Briefings in Bioinformatics, 2015, 1–11

doi: 10.1093/bib/bbv031
Paper

Revealing protein–lncRNA interaction

Fabrizio Ferrè, Alessio Colantoni and Manuela Helmer-Citterich

Corresponding author. Fabrizio Ferrè, University of Bologna Alma Mater, Department of Pharmacy and Biotechnology (FaBiT), Via Belmeloro 6, 40126 Bologna, Italy. E-mail: fabrizio.ferre@unibo.it

Abstract

Long non-coding RNAs (lncRNAs) are associated to a *plethora* of cellular functions, most of which require the interaction with one or more RNA-binding proteins (RBPs); similarly, RBPs are often able to bind a large number of different RNAs. The currently available knowledge is already drawing an intricate network of interactions, whose deregulation is frequently associated to pathological states. Several different techniques were developed in the past years to obtain protein–RNA binding data in a high-throughput fashion. In parallel, *in silico* inference methods were developed for the accurate computational prediction of the interaction of RBP–lncRNA pairs. The field is growing rapidly, and it is foreseeable that in the near future, the protein–lncRNA interaction network will rise, offering essential clues for a better understanding of lncRNA cellular mechanisms and their disease-associated perturbations.

Key words: long non-coding RNAs; protein–RNA interactions; high-throughput sequencing; co-immunoprecipitation

Introduction

Protein–RNA interactions are key aspects of many cellular processes that go beyond the already established steps of the mRNA production and usage as information carriers, e.g. splicing, polyadenylation, transport, stability and translation [1–4]. A growing knowledge of RNA-binding proteins (RBPs) targets is shifting the attention towards non-coding RNAs, from RNAs involved in the translation machinery and its regulation (rRNAs, tRNAs, small interfering RNAs and miRNAs) to the large and heterogeneous class of long non-coding RNAs (lncRNAs).

Only a small number of lncRNAs have been functionally well-characterized. However, they are involved in a wide range of biological functions through diverse molecular mechanisms often including the interaction with one or more protein partners [5]. Some of them remain linked to their transcription site, and interact with proteins to regulate the expression of genes in *cis* [6]. Others function as molecular decoys, binding to specific transcription factors to prevent their association with DNA [7]. LncRNAs can also interact with chromatin-modifying complexes and lead them to their genomic target in *trans* [8]. They can function as sponges for miRNA [9] or bind to enhancers and help them in their activity, for example by promoting the formation of chromatin loops and the recruitment of remodelling complexes [10]. Moreover, they can bind antisense mRNAs and regulate them post-transcriptionally [11], or function as scaffold for the assembly of macromolecular complexes [12].

According to the pervasiveness of protein–RNA interactions, many reports underline how their perturbation is linked to pathologies, including autoimmune and metabolic diseases, neurological and muscular disorders and cancer [2, 13]. Many proteins implicated in different cancer stages, such as DNA methyltransferases (DNMTs), heterochromatin protein 1, MOF, MSL, DDP1, Polycomb-group and Trithorax-group proteins, are RBPs that are able to bind lncRNAs [14–16]. Expression levels of RBPs are significantly fluctuating in tumour samples, and they might provide clues for prognosis [17]. HOTAIR, one of the first

Fabrizio Ferrè is Associate Professor of Molecular Biology at the Department of Pharmacy and Biotechnology (FaBiT) at the University of Bologna Alma Mater (Italy). His research topics include computational and functional genomics.
Alessio Colantoni worked at this project during his PhD at the Department of Biology in the University of Rome Tor Vergata (Italy). He is now a postdoctoral researcher at the Department of Biology and Biotechnology Charles Darwin at the University of Rome La Sapienza.
Manuela Helmer-Citterich is full Professor of Molecular Biology and Bioinformatics in the Department of Biology of the University of Rome Tor Vergata (Italy). Her research interests include structural bioinformatics, molecular interactions and RNA analysis.
Submitted: 18 February 2015; **Received (in revised form):** 30 April 2015

known lncRNAs and the first one for which a relation with cancer was demonstrated, is deregulated in a number of tumours including breast cancer and hepatocellular carcinoma, and participates in chromatin modification complexes by the interaction with the polycomb group protein PRC2, which is a histone methyltransferase, and with LSD1, which is a histone demethylase [18, 19]. MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) is an lncRNA that is up-regulated in breast, prostate, colon, liver and uterus cancers, and has been shown to interact with members of the SR protein family of splicing regulators [20]. CCND1/Cyclin D1 is an lncRNA which is transcribed from the promoter region of the Cyclin D1 gene; it is a cell cycle regulator involved in many cancer types, that can interact with the TLS protein, which is a sensor of DNA damage [21]. ANRIL is an antisense lncRNA transcribed from the INK4 locus that is up-regulated in prostate cancer. ANRIL can interact with the chromobox 7 (CBX7) protein, which is part of the polycomb group PRC1 protein complex [22]. Several other examples can be found in recent reviews [23–27].

The paucity of information that we have about the functions of lncRNAs, and even more, about the specific sequences involved in carrying out these functions, prevents us to better rationalize their involvement in cellular processes and in disease. On the other hand, experimental and computational techniques are available to analyse, in high-throughput settings and at high resolution, protein–RNA interactions, allowing the identification of binding partners, binding sites and interaction determinants. While these methods were mostly applied for protein-coding RNAs (i.e. mRNA) analysis, they can all in principle be used for lncRNAs as well, and a growing amount of data depicting protein–lncRNA interactions is becoming available, shedding light to this heterogeneous class of cellular regulators.

Detection of protein–lncRNA interactions

A number of methods to uncover the interaction between proteins and RNAs were developed in the past decades. The first proposed methods were low-throughput procedures able to identify only one or few RNAs linked to a protein: these include the RNA electrophoretic mobility shift assay [28], RNA pull-down assay [29], oligonucleotide-targeted RNase H protection assay [30] and FISH co-localization [31].

More recent methods provide high-throughput transcriptome- or proteome-wide overviews of protein–RNA binding (the general classification of these methods followed in this review is shown schematically in Table 1). This is of particular importance, as it appears that relationships between proteins and RNAs are many-to-many, meaning that each RBP is able to bind several RNAs, and a given RNA often interacts with more than one RBP [32]. These methods can be divided in protein-focused and RNA-focused [33, 34]. The goal of protein-focused approaches is the identification of RNAs bound by a protein of interest. These methods can be further classified in *in vitro* and *in vivo* [35]. In *in vitro* technologies, RNA libraries are tested against a protein, and high-affinity RNAs are isolated after rounds of stringent selection. In *in vivo* methods, RNAs bound to the protein of interest in a sample are pulled down using variants of immunoprecipitation techniques. In RNA-focused approaches, the goal is the opposite: to identify all proteins bound to an RNA of interest. Finally, *in silico* inference can be used to predict interactions, usually starting from experimental evidence used to train interaction models.

Table 1. Classification of methods for the depiction of protein–lncRNA interactions followed in this review

Class	Sub-class	Method
Protein-focused	*In vitro*	SELEX-based (HT-SELEX, SEQRS, RAPID-SELEX)
		RNAcompete
		RNA Bin-n-Seq
		RNA-MaP
	In vivo	RIP-Chip
		RIP-Seq
		CLIP (HITS-CLIP, PAR-CLIP, iCLIP)
		CRAC
RNA-focused	RNA-based capture and MS	MS2 trapping
		SILAC-based
	Selection on protein libraries	Phage display
		Protein arrays
In silico		RPISeq
		catRAPID
		Wang's method
		lncPro
		Oli suite
		RPI-Pred
		PRIPU
Other	RNA interaction with chromatin	ChIRP
		CHART
		RAP
	RNA–RNA interaction	CLASH

Protein-focused *in vitro* approaches

In vitro approaches provide insights on binding preferences of target proteins by the isolation of RNAs, generally called aptamers, from synthetic libraries showing affinity for the RBP of interest (Figure 1). In the SELEX (Systematic Evolution of Ligands by EXponential enrichment) technology [36, 37], RNAs having high affinity for a purified protein are selected from a random RNA oligonucleotide library in a number of cycles of selection and PCR amplification, and then cloned and sequenced by the Sanger method. Extensions of SELEX to high-throughput sequencing, e.g. HT-SELEX [38], SEQRS [39] and RAPID-SELEX [40], require less (sometimes only one) selection rounds and as such are able to identify a larger number of bound RNAs having a wider affinity spectrum. Another *in vitro* method is RNAcompete [41] that uses smaller RNA libraries whose design is aimed at creating short (often seven nucleotides long) sequences embedded in a single-stranded or weakly paired context. Using a microarray, the enrichment of each oligonucleotide in the library after selection can be quantitatively measured. An additional recent variant is RNA Bind-n-Seq (RBNS) [42], in which RNAs from a random library bound to an RBP of interest are retro-transcribed and deep-sequenced. The major novelties of the method consist of the usage of different RBP concentrations in the selection step (allowing the statistical modelling and estimation of the dissociation constants between RBP and bound RNA), and on using libraries composed of longer RNAs (40 bp) than in other similar methods (that allows a more reliable RNA fold prediction and a better identification of the binding structural determinants). Motifs identified by RBNS for a number of RBPs recapitulate well those identified by *in vivo* methods such as iCLIP (described in the next section)

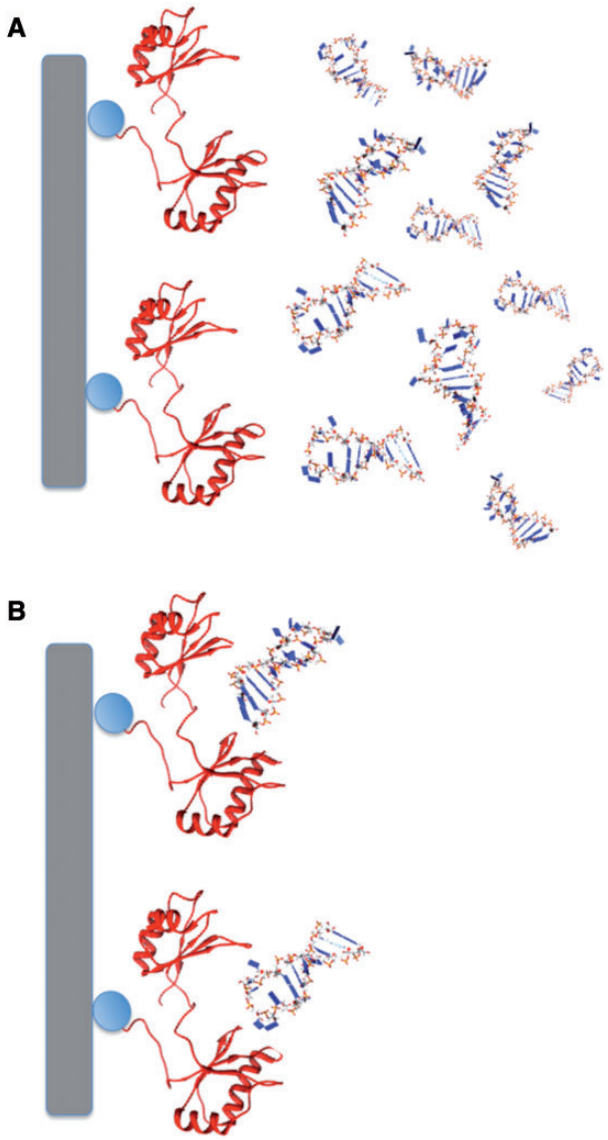

Figure 1. Schematic description of the *in vitro* protein-focused approaches. (**A**) Protein baits (coloured in red) are immobilized to a support (shown here as a grey bar), and exposed to RNAs from a library or from transcriptomic fragments; (**B**) after cycles of selection and amplification, RNAs showing affinity towards the immobilized protein can be isolated and sequenced. A colour version of this figure is available at BIB online: http://bib.oxfordjournals.org.

[42]. A technique (RNA-MaP) using an Illumina sequencer was used to accurately measure the binding affinity between a protein (the bacteriophage MS2 coat protein) and a large library of variants of the hairpin that this protein naturally binds [43], by converting the sequencer flow cell into a high-density RNA array. This method holds promises for the detailed analysis of the kinetics of RNA binding, and the parallel analysis of variations of a given RNA motif allows the identification of the sequence and structural contribution to the binding affinity. Finally, recent developments of the SELEX technology allow the usage of genomic or transcriptomic fragments instead of synthetic RNA libraries, and they have been applied for the analysis of protein–RNA binding [44–46].

The RNA sequences collected by these *in vitro* methods do not necessarily correspond to known RNAs; consequently, motif finding procedures should be used to determine recognition determinants and to extend them to known RNAs. Motif finding algorithms such as MEME [47] or GLAM2 [48] are often used to identify primary sequence preferences of binding. Other methods take advantage of structural properties of the identified RNAs. MEMERIS [49] is an extension of the MEME algorithm that incorporates RNA structure predictions to identify single-stranded motifs, and it has been successfully applied to SELEX data. Similarly, Aptamotif [50] uses an ensemble of suboptimal RNA folds to identify recurrent single-stranded regions in a collection of aptamers.

The inherent limit of all techniques based on random libraries is that the detection of *in vitro* RNA binding ability for a given target protein does not necessarily prove that the protein is an RBP. Even if selection cycles are performed using stringent conditions, a non-RBP might show affinity for some RNAs, for example because these RNAs are mimicking its natural ligands. Detection of biologically non-relevant interactions can be avoided only by using RNA libraries derived from cellular transcriptomes, or by using genomic fragments. Besides, even if the target protein is a *bona fide* RBP, the *in vitro* efficient binding of some RNAs in the library does not necessarily reflect the *in vivo* possibility of interaction, where the specific cellular context might affect site availability; the secondary and tertiary structure of the RNA binding partner might dictate which sites are available and which are not; moreover, binding of sites with lower affinities might be preferred *in vivo* to increase the ability of modulation. Nevertheless, motifs identified *in vitro* for some RBPs closely resemble those retrieved in *in vivo* studies, validating the usefulness of these methods [20].

As of today, *in vitro* methods have not been extensively used for investigation of ncRNA binding by proteins, for which *in vivo* methods, described later, are becoming prevalent, yet a number of examples can be found; for example, Wu and coworkers [51] screened a nuclear RNA repertoire of a melanoma cell line for the ability of binding to the polypyrimidine tract-binding protein-associated splicing factor, identifying a previously unreported lncRNA (Llme23), and discussed its role in maintaining the malignant properties of the melanoma cells. In [52], binding partners of the *Escherichia coli* global regulator protein Hfq were screened through genomic SELEX coupled with high-throughput sequencing, identifying a large number of antisense ncRNAs regulating gene expression *via* binding in *cis*. Public data sets are also available from which RBP–lncRNA interactions detected *in vitro* can be retrieved. RBPDB [53] is a database of binding determinants for a collection of RBPs, some of which can bind lncRNAs. Most data derive from SELEX approaches, but also a number of *in vivo* experiments are included. In a comprehensive analysis of the human RNA-binding proteome, RNAcompete was used to elucidate binding preferences for a large collection of RBPs, including some proteins known to bind also lncRNAs [54].

Protein-focused *in vivo* approaches

The possibility of pulling down an RBP with a specific antibody is the basis of the *in vivo* approaches (Figure 2). RIP-Chip [55] and RIP-Seq [56] are high-throughput antibody-based techniques in which bound RNA is obtained through immunoprecipitation of its protein partner; then, the identification of the bound transcripts is performed via microarray (RIP-chip) or RNA-Seq (RIP-Seq). Such methods allow the identification of bound transcripts, but do not provide direct information about the localization of the binding site. While most RIP-Chip experiments were performed not specifically for the analysis of lncRNA binding,

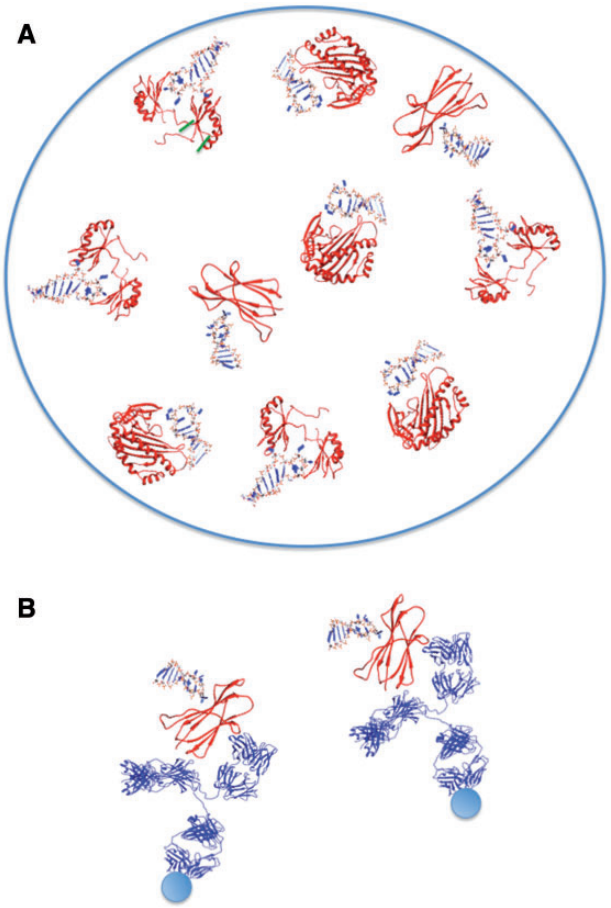

Figure 2. Schematic description of the *in vivo* protein-focused approaches. Protein-RNA complexes in the cells of a sample (**A**), either stabilized by cross-linking or not, are isolated after cell lysis by immunoprecipitation using an antibody specific for an RBP, here shown in blue; the light blue dot indicates the solid-phase support for the immunoprecipitation, often beads of various materials (not shown in scale); (**B**) the isolated RNAs are then sequenced. A colour version of this figure is available at BIB online: http://bib.oxfordjournals.org.

some panels can nevertheless provide this kind of information, depending on the array design. For example, the platform used to detect ELAVL1 RNA interactors [57] is especially rich in lncRNAs, as shown by Cao and coworkers [58].

The Cross-Linking ImmunoPrecipitation (CLIP) [59] procedure takes advantage of the ability of 254 nm ultraviolet (UV) light to induce the *in vivo* formation of covalent bonds between RNA nucleotides and proximal RBP amino acids at the binding site; then, immunoprecipitation allows for the isolation of the protein–RNA complex of interest. By combining CLIP with high-throughput sequencing (HITS-CLIP), it is possible to identify transcriptome-wide the protein-bound RNAs [60]. The UV-induced covalent bond being irreversible, the cross-linked RBP is digested with proteinase, which might not completely detach the cross-linked amino acids from the RNA. It has been observed that, during the conversion to cDNA, these amino acids can create an obstacle for the reverse transcriptase, leading to the introduction of a mutation (often a deletion, but it depends on the protein) at the cross-linking site (cross-linking induced mutations, or CIMS) [61]. These diagnostic mutations can be used to map protein–RNA interactions at single-nucleotide resolution. PAR-CLIP is a variant of HITS-CLIP, in which living cells are provided with photoreactive ribonucleoside analogues, such as 4-thiouridine (4-SU) and 6-thioguanosine (6-SG), that are incorporated into nascent transcripts [62]. This allows the use of UV light of 365 nm for a more efficient cross-linking; in addition, the cross-linking induces T->C (4-SU) or G->A (6-SG) transitions, which can be used to identify the precise position of cross-linking and to better discriminate between cross-linked RNAs and abundant cellular RNAs. This notwithstanding, PAR-CLIP technology has a number of drawbacks, for example it is limited to cultured cells, and the nucleoside analogue uptake can be in some cases not efficient. Another variant of the CLIP strategy is the individual-nucleotide resolution CLIP (iCLIP), which is based on the observation that the reverse transcriptase often truncates prematurely cDNAs at the cross-linking nucleotide, and hence allows mapping the binding sites with great accuracy [63]. A protein-focused method not based on antibodies directed towards the RBPs is the cross-linking and cDNA analysis (CRAC) [64], in which RBPs are tagged to allow tandem affinity purification. Complexes are stabilized *in vivo* by UV irradiation, and, after immobilized metal ion affinity chromatography and proteinase treatment, isolated RNAs are amplified and sequenced.

Analysis of protein–RNA interaction raw data produced in high-throughput settings requires complex bioinformatics procedures that borrow approaches and tools from the investigation of NGS data and particularly from other immunoprecipitation procedures aimed at protein-bound genomic DNA exploration, such as ChIP-Seq [65]. The required pipelines can be summarized in three major steps: (i) mapping the reads on the reference genome/transcriptome, (ii) identification of clusters of reads indicating putative RNA targets and binding sites and (iii) inference of the interaction determinants. As with all NGS data, particular care must be dedicated to read mapping to the reference genome, using an algorithm able to handle reads spanning exon-exon junctions that will align to the genome in two separate fragments. Alternatively, reads can be mapped to the transcriptome. A beneficial procedure for HITS-CLIP and PAR-CLIP experiments is the identification and collapse of duplicate reads, i.e. reads that have the same mapping coordinates (including strand) and are likely to represent artifacts introduced by preferential PCR amplification of particular sequences [66]. Read mapping must also take into account the presence of cross-linking induced mutations and distinguish them from sequencing errors or genomic variations between the sample donor genome and the reference.

The critical step is the identification of genomic regions encoding for the RBP interaction sites, which can be achieved at different degrees of resolution, from a coarse-grained target RNA inference to a single-nucleotide level binding site definition. A number of *caveats* intrinsic for the applied technologies can have a profound effect on the data analysis outcome. Background noise can arise in several ways and must be taken into account. Antibody cross-reactivity with a protein different from the intended RBP, or RNAs unspecifically pulled down can contaminate the sample. The usage of control data can be highly beneficial. Performing a CLIP-Seq or RIP-Seq run using an antibody targeting a protein known to be unable to bind RNA can provide an overview of the unspecific RNA pull-down, while an RNA-Seq run can give estimates of the transcripts abundance. Nevertheless, not all the currently available analysis procedures are able to take advantage of such supplementary data.

The identification of the bound RNAs and of the RNA region interacting with the examined RBP was performed initially simply seeking read clusters by setting *ad hoc* cutoffs defining the extension of the minimum read overlap and the cluster

Table 2. Methods for read cluster identification in CLIP or RIP-Seq experiments

Method	Data	Input format	Implementation
PARalyzer	PAR-CLIP	*SAM* or *BAM*	Stand-alone
Pyicoclip	HITS-CLIP, PAR-CLIP, iCLIP	*SAM*, *BAM* or *BED*	Stand-alone
Piranha	HITS-CLIP, PAR-CLIP, iCLIP, RIP-Seq	*BAM* or *BED*	Stand-alone
wavClusteR	PAR-CLIP	*BAM*	Stand-alone
RIPSeeker	RIP-Seq	*SAM*, *BAM* or *BED*	Stand-alone
PIPE-CLIP	HITS-CLIP, PAR-CLIP, iCLIP	*SAM* or *BAM*	Web-based
MiClip	HITS-CLIP, PAR-CLIP	*SAM* or *BAM*	Stand-alone, Web-based

amplitude. Verifying the presence of a sufficient number of CIMS or PAR-CLIP transitions in each cluster reduced the number of false hits. A number of sophisticated algorithms have been developed, e.g. PARalyzer [67], Piranha [68], wavClusteR [69], RIPSeeker [70], MiClip [71], PIPE-CLIP [72] and the Pyicoclip module of the Pyicoteo toolkit (previously called Pyicos) [73]. While different in implementation, all these procedures can be summarized in the same two steps: (i) clusters identification from the reads genomic alignment; and (ii) identification and ranking of binding sites within enriched clusters using, whenever possible, diagnostic mutations to prioritize more reliable sites. Table 2 reports a list of methods for read cluster identification, specifying the experimental protocols the method is designed for, the input data format and the availability (as stand-alone software or through a Web-based interface).

Genomic coordinates of genes, transcripts and exons allow linking the identified clusters to the bound transcript identity. Normally used gene-sets from common public repositories might not contain the most updated annotations of lncRNAs, and specialized data sets can provide more recent and exhaustive collections, for example the GENCODE lncRNA catalogue [74], NONCODE [75] or LNCipedia [76]. Other useful resources are the lncRNAdb of functionally annotated lncRNAs [77] and the NRED database of lncRNA expression [78].

The extraction of the binding determinants from the list of identified binding sites is a challenging step for which no tool is currently able to model accurately all possible cases. The reason is that RBP–RNA binding is heterogeneous in nature and different RBP domains are governed by different rules. Generally, sequence-level preferences are often found, allowing the definition of sequence motifs. Tools such as MEME [47] or cERMIT [79] have been successfully applied to the analysis of CLIP-Seq data. Yet, these sequence motifs often must be embedded in a specific secondary structure context [49, 80, 81]. In other cases, RNA secondary structure dictates the interaction: proteins tend to recognize complex secondary structure elements such as stem-loops and bulges [82]. A more extensive review of the influence of RNA structural constraints in the context of motif discovery at RBP binding sites can be found in [83]. The already mentioned MEMERIS [49], CMfinder [84], RNApromo [85], StructRED [86], the algorithm by Li and collaborators [87], RNAcontext and RBPmotif [88, 89], GraphProt [90] and Zagros [91] are all tools that use, with different strategies and to different extent, secondary structure information to determine more specific binding motifs.

The application of these technologies revealed the sometimes unexpected lncRNA-binding ability of several RBPs. A HITS-Clip study [92] revealed novel roles for the Microprocessor complex in the maturation and expression control of lncRNAs. Several lncRNAs were identified in PAR-CLIP analysis of binding ability of AUF1 [93], a protein linked to ageing and cancer. TDP-43 is a nuclear RBP involved in amyotrophic lateral sclerosis; iCLIP analysis [94] revealed the TDP-43 binding with the MALAT1 and NEAT1 lncRNAs. A CRAC-based approach was used to depict the binding ability of a panel of 13 *Saccharomyces cerevisiae* proteins involved in different steps of RNA maturation [95], highlighting differences and similarities between the maturation pathways of lncRNAs and mRNAs in yeast. JARID2 is a DNA-binding protein that acts as a transcriptional repressor by interacting with the PRC2 complex; PAR-Clip analysis [96] revealed the ability of JARID2 in binding lncRNAs, and that these interactions are essential for the recruitment of PRC2 to the chromatin. RIP-Seq was used in [97] to identify a ncRNA binding to DNA (cytosine-5)-methyltransferase 1 (DNMT1), which is a regulator of patterns of cytosine methylation. Interaction with this ncRNA prevents the methylation of the *locus* from which the ncRNA is transcribed. The authors identified several other possible mechanisms of this nature, and proposed this as a widespread methylation regulation managed by ncRNAs.

RNA-focused approaches

In this class of methods, an RNA of interest is purified and used as bait to isolate RBPs bound to it, and then identified using mass spectrometry (MS), protein arrays or other techniques (Figure 3). In RNA-focused *in vitro* methods, the RNA of interest is immobilized to a solid support, and then it is exposed to proteins from a cellular lysate or from protein libraries generated with various means and purposes. After washing and elution, proteins bound to the immobilized RNA are identified by MS. In *in vivo* methods, cross-linking between proteins and RNAs is induced by UV or formaldehyde to stabilize physiological interactions, then cells are lysed and the RNA of interest is captured and its bound proteins detected. The various experimental strategies and their technical aspects for both general approaches have been recently reviewed [33].

A number of applications of RNA-focused strategies for the analysis of the lncRNA-bound proteome can be found in the literature. In 2012, Gong *et al.* described the usage of a stem-loop structure of viral origin, inserted at the 3′ end of the lncRNA of interest, which can be bound by the MS2 bacteriophage coat protein [98]. A vector carrying the lncRNA with the inserted stem-loops in multiple copies, and a second one carrying the coat protein were transfected into cultured cells, which were then treated with formaldehyde. An antibody was used to immunoprecipitate the coat protein, pulling down the lncRNA and its bound RBPs. This strategy was used in [99] to identify a ribonucleoproteic complex involving the translational regulatory lncRNA and to elucidate its role in metastasis progression. SILAC (stable isotope labelling with amino acids in cell culture) quantitative MS was used to identify interactors of the telomeric repeat-containing ncRNA TERRA [100]. Oligonucleotides containing TERRA repeats, and shuffled control repeats were synthesized and tagged with biotin at the 3′ end, and incubated with cellular extracts. SILAC allowed the identification of 115 enriched proteins in TERRA repeats versus the control, including proteins involved in chromatin remodelling, DNA replication, RNA degradation, transcription and translation. Other recent findings include the binding of the SPRY4-IT1 lncRNA

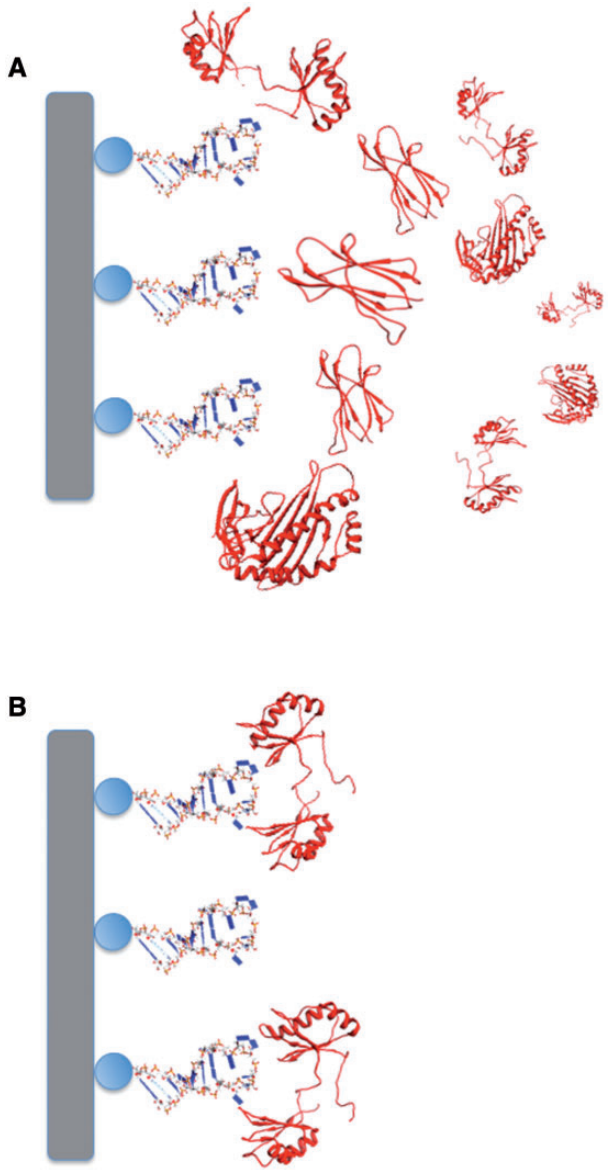

Figure 3. Schematic description of the *in vitro* RNA-focused approaches. (**A**) RNA baits are immobilized to a support (shown here as a grey bar), and exposed to proteins (coloured in red) contained in a cellular lysate or from a protein library; (**B**) after washing and elution, proteins showing affinity towards the immobilized RNAs can be isolated, and then identified (e.g. by mass spectrometry). A colour version of this figure is available at BIB online: http://bib.oxfordjournals.org.

and lipin-2 in melanoma cells, and its role in apoptosis regulation [101], and the interaction of the HULC lncRNA with the IGF2BP protein family and its relationship with translational control in hepatocellular carcinoma [102].

The RNA bait is not necessarily limited to one specific RNA, but can also represent an entire class of RNAs. In [103] UV cross-linking and oligo(dT) affinity, purification followed by quantitative MS was used to identify the entire polyadenylated RNA-bound proteome in HeLa cells revealing around 860 proteins, a large number of which do not seem to contain a recognizable RNA-binding domain. Similarly, Baltz and collaborators [104] analysed the repertoire of RBPs bound to polyadenylated RNAs in an embryonic kidney cell line, while Kwon and coworkers [105] delineated the RNA-bound proteome in mouse embryonic stem cells. In all these cases, several proteins with no known RNA binding ability were retrieved, suggesting that protein–RNA interactions can go well beyond those mediated by known RNA-binding domains, and revealing the intricacies of the protein–RNA network. As many lncRNAs are polyadenylated, these studies are not limited to mRNAs.

While powerful, these techniques are still technically challenging [33, 34]. For *in vivo* procedures especially, the amount of purified proteins might not be sufficient for MS analysis. For this reason, most examples found in the literature used RNAs having high expression levels, which hinders their application to lncRNAs; cross-linking-based strategies can nevertheless extend the identification to less abundant and/or more transient interactions.

In silico methods

The growing wealth of public data describing protein–RNA binding allows the training of computational models that can be used for inference of novel interactions. Most of these methods were developed primarily for the prediction of interaction between proteins and lncRNAs, the assumption being that the binding determinants might be similar regardless of the RNA type. Hence, *in silico* methods can be of primary importance for the characterization of lncRNAs, for which experimental data are less abundant and often technically challenging.

Most algorithms use as training examples, the three-dimensional structures of proteins bound with an RNA or RNA fragment mined from the PDB [106] from which a number of physicochemical features (e.g. propensities for hydrogen bonding, van der Waals interaction and secondary structure) can be extracted to describe each protein–RNA pair. Feature vectors of known interactions are compared to those computed on control sets of non-interacting pairs, and statistical methods, machine learning algorithms or *ad hoc* scoring systems are used for evaluating the binding potential of a novel pair. RPISeq [107], catRAPID [108], the method described in Wang *et al.* [109], lncPro [110] and RPI-Pred [111] all follows these lines. The recent PRIPU method [112] differs from the others by using statistical learning methods trained on only positive examples. The Oli suite [113] trains on high-throughput evidence including PAR-CLIP experiments, and used features include sequence composition, predicted secondary structure and presence of motifs.

The main concern of computational methods is the usage of descriptors that can be easily applied to cases that are not present in the training set. For example, one of the first methods for computational prediction of RNA–protein interactions [114] used features, such as gene ontology terms, protein localization and genetic interactions, that might not be available or be difficult to compute for some RNAs or proteins. In fact, the performance of the method on RBPs not included in the training set was reported as very variable.

Many methods were tested on protein–lncRNA interactions. catRAPID reached 89% prediction accuracy on the NPInter [115] data set. The RPISeq-predicted interaction between linc-UBC1 RNA and PRC2 was experimentally validated [116]. Testing lncPro on NPInter provided good inference accuracy, and the application of lncPro on the entire human proteome against a collection of lncRNAs retrieved a significant enrichment in nuclear proteins, coherently with the observation that many lncRNAs reside in the nucleus [117]. The method of Wang and coworkers was used on a data set of *Caenorhabditis elegans* ncRNAs and proteins, confirming some of their predictions in pull-down experiments [109].

Table 3. Features and drawbacks of protein-focused and RNA-focused approaches

Class of methods	Features	Drawbacks
Protein-focused *in vitro* methods	Not dependent on which RNAs are expressed in a sample; can be applied to organisms for which no genome assembly is available; some variants allow the estimation of the binding affinity.	The identified binding sequences might not correspond to any known RNAs; favour RNAs binding with high affinity.
Protein-focused *in vivo* methods	Permit the analysis of entire transcriptomes; permit to study interactions in physiological conditions; can retrieve low-affinity binding; some variants allow the identification of the binding sites at single-nucleotide resolution.	Depend on a good reference genome assembly; can only detect binding to RNAs expressed in the analysed sample; can suffer from some biases introduced by the cross-linking and sequencing strategies; the bioinformatics analysis is relatively complex; it is difficult to estimate binding affinities.
RNA-focused methods	Permit the analysis of entire proteomes; can be applied to entire RNA classes (e.g. all the polyadenylated RNAs).	Technically more challenging; the throughput is lower than protein-focused methods; favour relatively abundant proteins.

Remarks and perspectives

The growing interest in the cellular role of protein–lncRNA interactions is reflected by the rising availability of data. As described in this review, a large and growing number of methods are available, each providing unique features but also some drawbacks (as summarized in Table 3). A growing wealth of CLIP-Seq and RIP-Seq data sets is hosted in databases such as Gene Expression Omnibus (GEO) [118] or ArrayExpress [119]. The number of CLIP-Seq data sets in GEO alone is reported at more than 400 [120]. Often the original purpose of the experiment was not focused on the detection of lncRNAs; however, these transcriptome-wide techniques can capture this class of RNAs as well; nevertheless, a re-analysis might be needed to fully exploit public data sets to this purpose, for example providing to the analysis algorithms the quickly changing lncRNA collections. A number of databases such as CLIPZ [121], starBase [122], doRiNA [120] and NPInter [115] offer protein–RNA interactions retrieved from the literature or by the analysis of public data. CLIPZ, in addition to providing access to some CLIP-Seq experiments, also offers an analysis environment for user-uploaded data sets. NPInter is a manually curated database that stores more than 200,000 functional interactions (i.e. not necessarily physical interactions) extracted from the literature. StarBase focuses on CLIP-Seq data, reporting in the v2.0 interaction results from 108 experiments; it also provides data for miRNA binding to mRNAs and lncRNAs, and has an ample section dedicated to cancer samples. DoRiNA includes 136 CLIP-Seq data sets, and also known and predicted interactions with miRNAs. Another resource, RBPmap [123], contains more than 100 RBP-binding motifs extracted from the literature and represented as a *consensus* sequence or a position-specific scoring matrix. Given as input one or more RNA sequences, the server estimates binding by an RBP against a background model.

Taken together, these data are paving the road for a comprehensive reconstruction of the protein–lncRNA interaction network, and the first integrative studies where protein–RNA binding data are augmented with genome-wide information of different kinds are appearing [124, 125]. The ChIPBase database [126] is one example of integrative resource in which regulatory mechanisms formed between lncRNAs, microRNAs and transcription factors are reconstructed by a large-scale analysis of ChIP-Seq data sets. In a recent work [127], the StarBase authors integrated protein–lncRNA interactions with expression data in several cancer types and single-nucleotide polymorphisms, reconstructing regulatory circuits and how they might be affected by mutations in pathological conditions.

As illustrated in this review, many experimental strategies are available, all having advantages and drawbacks with respect to the others. Being based on oftentimes radically different features, different techniques can allow looking at the same problem from different angles. A comparative PAR-CLIP and SILAC-based RNA pull-down study on four different RBPs [128] highlighted how the findings that are common to the two approaches were in very good agreement, while each method provided unique evidence, and they can, therefore, complement each other. Nevertheless, it is still unclear how many unspecific interactions are detected by all the presented techniques, as well as how to better define the biological meaning of the detected interactions; post-processing procedures and complementation with evidence of different nature can reduce the number of false-positive hits and help in the delineation of regulatory mechanisms. CLIP-Seq and RIP-Seq experiments are often accompanied by a standard RNA-Seq analysis of gene expression levels to filter by those genes actually expressed in the considered sample. Methods that depict chromatin states and epigenetic regulations might also help in drawing a more complete and accurate picture of the regulatory circuits created by the interactions between RBPs and lncRNAs. Finally, the role of RNA editing and RNA methylation in promoting and/or inhibiting the interaction with proteins is still unclear, and could be explored by integrative analyses.

Other techniques have been developed to provide high-throughput functional characterization of lncRNAs not directly focused on the interaction with RBPs. A growing interest on the lncRNA ability of interacting with chromatin has led to the development of methods for depicting the genomic regions to which a given ncRNA is bound, either binding the genomic DNA or through interaction with chromatin proteic components. Chromatin Isolation by RNA Purification (ChIRP) [129], Capture hybridization analysis of RNA targets (CHART) [130] and RNA antisense purification (RAP) [131] provide high-throughput ways to identify genomic binding sites of a given lncRNA, and can offer detailed perspective on how lncRNA can form ribonucleoproteic complexes at specific *loci* to exert regulative roles. In the ChIRP technique, after cross-linking and sonication, a number of biotinylated DNA oligonucleotides antisense to a target RNA are hybridized to chromatin fragments carrying that RNA, and then recovered using beads coated with streptavidin. Genomic regions bound to the target RNA can be identified by high-throughput DNA sequencing, as well as the proteins bound to either the RNA or the genomic DNA by MS. The RAP and CHART methods are similar to ChIRP, differing mostly in

the design strategy of the antisense oligonucleotides and in the cross-linking protocols. While these methods are offering important evidence for the involvement of lncRNAs in gene expression regulation and chromatin remodelling [132], it should be noted that they cannot prove direct lncRNA–protein binding. Finally, the cross-linking, ligation and sequencing of hybrids technique [133] (CLASH) was developed for the detection of RNA–RNA interactions. It was applied up to now for the high-throughput depiction of the miRNA–target RNA interaction mediated by the AGO protein, but it can in principle be extended to other RNA–RNA interactions involving other proteins [134].

A number of technical issues can limit current technologies, and overcoming them will afford more detailed and accurate views. For example, CLIP-based technologies currently provide only a qualitative view of protein–RNA interactions. To move towards quantitative estimates, specifically developed normalization procedures need to be introduced that should take into account transcript expression levels, cross-linking efficiency and all potential biases introduced in the various steps of these procedures. Another overlooked issue concerns the RNA structural features involved in the RBP interaction. While the structural constraints for RNA binding by an RBP can vary or be unessential, in many cases it is known that the recognized binding motif must have specific structural characteristics. Current methods for binding site identification in CLIP-Seq usually do not take advantage of these potential features, which can reduce the number of false hits. Structural constraints are currently introduced only in the successive motif identification steps, while they could be beneficial also throughout the whole pipelines. This is not a trivial task given the inherent complexity of handling efficiently the RNA structure representations. Novel RNA secondary structure encodings could be beneficial for the integration of structural information in the current pipelines [135]. In addition, multiple proteins can bind the same RNAs in cooperative or competitive ways, and taking into account these aspects can give a more realistic view of RBP binding in the crowded cellular environment.

Finally, it is not clear whether well-known RBP families compose the lncRNA-bound proteome, or if still uncharacterized protein domains and architectures are involved. As discussed previously, detection of proteins bound to the whole set of polyadenylated RNAs revealed many proteins escaping the usual definitions of RBPs. RNA-focused strategies can offer an unbiased way to explore these interactions. Unfortunately, as of today, these approaches are still challenging, especially for lncRNAs that have low expression levels.

Key Points

- Protein–RNA interactions are keys to a host of cellular processes, and their deregulation is implicated in pathologies.
- Long non-coding RNAs often exert their functions by binding one or more proteins; conversely, RNA-binding proteins are often able to bind several lncRNAs.
- A large number of different experimental techniques have been developed for the high-throughput determination of protein–lncRNA interactions, each having advantages and caveats.
- Accurate computational inference, for which a number of tools are currently available, can provide additional interaction evidence.
- A quickly growing body of knowledge depicting protein–lncRNA binding will allow in the near future an *ensemble* outlook of the protein–lncRNA interaction network, leading to a systems view of lncRNA functions.

Funding

This work was supported by Programmi di Ricerca di rilevante Interesse Nazionale PRIN 2010 (prot. 20108XYHJS_006 to M.H.C.).

References

1. Singh, R. RNA-protein interactions that regulate pre-mRNA splicing. *Gene Expr* 2002;**10**:79–92.
2. Lukong KE, Chang KW, Khandjian EW, *et al*. RNA-binding proteins in human genetic disease. *Trends Genet* 2008;**24**(8):416–25.
3. Kishore S, Luber S, Zavolan M. Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. *Brief Funct Genomics* 2010;**9**:391–404.
4. Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet* 2010;**11**:75–87.
5. Zhu J, Fu H, Wu Y, *et al*. Function of lncRNAs and approaches to lncRNA-protein interactions. *Sci China Life Sci* 2013;**56**:876–85.
6. Wang X, Arai S, Song X, *et al*. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 2008;**454**:126–30.
7. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 2013;**14**(11):699–712.
8. 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.
9. Cesana M, Cacchiarelli D, Legnini I, *et al*. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 2011;**147**:358–69.
10. Wang KC, Yang YW, Liu B, *et al*. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 2011;**472**:120–24.
11. Carrieri C, Cimatti L, Biagioli M, *et al*. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 2012;**491**:454–57.
12. 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.
13. Castello A, Fischer B, Hentze MW, *et al*. RNA-binding proteins in Mendelian disease. *Trends Genet* 2013;**29**(5):318–27.
14. Jeffery L, Nakielny S. Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J Biol Chem* 2004;**279**(47):49479–87.
15. Bernstein E, Allis CD. RNA meets chromatin. *Genes Dev* 2005;**19**(14):1635–55.
16. Barrandon C, Spiluttini B, Bensaude O. Non-coding RNAs regulating the transcriptional machinery. *Biol Cell* 2008;**100**(2):83–95.
17. Kechavarzi B, Janga SC. Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biol* 2014;**15**(1):R14.

18. Wu Y, Zhang L, Wang Y, *et al*. Long noncoding RNA HOTAIR involvement in cancer. *Tumour Biol* 2014;**35**(10):9531–38.
19. Cai B, Song XQ, Cai JP, *et al*. HOTAIR: a cancer-related long non-coding RNA. *Neoplasma* 2014;**61**(4):379–91.
20. Ankö ML, Neugebauer KM. RNA-protein interactions *in vivo*: global gets specific. *Trends Biochem Sci* 2012;**37**(7):255–62
21. Kurokawa R. Promoter-associated long noncoding RNAs repress transcription through a RNA binding protein TLS. *Adv Exp Med Biol* 2011;**722**:196–208.
22. Yap KL, Li S, Muñoz-Cabello AM, *et al*. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* 2010;**38**(5):662–74.
23. Huarte M, Rinn JL. Large non-coding RNAs: missing links in cancer? *Hum Mol Genet* 2010;**19**(R2):R152–61.
24. Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. *Mol Cancer* 2011;**10**:38.
25. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discov* 2011;**1**(5):391–407.
26. Tsai MC, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res* 2011;**71**(1): 3–7.
27. Nie L, Wu HJ, Hsu JM, *et al*. Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer. *Am J Transl Res* 2012;**4**(2):127–50.
28. Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc* 2007;**2**:1849–1861.
29. Wang W, Caldwell MC, Lin S, *et al*. HuR regulates cyclin A and cyclin B1 mRNA stability during cell proliferation. *EMBO J* 2000;**19**:2340–50.
30. Günzl A, Palfi Z, Bindereif A. Analysis of RNA-protein complexes by oligonucleotide-targeted RNase H digestion. *Methods* 2002;**26**:162–69.
31. Shih JD, Waks Z, Kedersha N, *et al*. Visualization of single mRNAs reveals temporal association of proteins with microRNA-regulated mRNA. *Nucleic Acids Res* 2011;**39**: 7740–49.
32. Hogan DJ, Riordan DP, Gerber AP, *et al*. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol* 2008;**6**(10): e255.
33. Faoro C, Ataide SF. Ribonomic approaches to study the RNA-binding proteome. *FEBS Lett* 2014;**588**(20):3649–64.
34. McHugh CA, Russell P, Guttman M. Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biol* 2014;**15**(1):203
35. Cook KB, Hughes TR, Morris QD. High-throughput characterization of protein-RNA interactions. *Brief Funct Genomics* 2015;**14**(1):74–89.
36. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 1990;**249**(4968):505–10.
37. Ellington AD, Szostak JW. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* 1990;**346**(6287):818–22.
38. Zhao Y, Granas D, Stormo GD. Inferring binding energies from selected binding sites. *PLoS Comput Biol* 2009;**5**(12): e1000590.
39. Campbell ZT, Bhimsaria D, Valley CT, *et al*. Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep* 2012;**1**(5):570–81.
40. Szeto K, Latulippe DR, Ozer A, *et al*. RAPID-SELEX for RNA aptamers. *PLoS One* 2013;**8**(12):e82667.
41. Ray D, Kazan H, Chan ET, *et al*. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol* 2009;**27**(7):667–70.
42. Lambert N, Robertson A, Jangi M, *et al*. RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* 2014;**54**(5):887–900.
43. Buenrostro JD, Araya CL, Chircus LM, *et al*. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat Biotechnol* 2014;**32**(6):562–8.
44. Lorenz C, von Pelchrzim F, Schroeder R. Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat Protoc* 2006;**1**(5): 2204–12.
45. Zimmermann B, Bilusic I, Lorenz C, *et al*. Genomic SELEX: a discovery tool for genomic aptamers. *Methods* 2010;**52**(2): 125–32.
46. Bannikova O, Zywicki M, Marquez Y, *et al*. Identification of RNA targets for the nuclear multidomain cyclophilin atCyp59 and their effect on PPIase activity. *Nucleic Acids Res* 2013;**41**(3):1783–96.
47. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994;**2**:28–36.
48. Frith MC, Saunders NF, Kobe B, *et al*. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol* 2008;**4**(4):e1000071.
49. Hiller M, Pudimat R, Busch A, *et al*. Using RNA secondary structures to guide sequence motif finding towards single-stranded regions. *Nucleic Acids Res* 2006;**34**(17):e117.
50. Hoinka J, Zotenko E, Friedman A, *et al*. Identification of sequence-structure RNA binding motifs for SELEX-derived aptamers. *Bioinformatics* 2012;**28**(12):i215–23.
51. Wu CF, Tan GH, Ma CC, *et al*. The non-coding RNA llme23 drives the malignant property of human melanoma cells. *J Genet Genomics* 2013;**40**(4):179–88.
52. Lorenz C, Gesell T, Zimmermann B, *et al*. Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res* 2010;**38**(11): 3794–808.
53. Cook KB, Kazan H, Zuberi K, *et al*. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res* 2011; **39**:D301–8.
54. Ray D, Kazan H, Cook KB, *et al*. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 2013; **499**(7457):172–77.
55. Keene JD, Komisarow JM, Friedersdorf MB. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* 2006;**1**:302–07.
56. Zhao J, Ohsumi TK, Kung JT, *et al*. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* 2010; **40**:939–53.
57. Mukherjee N, Corcoran DL, Nusbaum JD, *et al*. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell* 2010;**43**(3):327–39.
58. Cao WJ, Wu HL, He BS, *et al*. Analysis of long non-coding RNA expression profiles in gastric cancer. *World J Gastroenterol* 2013;**19**:3658–64.

59. Ule J, Jensen K, Mele A, *et al*. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* 2005; **37**:376–386.
60. Licatalosi DD, Mele A, Fak JJ, *et al*. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 2008;**456**:464–69.
61. Zhang C, Darnell RB. Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* 2011;**29**:607–14.
62. Hafner M, Landthaler M, Burger L, *et al*. PAR-CliP–a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J Vis Exp* 2010;**41**:2034.
63. Konig J, Zarnack K, Rot G, *et al*. iCLIP–transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp* 2011;**50**:2638.
64. Granneman S, Kudla G, Petfalski E, *et al*. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci USA* 2009;**106**(24):9613–18.
65. Kloetgen A, Münch PC, Borkhardt A, *et al*. Biochemical and bioinformatic methods for elucidating the role of RNA-protein interactions in posttranscriptional regulation. *Brief Funct Genomics* 2014;pii:elu020.
66. Chou CH, Lin FM, Chou MT, *et al*. A computational approach for identifying microRNA-target interactions using high-throughput CLIP and PAR-CLIP sequencing. BMC Genomics 2013;**14** (Suppl 1), S2.
67. Corcoran DL, Georgiev S, Mukherjee N, *et al*. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* 2011;**12**:R79.
68. Uren PJ, Bahrami-Samani E, Burns SC, *et al*. Site identification in high-throughput RNA-protein interaction data. *Bioinformatics* 2012;**28**(23):3013–20.
69. Sievers C, Schlumpf T, Sawarkar R, *et al*. Mixture models and wavelet transforms reveal high confidence RNA-protein interaction sites in MOV10 PAR-CLIP data. *Nucleic Acids Res* 2012;**40**(20):e160.
70. Li Y, Zhao DY, Greenblatt JF, *et al*. RIPSeeker: a statistical package for identifying protein-associated transcripts from RIP-seq experiments. *Nucleic Acids Res* 2013; **41**(8):e94.
71. Wang T, Chen B, Kim M, *et al*. A model-based approach to identify binding sites in CLIP-Seq data. *PLoS One* 2014;**9**: e93248.
72. Chen B, Yun J, Kim MS, *et al*. PIPE-CLIP: a comprehensive online tool for CLIP-seq data analysis. *Genome Biol* 2014;**15**(1): R18.
73. Althammer S, González-Vallinas J, Ballaré C, *et al*. Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics* 2011;**27**(24):3333–40.
74. 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**(9): 1775–89.
75. Xie C, Yuan J, Li H, *et al*. NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res* 2014:**42**:D98–103.
76. Volders PJ, Verheggen K, Menschaert G, *et al*. An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res* 2015;**43**:D174–80
77. Amaral PP, Clark MB, Gascoigne DK, *et al*. lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res* 2011;**39**:D146–51.
78. Dinger ME, Pang KC, Mercer TR, *et al*. NRED: a database of long noncoding RNA expression. *Nucleic Acids Res* 2009;**37**: D122–6.
79. Georgiev S, Boyle AP, Jayasurya K, *et al*. Evidence-ranked motif identification. *Genome Biol* 2010;**11**(2):R19.
80. Buckanovich RJ, Darnell RB. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets *in vitro* and in vivo. *Mol Cell Biol* 1997;**17**:3194–201.
81. Hackermüller J, Meisner NC, Auer M, *et al*. The effect of RNA secondary structures on RNA-ligand binding and the modifier RNA mechanism: a quantitative model. *Gene* 2005;**345**:3–12.
82. Cusack S. RNA-protein complexes. *Curr Opin Struct Biol* 1999; **9**:66–73.
83. Li X, Kazan H, Lipshitz HD, *et al*. Finding the target sites of RNA-binding proteins. *Wiley Interdiscip Rev RNA* 2014;**5**(1): 111–30.
84. Yao Z, Weinberg Z, Ruzzo WL. CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics* 2006;**22**(4):445–52.
85. Rabani M, Kertesz M, Segal E. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. *Proc Natl Acad Sci USA* 2008;**105**(39):14885–90.
86. Foat BC, Stormo GD. Discovering structural cis-regulatory elements by modeling the behaviors of mRNAs. *Mol Syst Biol* 2009;**5**:268.
87. Li X, Quon G, Lipshitz HD, *et al*. Predicting *in vivo* binding sites of RNA-binding proteins using mRNA secondary structure. *RNA* 2010;**16**(6):1096–107.
88. Kazan H, Ray D, Chan ET, *et al*. RNAcontext: a new method for learning the sequence and structure binding preferences of RNA-binding proteins. *PLoS Comput Biol* 2010;**6**:e1000832.
89. Kazan H, Morris Q. RBPmotif: a web server for the discovery of sequence and structure preferences of RNA-binding proteins. *Nucleic Acids Res* 2013;**41**:W180–6.
90. Maticzka D, Lange SJ, Costa F, *et al*. GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biol* 2014;**15**(1):R17.
91. Bahrami-Samani E, Penalva LO, Smith AD, *et al*. Leveraging cross-link modification events in CLIP-seq for motif discovery. *Nucleic Acids Res* 2015;**43**(1):95–103.
92. Macias S, Plass M, Stajuda A, *et al*. DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nat Struct Mol Biol* 2012;**19**(8):760–6.
93. Yoon JH, De S, Srikantan S, *et al*. PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nat Commun* 2014;**5**:5248.
94. Tollervey JR, Curk T, Rogelj B, *et al*. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat Neurosci* 2011;**14**(4):452–8.
95. Tuck AC, Tollervey D. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell* 2013;**154**(5):996–1009.
96. Kaneko S, Bonasio R, Saldaña-Meyer R, *et al*. Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell* 2014;**53**(2):290–300.
97. Di Ruscio A, Ebralidze AK, Benoukraf T, *et al*. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 2013;**503**(7476):371–6.
98. Gong C, Popp MW, Maquat LE. Biochemical analysis of long non-coding RNA-containing ribonucleoprotein complexes. *Methods* 2012;**58**(2):88–93.

99. Gumireddy K, Li A, Yan J, *et al*. Identification of a long non-coding RNA-associated RNP complex regulating metastasis at the translational step. *EMBO J* 2013;**32**(20):2672–84.
100. Scheibe M, Arnoult N, Kappei D, *et al*. Quantitative interaction screen of telomeric repeat-containing RNA reveals novel TERRA regulators. *Genome Res* 2013;**23**(12):2149–57.
101. Mazar J, Zhao W, Khalil AM, *et al*. The functional characterization of long noncoding RNA SPRY4-IT1 in human melanoma cells. *Oncotarget* 2014;**5**(19):8959–69.
102. Hämmerle M, Gutschner T, Uckelmann H, *et al*. Posttranscriptional destabilization of the liver-specific long noncoding RNA HULC by the IGF2 mRNA-binding protein 1 (IGF2BP1). *Hepatology* 2013;**58**(5):1703–12.
103. Castello A, Fischer B, Eichelbaum K, *et al*. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 2012;**149**(6):1393–406.
104. Baltz AG, Munschauer M, Schwanhsser B, *et al*. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* 2012;**46**(5):674–90.
105. Kwon SC, Yi H, Eichelbaum K, *et al*. The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol* 2013;**20**(9):1122–30.
106. Berman HM, Westbrook J, Feng Z, *et al*. The Protein Data Bank. *Nucleic Acids Res* 2000;**28**(1):235–42.
107. Muppirala UK, Honavar VG, Dobbs D. Predicting RNA-protein interactions using only sequence information. *BMC Bioinformatics* 2011;**12**:489.
108. Bellucci M, Agostini F, Masin M, *et al*. Predicting protein associations with long noncoding RNAs. *Nat Methods* 2011;**8**(6):444–5.
109. Wang Y, Chen X, Liu ZP, *et al*. *De novo* prediction of RNA-protein interactions from sequence information. *Mol Biosyst* 2013;**9**(1):133–42.
110. Lu Q, Ren S, Lu M, *et al*. Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics* 2013;**14**:651.
111. Suresh V, Liu L, Adjeroh D, *et al*. RPI-Pred: predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Res* 2015;**43**(3):1370–9.
112. Cheng Z, Zhou S, Guan J. Computationally predicting protein-RNA interactions using only positive and unlabeled examples. *J Bioinform Comput Biol* 2015;1541005.
113. Livi CM, Blanzieri E. Protein-specific prediction of mRNA binding using RNA sequences, binding motifs and predicted secondary structures. *BMC Bioinformatics* 2014;**15**:123.
114. Pancaldi V, Bähler J. In silico characterization and prediction of global protein-mRNA interactions in yeast. *Nucleic Acids Res* 2011;**39**(14):5826–36.
115. Yuan J, Wu W, Xie C, *et al*. NPInter v2.0: an updated database of ncRNA interactions. *Nucleic Acids Res* 2014;**42**:D104–8
116. He W, Cai Q, Sun F, *et al*. linc-UBC1 physically associates with polycomb repressive complex 2 (PRC2) and acts as a negative prognostic factor for lymph node metastasis and survival in bladder cancer. *Biochim Biophys Acta* 2013;**1832**(10):1528–37.
117. Djebali S, Davis CA, Merkel A, *et al*. Landscape of transcription in human cells. *Nature* 2012;**489**(7414):101–8.
118. Barrett T, Wilhite SE, Ledoux P, *et al*. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* 2013;**41**:D991–5.
119. Kolesnikov N, Hastings E, Keays M, *et al*. ArrayExpress update-simplifying data submissions. *Nucleic Acids Res* 2015:**43**:D1113–6
120. Blin K, Dieterich C, Wurmus R, *et al*. DoRiNA 2.0-upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res* 2015:**43**:D160–7.
121. Khorshid M, Rodak C, Zavolan M. CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res* 2011;**39**:D245–52.
122. Li JH, Liu S, Zhou H, *et al*. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 2014;**42**:D92–7.
123. Paz I, Kosti I, Ares M Jr, *et al*. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res* 2014;**42**:W361–7.
124. Bonnici V, Russo F, Bombieri N, *et al*. Comprehensive reconstruction and visualization of non-coding regulatory networks in human. *Front Bioeng Biotechnol* 2014:**2**:69.
125. Shang D, Yang H, Xu Y, *et al*. A global view of network of lncRNAs and their binding proteins. *Mol Biosyst* 2015;**11**(2):656–63.
126. Yang JH, Li JH, Jiang S, *et al*. ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res* 2013;**41**:D177–87
127. Li JH, Liu S, Zheng LL, *et al*. Discovery of Protein-lncRNA Interactions by Integrating Large-Scale CLIP-Seq and RNA-Seq Datasets. *Front Bioeng Biotechnol* 2015;**2**:88.
128. Scheibe M, Butter F, Hafner M, *et al*. Quantitative mass spectrometry and PAR-CLIP to identify RNA-protein interactions. *Nucleic Acids Res* 2012;**40**(19):9897–902.
129. Chu C, Qu K, Zhong FL, *et al*. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell* 2011;**44**(4):667–78.
130. Simon MD, Wang CI, Kharchenko PV, *et al*. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci USA* 2011;**108**(51):20497–502.
131. Engreitz J, Lander ES, Guttman M. RNA Antisense Purification (RAP) for Mapping RNA Interactions with Chromatin. *Methods Mol Biol* 2015;**1262**:183–97.
132. Chu C, Spitale RC, Chang HY. Technologies to probe functions and mechanisms of long noncoding RNAs. *Nat Struct Mol Biol* 2015;**22**(1):29–35.
133. 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 USA* 2011;**108**(24):10010–5.
134. Helwak A, Kudla G, Dudnakova T, *et al*. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 2013;**153**(3):654–65.
135. Mattei E, Ausiello G, Ferrè F, *et al*. A novel approach to represent and compare RNA secondary structures. *Nucleic Acids Res* 2014;**42**(10):6146–57.