

The Role of Long Noncoding RNAs in Gene Expression Regulation

Zhijin Li, Weiling Zhao, Maode Wang and Xiaobo Zhou

Abstract

Accumulating evidence highlights that noncoding RNAs, especially the long noncoding RNAs (lncRNAs), are critical regulators of gene expression in development, differentiation, and human diseases, such as cancers and heart diseases. The regulatory mechanisms of lncRNAs have been categorized into four major archetypes: signals, decoys, scaffolds, and guides. Increasing evidence points that lncRNAs are able to regulate almost every cellular process by their binding to proteins, mRNAs, miRNA, and/or DNAs. In this review, we present the recent research advances about the regulatory mechanisms of lncRNA in gene expression at various levels, including pretranscription, transcription regulation, and posttranscription regulation. We also introduce the interaction between lncRNA and DNA, RNA and protein, and the bioinformatics applications on lncRNA research.

Keywords: long noncoding RNAs, gene regulation, lncRNA binding, bioinformatics

1. Introduction

It was estimated that there are approximately 20,500 protein coding genes that account for 2% of the genome [1], and another 98% of the genome were considered as “DNA junks” previously due to their disability in coding proteins. The application of high-throughput next generation sequencing (NGS) technology has changed our view of the genome, about 90% of the human genome can be transcribed into RNA transcripts [2–5]. Except the small portion of transcripts encoding for proteins, the majority of RNA transcripts in the type have been grouped into noncoding RNAs (ncRNAs), including transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small ncRNAs such as micro RNAs (miRNAs), small interfering RNAs (siRNAs), small nuclear RNAs (snRNA), circular RNA, as well as long ncRNAs (lncRNAs). The rRNA and tRNA are two basic and most abundant RNAs that play important roles in mRNA translation. The miRNAs are short single strand RNAs of 20–22 base with the role of promoting mRNA degradation. siRNA is a class of double strand RNA with a length of 20–25 base pairs, which interferes target gene expression by degrading mRNA or preventing translation. The snRNA, with an average length of 150 base, is a class of small RNA located in nucleus speckles. The primary function of snRNA is in the processing of pre-mRNA splicing. Circular RNAs is a very special class of ncRNA. The 3' and 5' ends of the circular RNA join together to form a covalently closed continuous loop. The

function of the circular RNA is not clear. LncRNAs are defined as the noncoding linear transcripts longer than 200 nucleotides, which have different features with other ncRNA listed above. In general, they share common characteristics with mRNAs. The majority of LncRNAs are usually transcribed by RNA polymerase II, capped at 5' end, and spliced; most of them are also polyadenylated at the 3' end and have promoter regions. Compared to the protein coding gene mRNA, lncRNAs lack of open reading frame (ORF), contain fewer exons (~2.8 exons in lncRNAs compared to 11 exons for protein coding gene), are expressed in low abundance and more tissue-specific [6]. Some of them have no polyadenylation tails. The lncRNAs account for the major class of the ncRNAs in the genome. There are ~30,000 high-confidence transcripts of lncRNAs in human according to GENCODE reference genome annotation [7], and more and more new lncRNAs are coming into the light. The database LNCipedia has collected 127,802 transcripts from 56,946 long noncoding genes in human [8, 9]. Many of the ncRNAs have been confirmed as playing crucial regulatory roles in diverse biological processes and tumorigenesis. Increasing evidence indicates that lncRNAs play important roles in various cellular processes, such as DNA repair [10], proliferation [11], epithelial-mesenchymal transition (EMT) [12] by regulating various aspects of the related gene expression. LncRNAs have been associated with various diseases [13–16] and identified as potential biomarkers in some diseases, such as cancers, cardiovascular diseases, nervous system diseases, etc. [17–19]. LncRNAs could regulate gene expression by serving as molecular signals, guides, decoys and/or scaffolds [20]. In this review, we will present the recent research advances about the regulatory mechanisms of lncRNAs in gene expression at various levels, including pretranscription, transcription regulation, and post transcription regulation. We will also discuss the interaction between lncRNA and DNA, RNA and protein, as well as the applications of bioinformatics in lncRNA-related research.

2. The role of lncRNA in pretranscription regulation

Gene expression is regulated at many levels, such as epigenetic, transcriptional, post-transcriptional, translational, and post-translational. In order to be transcribed, changes in the chromatin structure of a gene must take place to make the chromatin open to polymerases and transcriptional factors (TFs). Modification on chromatin DNA and histones affects gene accessibility and is associated with distinct transcription states. For example, H3 hyperacetylation or methylation at lysine 4 often makes the gene easily accessible, thereby actively transcribed. In contrast, histone methylation at lysine 9 results the assembly of compact or closed chromatin around the DNA, leading to transcription silence. A number of epigenetic control factors have been identified to modify histones. Some of them can facilitate transcriptional activation, such as p300/CBP, Esa1, and TAF1, and others participate in transcriptional silencing, such as EZH2 and Ubc9. However, the majority of epigenetic factors, such as DNA methyltransferases/demethylases and histone modification enzymes may not efficiently recognize specific DNA sequences. Emerging studies show that lncRNAs can act as signals, guides or scaffolds at chromatin level to regulate gene expression [21–23]. LncRNAs have been reported to participate in the methylation processes. For example, Wang et al. found that lncRNA Dum (developmental pluripotency-associated 2 (Dppa2) Upstream binding Muscle lncRNA) regulated DPPa2 expression by affecting DNA methylation [24]. Dnmt proteins are known as DNA methyltransferases. Dum promoted DNA methylation of Dppa2 promoter by recruiting Dnmt1, Dnmt3a, and Dnmt3b to its promoter site,

thereby silencing Dppa2 expression in cis and stimulating myogenic differentiation. Studies show that lncRNA HOTAIR (Hox transcript antisense intergenic RNA), transcribed from chromosome 12, can coordinate histone modification by binding to histone modifiers [25, 26]. Rinn et al. found that knock-down of HOTAIR led to transcriptional activation of HOXD locus genes present in the chromosome 2 and HOTAIR binding is required to guide polycomb repressive complex 2 (PRC2) to the HOXD locus [26]. PRC2 is epigenetic factor and can catalyze methylation on lysine 27 of histone H3 (H3K27). PRC2 and LSD1 (lysine-specific demethylase 1) bind to the 5' and 3' domains of HOTAIR, respectively. The HOTAIR-PRC2-LSD1 complex then targets the HOXD locus on the chromosome 2, silencing the genes involved in the suppression of metastasis. LncRNA HOTTIP (HOXA transcript at the distal tip) binds to and targets WDR5-MLL complexes to the 5' HOXA locus, mediating the transcriptional activation of HOXA via driving H3K4 methylation [27]. LncRNA Evf-2 interacts with methyl-CpG binding protein 2 (MECP2), inhibiting the methylation at DLX5/6 enhancer [28]. Similarly, LncRNA GClnc1 (gastric cancer-associated lncRNA 1) promotes gastric carcinogenesis by acting as a modular scaffold of WDR5 and KAT2A complexes to specify the histone modification pattern on superoxide dismutase 2 [29]. Zhao et al. showed that lncRNA PAPAS (promoter and pre-rRNA antisense) guided CHD4/NuRD (nucleosome remodeling and deacetylation) to the rDNA promoter by forming a DNA-RNA triplex structure at the enhancer region of rDNA [30]. Other studies have also explored the role of lncRNAs in epigenetic regulation of transcription [31, 32].

3. The role of lncRNA in transcription regulation

Transcription begins with the binding of RNA polymerase II to the promoter region of a gene with the support of general transcription factors (GTFs). Other transcription factors (TFs) bind to the enhancer region accelerate transcription. The transcription is terminated when the polymerase meets to the terminator. LncRNAs can regulate gene expression by direct binding with TFs or PNA Pol II, or interfering the binding polymerase with promoter. For example, lncMyoD is an lncRNA activated by myogenic differentiation (MyoD) during myogenesis. lncMyoD can directly binds to IGF2-mRNA-binding protein 2 (IMP2) and negatively regulates IMP2-mediated translation of proliferation genes such as N-Ras and c-Myc, which create a permissive state for differentiation [33]. lncRNA Gas5 (growth arrest-specific 5) attenuates some of GR positive related gene expression by binding to glucose receptors (GR) [24]. lncHIFCAR (long noncoding HIF-1 α co-activating RNA) level is upregulated in oral carcinoma. Shih et al. found that lncHIFCAR acted as a HIF-1 α co-activator driving oral cancer progression [34]. lncHIFCAR formed a complex with HIF-1 α via directly binding and facilitates the recruitment of HIF-1 α and p300 cofactor to the target promoters. In addition, lncRNAs can guide RNA polymerase II to bind to the promoter of specific genes. Miao et al. found that lncRNA LEENE guided and facilitated the recruitment of RNA Pol II to the eNOS promoter to upregulate eNOS RNA transcription [35]. In addition, recent studies indicated that lncRNA gene promoter could compete for enhancer with protein coding gene promoter. Enhancer is the cis-acting DNA sequence that can enhance the transcription of an associated gene, when bound by specific transcription factors. Cho et al. found that the lncRNA PVT1 promoter has a tumor-suppressor function that is independent of PVT1 gene [36]. The promoter of lncRNA PVT1 competes with the Myc promoter for engagement to four intra-genic enhancers, thereby inhibiting the expression of Myc gene.

4. lncRNA on posttranscription regulation

After transcription, the pre-mRNAs are regulated by various RNA-binding proteins (RBPs). The pre-mRNAs are capped, polyadenylated, spliced, edited and transferred from nucleus to cytoplasm. The stability of mRNA is also an important aspect for translation. There are evidence showing the role of lncRNAs in mRNA splicing, editing, transporting, mRNA stability, and mRNA translation. In addition, lncRNAs can regulate mRNA expression indirectly by acting as competing endogenous RNAs.

4.1 lncRNA and alternative splicing

Alternative splicing is a regulatory process during gene expression that enables a single gene coding for multiple proteins. Recent studies indicate that lncRNA can regulate alternative splicing through two main mechanisms. LncRNAs can interact with specific splicing factors or form RNA-RNA duplexes with pre-mRNAs. SR (splicing factor) proteins, such as SRSF1, are a conserved family of proteins involved in RNA splicing regulation in a concentration- and phosphorylation-dependent manner. MALAT1 is a highly conserved lncRNA among mammals and predominantly localizes to nuclear speckles. Tripathi et al. [37] showed that MALAT1 acts as a molecular sponge to titrate the cellular pool of SR splicing factors, affecting the distribution of splicing factors in nuclear speckles where the alternative splicing occurs and ultimately controlling alternative splicing. The 5' region of MALAT1 can also bind to the serine/arginine domain of SRSF1 and regulates its cellular levels of the phosphorylated forms. SORL1 (Sorting Protein-Related Receptor Containing LDLR Class A Repeats), a sorting receptor for amyloid precursor protein (APP), can interact with amyloid APP and affect its transport and process in brain. Downregulation of SORL1 expression increases APP secretion and subsequently A β formation. LncRNA 51A, an antisense mapping to the intron 1 of the SORL1 gene, masked canonical splicing sites by pairing with SORL1 pre-mRNA, driving a splicing shift of SORL1 from the canonical long protein variant A to an alternatively spliced protein form B [38].

4.2 The regulation of lncRNA on mRNA stability

Different mRNAs have different lifespans, even in a single cell. The greater the stability of an mRNA molecule is, the more proteins may be produced by the mRNA molecule. The steady-state level of a mRNA is determined by the rate of synthesis and degradation. Modulation of mRNA degradation is an important control point in gene expression to regulate protein synthesis in response to physiological needs and environmental signals. Studies have shown that the role of lncRNA in the regulation of mRNA stability. For example, Cao et al. found that lncRNA LAST (LncRNA-Assisted Stabilization of Transcripts) acted as a mRNA stabilizer by cooperating with CNBP (CCHC-type zinc finger nucleic acid binding protein) to promote Cyclin D1 mRNA stability [39]. Antisense lncRNAs are transcripts emerging from the opposite strand of a coding-RNA region. β -site amyloid precursor protein cleaving enzyme (BACE1) is involved in the production of the amyloid- β (A β) peptides that form plaques in the brains of individuals with AD. BACE1-AS expression is elevated in the brain of Alzheimer mouse model. Faghihi et al. found that BACE1-AS increased the stability of BACE1 mRNA and upregulated the BACE1 protein by forming RNA duplex with BACE1 mRNA, which masked the binding site for miR-485-5p and thereby increase the BACE1 mRNA stability [40, 41]. Matsui et al. found that iNOS antisense transcript stabilized iNOS mRNA through interaction

with AU-rich element-binding HuR protein [42]. lncRNAs can also reduce the stability of mRNA by making the transcript prone to degradation. aHIF is a natural antisense transcript of hypoxia-inducible factor 1alpha (HIF-1 α). Rossignol et al. reported that aHIF could expose AU-riches elements present in the HIF-1 α mRNA 3' UTR, thus increasing the degradation speed of HIF-1 α mRNA [43].

4.3 lncRNA and protein stability

lncRNA can also directly interact with proteins and regulate their stability by retarding protein ubiquitination and degeneration. Androgen receptor (AR) is a critical risk factor in castration-resistant prostate cancer. Zhang et al. shown that lncRNA HOTAIR bound to AR protein to block AR interaction with E3 ubiquitin ligase MDM2, thereby preventing AR ubiquitination and AR protein degradation [44]. Liu et al. identified lncRNA MT1JP as a critical factor in restraining cell transformation by modulating p53 translation through binding and stabilizing the RNA binding protein TIAR [45]. LincRNA-p21 is a hypoxia-responsive lncRNA. LncRNA-p21 can bind to HIF-1 at its VHL binding region and attenuate VHL-mediated HIF-1 α ubiquitination, leading to HIF-1 α accumulation [46].

4.4 lncRNA regulates protein translation

Transcription and translation are two main stages in gene expression. In translation, the ribosomal preinitiation complex, consisting of eukaryotic initiation factors (eIFs) and ribosomes, is positioned the start codon of the target RNA. With the help of tRNA, the mRNA is decoded to produce peptide chains. It has been reported that lncRNAs participate in protein translation by interaction with rRNA, ribosome or eIFs. Li et al. reported that a nucleolar-specific lncRNA, LoNA reduced rRNA production and ribosome biosynthesis [47]. The 5' region of LoNA bound to and sequestered nucleolin to suppress rRNA transcription and the 3' end recruited and diminishes fibrillarin activity to reduce rRNA methylation [47]. Tran et al. found that lncRNA AS-RBM15, the antisense of RNA binding motif protein 15, overlapped with the 5' UTR of RBM15. As-RBM15 enhanced RBM15 protein translation via incorporation into the RBM15 mRNA-containing polyribosome in a CAP-dependent manner [48]. The antisense of UCHL1 binds the 5' UTR of UCHL1 mRNA to active polysomes for UCHL1 translation [49]. lncRNA Gas5 (growth arrest specific 5) has been reported to interact and cooperate with eIF4E to regulate c-MYC translation [50].

4.5 lncRNAs act as microRNA sponges

Micro RNAs (miRNAs), short ncRNAs, can bind to the complementary regions of mRNAs to post-translationally regulate the expression of target genes. The subcellular localization of lncRNAs is associated with their functions. Studies have shown that lncRNAs can act as miRNA sponges to inhibit the binding of miRNAs to their mRNA targets, thereby stabilizing the target mRNAs and regulating the corresponding protein expression [51]. Hu et al. found that p53-responding lncRNA GUARDIN is important for maintaining genomic integrity under exogenous genotoxic stress. GUARDIN contains eight regions complementary to the 'seed' region of miR-23a. Telomeric repeat-binding factor-2 (TRF2), a critical component of the shelterin complex, is one of the targets of miR-23a. GUARDIN maintained the expression of TRF2 by sequestering miR-23a. Zhou et al. observed that lncRNA H19 acted as a sponge for the miRNAs miR-200b/c and Let-7b to promote an epithelial or mesenchymal switch in tumor cells [52]. In the epithelial-like tumor cells, H19 inhibited the migration-related protein ARF by sequestering miR-200b/c. In

contrast, H19 activated ARF by sequestering Let-7b in the mesenchymal-like tumor cells. H19 stimulated HMGA2-mediated EMT by sequestering Let-7 to promote pancreatic ductal adenocarcinoma cell invasion and migration [52, 53].

4.6 lncRNA sequesters proteins

lncRNA can act as protein molecular decoy by binding and sequestering proteins, thereby inhibiting their functions [54]. Lee et al. identified a lncRNA named noncoding RNA activated by DNA damage (NORAD) [55] and found that NORAD played an important role in maintaining genomic stability by binding and sequestering PUMILIO proteins. In the absence of NORAD, PUMILIO proteins drove chromosomal instability by acting as negative regulators of gene expression involved in DNA damage repair and mitosis. S-adenosyl-L-homocysteine hydrolase (SAHH) is the only mammalian enzyme capable of catalyzing the hydrolysis of S-adenosyl-L-homocysteine (SAH), which is an inhibitor of S-adenosylmethionine (SAM)-dependent methyltransferases. Zhou et al. reported that lncRNA H19 bound to SAHH and suppressed its enzymatic activity, leading to an increased accumulation of SAH [56]. lncRNA Gas5 is induced by stress like starvation. Glucocorticoid receptor (GR) plays important role in regulating genes associated with metabolism, development and immune response. Kino et al. showed that lncRNA Gas5 bound to the DNA-binding domain of glucocorticoid receptor (GR), thereby, inhibiting the ability of GR to regulate the related gene expression [57].

5. The lncRNA interaction with other molecules

As shown above, lncRNAs regulate gene expression at pretranscriptional, transcriptional and posttranscriptional levels by interacting with DNAs, mRNAs, miRNAs, and proteins. Here we will give a brief introduction of the interactions of lncRNA with other molecules.

5.1 lncRNA-DNA

lncRNAs regulate gene expression through various complicated mechanisms, one of the them is binding to the DNA or forming RNA-dsDNA triplexes by targeting specific DNA sequences. The lncRNA as a third strand can inserted into the major groove of the DNA duplex by Hoogsteen hydrogen binding [58]. The Hoogsteen hydrogen binding by a third strand is usually weaker than the Watson-Crick hydrogen bonding. Studies from Roberts et al. indicate RNA may form more stable triplexes than their DNA counterparts [59]. Various methods have been used to investigate the interaction of lncRNA and DNA. The formation of lncRNA DHFR and dsDNA triplex was demonstrated using electrophoresis mobility shift assay [60] and the binding of lncRNA Fendrr to dsDNA was determined by *in vitro* pull-down experiments [61, 62]. High throughput methods are also applied to investigate the genomic binding sites of lncRNAs, such as capture hybridization analysis of RNA targets sequencing (CHART-seq) [63]. Although the mechanism of RNA-dsDNA triplexes formation is still not well illustrated, it is clear that the lncRNA-DNA interaction offers a potent mechanism for gene regulation. lncRNAs can bind DNAs acting as scaffolds for introducing proteins into the gene loci. When the proteins introduced by lncRNAs are methylation-related enzymes, these enzymes can induce promotor CpG island methylation or demethylation (**Figure 1A**). When the proteins imported by lncRNAs are histone modifier enzymes (**Figure 1A**), histone modifications can result gene expression, transcriptional silencing, or DNA

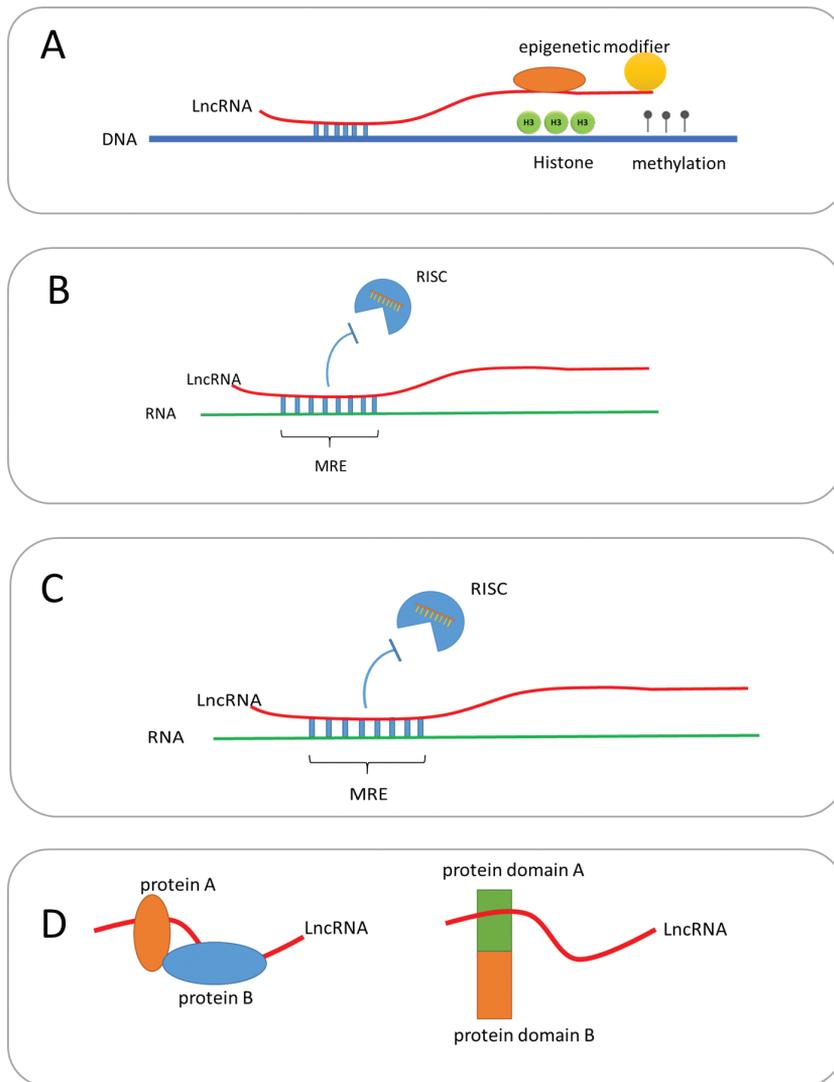


Figure 1.
 The mechanisms of lncRNA regulating gene expression by interacting with other molecule. (A) lncRNA acting as scaffold binds to chromatin and epigenetic modifier, guide the epigenetic modifier to gene promotor. (B) lncRNA acting as miRNA sponges attenuates the miRNA' effect on downregulating mRNA expression. (C) lncRNA can bind to RNA' special region such as miRNA response element, thus masking the region. (D) lncRNA can act as scaffold for two or more protein, these proteins will act coordinately or act as a complex (left). lncRNA can bind to protein, the bonded protein domain may change its function (right).

repair and genomic imprinting. LncRNA can regulate either neighboring (cis) or distal (trans) protein coding genes. LncRNAs derived from one chromatin can bind to another chromatin, such as LncRNA HOTAIR, which is transcribed from the HOXC locus on the chromatin 12 and represses transcription in trans across 40 kb of the HOXD locus [26].

5.2 lncRNA-miRNA

The competitive endogenous RNA (ceRNA) hypothesis proposes that RNA transcripts, including both coding and noncoding RNAs, compete for post-translational regulation with shared miRNA binding sites (i.e. miRNA response elements) [64]. Cytoplasmic lncRNA can act as a molecular sponge of miRNA to regulate the rate

of translation and degradation of mRNA (**Figure 1B**). RNA interference is a powerful mechanism of gene silencing and mediated by RNA-induced silencing complexes (RISC). RISC is ribonucleoprotein complex, which incorporates on strand of a single-stranded RNA fragment, such as miRNA [65]. Once a mRNA is transcribed and exported to the cytoplasm, it can be targeted by the RISC, resulting in the accelerated degradation of the mRNA, or blocked translation. lncRNAs with miRNA response element can compete for miRNA targeting and binding to RISC, leading to sequestration of the microRNA-RISC and preventing RISC-mediated degradation of mRNAs and increasing mRNA expression. Using AGO-CLIP-seq high throughput method, thousands of miRNA-target interaction are estimated [66, 67]. We have shown studies that lncRNAs act as miRNA spongers above. Except lncRNAs, mRNA, circular RNAs and pseudogenes can also act as the ceRNA. The most widely used miRNA target prediction rule is the 6-nucleotide interactions between 5' ends of the microRNA which is called "seed region" [68]. Note that one microRNA can bind to hundreds of RNAs and one RNA molecule may also bind to diverse microRNA with different affinity, which is difficult to quantify.

5.3 lncRNA-mRNA

As described above, lncRNAs are able to regulate pre-mRNA splicing and mRNA stability. Some antisense lncRNAs can bind to the homologous mRNA at the splice site, thereby masking the splice site and blocking spliceosome assembly [69]. lncRNAs can also mask miRNA target sites of an mRNA by binding to the mRNA at the miRNA response elements (**Figure 1C**), resulting in elevated stability and expression of the mRNA [40]. lncRNAs could bind to protein and mRNA simultaneously. Gong et al. reported that lncRNAs bound to the Alu region at the 3' UTR of mRNA and the looped lncRNAs bound to STAU1 protein (Double-Stranded RNA-Binding Protein Staufen Homolog 1), leading to the activation of the Staufen-mediated decay pathway [70]. Recently, computational and experimental methods have been developed to determine RNA-RNA interaction. For example, Sharma et al. [71] developed a high throughput sequencing method, named "LIGATION of interacting RNA followed by high-throughput sequencing" (LIGR-seq), to reveal a remarkable landscape of RNA-RNA interactions involving the major classes of ncRNA and mRNA. Nguyen developed MARIO (Mapping RNA interactome in vivo) [72] approach to map tens of thousands of endogenous RNA-RNA interactions.

5.4 lncRNA-protein

There is no doubt that RNA-protein interactions play a crucial role in fundamental cellular processes. lncRNAs can function as protein decoys recruiting or sequester proteins, or act as scaffolds linking different proteins, which may act coordinately or act as a complex (**Figure 1D** left). Several models have been established to understand how an lncRNA regulates gene expression by protein binding. lncRNAs are able to recruit chromatin modifier to achieve chromatin modification [73]. lncRNA can bind and stabilize a protein by masking its ubiquitination site, inducing the accumulation of the target protein [44]. lncRNA can bind to protein, the bonded domain may change its function (**Figure 1D** right) [74]. lncRNAs can bind to TFs to mask their DNA-binding sequences or stabilize TFs [33]. lncRNAs are also able to bind to functional enzymes to inhibit their activities [33], resulting elevated levels of substrate proteins. The high throughput experimental methods, such as CLIP-seq and RIP-seq, have been used to study RNA-protein interactions [74].

5.5 The lncRNA location and its role

Most of lncRNAs are located exclusively in the nucleus, but some of them are located in the cytoplasm or in both nucleus and cytoplasm. Increasing evidence reveals that RNA subcellular location is a very important feature in understanding lncRNA functions. The nuclear function of lncRNAs are apt to regulate gene expression *in cis* or *in trans*. In the nucleus, a lncRNA can accumulate at its transcription site and recruit transcription factors or chromatin modifiers. lncRNAs in the nucleus also can regulate gene expression *in trans* by binding to a remote genome sites. In addition, the effect of lncRNA on alternative splicing usually occurs in the nucleus. The lncRNAs exported to cytoplasm intend to interfere translation or sequester proteins/miRNA. Currently, the subcellular locations of known lncRNAs are mainly determined by biological experiments. Most recently, Zhang et al. built a web-accessible database (RNALocate) to provide a high-quality RNA subcellular localization resource and facilitate future researches on RNA functions or structures based on the experimental data [75]. Su et al. developed a sequence-based bioinformatics tool (iLoc-lncRNA) to annotate and predict the subcellular locations of lncRNAs by binomial distribution of the 8-tuple nucleotide signatures into the general pseudo K-tuple nucleotide composition [76].

6. The application of bioinformatics in lncRNA studies

The research on lncRNAs research has increased rapidly in recent years. The bioinformaticians have developed various approaches for identification of lncRNA from the genome and transcriptome data, prediction of lncRNA structures, investigation of lncRNA functions and construction of lncRNA-associated regulatory networks. lncRNA databases have been established from some giant studies or by integrating different studies. The association analyses based methods, such as gene set enrichment analysis (GSEA) and weighted gene co-expression network analysis (WGCNA) make it easier for biologists to research the lncRNA. As described above, lncRNA can bind with single or multiple different types of molecules to regulate gene expression in the different level directly or indirectly. Therefore, the databases and tools are important for investigating lncRNA bindings with other molecules. The binding pattern is usually determined by the molecular sequences (nucleotide or amino acid) and the structures. Accumulated high throughput experimental methods have been developed for the identification of the lncRNA-DNA, lncRNA-RNA, and lncRNA-protein bindings. Based on these existing data, bioinformatics approaches, such as deep learning, can be used to identify the binding rules, which can be used to predict the potential molecule that could bind to given lncRNA based on the DNA/RNA sequences and/or structure. Note that the bioinformatics tools are useful, but one also should keep in mind that the prediction and associated analysis is a good screening process. Validation of the predicted results with experimental methods are required. Here, we briefly reviewed the bioinformatics tools for investigating the interaction of lncRNAs with other molecules.

The formation of RNA-dsDNA triplex is sequence-specific. Some motifs are easier to form triplex [77]. Therefore, the bioinformatics method can be used to predict the binding sites. Triplexator [78] and Longtarget [79] are two bioinformatics tools that have been used to predict lncRNA-dsDNA bindings. There are a number of bioinformatics tools to predict RNA-RNA interaction, such as RNAup [80], intaRNA [81], RNAPlex [82], etc. RNAup calculates the thermodynamics of RNA-RNA interactions. IntaRNA is a fast approach for predicting RNA-RNA

interaction incorporating accessibility and seeding of interaction sites. Recently, Gawronski et al. proposed a pipeline (MechRNA) to predict lncRNA interactions with target RNA using IntaRNA2, as well as the protein bindings with other tools [83]. The performance of various RNA-RNA interaction tools has been summarized previously [84].

Based on ceRNA hypothesis and seed region match rule, many databases and bioinformatics tools have been developed. Targetscan [68], starBase [85], miRTarBase [86], Pictar [87], miRanda [88] and other bioinformatics tools have been used for the prediction of miRNA targets. Our group developed a novel network-based method previously to integrate the correlations between lncRNA, protein coding genes and miRNA [71]. We also designed a model to assess the combined impact of mRNAs, lncRNAs and miRNAs on cellular signaling transduction networks recently [71]. Tong et al. developed tools and a server to enable validating predicted lncRNA-miRNA-mRNA regulations from TCGA RNA-seq data and identifying miRNA-associated cancer signaling pathways and related lncRNA sponges [89, 90]. However, because of the complicate reticular networks, it is difficult to evaluate the effect of individual ceRNAs based solely on the bioinformatics analysis [91, 92].

Bioinformatics sciences have developed various tools for predicting ncRNA-protein interaction. Most of the existing methods are based on the sequences of either proteins or RNAs. Some of them investigate the associations between proteins and RNAs, such as RPI-Pred [93], RPIseq [94], and lncPro [95]. Some tools are used to predict binding sites of RNAs or proteins, such as BindN [96], RNABindR [97], RNAProB [98], PPRint [99], PRINTR [100], PRBR [101], SRCPred [102], RNABindRPlus [103] and RBRIdent [104]. The CatRAPID [105] method is specially designed to determine residue-nucleotide interactions. We developed a three-step prediction model called RPI-Bind for the identification of RNA-protein binding regions based on both sequences and structures of proteins and RNAs. These three steps include: (1) prediction of RNA binding regions on proteins, (2) prediction of protein binding regions on RNA, and (3) simultaneous prediction of interaction regions on RNA and proteins [106]. Suresh et al. developed a RNA-protein interaction predictor (RPI-Pred) using a new support-vector machine-based method to predict protein-RNA interaction pairs, based on both the sequences and structures [93]. The usage of machine learning algorithms have improved the prediction accuracies [93, 106].

7. Conclusions

Advances in high throughput technologies results in a rapid identification of a large amount of lncRNAs and expands our understanding of the mechanisms how lncRNAs function. Accumulating evidence indicates the complicated roles of lncRNAs. In order to gain deep insight into the interaction of lncRNA with other molecules, a number of bioinformatics tools have been developed. It is undeniable that our understanding of gene regulation by lncRNAs is still in the early stages. Given the growing numbers of lncRNA studies, we can anticipate that the detailed roles of novel lncRNAs will be addressed in the near future.

Acknowledgements

This work was partially supported by National Institutes of Health grants [NIH R01GM123037, U01AR069395-01A1, and U01CA166886] to X. Zhou.

Author details

Zhijin Li^{1,2}, Weiling Zhao¹, Maode Wang² and Xiaobo Zhou^{1,3,4*}

1 Center for Computational Systems Medicine, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA

2 Department of Neurosurgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China

3 McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA

4 School of Dentistry, The University of Texas Health Science Center at Houston, Houston, TX, USA

*Address all correspondence to: xiaobo.zhou@uth.tmc.edu

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Green ED, Watson JD, Collins FS. Human genome project: Twenty-five years of big biology. *Nature*. 2015;**526**(7571):29-31
- [2] Hangauer MJ, Vaughn IW, McManus MT. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genetics*. 2013;**9**(6):e1003569
- [3] Mortazavi A et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*. 2008;**5**(7):621-628
- [4] Djebali S et al. Landscape of transcription in human cells. *Nature*. 2012;**489**(7414):101-108
- [5] Birney E et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;**447**(7146):799-816
- [6] Derrien T et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*. 2012;**22**(9):1775-1789
- [7] Harrow J et al. GENCODE: The reference human genome annotation for the ENCODE project. *Genome Research*. 2012;**22**(9):1760-1774
- [8] Volders PJ et al. An update on LNCipedia: A database for annotated human lncRNA sequences. *Nucleic Acids Research*. 2015;**43**(Database issue):D174-D180
- [9] Volders PJ et al. LNCipedia: A database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Research*. 2013;**41**(Database issue):D246-D251
- [10] Dianatpour A, Ghafouri-Fard S. The role of long non coding RNAs in the repair of DNA double strand breaks. *International Journal of Molecular and Cellular Medicine*. 2017;**6**(1):1-12
- [11] Chen J, Liu S, Hu X. Long non-coding RNAs: Crucial regulators of gastrointestinal cancer cell proliferation. *Cell Death Discovery*. 2018;**4**:50
- [12] Wang L et al. Missing links in epithelial-mesenchymal transition: Long non-coding RNAs enter the arena. *Cellular Physiology and Biochemistry*. 2017;**44**(4):1665-1680
- [13] Xu T et al. Pathological bases and clinical impact of long noncoding RNAs in prostate cancer: A new budding star. *Molecular Cancer*. 2018;**17**(1):103
- [14] Xie H et al. Long non-coding RNA CRNDE in cancer prognosis: Review and meta-analysis. *Clinica Chimica Acta*. 2018;**485**:262-271
- [15] Huang H et al. Long noncoding RNAs and their epigenetic function in hematological diseases. 2018. DOI: 10.1002/hon.2534
- [16] Archer K et al. Long non-coding RNAs as master regulators in cardiovascular diseases. *International Journal of Molecular Sciences*. 2015;**16**(10):23651-23667
- [17] Beck D et al. A four-gene LincRNA expression signature predicts risk in multiple cohorts of acute myeloid leukemia patients. 2018;**32**(2):263-272
- [18] Mou Y et al. Identification of long noncoding RNAs biomarkers in patients with hepatitis B virus-associated hepatocellular carcinoma. *Cancer Biomarkers*, 2018;**23**(1):95-106
- [19] Jiang X, Lei R, Ning Q. Circulating long noncoding RNAs

as novel biomarkers of human diseases. *Biomarkers in Medicine*. 2016;**10**(7):757-769

[20] Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Molecular Cell*. 2011;**43**(6):904-914

[21] Brockdorff N. Noncoding RNA and Polycomb recruitment. *RNA*. 2013;**19**(4):429-442

[22] Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural & Molecular Biology*. 2013;**20**(3):300-307

[23] Ulitsky I, Bartel DP. lincRNAs: Genomics, evolution, and mechanisms. *Cell*. 2013;**154**(1):26-46

[24] Wang L et al. LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. *Cell Research*. 2015;**25**(3):335-350

[25] Tsai MC et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. 2010;**329**(5992):689-693

[26] Rinn JL et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007;**129**(7):1311-1323

[27] Wang KC et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011;**472**(7341):120-124

[28] Berghoff EG et al. Evi2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development*. 2013;**140**(21):4407-4416

[29] Sun TT et al. LncRNA GClncl promotes gastric carcinogenesis and

may act as a modular scaffold of WDR5 and KAT2A complexes to specify the histone modification pattern. *Cancer Discovery*. 2016;**6**(7):784-801

[30] Zhao Z et al. lncRNA PAPAS tethered to the rDNA enhancer recruits hypophosphorylated CHD4/NuRD to repress rRNA synthesis at elevated temperatures. *Genes & Development*. 2018;**32**(11-12):836-848

[31] Zhou Y et al. Activation of p53 by MEG3 non-coding RNA. *The Journal of Biological Chemistry*. 2007;**282**(34):24731-24742

[32] Cabianca DS et al. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell*. 2012;**149**(4):819-831

[33] Gong C et al. A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Developmental Cell*. 2015;**34**(2):181-191

[34] Shih JW et al. Long noncoding RNA LncHIFCAR/MIR31HG is a HIF-1alpha co-activator driving oral cancer progression. *Nature Communications*. 2017;**8**:15874

[35] Miao Y et al. Enhancer-associated long non-coding RNA LEENE regulates endothelial nitric oxide synthase and endothelial function. 2018;**9**(1):292

[36] Cho SW et al. Promoter of lncRNA gene PVT1 is a tumor-suppressor DNA boundary element. *Cell*. 2018;**173**(6):1398-1412.e22

[37] Tripathi V et al. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genetics*. 2013;**9**(3):e1003368

- [38] Ciarlo E et al. An intronic ncRNA-dependent regulation of SORL1 expression affecting Abeta formation is upregulated in post-mortem Alzheimer's disease brain samples. *Disease Models & Mechanisms*. 2013;**6**(2):424-433
- [39] Cao L et al. LAST, a c-Myc-inducible long noncoding RNA, cooperates with CNBP to promote CCND1 mRNA stability in human cells. *eLife*. 2017;**6**:e30433
- [40] Faghihi MA et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biology*. 2010;**11**(5):R56
- [41] Faghihi MA et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nature Medicine*. 2008;**14**(7):723-730
- [42] Matsui K et al. Natural antisense transcript stabilizes inducible nitric oxide synthase messenger RNA in rat hepatocytes. *Hepatology*. 2008;**47**(2):686-697
- [43] Rossignol F, Vache C, Clottes E. Natural antisense transcripts of hypoxia-inducible factor 1alpha are detected in different normal and tumour human tissues. *Gene*. 2002;**299**(1-2):135-140
- [44] Zhang A et al. LncRNA HOTAIR enhances the androgen-receptor-mediated transcriptional program and drives castration-resistant prostate cancer. *Cell Reports*. 2015;**13**(1):209-221
- [45] Liu L et al. LncRNA MT1JP functions as a tumor suppressor by interacting with TIAR to modulate the p53 pathway. *Oncotarget*. 2016;**7**(13):15787-15800
- [46] Yang F et al. Reciprocal regulation of HIF-1alpha and lincRNA-p21 modulates the Warburg effect. *Molecular Cell*. 2014;**53**(1):88-100
- [47] Li D et al. Activity dependent LoNA regulates translation by coordinating rRNA transcription and methylation. *Nature Communications*. 2018;**9**(1):1726
- [48] Tran NT et al. The AS-RBM15 lncRNA enhances RBM15 protein translation during megakaryocyte differentiation. 2016;**17**(6):887-900
- [49] Carrieri C et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*. 2012;**491**(7424):454-457
- [50] Hu G, Lou Z, Gupta M. The long non-coding RNA GAS5 cooperates with the eukaryotic translation initiation factor 4E to regulate c-Myc translation. *PLoS One*. 2014;**9**(9):e107016
- [51] Yamamura S et al. Interaction and cross-talk between non-coding RNAs. *Cellular and Molecular Life Sciences*. 2018;**75**(3):467-484
- [52] Zhou W et al. The lncRNA H19 mediates breast cancer cell plasticity during EMT and MET plasticity by differentially sponging miR-200b/c and let-7b. 2017;**10**(483). DOI: 10.1126/scisignal.aak9557
- [53] Ma C et al. H19 promotes pancreatic cancer metastasis by derepressing let-7's suppression on its target HMGA2-mediated EMT. *Tumour Biology*. 2014;**35**(9):9163-9169
- [54] Morriss GR, Cooper TA. Protein sequestration as a normal function of long noncoding RNAs and a pathogenic mechanism of RNAs containing nucleotide repeat expansions. *Human Genetics*. 2017;**136**(9):1247-1263
- [55] Lee S et al. Noncoding RNA NORAD regulates genomic stability by

- sequestering PUMILIO proteins. *Cell*. 2016;**164**(1-2):69-80
- [56] Zhou J et al. H19 lncRNA alters DNA methylation genome wide by regulating S-adenosylhomocysteine hydrolase. 2015;**6**:10221
- [57] Kino T et al. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Science Signaling*. 2010;**3**(107):ra8
- [58] Duca M et al. The triple helix: 50 years later, the outcome. *Nucleic Acids Research*. 2008;**36**(16):5123-5138
- [59] Roberts RW, Crothers DM. Stability and properties of double and triple helices: Dramatic effects of RNA or DNA backbone composition. *Science*. 1992;**258**(5087):1463-1466
- [60] Martianov I et al. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*. 2007;**445**(7128):666-670
- [61] Grote P, Herrmann BG. The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. *RNA Biology*. 2013;**10**(10):1579-1585
- [62] Grote P et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental Cell*. 2013;**24**(2):206-214
- [63] Simon MD et al. The genomic binding sites of a noncoding RNA. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;**108**(51):20497-20502
- [64] Salmena L et al. A ceRNA hypothesis: The Rosetta stone of a hidden RNA language? *Cell*. 2011;**146**(3):353-358
- [65] Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? *Nature Reviews. Genetics*. 2008;**9**(2):102-114
- [66] Ahadi A, Sablok G, Hutvagner G. miRTar2GO: A novel rule-based model learning method for cell line specific microRNA target prediction that integrates Ago2 CLIP-Seq and validated microRNA-target interaction data. *Nucleic Acids Research*. 2017;**45**(6):e42
- [67] Zhang XQ, Yang JH. Discovering circRNA-microRNA interactions from CLIP-Seq data. *Methods in Molecular Biology*. 2018;**1724**:193-207
- [68] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;**120**(1):15-20
- [69] Beltran M et al. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes & Development*. 2008;**22**(6):756-769
- [70] Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature*. 2011;**470**(7333):284-288
- [71] Sharma E et al. Global mapping of human RNA-RNA interactions. *Molecular Cell*. 2016;**62**(4):618-626
- [72] Nguyen TC et al. Mapping RNA-RNA interactome and RNA structure in vivo by MARIO. *Nature Communications*. 2016;**7**:12023
- [73] Wang C et al. lncRNA structural characteristics in epigenetic regulation. *International Journal of Molecular Sciences*. 2017;**18**(12). DOI: 10.3390/ijms18122659

- [74] Ferre F, Colantoni A, Helmer-Citterich M. Revealing protein-lncRNA interaction. *Briefings in Bioinformatics*. 2016;**17**(1):106-116
- [75] Zhang T, Tan P, Wang L. RNALocate: A resource for RNA subcellular localizations. 2017;**45**(D1):D135-d138
- [76] Su ZD et al. iLoc-lncRNA: Predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics*. 2018. DOI: 10.1093/bioinformatics/bty508
- [77] Li Y, Syed J, Sugiyama H. RNA-DNA triplex formation by long noncoding RNAs. *Cell Chemical Biology*. 2016;**23**(11):1325-1333
- [78] Buske FA et al. Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Research*. 2012;**22**(7):1372-1381
- [79] He S et al. LongTarget: A tool to predict lncRNA DNA-binding motifs and binding sites via Hoogsteen base-pairing analysis. *Bioinformatics*. 2015;**31**(2):178-186
- [80] Muckstein U et al. Thermodynamics of RNA-RNA binding. *Bioinformatics*. 2006;**22**(10):1177-1182
- [81] Busch A, Richter AS, Backofen R. IntaRNA: Efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*. 2008;**24**(24):2849-2856
- [82] Tafer H, Hofacker IL. RNAplex: A fast tool for RNA-RNA interaction search. *Bioinformatics*. 2008;**24**(22):2657-2663
- [83] Gawronski AR et al. MechRNA: Prediction of lncRNA mechanisms from RNA-RNA and RNA-protein interactions. *Bioinformatics*, 2018; **34**(18):3101-3110
- [84] Umu SU, Gardner PP. A comprehensive benchmark of RNA-RNA interaction prediction tools for all domains of life. *Bioinformatics*. 2017;**33**(7):988-996
- [85] Li JH et al. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*. 2014;**42**(Database issue):D92-D97
- [86] Hsu SD et al. miRTarBase: A database curates experimentally validated microRNA-target interactions. *Nucleic Acids Research*. 2011;**39**(Database issue):D163-D169
- [87] Krek A et al. Combinatorial microRNA target predictions. *Nature Genetics*. 2005;**37**(5):495-500
- [88] John B et al. Human microRNA targets. *PLoS Biology*. 2004;**2**(11):e363
- [89] Liu K et al. Annotating function to differentially expressed lincRNAs in myelodysplastic syndrome using a network-based method. *Bioinformatics*. 2017;**33**(17):2622-2630
- [90] Tong Y, Ru B, Zhang J. miRNACancerMAP: an integrative web server inferring miRNA regulation network for cancer. *Bioinformatics*, 2018;**34**(18):3211-3213
- [91] Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature*. 2014;**505**(7483):344-352
- [92] Thomson DW, Dinger ME. Endogenous microRNA sponges: Evidence and controversy. *Nature Reviews. Genetics*. 2016;**17**(5):272-283
- [93] Suresh V et al. RPI-Pred: Predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Research*. 2015;**43**(3):1370-1379

- [94] Livi CM, Blanzieri E. Protein-specific prediction of mRNA binding using RNA sequences, binding motifs and predicted secondary structures. *BMC Bioinformatics*. 2014;**15**:123
- [95] Lu Q et al. Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics*. 2013;**14**:651
- [96] Wang L, Brown SJ. BindN: A web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucleic Acids Research*. 2006;**34**(Web Server issue):W243-W248
- [97] Terribilini M et al. RNABindR: A server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Research*. 2007;**35**(Web Server issue):W578-W584
- [98] Liu ZP et al. Prediction of protein-RNA binding sites by a random forest method with combined features. *Bioinformatics*. 2010;**26**(13):1616-1622
- [99] Kumar M, Gromiha MM, Raghava GP. Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins*. 2008;**71**(1):189-194
- [100] Wang Y et al. PRINTR: Prediction of RNA binding sites in proteins using SVM and profiles. *Amino Acids*. 2008;**35**(2):295-302
- [101] Ma X et al. Prediction of RNA-binding residues in proteins from primary sequence using an enriched random forest model with a novel hybrid feature. *Proteins*. 2011;**79**(4):1230-1239
- [102] Fernandez M et al. Prediction of dinucleotide-specific RNA-binding sites in proteins. *BMC Bioinformatics*. 2011;**12**(Suppl. 13):S5
- [103] Walia RR et al. RNABindRPlus: A predictor that combines machine learning and sequence homology-based methods to improve the reliability of predicted RNA-binding residues in proteins. *PLoS One*. 2014;**9**(5):e97725
- [104] Xiong D, Zeng J, Gong H. RBRIdent: An algorithm for improved identification of RNA-binding residues in proteins from primary sequences. *Proteins*. 2015;**83**(6):1068-1077
- [105] Livi CM et al. catRAPID signature: Identification of ribonucleoproteins and RNA-binding regions. *Bioinformatics*. 2016;**32**(5):773-775
- [106] Luo J et al. RPI-bind: A structure-based method for accurate identification of RNA-protein binding sites. *Scientific Reports*. 2017;**7**(1):614