., Wilson, G. M., and Gorospe, M. (2013) Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination, *Nature communications* 4, 2939.
127. Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E., and Chang, H. Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes, *Science* 329, 689-693.

128. Livengood, A. J., Zaug, A. J., and Cech, T. R. (2002) Essential regions of *Saccharomyces cerevisiae* telomerase RNA: separate elements for Est1p and Est2p interaction, *Molecular and cellular biology* 22, 2366-2374.
129. Stuckenholz, C., Meller, V. H., and Kuroda, M. I. (2003) Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in *Drosophila melanogaster*, *Genetics* 164, 1003-1014.
130. Fortini, E., Li, R., and Fox, A. (2013) Long Non-coding RNAs and Nuclear Body Formation and Function, In *Molecular Biology of Long Non-coding RNAs* (Khalil, A. M., and Collier, J., Eds.), Springer Science.
131. Ip, J. Y., and Nakagawa, S. (2012) Long non-coding RNAs in nuclear bodies, *Dev Growth Differ* 54, 44-54.
132. Mao, Y. S., Zhang, B., and Spector, D. L. (2011) Biogenesis and function of nuclear bodies, *Trends in genetics : TIG* 27, 295-306.
133. Shevtsov, S. P., and Dundr, M. (2011) Nucleation of nuclear bodies by RNA, *Nature cell biology* 13, 167-173.
134. Kawaguchi, T., Tanigawa, A., Naganuma, T., Ohkawa, Y., Souquere, S., Pierron, G., and Hirose, T. (2015) SWI/SNF chromatin-remodeling complexes function in noncoding RNA-dependent assembly of nuclear bodies, *Proc Natl Acad Sci U S A*.
135. Ishizuka, A., Hasegawa, Y., Ishida, K., Yanaka, K., and Nakagawa, S. (2014) Formation of nuclear bodies by the lncRNA Gomafu-associating proteins Celf3 and SF1, *Genes to cells : devoted to molecular & cellular mechanisms* 19, 704-721.
136. Mao, Y. S., Sunwoo, H., Zhang, B., and Spector, D. L. (2011) Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs, *Nature cell biology* 13, 95-101.

137. Souquere, S., Beauclair, G., Harper, F., Fox, A., and Pierron, G. (2010) Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies, *Mol Biol Cell* 21, 4020-4027.
138. Fox, A. H., Bond, C. S., and Lamond, A. I. (2005) P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner, *Mol Biol Cell* 16, 5304-5315.
139. Naganuma, T., and Hirose, T. (2013) Paraspeckle formation during the biogenesis of long non-coding RNAs, *RNA Biol* 10, 456-461.
140. Daugherty, M. D., Booth, D. S., Jayaraman, B., Cheng, Y., and Frankel, A. D. (2010) HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes, *Proc Natl Acad Sci U S A* 107, 12481-12486.
141. Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., Grishin, N. V., Frantz, D. E., Schneider, J. W., Chen, S., Li, L., Sawaya, M. R., Eisenberg, D., Tycko, R., and McKnight, S. L. (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels, *Cell* 149, 753-767.
142. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., Thomas, M., Davis, C. A., Shiekhata, R., Gingeras, T. R., Hubbard, T. J., Notredame, C., Harrow, J., and Guigo, R. (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression, *Genome Res* 22, 1775-1789.
143. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R., and Schneider, B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids, *Biophysical journal* 63, 751-759.

144. Coimbatore Narayanan, B., Westbrook, J., Ghosh, S., Petrov, A. I., Sweeney, B., Zirbel, C. L., Leontis, N. B., and Berman, H. M. (2014) The Nucleic Acid Database: new features and capabilities, *Nucleic Acids Res* 42, D114-122.
145. Waugh, A., Gendron, P., Altman, R., Brown, J. W., Case, D., Gautheret, D., Harvey, S. C., Leontis, N., Westbrook, J., Westhof, E., Zuker, M., and Major, F. (2002) RNAML: a standard syntax for exchanging RNA information, *RNA (New York, N.Y.)* 8, 707-717.
146. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002) The human genome browser at UCSC, *Genome Res* 12, 996-1006.
147. Ramakrishnan, V. (2014) The ribosome emerges from a black box, *Cell* 159, 979-984.
148. Mitton-Fry, R. M., DeGregorio, S. J., Wang, J., Steitz, T. A., and Steitz, J. A. (2010) Poly(A) tail recognition by a viral RNA element through assembly of a triple helix, *Science* 330, 1244-1247.
149. Bouchard, P., and Legault, P. (2014) Structural insights into substrate recognition by the *Neurospora Varkud* satellite ribozyme: importance of U-turns at the kissing-loop junction, *Biochemistry* 53, 258-269.
150. Kolk, M. H., van der Graaf, M., Wijmenga, S. S., Pleij, C. W., Heus, H. A., and Hilbers, C. W. (1998) NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA, *Science* 280, 434-438.
151. Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Johnson, J., de Bear, J. S., Gough, G. R., Gilham, P. T., and Moras, D. (1989) Crystallographic structure of an RNA helix: [U(UA)6A]2, *J Mol Biol* 209, 459-474.
152. Auweter, S. D., Fasan, R., Reymond, L., Underwood, J. G., Black, D. L., Pitsch, S., and Allain, F. H. (2006) Molecular basis of RNA recognition by the human alternative splicing factor Fox-1, *EMBO J* 25, 163-173.

153. Loughlin, F. E., Mansfield, R. E., Vaz, P. M., McGrath, A. P., Setiyaputra, S., Gamsjaeger, R., Chen, E. S., Morris, B. J., Guss, J. M., and Mackay, J. P. (2009) The zinc fingers of the SR-like protein ZRANB2 are single-stranded RNA-binding domains that recognize 5' splice site-like sequences, *Proc Natl Acad Sci U S A* 106, 5581-5586.
154. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., and Burley, S. K. (2000) Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome, *Cell* 100, 323-332.
155. Stefl, R., Oberstrass, F. C., Hood, J. L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R. B., and Allain, F. H. (2010) The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove, *Cell* 143, 225-237.
156. Jayaraman, B., Crosby, D. C., Homer, C., Ribeiro, I., Mavor, D., and Frankel, A. D. (2014) RNA-directed remodeling of the HIV-1 protein Rev orchestrates assembly of the Rev-Rev response element complex, *eLife* 3, e04120.

Figure legends

Figure 1. Common secondary and tertiary structures of RNA. **A.** The dot-bracket notation and 2-D map of a small region from the 5' end of B2-SINE lncRNA showing common secondary structure elements. **B.** 2-D and 3-D representations of RNA tertiary structures. A double-stranded region of the expression and nuclear retention element (ENE) sequesters the RNA poly (A) tail (magenta) from Kaposi's sarcoma-associated herpesvirus using Watson-Crick and Hoogsteen interactions to form a stabilising triple helix (PDB entry 3P22¹⁴⁸). A kissing loop between the stem-loop substrate and stem-loop catalytic domain by the *Neurospora Varkud* satellite ribozyme (PDB entry 2MI0¹⁴⁹). The three Watson-Crick base-pairs responsible for the interaction between the two loops are indicated with a dotted line. A classical pseudoknot from the turnip yellow mosaic virus showing the reciprocal interaction between two stemloops where the loop of each stemloop forms the stem of the other (PDB entry 1A60¹⁵⁰). All tertiary structures are shown as cartoon representations coloured by element. Wobble and non-Watson-Crick base-pairs are shown in the 2-D maps as unfilled and filled circles respectively.

Figure 2. Atomic resolution structures of lncRNA fragments. **A.** The NMR structure of a 32-nucleotide (nt) hairpin from the human telomerase lncRNA adopts a canonical A-form RNA structure within double helical stem region capped by a typical UUCG tetraloop (PDB entry 1Z31⁸²). The hairpin contains an asymmetric internal loop introducing a twist in the RNA structure. The loop provides a key functional feature through the generation of a docking site for the enzymatic protein component of the telomerase, hTERT. **B.** A novel AUCG tetraloop in a hairpin from the lncRNA *Xist* is an important functional element (PDB entry 2Y95⁸³). **C.** A double stranded helix from the lncRNA *Gas5* (PDB entry 4MCE⁸⁴)

contains a widened major groove width of 15 Å in comparison to a typical A-form RNA helix that has a major groove width of 3.4 Å (PDB entry 1RNA¹⁵¹). The widened major groove of Gas5 mimics DNA to compete for the binding of steroid receptors. **D.** The crystal structure of a triple helix element at the 3' end of the lncRNA MALAT1 is critical for the prevention of nuclear decay (PDB entry 4PLX⁸⁵). All tertiary structures are shown as cartoon representations coloured by element. Wobble and non-Watson-Crick base-pairs are shown in the 2-D maps as unfilled and filled circles respectively.

Figure 3. Atomic resolution structures of large RNAs. **A.** The secondary and NMR structure of the HIV-1 RNA packaging signal (PDB entry 2N1Q⁸⁶) forming a tandem three-way junction. Brackets indicate each junction and guanosine residues exposed by the junction are coloured red. **B.** The crystal structure of a bacterial RNase P (PDB entry 3Q1R⁸⁷). The metal ions are indicated and shown as grey spheres and the tRNA substrate is coloured magenta. **C.** The crystal structure of the 30S subunit of the ribosome from *Thermus thermophilus* (PDB entry 1J5E⁸⁸). RNA tertiary structures are shown as cartoon representations coloured by element, and the surface representations of protein structures shown in light blue. Wobble and non-Watson-Crick base-pairs are shown in the 2-D map as unfilled and filled circles respectively.

Figure 4. Structures of RNA binding motifs bound to their cognate RNA. Nucleic acids are coloured by element and shown in the cartoon representation. Both the surface and cartoon representation of the proteins are shown (light blue) with the nucleic acid interacting residues coloured purple and shown as sticks. **A.** The RRM of Human Fox-1 binds to UGCAUGU in a canonical way through three conserved aromatic residues on the four-stranded B-sheet surface. (PDB entry 2ERR¹⁵²) **B.** The dsRBM of ADAR2

recognises double stranded RNA in the form of a stem-loop (PDB entry 2L3C¹⁵⁵). **C.** The KH domain of the Nova protein recognises an RNA stem loop (PDB entry 1EC6¹⁵⁴). **D.** The zinc finger of ZRANB2 binds to an AGGUAA motif through side-chain hydrogen bonds and the formation of a guanine-tryptophan-guanine “ladder” (PDB entry 3G9Y¹⁵³)

Figure 5. Structures of lncRNA-protein interactions. Nucleic acids are coloured by element and shown in the cartoon representation. Both the surface and cartoon representation of the proteins are shown (light blue) with the nucleic acid interacting residues coloured purple and shown as sticks. **A.** The La-domain and RRM1 of LARP interaction with a U-rich hairpin from 7SK lncRNA (PDB entry 4WKR¹⁰³). Amino acids from the RRM that are involved in RNA interactions are shown as sticks. **B.** The ARMs in the viral Rev protein homo-dimer bind to the Rev-response element (RRE) lncRNA (PDB entry 4PMI¹⁵⁶).

Figure 6. Model for the assembly of a paraspeckle. NEAT1_v1 and the 5' and 3' ends of NEAT1_v2 localise to the periphery of the paraspeckle with the central sequence localising to the inner core. DBHS proteins oligomerise along NEAT1 through protein-protein and protein-RNA interactions while the other paraspeckle proteins may be bridging protein-protein or protein-RNA interactions to form a paraspeckle.