

State of the art technologies to explore long non-coding RNAs in cancer

Saeede Salehi^{a, b, c, #}, Mohammad Naser Taheri^{a, b, c, #}, Negar Azarpira^d, Abdolhossein Zare^d, Abbas Behzad-Behbahani^{a, b, *} 

^a Diagnostic Laboratory Sciences and Technology Research Center, School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

^b Department of Medical Biotechnology, School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

^c Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran

^d Transplant Research Center, Namazi Hospital, Shiraz University of Medical Sciences, Shiraz, Iran

Received: August 12, 2016; Accepted: April 17, 2017

• Introduction

• LncRNAs: the emerging master regulators of gene expression and cell fate

• LncRNAs: key players in cancer biology and diagnosis

• Methods used to localize lncRNAs inside the cell

– Single-molecule RNA *in situ* hybridization (sm-RNA FISH)

– RNA Mimics of GFP and RNA visualization

– Clustered regularly interspaced short palindromic repeats (CRISPR) RNA Tracking

• Methods used to predict lncRNA structure

– Selective 2'-hydroxyl analysed by primer extension (SHAPE)

– Parallel analysis of RNA structure (PARS)

– Domain-specific chromatin isolation by RNA purification (dChIRP)

• Methods utilized for determination of lncRNA function

– Knockout techniques

Clustered regularly interspaced short palindromic repeats (CRISPR) Cas technology

– Knock-down techniques

RNA interference (RNAi)

CRISPR Interference (CRISPRi)

Combined knock-down and localization analysis of non-coding RNAs (c-KLAN)

Antisense oligonucleotides (ASOs)

• Methods used for mapping lncRNA interactions

– Mapping lncRNA–DNA interactions

ChIRP, CHART and RAP as RNA-centric methods to map lncRNA genomic binding sites

– Mapping lncRNA–protein interactions by protein pull-down RNA immunoprecipitation (RIP)

Cross-linking and immunoprecipitation (CLIP)

• Next-generation sequencing-based methods for global investigation of lncRNAs

– Ligation of interacting RNA followed by high-throughput sequencing (LIGR-seq)

– Inosine Chemical Erasing Sequencing (ICE-seq)

– Global mapping of lncRNA–protein interactions based on RNA modification and RNA-seq

– Global run-on assay sequencing (Gro-seq)

– 5'-bromo-uridine immunoprecipitation chase–deep sequencing (BRIC-seq)

– Fluorescent *in situ* sequencing (FISSEQ)

• Conclusion

Abstract

Long non-coding RNAs (lncRNAs) comprise a vast repertoire of RNAs playing a wide variety of crucial roles in tissue physiology in a cell-specific manner. Despite being engaged in myriads of regulatory mechanisms, many lncRNAs have still remained to be assigned any functions. A constellation of experimental techniques including single-molecule RNA *in situ* hybridization (sm-RNA FISH), cross-linking and immunoprecipitation (CLIP), RNA interference (RNAi), Clustered regularly interspaced short palindromic repeats (CRISPR) and so forth has been employed to shed light on lncRNA cellular localization, structure, interaction networks and functions. Here, we review these and other experimental

[#]These authors contributed equally to this work.

*Correspondence to: Abbas BEHZAD-BEHBHANI

E-mails: behzadba@sums.ac.ir; behzadba@gmail.com

approaches in common use for identification and characterization of lncRNAs, particularly those involved in different types of cancer, with focus on merits and demerits of each technique.

Keywords: long non-coding RNAs (lncRNAs) • cancer • experimental techniques • interaction networks • lncRNA function

Introduction

In the post-genomic era, the 'junk' moniker is giving way to the understanding that most of the genome is fully functional [1]. We are facing an expanding RNA universe filled with transcripts of unknown functions (TUFs) [1]. The catalogue of non-coding RNAs (ncRNAs) is skyrocketing in an ever-increasing way and the part of the human genome (nearly two-thirds) transcribed into non-coding RNAs vastly outweighs the part apportioned to protein-coding genes [1]. Among recently highlighted ncRNAs are lncRNAs that surpass 200 base pairs in size [1]. According to a current GENCODE release, 9640 loci encoding 15512 lncRNA transcripts have been identified by now [2]. Parts of the human genome undergoing active transcription can be distinguished from other parts by specific signatures [3]. Important among such signatures are a stretch of trimethylation of histone at the lysine in position 4 (H3K4me3) associated with the promoter areas, followed by another stretch of trimethylation of histone at the lysine in position 36 (H3K36me3) spanning the whole transcribed region [3, 4]. Among transcripts encoded by these regions are long non-coding RNAs [5].

Most lncRNAs are transcribed by RNA polymerase II [1], and similar to protein-coding RNAs, can be subjected to post-transcriptional modifications, including 5'-end capping, 3'-end polyadenylation, splicing [6–11] N⁶ adenosine methylation (m⁶A) [12] and A-to-I editing [13]. Due to not having a comprehensive knowledge of their function, lncRNAs are categorized based on their location relative to the protein-coding genes as: sense (which are transcribed from sense strand of protein-coding genes and overlap one or more exons and introns), antisense (which are transcribed from the antisense strand of protein-coding genes), bidirectional (which are located 1 kb away from promoters of protein-coding genes in the opposite direction), intronic (which are located between exons of protein-coding genes) and intergenic (which are located between two protein-coding genes) [5, 7, 14, 15]. On the other hand, other lncRNAs including natural antisense transcripts (NATs) [16], circular RNAs (cRNAs) [17, 18], enhancer ncRNAs (eRNAs) [19–21], transcribed pseudogenes [22] and telomerase RNA component (TERC) [23] have been also identified.

lncRNAs: the emerging master regulators of gene expression and cell fate

lncRNAs play a diversity of regulatory functions due to their modular structures and RNA–RNA or RNA–DNA base pairing

ability [5]. Their modularity enables them to act as scaffolds, providing docking sites for proteins that function together in a given biological process [5].

By combining their modularity and base-pairing ability, lncRNAs act as guiding molecules, directing proteins to other RNA or DNA molecules [24]. For instance, the lncRNA HOTAIR was illustrated to modify the epigenetic state of the *HOXD* gene by binding two enzymes (PRC2 and LSD1) involved in chromatin modification [24]. In a series of HOTAIR deletion studies, the binding motifs for PRC2 and LSD1 were mapped to nucleotides 1–300 and 1500–2146, revealing two structural modules that bind and recruit the two proteins to the *HOXD* gene, where PRC2 deposits the H3K27me3 modification needed for repression and LSD1–CoREST–REST erases the H3K4me2 modification involved in activation [24]. As 'sponges', lncRNAs hybridize to miRNAs and prevent them from exerting their effects [25, 26]. The lncRNA, Cancer Susceptibility Candidate 2 (CASC2), sponges miR-18a and prevents its down-regulatory effect on PIAS3 which is an inhibitor of STAT pathway [25]. In another case, lncRNAs regulate mRNA functioning by interacting with them and blocking their translation or manipulating their splicing patterns [27]. Generally, the list of lncRNA regulatory functions at the molecular level includes chromatin modification [28], genomic imprinting [28], chromosomal dosage compensation [29] and alternative splicing [30] (Fig. 1). From a cellular perspective, lncRNAs play central regulatory roles in cell differentiation [1], cell cycle regulation [31], programming of stem cells [32] and apoptosis [33].

lncRNAs: key players in cancer biology and diagnosis

lncRNAs, now being emerged as master regulators of gene expression, have been shown to be involved in a variety of malignancies [34]. They can be an integral part at the frontline of research in cancer arena, as many of them are involved in tumour progression and aggressiveness by epigenetically silencing tumour suppressors [35]. HOTAIR, for instance, stimulates breast cancer metastasis by targeting PRC2 and LSD1–CoREST complexes to metastasis suppressor genes (such as *PCDH10*, *PCDH5* and *JAM2*) [36]. From another perspective, some lncRNAs serving as tumour suppressors, contribute to cancer progression, when down-regulated [37] Supplementary Table 1.

In addition, lncRNAs have been proven as important diagnostic and prognostic biomarkers of cancer and other disorders

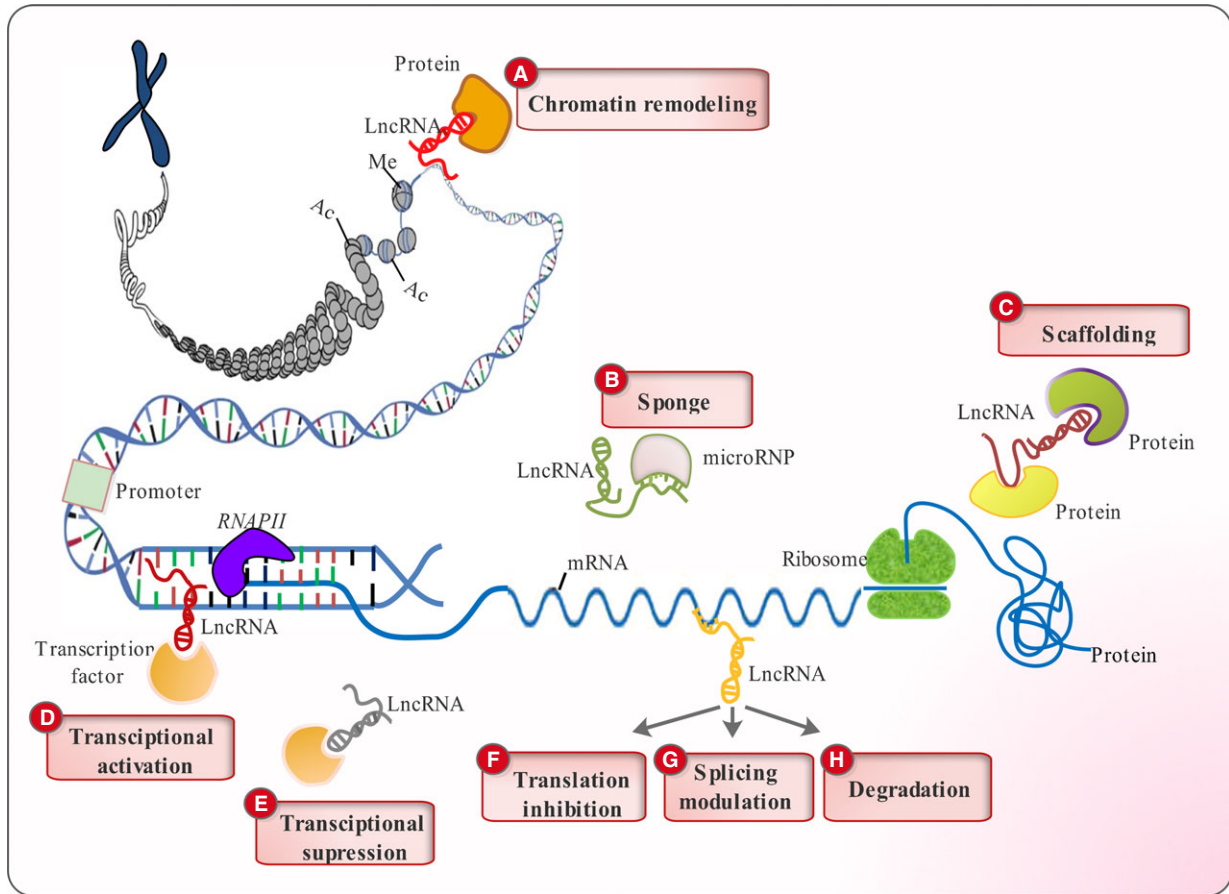


Fig. 1 Classification of lncRNA functions. **(A)** LncRNAs can recruit different protein components of the chromatin remodelling complex to change the chromatin organizational patterns. **(B)** They can act as ‘sponges’ by base pairing with their complementary miRNAs and reducing their effects. **(C)** LncRNAs can play scaffolding roles by providing docking sites for proteins that function together in the same biological pathway. **(D)** They activate transcription of certain genes by guiding transcription factors to their promoters. **(E)** LncRNAs are capable of suppressing transcription by sequestering transcription factors and keeping them away from their promoters. They can modulate mRNA functioning through base pairing with them and **(F)** inhibiting their translation **(G)** altering their splicing patterns and **(H)** subjecting them to degradative pathways.

[38]. They have tissue- and cancer-specific expression profiles and can be easily detected in blood and urine samples [38]. For instance, PCA3 is a prostate specific lncRNA, overexpressed in prostate cancer. Diagnostic tests have been developed for detection of PCA3 in urine samples as a biomarker for non-invasive diagnosis of prostate cancer [38, 39].

Thanks to advances in increasing the depth of RNA sequencing techniques, thousands of novel lncRNAs have been discovered in recent years. However, the task of characterizing newly discovered lncRNAs is an arduous one, contingent on the power of technologies at our disposal.

In this review, we strive to shed light on the ever-expanding toolkit of techniques employed in lncRNA research. The methods are classified based on methodology as techniques used to visualize lncRNAs, determine their structure, map their interactions with other cellular components and elucidate their functions. Each method is described

briefly, with their attributed pros and cons being discussed and summarized.

Methods used to localize lncRNAs inside the cell

Most of the genomic sequence is transcribed, and the majority of such transcripts do not code for proteins [40]. Unlike mRNAs whose sequence is a valuable source of insight into their functions, the sequence of lncRNAs usually do not provide much information about their function. Therefore, shedding light on the subcellular localization of lncRNAs can be a first step in revealing their functions in cells and malfunctioning in disease conditions including cancer [41]. Here, we review cutting-edge techniques, which can be employed in determining the localization of lncRNAs.

Table 1 Advantages and disadvantages of the techniques used to characterize lncRNAs

Approaches	Advantages	Disadvantages
Methods used to localize lncRNAs inside the cell		
1. sm-RNA FISH	- Visual localization of the intracellular location of lncRNAs	- The possibility of false positive detection due to similarity of true nuclear lncRNA signals to possible off-target nuclear signals
	- Detection of low abundance lncRNAs through probes that tile all of true nuclear lncRNA signals to possible off-target over the transcript length	
	- Quantitative measurement of lncRNA transcripts	
2. CRISPR RNA Tracking determining	- Does not require manipulation endogenous RNA molecules	- Three components need to transfected into the cell
	- Does not affect the abundance or expression profile of endogenous RNAs	- Nuclear localization of GFP-dCas9 may complicate the location of nuclear resident RNAs
Methods used to predict lncRNA structure		
1. SHAPE	- Capable of revealing single-stranded and double-stranded lncRNA motifs	- Inability to determine the range of base pairing interactions
		- Limited to <i>in vitro</i> experiments
2. PARS	- Capable of revealing the global structure of lncRNAs or the lncRNA structurome	- Limited to experiments done <i>in vitro</i> and cannot be done <i>in vivo</i>
	- Capable of revealing the global changes in lncRNA structure under different cellular conditions and presence of toxins or drugs	
3. dChIRP	- Increased signal to noise ratio due to recovery based on individual domains as compared to ChIRP	- Requires a prior knowledge of individual lncRNA domains
	- Revelation of lncRNA interactions with RNA, DNA and proteins in a single reaction unlike ChIRP and CLIP	
	- Characterization of individual lncRNA domains and their functions	
Methods utilized for determination of lncRNA function		
1. CRISPR/Cas9	- Less off-target effect in comparison to conventional knockout methods	- Leading to introduction of indels by NHEJ which cannot ablate non-coding genes
	- High recombination frequency	
	- Complete lncRNA perturbation in case of the dual gRNA system	
2. RNAi	- Low off-target effect if ESI RNAs are used	- Stable silencing of target transcripts is not possible if esi- or si-RNAs are used
	- Tighter intracellular level of silencing if ESI RNAs are used	- Inducible silencing is not possible if ESI or si-RNAs are used
	- Cost effective	
	- More resistant to nucleases compared to ASO	
	- Inducible and stable silencing is possible if shRNAs are used	

Table 1. Continued

Approaches	Advantages	Disadvantages
3. CRISPRi	- Acts on DNA and therefore has a broader target range	- Needs transfection of several components into the cell
	- Can be better applied to noncoding transcripts	
	- Can be adapted for multiple silencing of several genes	
	- Active in both cytoplasm and nucleus	
4. ASO	- Capable of silencing nuclear lncRNAs	- Being labile and sensitive to nucleases
	- The probe length can be fine-tuned to achieve the optimal specificity unlike RNAi probes which have a constant length	- Increased off-target effect in case of modified ASOs
Methods Used for mapping lncRNA interactions		
1. CHART	- Reduced background signal owing to fewer number of probes used	- Formaldehyde used as the cross-linking agent is not efficient
	- No prior knowledge of individual domains is required	- Probe synthesis is time consuming due to the RNase H assay required
2. ChIRP	- Revealing genome-wide lncRNA DNA binding sites	- Increased noise-to-signal ratio in comparison to dChIRP
	- Prior knowledge of individual domains is required by using probes that tile all over the length of lncRNAs	- Unable to reveal the function of individual lncRNA domains
	- Short and easy-to-synthesize probes used decrease the off-target effect	
3. RAP	- No prior knowledge of individual domains is required	- Long probes used which are difficult to synthesize and increase the off-target effect
	- Capable of detecting residual lncRNA fragments resulting from shearing	
Mapping lncRNA-protein interactions by protein pull-down		
1. RIP	- No prior knowledge of the RNA partners is required in case of RIP-seq	- Requires antibodies against known proteins
	- Genome-wide mapping of RNA-protein interactions	- Pull-down of off-targets due to nonspecific interactions
		- False negatives
2. CLIP	- Global screening of RNA-protein interactions in HITS-CLIP	- Pull-down of nonspecific proteins
	- No prior knowledge of RNA is required	- Inefficiency of UV cross-linking
	- Reduction of false negatives by using UV cross-linking	- False positive detections that may occur due to nonspecific interactions during cross-linking
	- Higher resolution in case of RNase digestion	
3. PAR CLIP	- Improvement of cross-linking due to 4-SU/6-SG labeling	- Limited to cell culture systems
	- More accurate mapping of RNA-protein interaction	- Pull-down of nonspecific proteins
4. iCLIP	- Nucleotide resolution mapping of lncRNA-protein interactions	- Nonspecific pull-down of proteins
	- Detection of rare events owing to amplification	- Bias due to nonlinear PCR amplification

Table 1. Continued

Approaches	Advantages	Disadvantages
Next generation sequencing based methods for global investigation of lncRNAs		
1. LIGR-seq:	- Maps RNA-RNA interactions	- Hybrid ligation may be difficult between short RNA fragments
	- Does not need prior knowledge of interacting protein partner	
	- Implemented <i>in vivo</i>	
2. TRIBE-seq	- Not dependent on antibodies for affinity purification	- Not capable of revealing exact protein-RNA interface
	- Needs a small number of cells	- Results may be complicated due to endogenous ADAR activity
	- Reveals protein-RNA interaction is a cell specific manner	
3. ICE-seq	- Reveals global A-to-I editing events	- Unable to detect completely edited sites
	- The results are not confounded by Gs occurring due to sequencing errors	
	- Individual variations due to SNPs do not confound the results	
4. GRO-seq	- Measurement of relative activity of sites undergoing transcription	- Introduction of artefacts during preparation of nuclei
	- Detection of divergent transcription	- New initiations that may occur ensuing abortive transcription
	- No prior knowledge of transcript initiation site is needed	- Limitation to cell culture for incorporation of labeled analogues
5. BRIC-seq	- Measurement of the half-life of lncRNAs	- Introduction of artefacts during preparation of nuclei
	- Revealing the identity of RNA molecules based on their half-life	- New initiations that may occur ensuing abortive transcription
		- Limitation to cell culture for incorporation of labeled analogues

Single-molecule RNA *in situ* hybridization (smRNA FISH)

As with other molecules, our knowledge of the subcellular localization of lncRNAs opens up new vistas to their potential cellular roles [42]. RNA FISH has been utilized for many years to localize RNA molecules inside the cells [43]. All FISH techniques are based on hybridization of precisely designed fluorescently labeled probes to their cellular targets. However, the conventional RNA FISH technique lacks enough sensitivity for low-abundance lncRNA molecules. smRNA-FISH was developed later to circumvent this problem, by making use of a pool of short probes designed to tile the entire length of lncRNA. smRNA-FISH is very sensitive in detecting low-abundance lncRNA transcripts, being quantitative and allowing absolute measurement of the transcripts [44, 45].

A common obstacle in revealing the subcellular localization of lncRNA molecules is the off-target hybridization of probes to nuclear aggregations, generating blob-like signals like that of true lncRNA signals. Cabili *et al.*, made use of two pools of probes labelled with different fluorophores and co-localizing the signals obtained using the two probe sets. If the signals co-localize, then they are construed as true ones indicating the lncRNA localizations, otherwise at least some of the signals are considered to be due to off-target hybridizations [46] (Table 1).

RNA FISH and immunofluorescence have been used in order to show the co-localization of Urothelial carcinoma associated 1 (UCA1 (and phosphorylated hnRNP1 in the perinuclear regions to confirm that the phosphorylated form of hnRNP1 is able to interact with UCA1 [47]. Dual RNA FISH and immunofluorescence using probes against MALAT1, U2-snRNA, and antibodies against SR splicing factor,

SRSF1, have been used to clarify the role of MALAT1 in RNA splicing. MALAT 1 was shown to be co-localized to nuclear speckles with SR splicing factors [48].

RNA Mimics of GFP and RNA visualization

RNA tracking methods have been developed that take advantage of green fluorescent protein fused to MS2 coat protein (MCP) and the MS2 binding sequence inserted into the RNA molecule to be analysed. The RNA molecule containing the MS2 binding site and the GFP-MCP fusion protein are expressed inside the cell. The localization of RNA molecules is revealed through GFP-MCP binding to the MS2 site [49]. However, as the GFP-MCP fusion protein contains nuclear localization signal sequences, it may affect the subcellular location of RNA molecules inside the cell. Furthermore, investigation of nuclear resident RNAs may become complicated due to accumulation of the GFP-MCP inside the nucleus and intense nuclear fluorescence [41]. These impediments have stimulated the development of newer and better techniques for locating RNA molecules.

The GFP fluorophore, 4-hydroxybenzylidene imidazolinone (HBI) does not show fluorescence in the absence or denaturation of the GFP protein. RNA mimics of GFP are aptamer molecules that can bind HBI or other GFP like fluorophores and enable them to fluoresce [42]. If inserted to 3' end of RNA molecules including lncRNAs, the localization of target RNAs can be revealed upon addition of the fluorophore to the cell. One RNA mimic of eGFP referred to as Spinach binds DFHBI [(Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one] was inserted to the 3' end of 5S rRNA and successfully revealed its subcellular distribution [42]. Therefore, the GFP RNA mimics can be applied to lncRNA research so as to enhance and simplify the attempt to map subcellular localization of these transcripts and how they might change in cancer disease.

Clustered regularly interspaced short palindromic repeats (CRISPR) RNA Tracking

The CRISPR/Cas9 system is considered as a type of adaptive immunity in bacteria and archaea [50]. The Cas9 protein is the effector nuclease in type II CRISPR systems, which targets and cleaves invading phage DNA in close proximity to a sequence called protospacer adjacent motif (PAM) [51]. The Cas9 protein from type II CRISPR system was believed to recognize only DNA but not RNA sequences [52]. However, a study performed by the Doudna laboratory in 2014 revealed the ability of Cas9 protein to target and cleave RNA sequences, therefore, expanding the biotechnological utility of this protein from DNA to RNA [52]. The Cas9 protein was shown to bind RNA in the presence of a guide RNA. However, no RNA cleavage activity was observed until addition of a short oligonucleotide sequence (called PAMmer) with an NGG PAM sequence at its 5' end. Furthermore, the binding affinity was shown to increase fivefold in the presence of the PAMmer [52].

The RNA-binding capability of Cas9 protein has spurred the development of technologies for tracking the subcellular localization of RNA molecules [53]. A nuclease null Cas9 protein fused to EGFP

(dCas9-EGFP) was shown to successfully recognize ACTB, CCNA2 and TFRC mRNAs only in the presence of gRNA [53]. However, dCas9 binding affinity to RNA was shown to be augmented in the presence of the PAMmer sequence [52, 53]. In the presence of gRNA and PAMmer components recognizing the 3' UTR of mRNA molecules, the nuclear localized dCas9-EGFP protein was found to be co-exported to the cytoplasm and reveal the subcellular localization of these RNA molecules, with comparable success to experiments performed with FISH on the same RNA molecules [53] (Fig. 2A). Furthermore, the CRISPR-based RNA tracking technology was also shown to be capable of locating the ACTB mRNA in stress granules, as both the dCas9-EGFP and Ras GTPase-activating protein-binding protein 1 were found to be merged in such granules [53].

Despite being tested on few mRNA molecules, the CRISPR-based RNA tracking technology has the potential to be utilized in studies aimed at revealing the subcellular localization of RNAs including lncRNAs.

Unlike RNA detection technologies based on RNA-protein interactions, the CRISPR-based RNA tracking technique does not require laborious genetic engineering procedures, as target recognition is based on RNA-RNA interactions [54]. In addition, unlike RNA tracking methods based on MCP protein (described in 1.2.), this system does not require modification of the target RNA molecules and has no significant effect on localization, abundance or expression profile of its target RNAs [54] (Table 1). Therefore, CRISPR RNA tracking can be a complement to current methods of RNA tracking, providing new opportunities for better tracking and visualization of lncRNAs localization and its likely aberrations involved in cancer disease.

Methods used to predict lncRNA structure

lncRNAs are capable of folding on themselves and therefore taking a wide variety of secondary and tertiary structures, enabling them to perform their functions [55]. Therefore, an understanding of structure is an indispensable part of lncRNA research [55]. There are several techniques available to allow the structural and conformational investigation of lncRNAs.

Selective 2'-hydroxyl analysed by primer extension (SHAPE)

SHAPE is among the key techniques commonly used to unravel the secondary structure of lncRNAs [56]. The technique makes use of the chemical N-methylisatoic anhydride (NMIA) that acylates single-stranded RNA and depends on the fact that 2' hydroxyl groups in single-stranded or flexible regions of RNA are highly reactive [56]. The chemical modification of single-stranded RNA hinders the reverse transcription process in flexible regions [56]. Therefore, the parts of RNA which could undergo reverse transcription are sequenced and considered as to be involved in secondary structures, while other parts are interpreted to be single-stranded [56] (Fig. 3A).

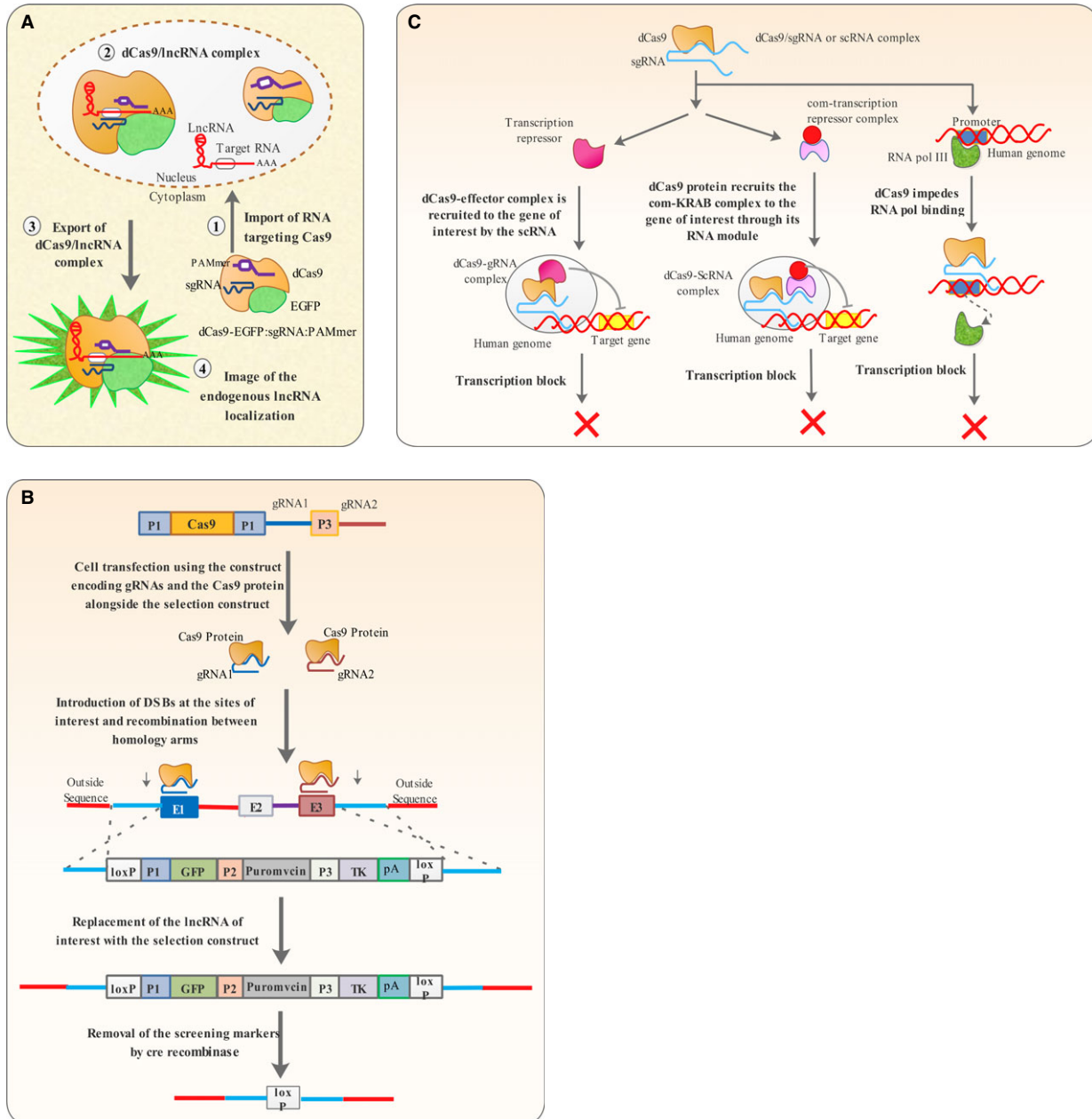


Fig. 2 CRISPR-based technologies. **(A)** CRISPR RNA tracking utilizes a catalytically dead Cas protein (dCas9) in the presence of gRNA and PAMmer oligonucleotide to visualize the subcellular location of RNA molecules. **(B)** The all-in-one system CRISPR knock-out strategy uses a two gRNAs that direct Cas9 nuclease activity to both ends of a given IncRNA, which is to be replaced by a construct comprising GFP, neomycin and tk, flanked by loxP sites. After screening using neomycin and GFP the DNA construct in excised out using cre and leaves a loxP site in the IncRNA locus, something confirmed by negative selection using tk. **(C)** CRISPRi suppresses the process of transcription by directing or recruiting chromatin modifying proteins to the gene of interest or by directly blocking RNA polymerase binding.

Despite being a powerful technique in unravelling the single and double-stranded regions of IncRNA molecules, SHAPE is not able to determine the range of such base pairing interactions [57]. The

shotgun secondary structure (3S) determination technique has been developed to overcome this problem by fragmenting RNA molecules before SHAPE is done [57]. Finally, the data from such fragmented

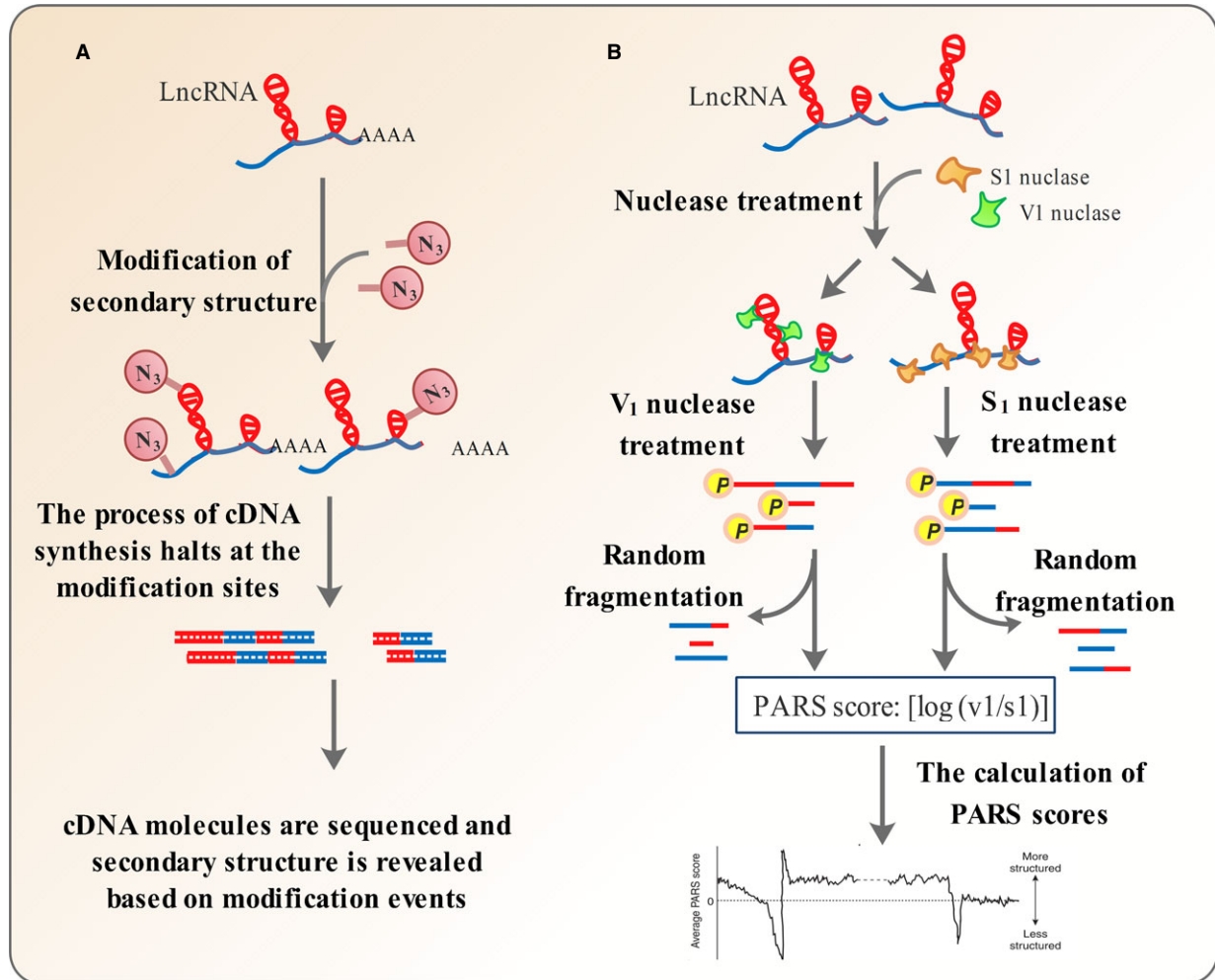


Fig. 3 Methods used to predict lncRNA structure. **(A)** In SHAPE, flexible regions of RNA sequence are modified using chemicals that react with 2' hydroxyl groups in single-stranded regions (Acylation). Then, cDNA synthesis is performed but halts at positions, which are chemically modified. Finally, reverse transcription stop positions are determined using deep sequencing to reveal RNA secondary structure. **(B)** In PARS, the lncRNA molecules are refolded *in vitro* and treated by V1 and S1 ribonucleases specific for double- or single-stranded RNA, respectively, in separate reactions. The resulting digested RNA molecules have phosphates at their 5' ends. The fragments are then ligated to adaptors at their 5' ends, subjected to random fragmentation and deep parallel sequencing. In each case, the RNA parts digested by the nucleases are the predominant parts, which could be ligated to adaptors and therefore sequenced. Finally, a PARS score is obtained as the ratio of double- versus single-strandedness. A higher PARS score indicates a greater probability of nucleotides to be in double-stranded regions, and vice versa.

transcripts are compared to whole transcript SHAPE data to provide insights into the low range and high range base pairings within the secondary structure [57]. Probably due to its short half-life in water and cross-reactivity with other nucleophiles abundant in the cell, the conventional electrophile used in SHAPE is not applicable to structural analysis of RNA molecules *in vivo* [58] (Table 1). Screening of a variety of electrophilic compounds led to the identification of 2-methyl-3-furoic acid imidazole (FAI) and 2-methylnicotinic acid imidazole (NAI), which could meet all the parameters needed for extending the applicability of SHAPE from *in vitro* to *in vivo* in icSHAPE (*in vivo* click SHAPE) [58].

An array of chemical probing techniques including SHAPE has been employed to characterize HOTAIR in structural terms [59]. The resulting structural map revealed that HOTAIR is made of four modules, with two of them serving as the protein-binding domains. This HOTAIR map also indicates several structural motifs conserved in the course of evolution [59]. The capability of SHAPE to reveal the single-stranded lncRNA parts can be served to clarify lncRNA–mRNA interactions [60]. SHAPE was used to show that the lncRNA, TINCR have single-stranded regions that are complementary to sequence motifs on its target mRNA, thereby indicating a case in which mRNAs can be regulated by an lncRNA through base pairing [60].

Parallel analysis of RNA structure (PARS)

Structural studies of ncRNAs in a single-molecule manner have been carried out successfully [61]. Two methods with similar mechanisms named PARS [62] and Frag-seq [63] were introduced recently, for global mapping of coding and non-coding RNA transcripts. The PARS technique utilizes RNase proteins specific for single- or double-stranded motifs in RNA molecules, leaving behind a 5' phosphate. RNA molecules are refolded *in vitro*, subjected to RNase treatment, ligated to adaptors and sequenced. Finally, a PARS score is obtained as the log ratio of in a double- or single-stranded conformation. The log ratio between the number of sequence reads obtained for each nucleotide in the V1 (cleaving double-stranded RNA) and S1 (cleaving single-stranded RNA) experiments. A nucleotide with a higher log ratio, or PARS score, thus indicates a higher probability of being in a double-stranded conformation (Fig. 3B). PARS is capable of expanding the boundaries of RNA structural analysis to the realm of high-throughput [62]. However, like many other techniques for analysing RNA structural features, PARS is limited to studies performed *in vitro* and cannot be employed *in vivo* [64] (Table 1).

PARS has been successfully applied to coding and non-coding RNA molecules, and structural evaluation of two domains from the HOTAIR lncRNA was successfully carried out using this technique [62]. The systematic probing of RNA structure using PARS under different chemical or physical conditions or in the absence or presence of other molecules can provide valuable insights into RNA structural changes that can affect cellular behaviour [64]. Therefore, these techniques have the potential to be used in evaluating the likely changes in lncRNA structure that may occur in cancer disease and screening of compounds that may restore such changes to the normal state.

Domain-specific chromatin isolation by RNA purification (dChIRP)

As pivotal regulators of gene expression, lncRNAs have been poorly characterized regarding their domain architecture. Domain-specific chromatin isolation by RNA purification (dChIRP) is a modification of ChIRP (described in 4.1.1) that allows simultaneous mapping of RNA–RNA, RNA–DNA and RNA–protein interactions at the level of individual domains [65].

In this technique, the cells are subjected to cross-linking agents to preserve protein–nucleic acid interactions and then sonicated to shear the nucleic acids. Specific biotinylated oligonucleotides are used to tile individual domains. Streptavidin magnetic beads are then added to capture specific interactions and non-specific interactions are washed away. The material recovered from dChIRP is finally fractionated into protein, DNA and RNA parts for extraction and analysis. The parts can be subjected to real-time PCR using primers for the lncRNA domain of interest and other RNAs suspected to be in its interaction. The protein content can be analysed by Western blotting or mass spectrometry to reveal any protein interacting with different domains. Recovered DNA can be also sequenced, revealing any

binding sites on genome that may become occupied by a given lncRNA domain [66].

Unlike other techniques used to map lncRNA interactome, dChIRP is capable of revealing lncRNA multivalent interactions with other RNA, DNA and protein molecules in a single reaction [65] (Table 1).

Using dChIRP, Rox lncRNA was indicated to have interaction with MSL proteins through its D1, D2 and D3 domains. The Rox binding sites on X-chromosome (dlg1, suv4-20 and u2af50) were also clearly determined. dChIRP has the potential to be applied to other lncRNAs with important implications in cancer, on the way to revealing their interactions in a domain by domain manner [66]. Furthermore, dChIRP performed on MDA-MB-231 breast cancer cells was indicated to be capable retrieving HOTAIR domain-specific RNAs, proving the generality of this technique [67]. These techniques can provide valuable insights into which lncRNA parts are the culprits in disrupting its overall function and narrow down the therapeutic strategies to focus on specific modules rather than the whole molecule.

Methods utilized for determination of lncRNA function

The traditional approach for gene discovery called forward genetics relies on finding aberrant phenotypes in a model organism and then unravelling the genes behind such phenotypes. In reverse genetics, however, a given gene of interest is perturbed and the resulting phenotypic change is observed, helping to find the gene function. Reverse genetics are particularly important in delineating the important roles of lncRNAs in cancer, as gene ablations can clearly depict their contributions to cancer phenotype or transformation of normal cells. Here, we look at the techniques utilized in lncRNA reverse genetics [68].

Knockout techniques

Clustered regularly interspaced short palindromic repeats (CRISPR) Cas technology

Several technologies for genome editing based on programmable proteins including Zinc-finger nucleases (ZFNs), transcription activator like effector nucleases (TALENs) and Cas9 have emerged over the past 10 years [69]. These proteins can be engineered to recognize specific DNA sequences and introduce double-stranded breaks. TALEN and ZFN specificities are based on protein–nucleic acid interaction [69]. The TALE domain of TALEN can be engineered to recognize different DNA sequences and fused to FokI restriction nuclease. The TALE domain then guides FokI to the site of interest where it induces a double-stranded break [69]. The Cas9 specificity is based on RNA–DNA interaction, and the protein itself has intrinsic nuclease activity [51]. Genome editing based on ZFN and TALEN is more cumbersome compared to Cas9 as it requires arduous genetic engineering [69]. Unlike coding genes, it is much more difficult to knock-down non-coding sequences, including lncRNAs [70]. Gene knockout based on programmable nucleases could not be applied to non-coding sequences efficiently, because lncRNAs can yet remain functional

even after insertions or deletions [71]. Ho and his colleagues developed an all-in-one system, comprising two gRNAs and the Cas9 protein to overcome this hurdle. The gRNAs are capable of directing the Cas9 endonuclease to the boundaries of lncRNA encoding sequence; therefore, the whole sequence will be replaced by a synthetic construct by homologous recombination [70] (Fig. 2B). Non-homologous end joining (NHEJ) is also an impediment towards perturbing non-coding sequences due to its competition with homologous recombination, therefore, leading to introduction of indels which are insufficient to completely disturb the function of lncRNAs. Coupling the all-in-one system strategy with siRNAs against Ku70, Lig4 and XRCC4 (proteins involved in NHEJ) can increase HR to NHEJ and therefore lead to complete knock-down of lncRNA of interest [70].

Making use of two gRNAs, the dual CRISPR system is suited for perturbing lncRNAs in contrast to conventional methods, which may lead to insertions or deletions, not enough to eliminate the function of lncRNAs [70, 72] (Table 1).

The applicability of the all-in-one system strategy in perturbing lncRNAs was illustrated by ablating urothelial carcinoma-associated 1 (UCA1), lncRNA-21A and AK023948 in various human cell lines [70]. In another study, PRNCR1 and PCGEM1 were perturbed and indicated to be involved in inhibiting PYGO2 function in AR-dependent promoter/enhancers in castration-resistant prostate cancer [73]. However, PRNCR1 and PCGEM1 were then shown not to be involved in castration-resistant prostate cancer [74]. A dual knockout system was also used to eliminate the promoter region of MALAT1 so as to block its transcription in a number of human cell lines [75]. Therefore, genome editing technologies including CRISPR/Cas9 are key tools in reverse genetics of lncRNAs and herald a new era in which many lncRNAs can be deleted systematically so as to provide a coherent picture of their role in cancer [76].

Knock-down techniques

RNA interference (RNAi)

While the knockout techniques eliminate gene expression by deleting the genes, RNAi suppresses expression by sequestering and then cleaving RNA molecules [77, 78]. Several knock-down strategies are in common use, utilizing either synthetic or enzymatically prepared RNAi libraries [79]. Endoribonuclease-prepared siRNAs are made by treatment of dsRNAs with RNase III or Dicer [79]. The resulting RNAs, called esi-RNAs, are delivered directly into the cytoplasm. Esi-RNAs show less off-target effect and trigger a tighter suppression of target RNAs compared to conventional methods [79]. The silencing response using esi-RNAs is transient [79–81]. Conventional siRNAs are prepared by chemical synthesis and delivered into cytoplasm directly [79]. Chemically synthesized siRNAs also bring about a tight silencing response, despite showing a more off-target effect compared to esi-RNAs. The response triggered by siRNAs is transient [79]. shRNAs are another class of RNAs used in RNAi. They are not delivered directly to the cytoplasm but are expressed inside the cell. The silencing response triggered by shRNAs may be transient or stable, and the off-target effect is greater compared to esi-RNAs [79] (Table 1).

Any RNAi experiment has to be accompanied by standard controls [82]. Oligonucleotides recognizing no targets inside the cell have to be used alongside the RNAi experiment so as to ensure that the resulting phenotypic change is not due to non-specific recognition of other targets [82]. If the control siRNA indicates no significant change in cell viability or phenotype compared to control, the resulting changes in cells treated with test siRNA are specific [82].

Involved in aggressive prostate cancer, second chromosome locus associated with prostate-1 (SchLAP1), is an lncRNA that inhibits the tumour-suppressive activity of the SWI/SNF complex. The role of SchLAP1 in contributing to aggressiveness of prostate cancer was assayed by performing siRNA-mediated knock-down experiments. Remarkably, the SchLAP1 knock-down dramatically decreased cell proliferation and invasiveness.[35]. siRNAs against UCA1 and hnRNP1 were used in different experiments to show that UCA1 knock-down is associated with increased p27 expression, while hnRNP1 depletion reduces p27 expression. These observations confirmed the oncogenic role of UCA1 in breast cancer through suppressing p27 expression.[47].

RNAi has been successfully applied to many loss-of-function studies involving lncRNAs [71]. However, when a nuclear lncRNA is to be targeted, RNAi is not efficient furthermore, although the RNAi machinery has been shown to be active in the nucleus [71, 83].

CRISPR Interference (CRISPRi)

The versatility of CRISPR/Cas9 has enabled repurposing of its utility from genome editing to genome regulation [84]. Recently, a modified CRISPR system called CRISPRi has been proven to be effective in down-regulating gene expression by blocking transcription [84]. A minimal CRISPRi machinery consists of a catalytically dead Cas9 protein lacking endonucleolytic activity (dCas9) and a gRNA recognizing the target gene [84]. gRNAs targeting the non-template DNA strand in promoter or -35 regions are more effective in gene down-regulation than those targeting template strand sequences or regions 100 bp upstream of the promoter [84] (Fig. 2C). Despite being quite efficient in repressing gene expression in bacteria, CRISPRi based on direct blocking of RNA polymerase was not as efficient when applied to eukaryotic cells [84]. This problem led researchers to design chimeric dCas9 proteins fused to chromatin modifying effectors, including KRAB (Krüppel-associated box) so as to bring about epigenetic silencing of gene expression [85]. This new CRISPRi platform was indicated to successfully repress the expression of *GFP* in HEK293 cells and endogenous genes including *CD71*, *CXCR4* in HeLa cells [85]. CRISPRi through epigenetic silencing has been further refined by making use of protein-binding RNA modules that bind and recruit the effector protein instead of direct fusion of effector to dCas9 [84]. Well-characterized protein-binding RNA modules including MS2, PP7, and com recognized by MCP, PCP, and com RNA-binding proteins can be fused to the 5' end of gRNA to make scaffold RNA (scRNA) [84]. Genes can be down-regulated by making use of an scRNA that can recruit any given RNA-binding protein fused to any chromatin modifier, com-KRAB fusion, for instance, to the site of interest. [84] In addition, simultaneous silencing of several genes can be brought about by utilizing several scRNAs consisting of com modules and gRNAs targeting different loci, therefore,

recruiting com-KRAB fusion to target sites and repressing their expression [84].

Acting at the DNA level, CRISPRi can be a useful tool to target non-coding transcripts including lncRNAs as well as microRNAs and products of RNA polymerase III [85]. The CRISPRi system has a greater target range compared to RNAi and does not compete with endogenous RNA machinery and therefore can complement the current tools at our disposal [85]. Knock-down studies are valuable resources in studying the role of different cellular components including lncRNAs and the CRISPRi holds a great promise for knock-down experiments aimed at revealing the role lncRNAs behind different cancers due to its capability to repress transcription at the DNA level and to repress multiple genes simultaneously [85] (Table 1).

Combined knock-down and localization analysis of non-coding RNAs (c-KLAN)

RNA-FISH is utilized to localize RNA transcripts inside the cell [43], whereas RNA interference is utilized in functional studies [78]. Despite being both highly useful and revealing, these techniques bear disadvantages including off-target silencing [79]. Endoribonuclease-prepared siRNAs have been developed to remove the drawbacks of off-target detection and difficulty of production associated with conventional RNAi [79]. esi-RNA production protocol can be also employed for production of labelled probes against lncRNA targets from cDNA pools [79]. c-KLAN combines RNA interference and localization techniques to shed light on both location and function of lncRNAs, paving the way to their robust characterization. cDNA is first amplified using PCR. Then, *in vitro* transcription in the presence of labelled UTPs can be performed for the generation of RNA FISH probes. *In vitro* transcription followed by RNase III digestion can be implemented for the production of esi-RNAs. c-KLAN combines the power of RNA FISH and RNAi for characterizing lncRNAs through loss-of-function studies and subcellular visualization [86].

c-KLAN has been efficiently utilized to reveal the lncRNAs involved in pluripotency [86]. Pluripotency-associated non-coding transcript 1 has been successfully found to be the major lncRNA involved in this state, using Oct4-GiP cells that express GFP under Oct4 promoter [86].

Antisense oligonucleotides (ASOs)

Despite being highly effective devices in loss-of-function studies, siRNA and shRNAs cannot be used effectively in depleting nuclear lncRNAs, although the RNAi machinery also exists in the nucleus [71, 87, 88].

As effective tools for silencing nuclear lncRNAs, ASOs are modified or unmodified single-stranded deoxyribonucleotides, with the ability to hybridize to their targets among a population of transcripts present in the cell. Knock-down experiments using ASOs are based on the ability of RNase H to degrade the RNA part of a DNA–RNA duplex [89]. Single-stranded antisense oligonucleotides are similar to siRNAs in that they interact with their targets through Watson–Crick base pairing [89]. However, unlike siRNAs, ASOs are labile inside the cell due to sensitivity of single-stranded DNA molecules to nucleases, a problem that led to the development of gapmers [89]. Gapmer oligonucleotides are modified ASOs harbouring 2–5 chemically

modified nucleotides (e.g. locked nucleic acid (LNA) or 2'-O-methoxyethylribose (2'-MOE)) on their termini, intervened by a central 'gap' of unmodified DNA, 8–10 bases in size [89]. The modified nucleotides confer nuclease resistance to gapmers, while the central DNA sequence allows them to bind their targets. However, modified antisense oligonucleotides cannot bind their targets efficiently, a problem that forces the use of a higher concentration of ASOs, leading to more off-target effects [89] (Table 1).

Several groups have used ASOs successfully in loss-of-function studies of nuclear lncRNAs. For instance, ASOs were utilized in an attempt to knock-down the nuclear-retained MALAT1, systemically. Two gapmers against different parts of MALAT1 were used, and systemic knock-downs using ASOs were shown to recapitulate the same results observed in knockout studies. Besides, the same reduction in metastasis happening upon genetic loss of *MALAT1* was also observed in systematic MALAT1 depletion using ASOs [90].

Methods used for mapping lncRNA interactions

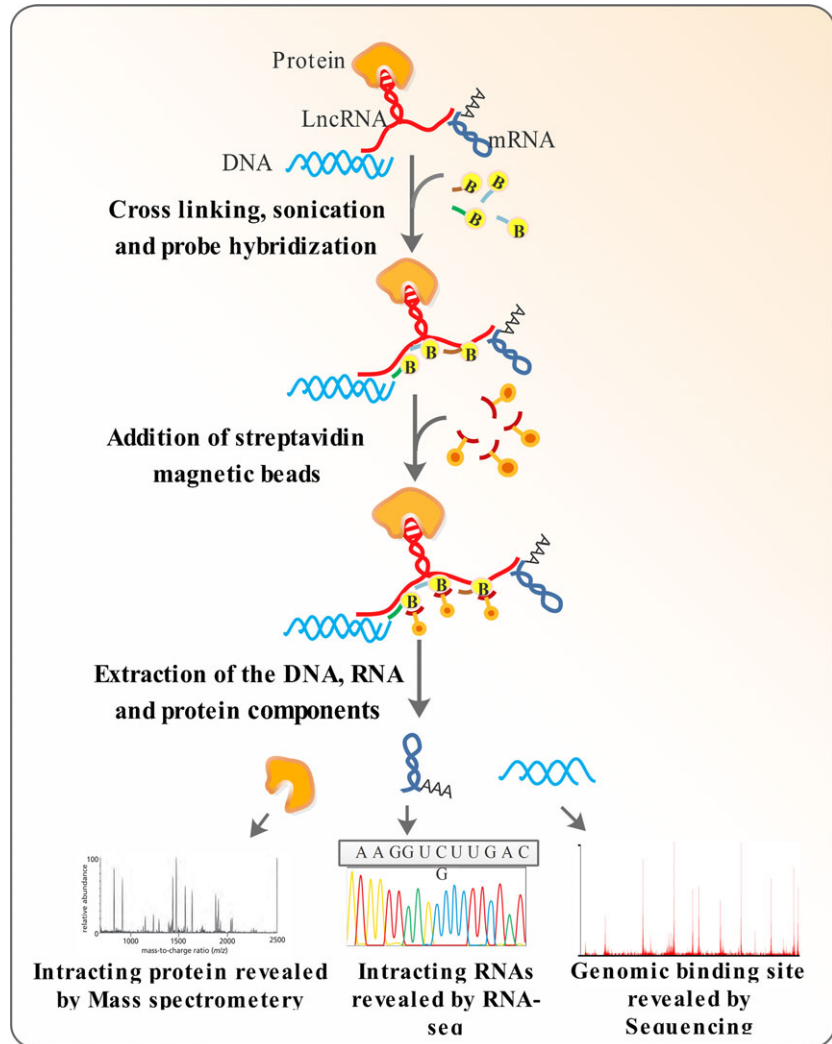
Cell function is governed by a complex network of interactions involving different cellular components [91]. Important among such components are lncRNAs, playing a wide variety of roles through their interaction with other RNA and protein molecules [91]. Any aberration in such interactions can have dire consequences including cancer [91]. Therefore, mapping lncRNA interaction networks is an indispensable part of any research in cancer biology. In this section, we summarize the techniques employed commonly in mapping lncRNA interactions.

Mapping lncRNA–DNA interactions

ChIRP, CHART and RAP as RNA-centric methods to map lncRNA genomic binding sites

Conventional methods of chromatin precipitation commonly make use of antibodies against proteins associated with DNA [92]. ChIRP (chromatin isolation by RNA purification) [93, 94], CHART (capture hybridization analysis of RNA targets) [95] and RAP (RNA antisense purification) [93] are RNA-centric techniques that benefit from probes against RNA molecules to precipitate chromatin to reveal the lncRNA genomic binding sites [96, 97]. The three techniques are similar regarding the overall procedure [96]. After cross-linking the cellular components, the lncRNAs are captured with biotinylated antisense probes. The RNA molecules are recovered using streptavidin-linked magnetic beads, and the precipitate is fractionated into protein, RNA and DNA parts. The RNA and DNA fractions are finally analysed using RT-PCR and high-throughput sequencing, revealing the genome-wide binding sites of the lncRNA of interest (Fig. 4) [96, 97]. The proteins associated to a given lncRNA molecule can be also identified by subjecting the protein fraction recovered to mass spectrometry in ChIRP-MS and RAP-MS [96]. However, the techniques have differences with

Fig. 4 Mapping lncRNA–DNA interactions. In ChIRP, the lncRNA molecules are cross-linked to their interacting partners including DNA, proteins and other RNA molecules. Biotinylated probes of about 20 bp in size are used to tile the entire length of the lncRNA. Streptavidin magnetic beads are added to pull-down the lncRNA molecules and their interacting components by binding to the biotinylated probes. The precipitated material is finally partitioned to DNA, RNA and protein fractions. ChIRP can be coupled to RNA-seq (ChIRP-seq) or mass spectrometry (ChIRP-MS) for analysing the nucleic acid or protein fractions, respectively. (RAP and CHART are similar to ChIRP in their overall procedure except for the probes of different size and different cross-linking agents used in these techniques).



respect to cross-linking mode, the size and design of probes used [96]. ChIRP and RAP make use of probes that cover all the length of lncRNA sequence, therefore, targeting all potential sites unmasked for hybridization [96]. Probes used in ChIRP are shorter (about 20 bp in size) and therefore easier to synthesize and potentially discriminate against off-targets, whereas longer probes costly to synthesize (about 120 bp in size) are used in RAP [96]. CHART is similar to ChIRP but makes use of an RNase H assay to find experimentally the best target site for designing probes (18–28 bp in size) [93]. The cross-linking mode is another point of difference between the three methods, with glutaraldehyde used in ChIRP, formaldehyde employed in CHART and a combination of them utilized in RAP [93, 96]. Longer cross-linking agents such as glutaraldehyde have been proven to be more efficient than the smaller ones such as formaldehyde [93] (Table 1).

ChIRP, which can be considered as an analogue of ChIP, has been used by different research groups hitherto to map the genomic

binding sites of different lncRNAs [93, 94, 96]. ChIRP, RAP and CHART have been employed in a wide variety of studies on lncRNAs involved in cancer (Supplementary Table 1). For instance, Chu and coworkers were the first group that used ChIRP to reveal the genome-wide DNA binding sites of HOTAIR lncRNA in MB-231 breast cancer cells. They found 832 HOTAIR occupancy sites throughout the genome, located on multiple chromosomes and enriched in regions notably annotated as promoters and enhancers. [98].

Mapping lncRNA–protein interactions by protein pull-down

RNA immunoprecipitation (RIP)

RNA immunoprecipitation (RIP) is a technique which can be employed to unravel all lncRNAs that interact with specific proteins of interest [29, 99]. The technique can be implemented with or without cross-linking [100]. UV irradiation can be also

used for detection of direct RNA–protein interactions by RIP [100]. The RNA–protein complexes are co-immunoprecipitated using antibodies against the protein component [100]. The RNA component is then recovered and subjected to sequencing, microarray or RT-PCR [100]. RIP combined with microarray or sequencing is referred to as RIP-chip or RIP-seq, respectively [48, 98, 100].

Like other protein-centric pull-down assays, RIP is dependent on antibodies previously produced against the suspected counterpart proteins [100] (Table 1).

In a study of the role of the lncRNA, UCA1, in breast cancer promotion, RIP was employed to indicate that UCA1 interacts with hnRNP1 and suppresses the p27 protein level by competitive inhibition [47]. Then, they performed RIP after UCA1 expression and showed a reduction in p27 mRNA pull-down level using antibodies against hnRNP1 [47].

To investigate the role of MALAT 1 in splicing of pre-mRNAs, antibodies against SRSF1 splicing factor, and primers specific to MALAT1 were used to reveal a specific interaction between them, using the RIP technique [48]. In addition, RIP has been also utilized to show SChLAP1 but not other cytoplasmic lncRNAs are immunoprecipitated by SWI/SNF complex [35].

Cross-linking and immunoprecipitation (CLIP)

Further refinement of RIP has led to cross-linking and immunoprecipitation (CLIP) technique, which is capable of revealing a more accurate picture of lncRNA–protein interfaces with single nucleotide resolution [101]. In this technique, the lncRNA and protein components are cross-linked; the complex is precipitated, and the RNA counterpart is identified [102]. There are several versions of the CLIP technique, including photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation CLIP (PAR-CLIP) [103] and individual-nucleotide-resolution CLIP (iCLIP) [101]. Both PAR-CLIP and iCLIP are based on the propensity of some cross-linked amino acids to remain associated with the RNA component after cross-linking and include a reverse transcription step in which the precipitated RNA is converted to cDNA [102] (Fig. 5). In PAR-CLIP, the RNA–protein interaction sites are revealed based on nucleotide misincorporation at the sites where such interactions occur [102]. However, detection of nucleotide misincorporation sites requires the process of reverse transcription to go into completion, something that does not occur efficiently as the reverse transcriptase halts before the misincorporation point [104]. To overcome this problem, size-selected cDNAs are circularized and sequenced in iCLIP, and the truncation points in cDNA synthesis reveal the RNA–DNA interfaces [101].

In contrast to other CLIP variants, which makes use of cross-linking with 254 nm UV light, PAR-CLIP takes advantage of 4-thiouridine to facilitate the cross-linking process, allowing it to occur at a higher UV wavelength (365 nm) [102] (Fig. 5).

irCLIP and eCLIP are newly introduced versions of CLIP techniques developed to increase reverse transcription efficiency and decrease RNA loss and sequencing backgrounds [105]. irCLIP makes use of a thermostable reverse transcriptase that functions at 60 °C to

avoid reverse transcription stoppage due to RNA secondary structures [105]. The utilization of CircLigase II in irCLIP has allowed the reverse transcription and circularization steps to be combined as a single step, enabling the method to be performed in less than half the time needed in iCLIP [105]. In eCLIP, a 5' adaptor has been utilized to avoid RNA loss and a 3' adaptor followed by a barcoded used to distinguish between unique reads and PCR duplicates [106].

The advantages and disadvantages of each of the CLIP techniques have been reviewed in (Table 1).

In a genome-wide analysis of lncRNA binding sites on different RBPs, CLIP-seq was employed to show the interaction of Ezh, a PRC2 complex subunit, with 35 lncRNAs including Tsix, Meg3, Rian and Pvt1 [107].

Using PAR-CLIP, TUG1, DLEU2 and GAS5 lncRNAs were also shown to bind at least two of the three RNA-binding proteins HuR, Ago2 and MOV10 at identical binding sites [107].

Moreover, CLIP has been employed to show air interactions with H3K9 histone methyl-transferase G9a [108].

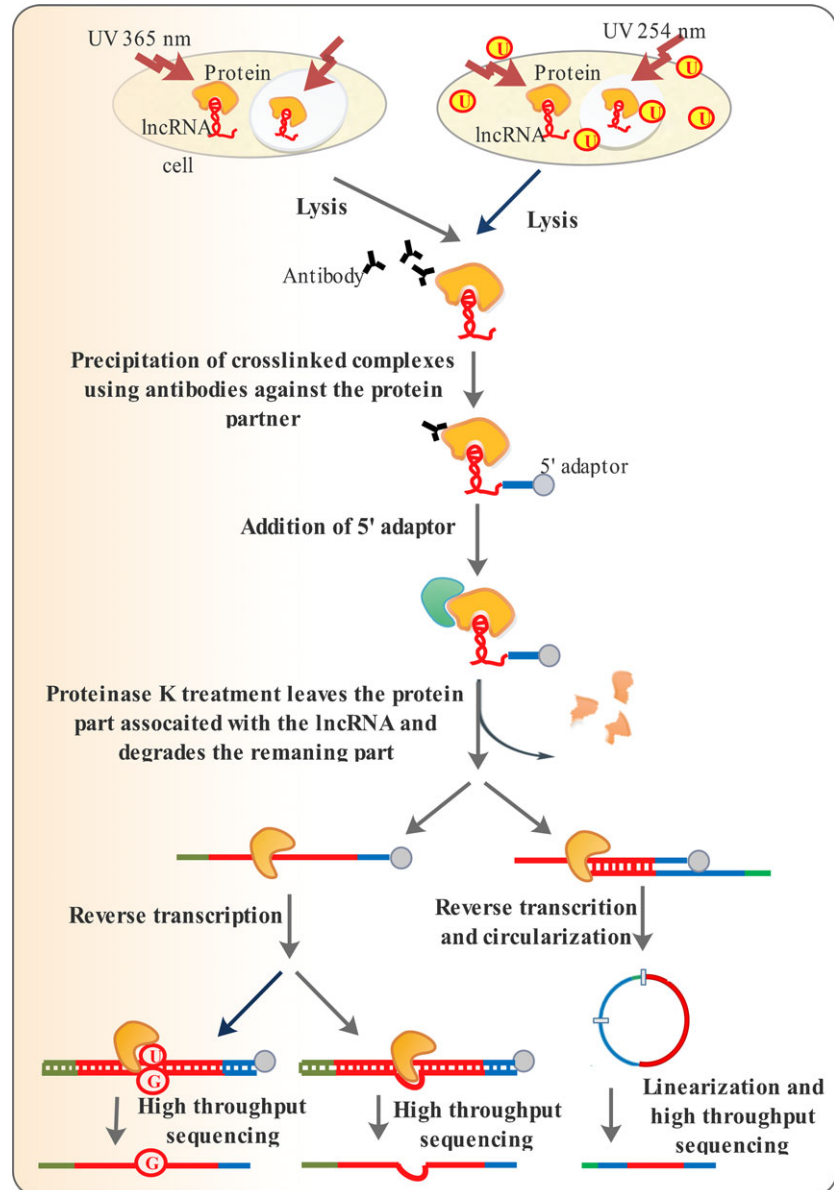
Next-generation sequencing-based methods for global investigation of lncRNAs

Over the past 20 years important efforts have been made allowing the research to be directed at transcriptome-wide analysis of RNAs and changes in their global patterns of expression [40]. Next-generation sequencing technologies lie at the heart of such efforts as they enable bulk RNA analysis through massive cDNA sequencing [109]. Transcriptomics strives to catalogue all RNA species inside the cells, delineate the splicing and other post-transcriptional modifications or editing events the transcripts may undergo and to provide a global picture of the changes in transcriptome during development or disease conditions [109]. Here, we will cover some of the technologies based on RNA-seq developed to tackle such transcriptome-wide questions.

Ligation of interacting RNA followed by high-throughput sequencing (LIGR-seq)

Several methods have been developed for global or single-molecule mapping of local primary and secondary structures of RNA molecules [56, 62]. However, these methods are not efficient in revealing the long-range tertiary structures and intermolecular RNA–RNA interactions [110]. LIGR-seq makes use of the modified psoralen derivative 4'-aminomethyltrioxalen (AMT), which intercalates into double-stranded RNA and joins them together upon irradiation at 365 nm [110]. Ensuuing cell lysis and partial digestion of single-stranded RNA, the enriched duplexes are ligated to each other by circRNA ligase. The cross-links are reversed by irradiation at 254 nm, and the RNA species are subjected to high-throughput sequencing to reveal the chimeras. To correct for background-ATM samples are used, and

Fig. 5 Mapping lncRNA–protein interactions by protein pull-down. High-throughput sequencing of RNA isolated by ultraviolet (UV) cross-linking and immunoprecipitation (HITS-CLIP), iCLIP and PAR-CLIP all make use of specific antibodies for affinity purification a protein of interest. To co-purify the associated RNA counterparts HITS-CLIP and iCLIP use irradiation at 254 nm, while PAR-CLIP employs irradiation at 365 nm using a photolabile 4-thiouridine. Treatment with proteinase K degrades all protein parts but leaves a tiny RNA-linked region. Following reverse transcription, the protein-linked RNA site leads to insertions in HITS-CLIP, results in truncation of cDNAs in iCLIP and causes base transition in PAR-CLIP because of 4-thiouridine analogue used to enhance cross-linking. Deep sequencing of the cDNAs following PCR allows identification of the protein-RNA interaction sites with single-base resolution.



unligated samples are utilized to normalize for false chimeric RNAs [110] (Fig. 6A).

Previous attempts have been made to map miRNA–mRNA interactions in RNA induced silencing complexes [111] or snoRNA–rRNA interactions in spliceosome [112]. However, all these methods require prior knowledge of the interacting protein partner for affinity purification and may lead to a high false-positive detections due to *in vitro* cross-linking [110].

Some lncRNAs are shown to function through interaction with other RNA molecules [113]. A global snapshot of the RNA–RNA interactome achieved by LIGR-seq revealed that at least one ncRNA is involved in the majority of stable interactions [110]. Therefore, LIGR-seq have the ability to reveal global maps of the

lncRNA–RNA interactions, and shedding light on how perturbation of such interaction may lead to disease states including cancer (Table 1).

Inosine chemical erasing sequencing (ICE-seq)

Deamination of Adenosine to inosine is a common RNA modification, which plays important regulatory roles and has been implicated in several disorders including cancer [114]. lncRNAs have been shown to undergo A-to-I editing with important effects on their secondary structures and interaction with other RNA species [13]. As deregulation of such editing events may perturb lncRNA

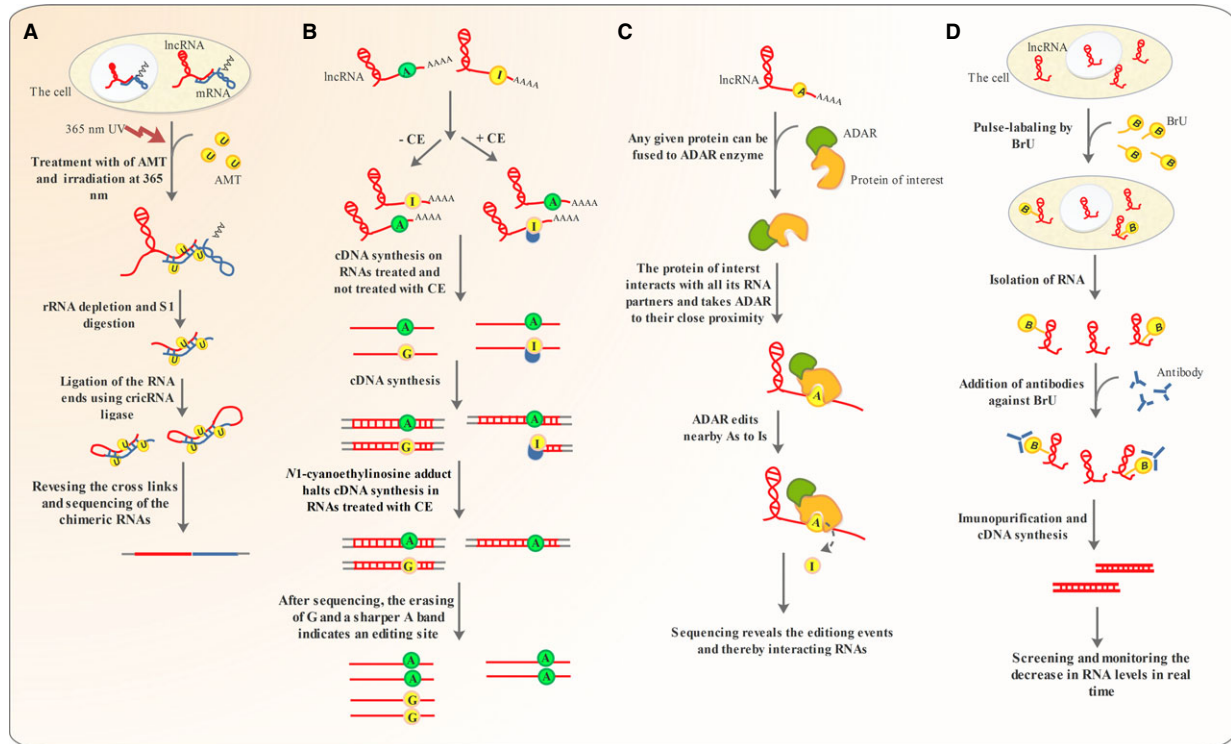


Fig. 6 NGS-based technologies used for global investigation of lncRNAs. **(A)** LIGR-seq makes use of an intercalating agent and UV irradiation to join interacting RNAs. The RNA ends are partially digested and ligated using cricRNA ligase. The cross-linking is reversed, and the chimeric RNAs are sequenced. **(B)** ICE-seq exploits cyanoethylation with Acrylonitrile to convert Inosine to *N1*-cyanoethylinosine, which halts cDNA synthesis. Following sequencing, the replacement of two A and G bands in the absence of CE with a sharper A band under CE treatment indicates a modification site. **(C)** In TRIBE-seq and PUP-2 RNA tagging, the protein of interest is fused to ADAR and PUP-2 enzymes. ADAR converts adenosine to Inosine and PUP-2 uridylyates any RNA molecules that interact with the protein of interest. If all the RNA species inside the cell are sequenced, the A-to-I modified and uridylyated RNAs indicate the ribonucleic acid partners of the protein of interest. **(D)** In BRIC-seq, the cells are treated with 5'-bromo-uridine, and the decrease in labelled RNA levels is screened in real time.

interaction networks and lead to disease conditions including cancer, a complete picture of normal or abnormal A-to-I editing is of paramount importance [115]. Conventionally, sequencing and aligning cDNA molecules to their genomic loci can reveal such editing events [116]. As Inosine base pairs with Cytosine, anywhere A has been replaced by G in the cDNA sequence information indicates an A-to-I conversion [116]. However, this approach has several limitations including the necessity to exclude individual SNP variations and the inability to detect true I-to-G modifications from Gs occurring due to sequencing errors [116].

ICE-seq makes use of Inosine cyanoethylation with Acrylonitrile. The resulting *N1*-cyanoethylinosine adduct halts the cDNA extension process and allows identification of A-to-I editing points. The procedure is carried out with or without CE. In CE⁻ conditions both I and A positions recognized in sequencing. However, under CE conditions the RNAs undergoing A-to-I modifications stop to extend beyond the edition point. Therefore, the G band (I is read G in sequencing process) is erased, and a sharper A band is detected and that point is identified as a true modification site [116] (Fig. 6B). Although not yet applied to

lncRNAs hitherto, ICE-seq has the potential to reveal the deregulations that may occur in A-to-I editing of lncRNAs and how it may contribute to cancer (Table 1).

Global mapping of lncRNA–protein interactions based on RNA modification and RNA-seq

Despite being effective in determining RNA–protein interactions, RIP and CLIP can only unravel protein–RNA interactions when specific antibodies against the protein of interest are already available [101]. Two methods have been recently developed, which allow mapping of protein–RNA interactions without the need to antibody affinity purification [117, 118]. In one of the methods referred to as RNA tagging, global cataloguing of RNA–protein interactions is accomplished based on the ability *Caenorhabditis elegans* poly (U) polymerase (PUP-2) [117]. PUP-2 does not possess any RNA-binding domain and therefore cannot uridylylate RNA independently. Any protein intended to be dissected regarding its universal interactions with RNA species inside the cell has to be fused with PUP-2 protein. Then upon its interaction

with any counterpart RNA, PUP-2 uridylylates the cognate RNA. The uridine-tagged RNAs are finally enriched using U-select primers and subjected to RNA sequencing, unravelling a global map of all RNAs interacting with the protein of interest [117]. However, Uridylation using PUP-2 can only reveal the interacting RNAs but not the exact RNA–protein interface [118] (Table 1). In another technique named targets of RNA-binding proteins identified by editing (TRIBE), the protein of interest is fused to *Drosophila* RNA-editing enzyme ADAR which converts adenosine to inosine [118]. Therefore, any RNA that interacts with the protein of interest is subjected to A-to-I editing and reveals the approximate interaction interface after next-generation sequencing (Fig. 6C). This technique, however, is not capable revealing the completely exact interaction site as in CLIP but circumvents the necessity to having a specific antibody against the protein whose interaction network is to be revealed [118]. Unlike CLIP which is commonly carried out on tissues, TRIBE-seq can be applied to a limited number of cells and be used to analyse RNA–protein interactions in homogenous cell populations (Table 1) [118].

These techniques have the potential to be extended to lncRNA research, shedding light on all the interactions a given protein of interest may have with the population of lncRNAs inside the cell and depicting probable interruptions in such interaction networks that may lead to cancer.

Global run-on assay sequencing (Gro-seq)

A global understanding of genomic regions undergoing active transcription can provide valuable information for discovery and functional analysis of lncRNAs. Global run-on assay sequencing (GRO-seq) is a technique based on NGS technologies that provides valuable insights into of the location, orientation and density transcripts undergoing active transcription by RNA polymerase II. In this technique, the RNA polymerase is allowed to function in the presence the UTP analogue Bromodeoxyuridine (BrdU). The nascent RNAs are captured using anti-BrdUTP antibodies and subjected to deep sequencing. The resulting sequence reads are aligned across the genome to map the genomic positions undergoing active transcription [119].

The technique, however, is limited to cell culture and other artificial systems due to its dependence on incorporation of Bromodeoxyuridine (BrdU). New initiation events that may happen and introduction of artefacts during preparation of nuclei are among other demerits of the technique [120] (Table 1).

Using this method, several studies have indicated that transcription by RNA pol II often occurs divergently and leads to production of lncRNAs [119, 121].

5'-bromo-uridine immunoprecipitation chase–deep sequencing (BRIC-seq)

The result of many studies has shown that there is an intimate relationship between mRNA half-life and its function, which implies an understanding the half-life of lncRNAs can provide useful information about their functions [122, 123]. Tani *et al.* [122], developed a

technique named 5'-bromo-uridine immunoprecipitation chase–deep sequencing analysis (BRIC-seq), to determine precise RNA half-life inside the cell.

The method involves pulse labelling of cells with 5'-bromo-uridine and then quantifying the decrease in labelled RNA levels that occurs over time making use of deep sequencing. The result of this study indicated that RNA molecules with half-lives <4 hrs include the majority of regulatory and lncRNAs [122] (Fig. 6D) (Table 1).

Fluorescent *in situ* sequencing (FISSEQ)

RNA sequencing provides valuable insights into global quantitative and qualitative changes in transcripts but does not picture the spatial distribution of transcripts [124]. On the other hand, *in situ* hybridization allows simultaneous visualization of cellular location of a small number of transcripts [124].

FISSEQ extends the applicability of RNA-seq beyond *in vitro* quantitative measurement of transcripts to *in vivo* and allows spatial organization of gene expression to be revealed globally [125]. In FISSEQ the cells are fixed, cDNA synthesis is carried out, and the cDNA molecules are circularized using CircLigase. To prevent diffusion, amino-lyl-dUTP is used during reverse transcription and primary amines are joined together using PEG. Finally, SOLiD sequencing chemistry is employed to provide a quantitative spatial vista of gene expression [125].

FISSEQ has the potential to reveal the global spatial changes in lncRNA expression in disease conditions including cancer. However, the technique needs to be refined regarding its read length, sequencing depth and coverage and library preparation [125].

Conclusion

The complexity of cells is better appreciated with this emerging view that most of the genome is functional and ubiquitously transcribed into a repertoire of coding and non-coding RNAs. lncRNAs are highlighted in the new transcriptional landscape replete with transcripts of unknown function. They play vital roles in various cellular processes and disease conditions including cancer.. Therefore, considering their important regulatory roles, lncRNAs are among the missing parts of the puzzle in the long-lasting fight against cancer and have to be scrutinized comprehensively so as to enhance our current understanding of this disease. The boundaries of our capability to understand how lncRNAs can contribute to cancer prevention or progression are contingent on power of the techniques available.

High-throughput sequencing has accelerated the pace at which new lncRNAs are being discovered and allows global analysis of lncRNAs. However, our ability to characterize the expanding catalogue of unknown lncRNAs is also dependent on other techniques enabling us to map lncRNA interactions, their structure and function. For instance, the subcellular location of lncRNAs can be determined using FISH and further clarified by the combination of FISH and esi-RNA technology (cKLAN). Compared to RNA FISH which can localize few transcripts in a single reaction, FISSEQ allows the research to be

directed at quantitative global and spatial analyses of lncRNA expression. New technologies developed based on CRISPR-Cas system can now be a complement to the current RNA visualization techniques as they may be better applied to non-coding transcripts. Structural studies of lncRNAs have become more feasible using SHAPE. But SHAPE is limited to single-molecule analyses carried out *in vitro*. To overcome the single-molecule limitation of SHAPE, global structural analysis of the complete RNA reservoir or structurome has been now possible by PARS and Fraug-seq. icSHAPE has extended the applicability of lncRNA structural studies by SHAPE to the realm of *in vivo*. lncRNA interaction networks have been clearly delineated using RNA-centric approaches comprising ChIRP, CHART, RAP and protein-centric methods including CLIP. dChIRP has enabled domain by domain interrogation of lncRNA interactions. CLIP has been refined regarding its throughput and resolution to iCLIP and PAR-CLIP. Recently, further refinement of CLIP has led to irCLIP and eCLIP which allow more accurate and rapid determination of lncRNA–protein interactions. RNAi has allowed the knock-down experiments employed in functional studies of lncRNAs and CRISPRi can increase the precision with which we can suppress lncRNA expression in reverse genetic investigations. TRIBE-seq enables global mapping of lncRNA–RNA interaction network, without any pre-existing antibody for affinity purification required in CLIP techniques. ICE-seq allows

the research to be directed at depicting a global picture of lncRNA epitranscriptome and is not confounded by individual variations due to SNPs. The half-life of lncRNAs can be screened in real time using BRIC-seq. GRO-seq provides a picture of transcriptionally active genomic regions as it can capture nascent RNA molecules.

As aberrations in abundance, subcellular localization and interaction networks of lncRNAs are among key contributors to cancer progression, the technological toolkit described throughout this review is central to a deeper understanding of lncRNAs not yet characterized, revealing their likely widespread roles in cancer disease and opening new perspective for therapy.

Acknowledgements

This work is based on thesis in medical biotechnology by Saeede Salehi (Project NO. 92-6889) supported by Shiraz University of Medical Sciences, Shiraz, Iran.

Conflicts of interest

The authors confirm that there are no conflicts of interest.

References

1. **Fatica A, Bozzoni I.** Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet.* 2014; 15: 7–21.
2. **Derrien T, Johnson R, Bussotti G, et al.** The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 2012; 22: 1775–89.
3. **Guttman M, Amit I, Garber M, et al.** Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature.* 2009; 458: 223–7.
4. **Benayoun BA, Pollina EA, Ucar D, et al.** H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell.* 2014; 158: 673–88.
5. **Guttman M, Rinn JL.** Modular regulatory principles of large non-coding RNAs. *Nature.* 2012; 482: 339–46.
6. **Yoon J-H, Abdelmohsen K, Gorospe M.** Posttranscriptional gene regulation by long noncoding RNA. *J Mol Biol.* 2013; 425: 3723–30.
7. **Fritah S, Niclou SP, Azuaje F.** Databases for lncRNAs: a comparative evaluation of emerging tools. *RNA.* 2014; 20: 1655–65.
8. **Mallory AC, Shkumatava A.** lncRNAs in vertebrates: advances and challenges. *Biochimie.* 2015; 117: 3–14.
9. **Ernst C, Morton CC.** Identification and function of long non-coding RNA. *Front Cell Neurosci.* 2013; 7: 168.
10. **Wang X, He C.** Dynamic RNA modifications in posttranscriptional regulation. *Mol Cell.* 2014; 56: 5–12.
11. **Zhang H, Chen Z, Wang X, et al.** Long non-coding RNA: a new player in cancer. *J Hematol Oncol.* 2013; 6: 1.
12. **Li X, Xiong X, Wang K, et al.** Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. *Nat Chem Biol.* 2016; 12: 311–316.
13. **Gong J, Liu C, Liu W, et al.** LNCediting: a database for functional effects of RNA editing in lncRNAs. *Nucleic Acids Res.* 2017; 45 (D1): D79–D84.
14. **Bhan A, Mandal SS.** Long noncoding RNAs: emerging stars in gene regulation, epigenetics and human disease. *ChemMedChem.* 2014; 9: 1932–56.
15. **Wang KC, Chang HY.** Molecular mechanisms of long noncoding RNAs. *Mol Cell.* 2011; 43: 904–14.
16. **Katayama S, Tomaru Y, Kasukawa T, et al.** Antisense transcription in the mammalian transcriptome. *Science.* 2005; 309: 1564–6.
17. **Holdt LM, Stahringer A, Sass K, et al.** Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun.* 2016; 7: 12429.
18. **Hansen TB, Jensen TI, Clausen BH, et al.** Natural RNA circles function as efficient microRNA sponges. *Nature.* 2013; 495: 384–8.
19. **Li W, Notani D, Rosenfeld MG.** Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet.* 2016; 17: 207–23.
20. **Pefanis E, Wang J, Rothschild G, et al.** RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell.* 2015; 161: 774–89.
21. **Ilott NE, Heward JA, Roux B, et al.** Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. *Nat Commun.* 2014; 5: 3979.
22. **Poliseno L, Salmena L, Zhang J, et al.** A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature.* 2010; 465: 1033–8.
23. **Quinn JJ, Chang HY.** Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet.* 2016; 17: 47–62.
24. **Tsai M-C, Manor O, Wan Y, et al.** Long noncoding RNA as modular scaffold of histone modification complexes. *Science.* 2010; 329: 689–93.

25. **Huang G, Wu X, Li S, et al.** The long non-coding RNA CASC2 functions as a competing endogenous RNA by sponging miR-18a in colorectal cancer. *Sci Rep.* 2016; 6: 26524.
26. **Bian Z, Jin L, Zhang J, et al.** LncRNA—UCA1 enhances cell proliferation and 5-fluorouracil resistance in colorectal cancer by inhibiting miR-204-5p. *Sci Rep.* 2016; 6: 23892.
27. **Gong C, Maquat LE.** lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3 [prime] UTRs via Alu elements. *Nature.* 2011; 470: 284–8.
28. **Böhmendorfer G, Wierzbicki AT.** Control of chromatin structure by long noncoding RNA. *Trends Cell Biol.* 2015; 25: 623–32.
29. **Zhao J, Sun BK, Erwin JA, et al.** Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science.* 2008; 322: 750–6.
30. **Gonzalez I, Munita R, Agirre E, et al.** A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nat Struct Mol Biol.* 2015; 22: 370–6.
31. **Hung T, Wang Y, Lin MF, et al.** Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet.* 2011; 43: 621–9.
32. **Flynn RA, Chang HY.** Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell.* 2014; 14: 752–61.
33. **Kotzin JJ, Spencer SP, McCright SJ, et al.** The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. *Nature.* 2016; 537: 239–43.
34. **Huarte M.** The emerging role of lncRNAs in cancer. *Nat Med.* 2015; 21: 1253–61.
35. **Prensner JR, Iyer MK, Sahu A, et al.** The long noncoding RNA SchLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet.* 2013; 45: 1392–8.
36. **Gupta RA, Shah N, Wang KC, et al.** Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature.* 2010; 464: 1071–6.
37. **Liu Q, Huang J, Zhou N, et al.** LncRNA loc285194 is a p53-regulated tumor suppressor. *Nucleic Acids Res.* 2013; 41: 4976–87.
38. **Prensner JR, Chinnaiyan AM.** The emergence of lncRNAs in cancer biology. *Cancer Discov.* 2011; 1: 391–407.
39. **Lee GL, Dobi A, Srivastava S.** Prostate cancer: diagnostic performance of the PCA3 urine test. *Nat Rev Urol.* 2011; 8: 123–4.
40. **Ozsolak F, Milos PM.** RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet.* 2011; 12: 87–98.
41. **You M, Jaffrey SR.** Structure and mechanism of RNA mimics of green fluorescent protein. *Annu Rev Biophys.* 2015; 44: 187–206.
42. **Paige JS, Wu KY, Jaffrey SR.** RNA mimics of green fluorescent protein. *Science.* 2011; 333: 642–6.
43. **Jansen R-P.** mRNA localization: message on the move. *Nat Rev Mol Cell Biol.* 2001; 2: 247–56.
44. **Mao YS, Sunwoo H, Zhang B, et al.** Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat Cell Biol.* 2011; 13: 95–101.
45. **Orjalo A Jr, Johansson HE, Ruth JL.** Stellaris™ fluorescence *in situ* hybridization (FISH) probes: a powerful tool for mRNA detection. *Nat Methods.* 2011; 1 8: 1548–7091.
46. **Cabili MN, Dunagin MC, McClanahan PD, et al.** Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* 2015; 16: 20.
47. **Huang J, Zhou N, Watabe K, et al.** Long non-coding RNA UCA1 promotes breast tumor growth by suppression of p27 (Kip1). *Cell Death Dis.* 2014; 5: e1008.
48. **Tripathi V, Ellis JD, Shen Z, et al.** The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell.* 2010; 39: 925–38.
49. **Bertrand E, Chartrand P, Schaefer M, et al.** Localization of ASH1 mRNA particles in living yeast. *Mol Cell.* 1998; 2: 437–45.
50. **Jinek M, Chylinski K, Fontana I, et al.** A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012; 337: 816–21.
51. **Wiedenheft B, Sternberg SH, Doudna JA.** RNA-guided genetic silencing systems in bacteria and archaea. *Nature.* 2012; 482: 331–8.
52. **O'Connell MR, Oakes BL, Sternberg SH, et al.** Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature.* 2014; 516: 263–6.
53. **Nelles DA, Fang MY, O'Connell MR, et al.** Programmable RNA tracking in live cells with CRISPR/Cas9. *Cell.* 2016; 165: 488–96.
54. **Nelles DA, Fang MY, Aigner S, et al.** Applications of Cas9 as an RNA-programmed RNA-binding protein. *BioEssays.* 2015; 37: 732–9.
55. **Lu Z, Chang HY.** Decoding the RNA structure. *Curr Opin Struct Biol.* 2016; 36: 142–8.
56. **Merino EJ, Wilkinson KA, Coughlan JL, et al.** RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc.* 2005; 127: 4223–31.
57. **Novikova IV, Dharap A, Hennesly SP, et al.** 3S: shotgun secondary structure determination of long non-coding RNAs. *Methods.* 2013; 63: 170–7.
58. **Spitale RC, Crisalli P, Flynn RA, et al.** RNA SHAPE analysis in living cells. *Nat Chem Biol.* 2013; 9: 18–20.
59. **Somarowthu S, Legiewicz M, Chillón I, et al.** HOTAIR forms an intricate and modular secondary structure. *Mol Cell.* 2015; 58: 353–61.
60. **Kretz M, Siprashvili Z, Chu C, et al.** Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature.* 2013; 493: 231–5.
61. **Novikova IV, Hennesly SP, Sanbonmatsu KY.** Tackling structures of long noncoding RNAs. *Int J Mol Sci.* 2013; 14: 23672–84.
62. **Kertesz M, Wan Y, Mazor E, et al.** Genome-wide measurement of RNA secondary structure in yeast. *Nature.* 2010; 467: 103–7.
63. **Underwood JG, Uzilov AV, Katzman S, et al.** FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat Methods.* 2010; 7: 995–1001.
64. **Mauger DM, Weeks KM.** Toward global RNA structure analysis. *Nat Biotechnol.* 2010; 28: 1178–9.
65. **Lau E.** Non-coding RNA: zooming in on lncRNA functions. *Nat Rev Genet.* 2014; 15: 574–5.
66. **Quinn JJ, Ilik IA, Qu K, et al.** Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat Biotechnol.* 2014; 32: 933–40.
67. **Quinn JJ, Ilik IA, Qu K, et al.** Domain ChIRP reveals the modularity of long non-coding RNA architecture, chromatin interactions, and function. *Nat Biotechnol.* 2014; 32: 933.
68. **Nakagawa S.** Lessons from reverse-genetic studies of lncRNAs. *Biochim Biophys Acta.* 2016; 1859: 177–83.
69. **Gaj T, Gersbach CA, Barbas CF.** ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 2013; 31: 397–405.

70. **Ho T-T, Zhou N, Huang J, et al.** Targeting non-coding RNAs with the CRISPR/Cas9 system in human cell lines. *Nucleic Acids Res.* 2015; 43(3): e17.
71. **Goyal A, Myacheva K, Groß M, et al.** Challenges of CRISPR/Cas9 applications for long non-coding RNA genes. *Nucleic Acids Res.* 2017; 45(3): e12.
72. **Sander JD, Joung JK.** CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol.* 2014; 32: 347–55.
73. **Yang L, Lin C, Jin C, et al.** lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature.* 2013; 500: 598–602.
74. **Prensner JR, Sahu A, Iyer MK, et al.** The lncRNAs PCGEM1 and PRNCR1 are not implicated in castration resistant prostate cancer. *Oncotarget.* 2014; 5: 1434–8.
75. **Aparicio-Prat E, Arnan C, Sala I, et al.** DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. *BMC Genom.* 2015; 16: 846.
76. **Zhu S, Li W, Liu J, et al.** Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat Biotechnol.* 2016; 34(12): 1279–86.
77. **Czech B, Hannon GJ.** Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet.* 2011; 12: 19–31.
78. **Castel SE, Martienssen RA.** RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet.* 2013; 14: 100–12.
79. **Buchholz F, Kittler R, Slabicki M, et al.** Enzymatically prepared RNAi libraries. *Nat Methods.* 2006; 3: 696–700.
80. **Kittler R, Surendranath V, Heninger A-K, et al.** Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nat Methods.* 2007; 4: 337–44.
81. **Kittler R, Putz G, Pelletier L, et al.** An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature.* 2004; 432: 1036–40.
82. **Hüttenhofer A, Schattner P, Hall J, et al.** Whither RNAi. *Nat Cell Biol.* 2003; 5: 489–90.
83. **Watts JK, Corey DR.** Silencing disease genes in the laboratory and the clinic. *J Pathol.* 2012; 226: 365–79.
84. **Qi LS, Larson MH, Gilbert LA, et al.** Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* 2013; 152: 1173–83.
85. **Gilbert LA, Larson MH, Morsut L, et al.** CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* 2013; 154: 442–51.
86. **Chakraborty D, Kappei D, Theis M, et al.** Combined RNAi and localization for functionally dissecting long noncoding RNAs. *Nat Methods.* 2012; 9: 360–2.
87. **Dias N, Stein C.** Antisense oligonucleotides: basic concepts and mechanisms. *Mol Cancer Ther.* 2002; 1: 347–55.
88. **Hall J.** Unravelling the general properties of siRNAs: strength in numbers and lessons from the past. *Nat Rev Genet.* 2004; 5: 552–7.
89. **Lennox KA, Behlke MA.** Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. *Nucleic Acids Res.* 2016; 44(2): 863–77.
90. **Arun G, Diermeier S, Akerman M, et al.** Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev.* 2016; 30: 34–51.
91. **Barabasi A-L, Oltvai ZN.** Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004; 5: 101–13.
92. **Milne TA, Zhao K, Hess JL.** Chromatin immunoprecipitation (ChIP) for analysis of histone modifications and chromatin-associated proteins. *Methods Mol Biol.* 2009; 538: 409–23.
93. **Vance KW, Ponting CP.** Transcriptional regulatory functions of nuclear long non-coding RNAs. *Trends Genet.* 2014; 30: 348–55.
94. **Chu C, Zhang QC, Da Rocha ST, et al.** Systematic discovery of Xist RNA binding proteins. *Cell.* 2015; 161: 404–16.
95. **West JA, Davis CP, Sunwoo H, et al.** The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell.* 2014; 55: 791–802.
96. **Roth A, Diederichs S.** Molecular biology: Rap and chirp about X inactivation. *Nature.* 2015; 521: 170–1.
97. **McHugh CA, Chen C-K, Chow A, et al.** The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature.* 2015; 521: 232–6.
98. **Chu C, Spitale RC, Chang HY.** Technologies to probe functions and mechanisms of long noncoding RNAs. *Nat Struct Mol Biol.* 2015; 22: 29–35.
99. **Huarte M, Guttman M, Feldser D, et al.** A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell.* 2010; 142: 409–19.
100. **Riley KJ, Steitz JA.** The “Observer Effect” in genome-wide surveys of protein-RNA interactions. *Mol Cell.* 2013; 49: 601–4.
101. **Huppertz I, Attig J, D’Ambrogio A, et al.** iCLIP: protein–RNA interactions at nucleotide resolution. *Methods.* 2014; 65: 274–87.
102. **König J, Zarnack K, Luscombe NM, et al.** Protein–RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet.* 2012; 13: 77–83.
103. **Yoon J-H, De S, Srikantan S, et al.** PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nat Commun.* 2014; 5: 5248.
104. **Sugimoto Y, König J, Hussain S, et al.** Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.* 2012; 13: 1.
105. **Zarnegar BJ, Flynn RA, Shen Y, et al.** irCLIP platform for efficient characterization of protein–RNA interactions. *Nat Methods.* 2016; 13: 489–92.
106. **Van Nostrand EL, Pratt GA, Shishkin AA, et al.** Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods.* 2016; 13: 508–14.
107. **Li J-H, Liu S, Zheng L-L, et al.** Discovery of protein–lncRNA interactions by integrating large-scale CLIP-Seq and RNA-Seq datasets. *Front Bioeng Biotechnol.* 2014; 2: 88.
108. **Nagano T, Mitchell JA, Sanz LA, et al.** The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science.* 2008; 322: 1717–20.
109. **Wang Z, Gerstein M, Snyder M.** RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009; 10: 57–63.
110. **Sharma E, Sterne-Weiler T, O’Hanlon D, et al.** Global mapping of human RNA-RNA interactions. *Mol Cell.* 2016; 62: 618–26.
111. **Grosswendt S, Filipchuk A, Manzano M, et al.** Unambiguous identification of miRNA: target site interactions by different types of ligation reactions. *Mol Cell.* 2014; 54: 1042–54.
112. **Kudla G, Granneman S, Hahn D, et al.** Cross-linking, ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast. *Proc Natl Acad Sci.* 2011; 108: 10010–5.
113. **Geisler S, Collier J.** RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol.* 2013; 14: 699–712.
114. **Bass BL.** RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem.* 2002; 71: 817.
115. **Paz N, Levanon EY, Amariglio N, et al.** Altered adenosine-to-inosine RNA editing

- in human cancer. *Genome Res.* 2007; 17: 1586–95.
116. **Sakurai M, Yano T, Kawabata H, et al.** Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. *Nat Chem Biol.* 2010; 6: 733–40.
 117. **Lapointe CP, Wilinski D, Saunders HA, et al.** Protein-RNA networks revealed through covalent RNA marks. *Nat Methods.* 2015; 12(12): 1163–70.
 118. **McMahon AC, Rahman R, Jin H, et al.** TRIBE: hijacking an RNA-editing enzyme to identify cell-specific targets of RNA-binding proteins. *Cell.* 2016; 165: 742–53.
 119. **Core LJ, Waterfall JJ, Lis JT.** Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science.* 2008; 322: 1845–8.
 120. **Adelman K, Lis JT.** Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet.* 2012; 13: 720–31.
 121. **Heinz S, Romanoski C, Benner C, et al.** Effect of natural genetic variation on enhancer selection and function. *Nature.* 2013; 503: 487–92.
 122. **Tani H, Mizutani R, Salam KA, et al.** Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* 2012; 22: 947–56.
 123. **Imamachi N, Tani H, Mizutani R, et al.** BRIC-seq: a genome-wide approach for determining RNA stability in mammalian cells. *Methods.* 2014; 67: 55–63.
 124. **Lee JH, Daugharthy ER, Scheiman J, et al.** Fluorescent *in situ* sequencing (FIS-SEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc.* 2015; 10: 442–58.
 125. **Lee JH, Daugharthy ER, Scheiman J, et al.** Highly multiplexed subcellular RNA sequencing *in situ*. *Science.* 2014; 343: 1360–3.
 126. **Fatima R, Akhade VS, Pal D, et al.** Long noncoding RNAs in development and cancer: potential biomarkers and therapeutic targets. *Mol Cell Ther.* 2015; 3: 1.
 127. **Rinn JL, Kertesz M, Wang JK, et al.** Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* 2007; 129: 1311–23.
 128. **Hajjari M, Salavaty A.** HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer Biol Med.* 2015; 12: 1–9.
 129. **Sahu A, Singhal U, Chinnaiyan AM.** Long noncoding RNAs in cancer: from function to translation. *Trends cancer.* 2015; 1: 93–109.
 130. **Lee B, Mazar J, Aftab MN, et al.** Long non-coding RNAs as putative biomarkers for prostate cancer detection. *J Mol Diagn.* 2014; 16: 615–26.
 131. **Salameh A, Lee AK, Cardó-Vila M, et al.** PRUNE2 is a human prostate cancer suppressor regulated by the intronic long non-coding RNA PCA3. *Proc Natl Acad Sci.* 2015; 112: 8403–8.
 132. **Wang J, Liu X, Wu H, et al.** CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res.* 2010; 38: 5366–83.
 133. **Ling H, Vincent K, Pichler M, et al.** Junk DNA and the long non-coding RNA twist in cancer genetics. *Oncogene.* 2015; 34: 5003–11.
 134. **Du Y, Kong G, You X, et al.** Elevation of highly up-regulated in liver cancer (HULC) by hepatitis B virus X protein promotes hepatoma cell proliferation *via* down-regulating p18. *J Biol Chem.* 2012; 287: 26302–11.
 135. **Clemson CM, Hutchinson JN, Sara SA, et al.** An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell.* 2009; 33: 717–26.
 136. **Chakravarty D, Sboner A, Nair SS, et al.** The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat Commun.* 2014; 5: 5383.
 137. **Adriaens C, Standaert L, Barra J, et al.** p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat Med.* 2016; 22(8): 861–8.
 138. **Prensner JR, Iyer MK, Balbin OA, et al.** Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol.* 2011; 29: 742–9.
 139. **Pickard M, Mourtada-Maarabouni M, Williams G.** Long non-coding RNA GAS5 regulates apoptosis in prostate cancer cell lines. *Biochim Biophys Acta.* 2013; 1832: 1613–23.
 140. **Renganathan A, Kresoja-Rakic J, Echeverry N, et al.** GAS5 long non-coding RNA in malignant pleural mesothelioma. *Mol Cancer.* 2014; 13: 1.
 141. **Mourtada-Maarabouni M, Pickard M, Hedge V, et al.** GAS5, a non-protein-coding RNA, controls apoptosis and is down-regulated in breast cancer. *Oncogene.* 2009; 28: 195–208.
 142. **Liu Y, Zhao J, Zhang W, et al.** lncRNA GAS5 enhances G1 cell cycle arrest *via* binding to YBX1 to regulate p21 expression in stomach cancer. *Sci Rep.* 2015; 5: 10159.
 143. **Xing Z, Lin A, Li C, et al.** lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell.* 2014; 159: 1110–25.
 144. **Li Z, Zhao X, Zhou Y, et al.** The long non-coding RNA HOTTIP promotes progression and gemcitabine resistance by regulating HOXA13 in pancreatic cancer. *J Transl Med.* 2015; 13: 1.