



## Review

## Ribonomic approaches to study the RNA-binding proteome

Camilla Faoro, Sandro F. Ataide\*

School of Molecular Biosciences, University of Sydney, NSW, Australia



## ARTICLE INFO

## Article history:

Received 22 May 2014

Revised 4 July 2014

Accepted 4 July 2014

Available online 19 August 2014

Edited by Maria Papatrifaftayllou

## Keywords:

RNA-binding proteins (RBPs)

Ribonucleoprotein particles (RNPs)

Ribonome

RNA–protein interactions

RNA-centric methods

Non-coding RNAs (ncRNAs)

## ABSTRACT

**Gene expression is controlled through a complex interplay among mRNAs, non-coding RNAs and RNA-binding proteins (RBPs), which all assemble along with other RNA-associated factors in dynamic and functional ribonucleoprotein complexes (RNPs). To date, our understanding of RBPs is largely limited to proteins with known or predicted RNA-binding domains. However, various methods have been recently developed to capture an RNA of interest and comprehensively identify its associated RBPs. In this review, we discuss the RNA-affinity purification methods followed by mass spectrometry analysis (AP-MS); RBP screening within protein libraries and computational methods that can be used to study the RNA-binding proteome (RBPome).**

© 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

### 1. Introduction

In recent years, our understanding of the emerging role of RNA–protein interactions in regulating and coordinating gene expression has substantially evolved. The outcome was a renewed interest in post-transcriptional regulation and elucidation of the components of ribonucleoprotein complexes (RNPs).

Ribonucleoprotein (RNP) particles are composed of one or more RNA molecules and at least one protein, and can vary in size from the large ribosome to the small nuclear RNPs (snRNPs). Some form stable functional structures, whereas others, such as the eukaryotic spliceosome [1], are assembling and disassembling in a spectacularly dynamic manner during their functional cycle. The collection of RNPs within a cell composes the ribonome, which can be described as a highly sophisticated, self-sustaining and self-limiting regulatory system that inextricably interconnects the transcriptome and the proteome [2]. The ribonome, in mammalian cells, consists of thousands messenger RNPs (mRNPs) that contain mRNAs and their associated non-coding RNAs along with RNA-binding proteins (RBPs) and auxiliary proteins.

Post-transcriptional events such as RNA metabolism, processing, transport, translation and storage are regulated by RBPs [3–5]. Eukaryotic RNAs are dynamically organized into different RNP structures, and all these structures are needed for RNA

transcription, processing and function. For example, pre-mRNA processing steps involving splicing, editing and polyadenylation are mediated by RBPs as soon as pre-mRNAs emerge from the RNA polymerase [6–8]. After these processing steps, the mRNAs are exported to the cytoplasm by nuclear pores through the formation in the nucleus of an mRNP export complex that is capable of shuttling back and forth through the nuclear pores [9]. Complexes consisting of motor proteins and RBPs, or even the signal recognition particle, may further contribute to localization of RNAs to specific subcellular regions [10,11]. Thereby, transport of mRNAs has to be accompanied by translational repression, which is mediated by certain RBPs [12,13].

Quantitative genomic, proteomic and microscopy analyses have recently given rise to the field of RNP biology. The functions of some RBPs have been characterized on the basis of RNA-binding and regulatory elements in their sequence, as well as through the analysis of their associated tissue-specific expression profiles. These approaches have revealed protein–nucleic acid and protein–protein interactions within RNPs, as well as mRNP protein localization, dynamics and processing. However, very little is known about how RBPs control constitutive expression patterns. For example, the study of RBPs involved in large RNA–protein complexes that regulate mRNA metabolism is difficult owing to the plasticity and complexity of these systems.

Currently, only a few RNPs have been studied extensively and the function of most RBPs can only be predicted on the basis of sequence similarity. A fundamental question has been how to determine the role of each RNP component. Elucidation of these

\* Corresponding author. Address: School of Molecular Bioscience, The University of Sydney, Rm 672, Bldg G08, NSW 2006, Australia. Fax: + 61 2 9351 5858.

E-mail address: [sandro.ataide@sydney.edu.au](mailto:sandro.ataide@sydney.edu.au) (S.F. Ataide).

components involves two complementary approaches. In protein-centric methods the goal is the identification of RNAs that are bound to a known protein, whereas in RNA-centric methods the aim is the identification of proteins associated with an RNA of interest.

In this review, we overview the biochemical and genetic strategies that have been developed in recent years to identify proteins that bind to a specific RNA target in a genomic scale, without prior knowledge of protein candidates. In particular, we consider the advantages and disadvantages of the main methods described, focusing on the subtypes of RNA that can be investigated. This review does not cover methods implicated in the analysis of RNAs associated to a precise RNP; excellent reviews of these topics can be found in literature [14–16].

## 2. Focus on the RBPome

The RBPome can be defined as the interface where RNA and RBPs meet, for controlling many aspects of gene regulation [8,17]. RNA–protein complexes are very dynamic and can undergo extensive remodelling. Therefore, they can control the pattern and spatiotemporal regulation of various sets of genes, including genes that are involved in cell cycle progression, cell differentiation, organ morphogenesis and embryonic development. Consistent with their physiologic importance, perturbation of RBP expression or function has been linked to several diseases, including metabolic and neurological disorders, muscular atrophies, fragile X syndrome, autoimmune pathologies and cancer [18,19].

### 2.1. Conserved RBP domains

The number and variety of RBPs, especially of metazoan RBPs, that are reported in the literature are rapidly expanding [20,21]. Eukaryotic cells encode a large number of RBPs that allows them to combine with each RNA in unique forms to generate a vast plethora of RNPs [22].

The majority of eukaryotic RBPs contain unique RNA-binding and protein–protein interaction domains. RBPs bind RNA through a large set of protein domains, including RNA recognition motifs (RRMs); zinc fingers; K homology (KH) domains; serine arginine (SR) domains; Sm domains; double stranded RNA-binding domains (dsRBDs); DEAD/DEAH-box helicase (DEAD-box) motifs; and less common domains, such as Piwi/Argonaute/Zwille (PAZ) and Pumilio/FBF (PUF) domains. Thus, a large number of proteins have been predicted to be RBPs based on the presence of these commonly occurring RNA-binding domains. Accordingly, in the nematode, *Caenorhabditis elegans*, it has been predicted that as many as 887 genes may encode RBPs [23]. In the yeast, *Saccharomyces cerevisiae*, approximately 600 transcripts are reported to code for putative RBPs, as well as almost one quarter of annotated human genes [24,25].

Although domain conservation has been used to identify several RBPs in different organisms, targets of these RBPs are poorly understood. A high degree of modularity exist at the structural level, in RBPs, as most contain multi-domains composed of at least one RNA-binding and auxiliary domains that can promote RBP interactions with other proteins. This modularity generates both multivalent RNA specificity and functional diversity within the RBPs [26]. Indeed, evidence suggests that RBPs are multi-targeted [27–29].

### 2.2. RBP modularity

The multi-targeted function of RBPs has led to the introduction of a post-transcriptional RNA operon (PTRO) model for the regulation of eukaryotic gene expression [30]. The PTRO model proposes that RBPs in eukaryotes coordinate groups of mRNAs coding for

functionally related proteins. Correlation between polycistronic mRNAs from bacterial operons and the regulation of multiple monocistronic mRNAs by RBPs has led to this theory. The PTRO model proposes that RBPs in eukaryotes coordinate groups of mRNAs coding for functionally related proteins. These RNA regulons function on the basis of an RNP-driven process, in which multiple *trans*-acting factors, such as RBPs, non-coding RNAs (ncRNAs) and metabolites, bind *cis*-acting elements within their mRNA targets. Multiple *cis*-regulatory elements on each mRNA results in a modular USER (untranslated sequence element for regulation) code [22]. These USER codes determine the association of specific RNA-binding factors with each mRNA in order to cooperate or compete to the regulatory fate of that molecule. Within the ribonome, each mRNA exists in different forms dictated by the multiple activities and compositions of RNPs that controls the stability or translatability of the mRNA in response to cellular signals. Thus in a concerted manner the RBPs can regulate multiple mRNA targets as well as one another's mRNAs within the ribonome.

Studies on RNP within different species indicate that the same RBP can bind mRNAs encoding proteins with similar functions [31]. The most studied example of a post-transcriptional RNA operon was that originally described by Gerber et al. which examined five Pumilio RBP family members in yeast and found a sub-population of mRNAs encoding proteins with related functions bound by the RBPs [32,33].

Strong evidence for widespread regulation at the post-transcriptional level arises from selective binding of RBPs to sets of mRNAs encoding functionally related proteins. Similarly, an important role for cell maintenance can be hinted by RBPs function in creating and maintaining spatial organization in the cell, upon combining protein production and mRNA decay in the same location [34,35]. By the organization of monocistronic mRNA in functional groups, RNA regulons are responsible for the temporal and spatial coordination between the co-transcriptional and post-transcriptional regulation of gene expression. In this context, the post-transcriptional operon model can be used to explain the discrepancy that is often observed between mRNA transcript levels and the final amount of protein produced.

The only way to decipher the RNA operon code and the role of individual RBPs in post-transcriptional regulation will be through a mechanistic characterization of RBPs binding preferences.

Recent developments in high-throughput technologies, such as CLIP [36], RIP-CHIP [37] and the RNA*compete* assay [38] have allowed for the identification of RNA targets of RBPs in a genome-wide manner [39–41]. These methods work on a similar concept where the RBP complex together with its target RNAs is first extracted and then the target RNA identified by microarray or sequencing analysis. Computational approaches, such as RBPmap [42] and RNAcontext [43] have been developed for accurate prediction and mapping of transcriptome-wide RBP binding sites. RBPs typically bind RNAs in a sequence-specific manner, interacting with degenerate and/or short sequence motifs [44]. However, sequence analysis is not sufficient to predict the binding sites or RNA targets of RBPs, as the motifs do not contain enough information. Indeed, it is generally believed that other factors, such as the accessibility of binding sites within folded RNA molecules or the clustering of binding sites which may allow for binding of RBP multimers or multidomain RBPs contribute to their specificity [45,46]. In addition, some RBPs bind indiscriminately along transcripts, and this complicates the prediction of the binding specificity of RBPs.

Protein-centric approaches can identify most RNA targets of an RBP, independent of prior knowledge of phenotypes or expression profiles, and uncover mRNA targets from functionally paralogous genes co-regulated by an RBP [47]. Although useful information regarding the location of the RNA sequence interacting with

proteins can be achieved by protein-centric methods, identification of protein partners that bind to a specific RNA remains challenging.

### 3. RNA-centric approaches for RBPome analysis

The characterization of the RBPome associated to a given RNA of interest is a fundamental area of exploration in elucidating post-transcriptional networks. The same RNA can be bound and regulated by multiple factors as indicated by several studies comparing multiple RBPs and their targets [24,48]. Multiple RBPs can associate with a single RNA simultaneously or sequentially and a single RBP or non-coding RNA can potentially target a large number of different RNA species [49–52]. The cooperation and competition among these factors result in the combinatorial regulation of a given message. Understanding the remodelling of the RBPs proteome will provide great insight into the dynamic interplay of RNAs and RBPs during an mRNA life cycle.

RNA-centric approaches have been widely used to gain initial insights into RBPs that cooperate for functional regulation of a given RNA or class of RNAs, including mRNAs and ncRNAs. The overall aim of these approaches is to predict the functional consequences of numerous RNA–protein interactions that take place sequentially or simultaneously on a target RNA. Compared with protein-centric methods that employ top-down approaches to study RNA interactions of individual RBPs using the protein as the bait in enrichment steps, RNA-centric methods employ tagged RNA as the bait in order to identify proteins interacting with it. Moreover, these methods can be used to analyse the cooperative, independent or competitive association of different proteins with the same mRNA.

These methods can be sub-divided into two main categories: RNA-affinity purification followed by mass spectrometry analysis (RAP-MS) assays; and RBP screening within protein libraries (Table 1).

### 4. RNA-affinity purification assays

RNA-affinity-based RBP purification is a versatile strategy that allows for the identification of RNA–protein interactions in both subcellular compartments and in vivo systems. This technique is particularly suited for studying RNA–protein interactions that are dictated by the formation of multi-protein complexes, as it facilitates the study of individual interactions within these larger molecular frameworks. RNA-affinity purification is an excellent method to purify RNP complexes from mammalian cells, in which post-transcriptional modifications, such as phosphorylation, are often used by regulatory proteins to increase or decrease the affinity for their targets. Finally, coupled with cross-linking, RNA-affinity purification has the ability to capture both stable and transient interactions. One advantage of RNP cross-linking, used on in vivo or in vitro-assembled complexes, is the potential to maintain a stable composition of RNP complexes even under the stringent washing conditions that are required for reducing the isolation of non-specifically-binding proteins.

In RNA-affinity capture-based approaches, RNP complexes are assembled in vitro or in vivo, and the bait RNA from cell/tissue lysates is immobilized to a solid support either covalently or non-covalently. Then, following a series of washing steps to remove non-specifically bound proteins, the RNP complexes are released from the chromatographic matrix and subjected to mass spectrometry (MS) analysis (Fig. 1).

**Table 1**  
RNA-centric methods to identify novel RBPs.

Category	RNA affinity capture and MS analysis	RBP screening within protein libraries
Description	In vivo and in vitro approaches to identify RBPs associated to an RNA of interest through immobilization of the bait RNA into a chromatographic matrix. After a series of washing steps to remove non-specifically bound proteins, the RNP complexes are eluted from the solid support and subjected to mass spectrometry analysis	In vivo and in vivo systems to detect and analyse RBPs that interact with a known RNA. By providing a link between each protein of a cDNA library and its encoding gene, clones with desirable features are subjected to iterative rounds of selection and amplification
Protocols	<ul style="list-style-type: none"> <li>• TRAP</li> <li>• RAT</li> <li>• MS2-BioTRAP</li> <li>• RaPID</li> <li>• Csy4 system</li> <li>• Interactome capture</li> <li>• PAIR</li> <li>• RiboTrap</li> <li>• CHART</li> <li>• 4 × S1m</li> <li>• ChIRP</li> <li>• Affinity medium</li> <li>• MS2-TRAP</li> <li>• RNase-assisted</li> </ul>	<ul style="list-style-type: none"> <li>• Phage display</li> <li>• mRNA display</li> <li>• Bacterial antitermination assay</li> <li>• Northwestern</li> <li>• Y3H</li> <li>• Protein microarray</li> <li>• IVC</li> <li>• Tat-hybrid system</li> </ul>
Advantages	<ul style="list-style-type: none"> <li>• Easy purification steps</li> <li>• Applicable to mammalian cells</li> <li>• Possibility to study RNA–protein interaction under physiological conditions</li> <li>• Flexibility and versatility</li> </ul>	<ul style="list-style-type: none"> <li>• Some of these methods can be used to study interactions in vivo</li> <li>• High sensitivity</li> <li>• High-throughput</li> <li>• Minimal amounts of starting material required</li> <li>• Amplification of detection</li> </ul>
Disadvantages	<ul style="list-style-type: none"> <li>• High background noise</li> <li>• Perturbation of RNA folding through the insertion of tags</li> <li>• Involves challenging design of antisense oligonucleotide</li> <li>• Large amounts of starting material required</li> <li>• Time-consuming</li> <li>• Identification of only relatively abundant proteins</li> <li>• Limited throughput</li> <li>• Involves challenges associated with transfection of cells</li> </ul>	<ul style="list-style-type: none"> <li>• High number of false positives</li> <li>• Some of these methods are limited to binary interactions</li> <li>• No direct correlation between protein abundance and activity</li> <li>• With some methods, interactions are studied under artificial conditions</li> </ul>
References	[56,57,60,62,64,65,67–69,72,86–88,91–93]	[72,74,79,81,85,88,92]

**Abbreviations:** TRAP, Tandem RNA affinity purification; RAT, RNA affinity in tandem; MS2-BioTRAP, MS2 in vivo biotin tagged RNA affinity purification; RaPID, RNA-binding protein purification and identification; PAIR, Peptide nucleic acid-assisted identification of RBPs; CHART, Capture hybridization analysis of RNA targets; ChIRP, Chromatin isolation by RNA purification; MS2-TRAP, MS2-tagged RNA affinity purification; Y3H, Three-hybrid system; IVC, In vitro compartmentalization.

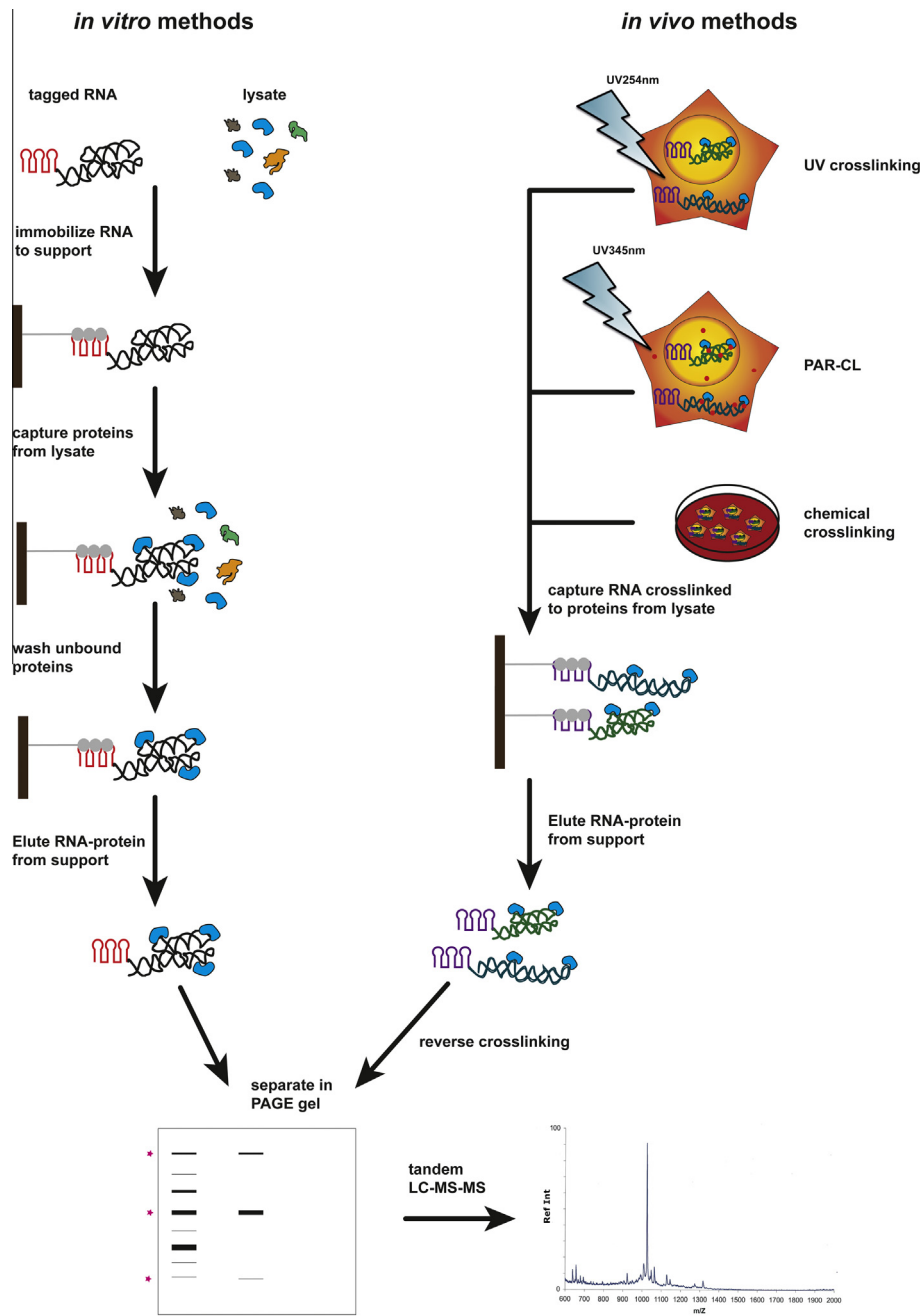
There are several variants for the affinity capture of a target RNA and these can be grouped into two major classes depending on the strategy used for RNA immobilization to chromatographic support: tagged RNA affinity purification; and antisense RNA-based affinity purification approaches.

#### 4.1. RNA-tag capture

In RNA-tag-mediated purification approaches, *in vitro*-synthesized RNA is chemically tagged through the incorporation of modified ribonucleotides that contain biotin, fluorescent dyes, digoxigenin or other compounds. The high affinity of biotin to

streptavidin has made this association a widely used tool for affinity chromatography chosen by many researchers [53,54]. However, one important drawback of chemical RNA labelling is that in some cases the chemical modifications can affect the secondary structure of RNA, leading to structural rearrangements that interfere with complex formation.

Alternatively, numerous natural or artificial aptamers can be incorporated within the RNA during *in vitro* or *in vivo* transcription. Aptamers are functional oligonucleotide sequences that have the ability to bind specifically and with high affinity to proteins, peptides and other small molecules [55]. S1 and D8 aptamers are among the most popular artificial aptamers used in ribonomic



**Fig. 1.** RNA affinity capture and proteomic analysis. *In vitro* approaches. Ribonucleoprotein complexes are assembled using cell lysates *in vitro* in the presence of the tagged RNA of interest, which has been immobilized on a solid support. After several washing steps that remove non-specific interactions, specific RBPs are released from the support and subjected to proteomic analysis. *In vivo* approaches. Cells are UV irradiated (c-CL (254 nm) or PAR-CL (365 nm)) or treated with formaldehyde in order to “freeze” physiological RNA–protein interactions. Cells are lysed and the ribonucleoprotein complexes are captured from solution. After washing to remove non-specific proteins, the bound complexes are eluted and released from the RNA (by heating the sample or by RNase digestion). Finally the proteins are subjected to proteomic analysis. In both approaches, samples are usually separated by gel electrophoresis and specific protein bands are excised, digested with proteases and identified by tandem LC–MS–MS.



assays. S1 and D8 were originally selected using SELEX procedures to obtain a sequence with high binding affinity to Sephadex and Streptavidin affinity matrices respectively [56,57]. Since then, the streptavidin aptamer D8 has been used to purify ribonuclease P (RNase P) from *S. cerevisiae*, and even the ribosome [58]. The Sephadex aptamer S1 has been used to study the *S. cerevisiae* RNase P with a differential protein composition between the precursor and mature forms of the RNP complex [59].

Well-characterized protein-binding RNA sequences can also be incorporated as aptamers within the RNA to isolate RBPs. The RNA tagging most widely used for identification of RNA–protein interaction is the bacteriophage MS2 coat protein and its cognate RNA [60,61]. The limiting step when using the MS2 RNA–protein interaction in the affinity purification is elution of the purified complex under native conditions due to the high binding affinity of the interaction. A way to overcome this restriction is to fuse to the MS2 coat protein to an additional peptide, such as the maltose-binding peptide (MBP) [62]. Alternatively, in some studies a protease cleavage site was inserted between the MS2 coat protein and the other protein [58].

However, there are some limitations to the broad use of aptamers. The incorporation of a foreign sequence to the RNA bait may alter the RNA structure and possibly the formation and composition of the RNP complex. In addition, if there is no structural information regarding the target RNA it can be difficult to predict the best position for insertion of the aptamer tag. Moreover, when used in a cellular context, efficiency of artificial aptamers decreases inevitably as their lifetime is drastically reduced owing to their degradation by nucleases. Thereby, the recovery yield of RBPs might be low but could be improved using nuclease-resistant aptamers or scaffold technology [63].

#### 4.2. Antisense RNA capture

Antisense RNA-based affinity purification approaches represent another powerful tool for analysing the protein components of RNPs. The structure and function of small nuclear ribonucleoprotein (snRNP) complexes have been studied using antisense oligonucleotides [64] and later they were adapted and modified for the purification of a variety of RNPs, including the telomerase [65] and more recently, the protein components of the small nuclear ribonucleoprotein (snoRNP) complex MBII-52 [66].

Affinity-tagged antisense oligonucleotides can be easily immobilized onto a chromatographic support, using the streptavidin–biotin interaction, to isolate the RNA bait and its associated proteins. In this case there is no requirement for chemical or sequence modification of the RNA of interest. The resulting complex can then be eluted either under denaturing conditions, or via a competitor oligonucleotide. This allows for RBP release under native conditions. However, the design of antisense oligonucleotides with high affinity to accessible single-stranded regions of highly structured RNAs is often a challenging step in this approach [67].

Depending on the type of detection and application required, each of these methods for the tagging and affinity purification of RNA molecules provides different advantages and limitations that must be considered. Recently developed affinity purification methods have provided a deep understanding of the higher-ordered structures of multi-subunit RNPs that are often low-abundance macromolecular complexes.

### 5. RNA affinity-based strategies for RNP characterization

One important variable in large-scale RNP analysis is the method used to define the RBP population that bound a given RNA. A number of different strategies have been developed to

isolate RNPs by using RNA-centric affinity purification methods. All existing protocols involve a similar concept but differ with regard to the specific procedure used for purifying RBP–RNA complexes and characterizing target RBPs. All strategies can be used to successfully isolate RBPs, but each method presents distinct difficulties that requires special consideration when designing the experiments.

#### 5.1. Tandem RNA affinity purification (TRAP)

Tandem RNA affinity purification (TRAP) tagging is a two-step affinity purification method developed by Krause and colleagues that facilitates purification of RNAs along with their associated proteins, RNAs and other small molecules [68]. The TRAP-tag method enables through the use of two RNA tags the purification of RBPs by approximately a million-fold, reducing significantly the chance for contaminants to be retained in the eluate. The sequence of the RNA of interest is tagged at its 5' or 3' end with two different tags, one sequence must interact with a ligand in a reversible manner without disrupting the RNA–protein complex formed on the target RNA sequence. Functional complexes containing the double-tagged RNA are often purified from cell extracts using a tandem affinity purification exploiting each of the tags.

#### 5.2. RNA affinity in tandem (RAT)

Hogg and Collins have characterized the composition of the 7SK RNP using an approach similar to TRAP, namely the RNA affinity in tandem (RAT) technique [69]. The high efficiency of the RAT is largely due to further optimization of affinity purification. The RAT tag consists of two hairpin stem-loops that interact specifically with *Pseudomonas aeruginosa* phage 7 (PP7) coat protein. Compared with the more commonly used MS2 coat protein, PP7 coat protein maintains a high affinity to its cognate RNA across a wider range of ionic strength and pH [70,71].

#### 5.3. $4 \times S1m$ -mediated RNA affinity chromatography

Leppek and Stoecklin generated an optimized streptavidin-binding RNA aptamer (S1 RNA9, which they termed S1modified ( $4 \times S1m$ )).  $4 \times S1m$  has a higher affinity for streptavidin compared with previous aptamers [72]. The optimization of structure and repeat conformation of the aptamer S1 RNA allowed an 15-fold increase in the recovery of a reporter mRNA expressed in cells, as compared with the established MS2 and PP7 systems.  $4 \times S1m$ -mediated RNA chromatography with a reporter mRNA containing the AU-rich element (ARE) of mouse tumor necrosis factor  $\alpha$  (TNF $\alpha$ ) has been used to purify ARE-binding proteins from cellular extracts. Accordingly, Roquin was identified as the major protein that interacts with the constitutive decay element (CDE), a stem-loop RNA degradation motif located downstream of the ARE at the 3' untranslated region (UTR) of the TNF $\alpha$  mRNA [73].

#### 5.4. MS2 in vivo biotin tagged RNA affinity purification (MS2-BioTRAP)

Tsai and co-workers have developed a strategy called MS2 in vivo biotin tagged RNA affinity purification (MS2-BioTRAP) to isolate in vivo-assembled RNP complexes [74]. In particular, this technique was used to identify the proteins bound to internal ribosome entry site (IRES)-containing mRNAs rather than to canonical mRNAs that undergo Cap-dependent translation. The principle of MS2-BioTRAP relies on the co-expression of two vectors: one vector codes for the RNA of interest tagged with a cluster of RNA stem-loops that can be recognized by the bacteriophage RBP MS2; and the second vector codes for the MS2 protein fused to an HB tag

[75]. The HB tag includes two hexahistidine tags – a TEV cleavage site and a signal sequence for *in vivo* biotinylation – and enables rapid and effective one-step purification of MS2 with its associated complexes of the stem-loop tagged RNA of interest. The biotinylated MS2 protein can then be removed from the RNP complexes using streptavidin-coated beads or Ni-coated solid support under native or denaturing conditions. Alternatively, RBPs can be directly eluted upon TEV cleavage. There are a few limitations with this method; first of all it is only applicable in easy-to-transfect cells. Second, the fact that one or two of the interacting molecules are over-expressed might lead to experimental artefacts, including mis-localization and the formation of spurious interactions or altered activities that modify the composition of RNP complexes. Despite its limitations, we predict that the MS2-BioTrap strategy will be widely used to analyse other RNP complexes [76–79].

### 5.5. RBP purification and identification (RaPID)

RBP purification and identification (RaPID) is an aptamer-based RNA affinity purification strategy that is similar to MS2-BioTrap [80]. The RaPID pull-down is a highly sensitive procedure that utilizes a MS2-GFP-SBP fusion protein. The streptavidin-binding protein (SBP) tag allows for the purification of RNA–protein complexes using streptavidin-conjugated beads, while the reporter tag allows for the localization of the RNP particle *in vivo* using fluorescence microscopy. By isolating tagged OXA1 mRNA using RaPID, Slobodin and colleagues identified Sec27, a yeast COPI subunit, as a candidate interacting protein [81]. In addition to identifying known and unknown proteins that interact with a specific mRNA, RaPID purification also enables *in vivo* visualization of the intracellular localization of the RNA of interest and quantification using western blot analysis of the fluorescent reporter. Moreover, the high-affinity interaction between SBP and streptavidin allows for stringent washing of the bound RNP complexes, improving the signal-to-noise ratio with a simple and specific elution step using biotin as a competitor. However, RaPID shares the same shortcomings of the previously mentioned aptamer-mediated strategy MS2-BioTRAP.

### 5.6. The CRISPR Csy4 – affinity system

Lee et al. generated a highly effective method of RNA–protein complex purification that is based on an engineered version of the clustered regularly interspaced short palindromic repeats (CRISPR) Csy4 endoribonuclease [82]. Csy4 recognizes a 16-nt hairpin sequence with a high affinity ( $K_d = 50$  pM), making its interaction with target RNAs one of the highest-affinity RNA–protein interactions of this size reported to date.

An inactivate biotinylated version of Cys4 binds irreversibly to the RNA of interest tagged with the Cys4 hairpin sequence at their 5' end. The cleavage activity of Cys4 can be inactivated without affecting substrate binding affinity or specificity and be rescued in the presence of imidazole. Once immobilized on a solid support Cys4 is activated to cleave the RNA removing the hairpin tag and releasing the RNA with its bound proteins. This specific elution via conditional enzymatic cleavage enables recovery of specific RBPs with few false-positives. Although the authors utilized this protocol to identify the RBPs associated to three pre-miRNAs (pre-let-7a, pre-miR-200a and pre-miR-342) Cys4 affinity purification is versatile and has a potential application in elucidating the RBPome associated to other classes of RNA.

### 5.7. Interactome capture

Recently, a protocol called interactome capture has been developed by Castello and colleagues for the identification of RBPs that

specifically associate with mRNAs in living cells, providing a complete picture of RBP activity [83].

Two recent studies revealed hundreds of novel RBPs generating a global mRNA interactome in HeLa [84] or HEK-293 [25] cell culture systems using the interactome capture technique. Using *in vivo* UV-crosslinking of RBPs to polyadenylated RNAs, covalently bound proteins are captured with oligo(dT) magnetic beads. Followed by purification under stringent conditions to eliminate contamination with non-crosslinked proteins and degradation of the cellular mRNA with the use of RNases, the polyA-bound interactomes are isolated.

This method presents some notable advantages over previous RBP identification methods. First, nucleic acid hybridization is stable in the presence of ionic detergents and in high-salt buffers allowing for efficient removal of polypeptides that are either associated non-covalently with the RNA, or involved in protein–protein interactions. Second, this protocol employs both conventional crosslinking (cCL–254 nm) and photoactivatable-ribonucleoside-enhanced crosslinking (PAR-CL–365 nm) in parallel taking advantage of the crosslinking chemistries of these techniques. Third, RNAs of interest do not require any chemical or sequence modification and the interacting molecules are expressed at endogenous levels. However, this approach fails to detect RBPs that are not expressed in the cell lines used or do not bind polyadenylated RNAs. Moreover, several RBPs that do not directly interact with nucleic acid bases but instead interact with other features, such as the sugar phosphate backbone or double stranded RNAs (dsRNAs), often lack the appropriate spatial arrangements between RNA bases and aromatic amino acids required for efficient UV cross-linking. Furthermore, because UV light only cross-links direct RNA–protein interactions, it cannot capture interactions that occur through a complex of multiple proteins. Finally, the major challenge of this method is the functional validation of all identified RBPs and putative RNA-binding domains.

### 5.8. Peptide nucleic acid (PNA)-assisted identification of RBPs (PAIR)

The Peptide nucleic acid (PNA)-assisted identification of RBPs (PAIR) is a procedure that has been used to identify the RBPs associated with ankylosis (ank) RNA, a panneuronal dendritically localized RNA [85]. This assay utilizes specific mRNA-binding probes (PNAs) that have the ability to cross the cell membrane of living cells and hybridize to complementary sequences on selected endogenous mRNAs. PNAs used in PAIR also contain the photoactivatable amino acid adduct *p*-benzophenylalanine (Bpa), which, after exposure of cells to UV light, can covalently cross-link with the RBPs that are associated (located at a distance  $\leq 4.5$  Å) to the RNA of interest. The PAIR technique allows for simultaneous, quantitative analysis of multiple RNAs and their associated RBPs in manner that reflects the interactions in living cells. Compared with the tag-mediated strategies cited before, PAIR offers the potential to analyse the endogenously formed RNP complexes freely of any fluctuation that could have been caused by the variable expression levels of transfected plasmids. Moreover, the PAIR technology can enable identification of RNPs in specific gene regions or exons *in vivo* [86]. This method has the potential to be applied to any mRNA expressed in living cell, although some mRNA sequences may not be accessible to PNA, as a result of RNA secondary structure formation.

### 5.9. Affinity medium

Another method for the direct isolation of RNA interacting proteins uses an affinity medium consisting of an artificial version of the bait RNA. At one end, the RNA of interest is modified with an

additional tail, which is annealed to a complementary DNA that is covalently linked on the surface of aminosilane glass powder. Thus the RNA is free and projected away from the matrix, facilitating its interactions with proteins. Using this method, three proteins specifically interacting with the C/EBP $\beta$ 30-untranslated region (30-UTR) RNA were isolated and identified [87].

#### 5.10. RiboTrap

RiboTrap is the first commercial kit developed by MBL International Corporation to isolate RBPs and other proteins that associate with any specific RNA from either the cytoplasmic or nuclear extract of cultured cells [88]. The RNA is *in vitro* transcribed with 5-bromo-UTP (BrUTP) and incubated with a cell lysate to form BrU-labeled RNA–RBP complexes. These complexes are then captured using an anti-BrdU monoclonal antibody. Differential washes allow for analysis of both weakly- and tightly-bound RBPs. In addition, the subpopulation of mRNAs that are present in an mRNP complex can be identified and examined for the presence of common sequence elements, such as 5' or 3' UTRs, or common functional features. However, the multi-component nature of mRNP complexes can interfere with efficient immunoprecipitation because of inaccessibility of reactive polypeptide epitopes.

The RiboTrap Kit can be also used to identify RBPs associated to ncRNAs. However, the BrU-labeled positions and numbers in the target sequence should be considered when performing RiboTrap using short-length RNAs.

#### 5.11. RNase-assisted RNA chromatography

Regardless of the purification methods used, the required chromatographic support imposes technical challenges as it often results in background noise associated with non-specific protein binding. Usually more stringent conditions, such as increased salt concentration in the washing step of the RNA pull-down procedure, may reduce the background noise, but at the same time these conditions can also result in the loss of important RBPs that are weakly bound to the RNA of interest. In order to eliminate non-specific selection, without increasing the buffer stringency, Michlewski and Cáceres introduced the RNase-assisted RNA chromatography method [89]. In RNase-assisted RNA chromatography, chromatographic support-coupled RNP complexes are treated with a high concentration of ribonucleases during the elution step to release only the proteins that are directly or indirectly bound to RNA. By contrast, contaminating proteins remain bound to the chromatographic support.

#### 5.12. lncRNA-containing RNP (lncRNP) capture

Presently, little information is available regarding the complete set of RBPs that are associated with a specific long non-coding RNA (lncRNA). The molecular mechanisms underlying the functions of lncRNAs are yet to be determined. However, many lncRNAs function through interactions with proteins, which implies the importance of identifying the RBPs associated with a specific lncRNA.

Gong and colleagues described a general procedure aimed at retrieving cytoplasmic and poly-adenylated lncRNA-containing RNP complexes from mammalian cells using affinity purification [90]. The procedure relies on the insertion of the twelve copies of the bacteriophage MS2 coat-protein binding sites into an expression vector coding for the lncRNA of interest. Cells are co-transfected with the expression plasmid and a second plasmid encoding a FLAG-tagged MS2 coat-protein. RNPs containing the hybrid lncRNA are then immunoprecipitated using an anti-FLAG antibody. Formaldehyde crosslinking is used to increase the specific recovery of lncRNA-MS2-containing RNP. By using a similar approach,

Gumireddy and co-workers identified that the regulatory lncRNA (treRNA) acts as a platform for a new ribonucleic protein complex that inhibits translation of E-cadherin and promotes cell invasion [91].

#### 5.13. MS2-tagged RNA affinity purification (MS2-TRAP)

A systematic approach, termed MS2-tagged RNA affinity purification (MS2-TRAP), allows for analysis of RBPs that are associated with a target RNA in the cell [92]. The method is based on the incorporation of MS2 RNA hairpin loops to a RNA of interest, which is then co-expressed with an MS2 protein fused to an affinity tag.

Although the authors illustrated the use of this methodology by identifying microRNAs associated with a long intergenic (li)ncRNA other interactions, including RNA–RNA and RNA–protein interactions, can be investigated using this strategy. The MS2-TRAP approach resembles the RiboTrap method that was developed by Keene and colleagues [88].

#### 5.14. Chromatin isolation by RNA purification (ChIRP)

Chromatin isolation by RNA purification (ChIRP) was developed by Chu and colleagues, and allows for unbiased high-throughput discovery of lncRNA-bound DNA and proteins [93]. This RNA-centric pull-down approach was applied for the investigation of lincRNAs and the lncRNA HOTAIR [94]. ChIRP combines oligo-based RNA pull-down with deep sequencing of attached DNA. In the ChIRP procedure, *in vivo* crosslinking of cultured cells is followed by extraction of chromatin and RNA pull-down using multiple biotinylated oligonucleotides covering the entire stretch of the RNA of interest. Following isolation of DNA from the purified complexes, the extremely high coverage of next-generation sequencing (NGS) facilitates the detection of even small amounts of DNA. This method is applicable to any lncRNA, without prior knowledge of its structure or function. The PCR amplification step prior to sequencing is responsible for the high degree of detection, which is likely to make this NGS-based approach advantageous compared to MS-based protein detection when using low amounts of starting material.

#### 5.15. Capture hybridization analysis of RNA targets (CHART)

To determine the location of lncRNA interaction with chromatin, Simon and colleagues developed the capture hybridization analysis of RNA targets (CHART) method, a hybridization-based technique that enriches endogenous RNAs and their targets within chromatin extracts [95]. Short biotinylated oligodeoxyribonucleotides that are complementary to the endogenous RNA of interest are used for the enrichment of RBPs within the cross-linked chromatin extracts. The RNA targets, proteins and DNA, can be identified upon their enrichment under these conditions. A genome-wide analysis can be performed by deep sequencing the enriched DNA.

CHART was used to identify DNA and protein associated with lncRNAs from flies and humans [96]. In addition to determining the genomic and proteomic targets of an RNA, CHART can be used to examine other RNAs. The enriched material can be used for an RNA-IP, as CHART involves reversible cross-linking. CHART allows for the enrichment of the RNA and examination of proteins associated with it instead of common pulling down protein and RNA identification.

Once the RBPs have been isolated from the RNA baits, the next step is to identify the low abundant proteins of interest in a complex protein solution that very often contain multiple highly abundant proteins.



## 6. Proteomic analysis of isolated RBPs

Mass spectrometry (MS) is a technique used for characterization of biological samples and its use has increased in omics studies. The improved sensitivity of MS technology has potentiated the unbiased systematic identification of RBPs after RNA-affinity capture.

### 6.1. Comparative proteomics

Comparative proteomic analysis represents the best choice to identify RBPs interacting with a precise RNA binding site. To discriminate between the unspecific proteins and the proteins of interest, captured by the RNA bait, a control condition must be used. This control condition must derive from a set up, where the proteins of interest are absent from the eluate. The quality of the data generated depends on the choice of the control condition that must replicate the biological question. In general, proteins purified from an RNA bait are compared with proteins captured by a control RNA sequence, ranging from totally unrelated sequences [97–99] to antisense sequences [100], coding sequences [101], or even closely related sequences with a few mutated nucleotides [102]. The closer the control and original sequence are, the more relevant the data will be in terms of sequence-specific interactions. However, this type of controls can be provided for only precisely-defined binding sites, but not uncharacterized long sequences.

Sample comparison can be either qualitative or quantitative. Gel-based separations are usually used for qualitative comparisons after RNA-affinity purification [103,104]. Prior to MS-based proteomic analysis, RBPs are generally separated by gel electrophoresis in order to reduce the complexity of samples. Subsequently, proteins are stained with MS compatible staining, such as Coomassie blue or the more sensitive Sypro Ruby staining, and protein bands of interest are excised from the gel and digested with proteases prior to MS analysis. Reducing the sample complexity is an important step to successfully identify low abundant sequence-specific interacting proteins, as it provides information about the molecular weight of the isolated RBPs and allows for comparison with RBPs captured by a control bait or just by the chromatographic support itself [105]. Although gel-based protein separation remains a standard tool for proteomic research it presents substantial analytical limitations when used for the system-wide analysis of complex protein mixtures. A first limitation relates to sample capacity and detection sensitivity. A broad range of proteins binds RNA in a non-specific manner. Although the affinity of these proteins to RNA is lower than that of sequence-specific RBPs, their abundance is higher, and this may often result in their capture by the RNA bait. The presence of highly abundant non-specific proteins might impair the identification of sequence-specific RBPs by MS when a gel-based separation step is used. A second limitation relates to the labour intensive nature of this method that limits its throughput in “omics” studies.

Quantitative proteomic approaches that allow for simultaneous quantification of proteins in the sample and the control have recently become popular. Cells are chemically or metabolically labelled to generate differentially tagged protein pools for MS analysis, in which the isotopes of the proteins are compared to provide direct quantification. In chemical labelling, such as ICAT (isotope coded affinity tag labelled peptides), proteins or peptides are tagged through a chemical reaction [106], whereas in metabolic labelling, such as SILAC (stable isotope labelling by amino acids in cell culture), the label is introduced to the whole cell through the growth medium [107]. The advantage of the quantitative approach is that the ratios of peptides from the experimental and control samples can be directly compared to separate true binding

partners from non-specific interactions. In this way, when using a system with high background a quantitative MS approach can provide increased ability to discriminate between specific and non-specific binders.

### 6.2. Non-comparative proteomics

When the goal is to identify all the proteins interacting with a specific RNA sequence and the key binding sites have not been defined the approach is different. Long lists of putative candidates are generated from MS analysis following large-scale unbiased long RNA affinity purification. The major difficulty is to determine sequence-specific and unspecific interacting proteins from the lists.

The long RNA bait may capture a very large number of different proteins, and the low abundant sequence-specific RBPs may become undetectable. Therefore, gel-free approaches are recommended, as they enable improved chromatographic separation of the complex peptide mixture prior to the MS analysis [108]. Such shotgun approaches include the reduction of sample complexity through liquid chromatography, multidimensional fractionation, or the multidimensional protein identification technology (MudPIT) combining strong cation exchange chromatography with reverse phase chromatography [109]. Although these novel approaches facilitate high-throughput analyses of the proteome of a cell type and provide an overview of the major protein constituents, they should be used as complements of gel-based approaches.

RAP-MS analysis has been used to identify the RNA-binding proteome that is associated with different classes of RNA, including mRNA and ncRNAs. There are several analytical challenges for identifying proteins associated with an RNA molecule by RAP-MS. First, RNA is unstable and flexible, which impedes not only the isolation of RNA-bound proteins but also the design of functional tags or antisense oligonucleotides. Second, RAP approaches require a significant amount of starting material to isolate enough proteins for detection. Third, the amino acid sequence of proteins that interact with bait RNA cannot be amplified, complicating the identification of binding partners of individual low abundance RNAs. Finally, one major concern when aiming to isolate and characterize endogenously assembled RNP complexes is the phenomenon of post-lysis re-organization [110]. As native RNA-protein interactions are expected to be dynamic and highly sensitive to conditions such as ionic strength, precise concentrations and sub-cellular localization, co-purification cannot be assumed to necessarily reflect an endogenous interaction [111]. To circumvent this problem, methods such as UV or chemical crosslinking should be used to stabilize endogenous RNP complexes prior to cell lysis.

Improvements in RNA-affinity purification and characterization of RBPs will facilitate the study of composition, stoichiometry and dynamic of RNA-interacting complexes. Although this represents the future directions in this field, they currently suffer from limitations in the sensitivity of the MS identification.

Simultaneous to the development in RAP strategies, new RNA-ligand assays have emerged in the literature.

## 7. RBP screening within protein libraries

In vitro or in vivo screening for RNA sequence-specific proteins within large libraries is a high-throughput approach to identify RBPs associated to an RNA of interest (Fig. 2).

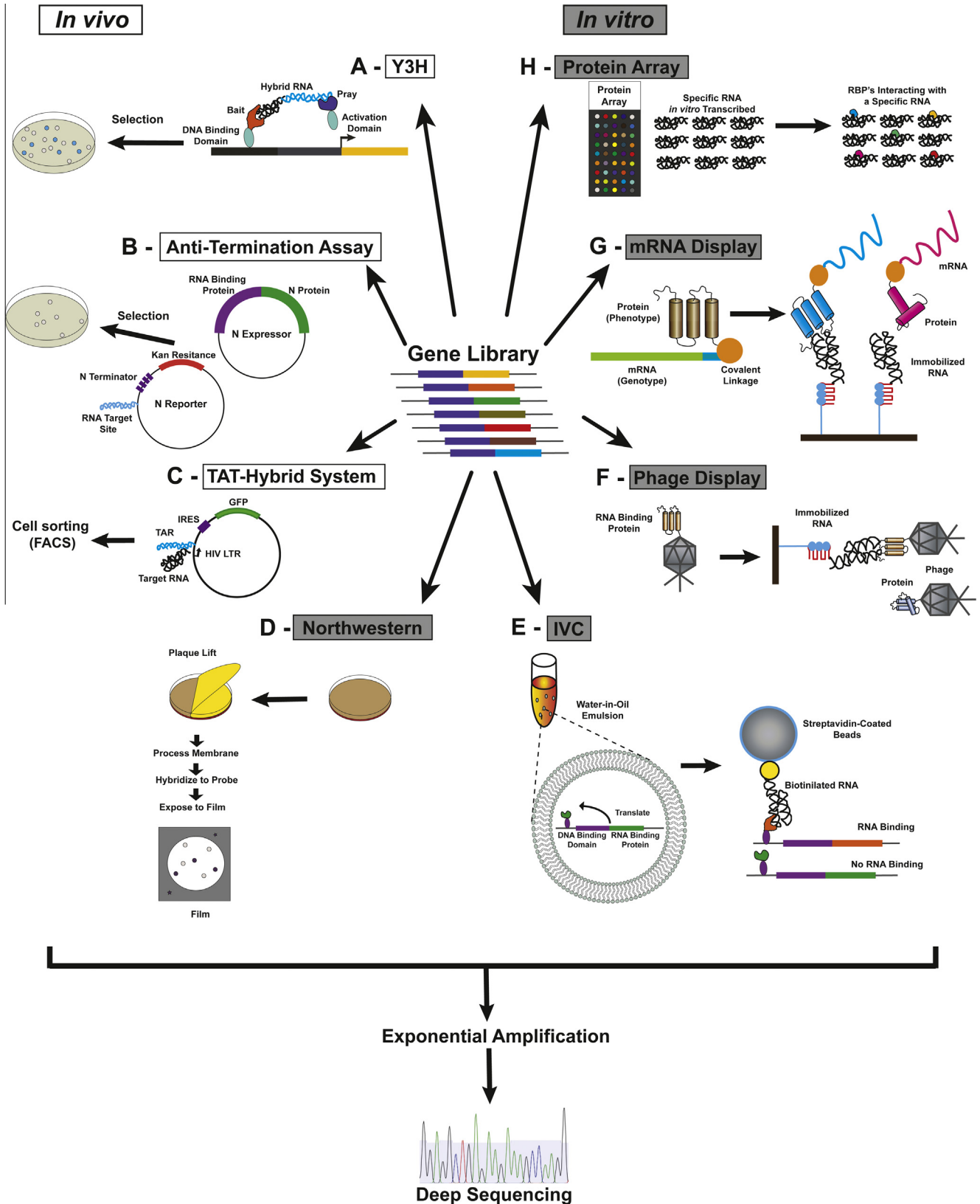
Three main steps are at the basis of all RBPs screening strategies: the construction of a gene library (genotype); the establishment of a linkage between each protein (phenotype) and its encoding gene (genotype); and the selection of desired proteins



(phenotype) from the library. Clones with desirable features, can be subject to iterative rounds of selection followed by amplification of the selected sub-library. As potential target proteins are overexpressed from expression vectors, these techniques can usually identify interacting proteins even when they are expressed at

very low endogenous levels and thus difficult to detect using traditional-biochemical methods.

*In vivo* systems can select for function in the context of cellular processes, whereas *in vitro* approaches have greater control over binding conditions upon sampling larger libraries and reduced



expression bias. The large library size increases the probability of identification of rare sequences, and also improves the diversity of the selected sequences. Finally, *in vitro* selection methods allow for *in vitro* mutagenesis [112] and recombination techniques during the process.

### 7.1. The yeast three-hybrid system

The yeast three-hybrid system is a useful tool for studying RNA–protein interactions in *S. cerevisiae* independent of the biological role of the RNA or protein [112]. This assay is a variation of the yeast two-hybrid system and has been successfully used to test candidate RNA–protein pairs, to identify RNA sequences that bind known RBPs, and to analyse RNP complexes. The system is based on the formation of a multi-subunit *trans*-activator complex that involves RNA–protein interactions, upstream of reporter genes in yeast cells (Fig. 2a). The reliability of this method originates from its ability to detect different strengths of RNA–protein interactions, based on the use of two reporter genes. One widespread use of the yeast three-hybrid system has been to discover proteins that bind to a given RNA sequence [113]. To date, nearly 40 RBPs have been successfully selected and characterized using this method.

One of the most pivotal advantages of this method is that the interactions can be monitored *in vivo* and the signal can be enhanced. However, some important limitations of the yeast three-hybrid system need to be taken into account. Some proteins require post-transcriptional modifications in the RNAs to be able to bind them. These RNAs and proteins are not suitable to be studied with the yeast three-hybrid system, as such modifications may not occur in the hybrid RNA. In addition, some RNAs can contain localization signals that can target it to outside the nucleus, preventing it from triggering the reporter gene activation. Finally, the biggest problem of the yeast three-hybrid screen is the selection of false positive clones that must be resolved upon additional time-consuming screening steps, limiting the ability to identify new targets on a genomic scale.

### 7.2. Bacterial assays

Bacterial assays exploiting RNA binding to cause transcriptional anti-termination or interference in translation have been recently

devised. A bacterial genetic assay for detecting polypeptide–RNA interactions was used to screen combinatorial libraries for novel arginine-rich peptides that bind to the Rev-response element (RRE) of human immunodeficiency virus (HIV) [114]. The method utilizes the transcription anti-termination activity of the bacteriophage  $\lambda$  N protein, which causes RNA polymerase to read through transcription termination sites by forming an anti-termination complex [115,116].

Peptide libraries fused to the N protein are encoded on one plasmid, and a second reporter plasmid engineered with an RNA site of interest allows for identification of those proteins that bind the RNA of interest (Fig. 2b).

The anti-termination system enables the identification of tight RNA-binding peptides from mixtures of peptides that have a wide range of affinities toward a particular RNA target. However, the anti-termination system produces a relatively high rate of false positives and to date it has used only short arginine-rich sequences as frameworks for generating protein libraries.

### 7.3. Mammalian Tat-hybrid system

In order to overcome the limitation of using artificial environments or microorganisms, a mammalian cell-based assay using the Tat-hybrid system has been recently developed to identify RBPs within cDNA libraries. The Tat-hybrid system is based on the transcriptional activation properties of the human immunodeficiency virus type 1 (HIV-1) *trans*-activator protein (Tat). Tat is a potent activator of viral gene transcription, and acts by binding to the *trans*-activating responsive (TAR) element, a long imperfect RNA hairpin placed at the 5' end of the RNA [117]. Tat can activate transcription through heterologous RNA–protein interactions when bound to the RNA [118]. Thus, by fusing a library to Tat, replacing TAR with bait RNA and using an HIV LTR reporter, it is possible to screen for RBPs by (Fig. 2c). Using this approach and a GFP reporter, Tan and Frankel identified novel arginine-rich peptides that bind to the Rev response element of HIV [119]. Moreover, Nakamura and colleagues identified several proteins that interact with an RNA hairpin, which is named Mason Pfizer monkey virus (MPMV) constitutive transport element (CTE) and is responsible for the transport of unspliced viral mRNA from the nucleus to

**Fig. 2.** Screening of RBPs within protein libraries. Schematic representation of the most popular strategies for screening RBPs within gene libraries. Three main steps are at the basis of all RBPs screening strategies: the construction of a gene library; the establishment of a linkage between each protein and its encoding gene and the selection of desired proteins from the library, followed by amplification and identification of the selected gene sub-library. *In vivo* approaches. (a) Yeast three-hybrid system. A hybrid protein containing a DNA-binding domain with RNA-binding domain 1 localizes to the promoter of an appropriate reporter gene. A second hybrid protein containing a transcriptional activation domain with RNA-binding domain 2 will activate transcription of the reporter gene when in close proximity to the gene's upstream regulatory sequences. A hybrid RNA containing sites recognized by the two RNA-binding proteins links the two hybrid proteins to one another, and the tripartite complex results in detectable expression of the reporter gene. Yeast triple transformants are assayed by assays of report genes. The identity of the RNA-dependent positives is determined by sequencing. (b) Bacterial anti-termination assay. Simplified overview of the two-plasmid system for detecting transcription anti-termination by heterologous RNA-binding polypeptides fused to the N protein. The ability to select for survival at different kanamycin concentrations allows for the identification of clones with varying RNA-binding affinities. (c) Tat-hybrid assay. A cDNA library is fused to the activation domain of Tat and the RNA-binding proteins are identified by their ability to activate an HIV-1 LTR reporter in which TAR is replaced by an RNA of interest. The Tat-fusion cDNA library is delivered in reporter cell lines. The interaction between the RNA bait and RBPs induces the expression of the reporter gene. For instance, by using a GFP reporter, GFP-expressing cells can be isolate by fluorescence activated cell sorting. Plasmids are subsequently isolated and individual clones are amplified and sequenced. *In vitro* approaches. (d) Northwestern screening. The screen utilizes short <sup>32</sup>P labelled RNA probes to isolate cDNA library clones expressing proteins that engage in sequence-specific interactions with the probe. Plaques of recombinant phage are allowed to form on a bacteria lawn. Then a nitrocellulose or nylon membrane is pressed against the surface thereby adsorbing phage from the plaques in a replica pattern. The membrane is then screened with the radiolabelled RNA probe. After hybridization, the probe solution is removed, and the filter is washed extensively, dried and submitted to autoradiography. The position of strong radioactive signals is related back to a master plate containing the original pattern of colonies, in order to identify positive colonies. The desired clones are then amplified and identified. (e) *In vitro* compartmentalization. Single genes are compartmentalized in water-in-oil emulsions and translated *in vitro*. RBPs are then selected by affinity purification. After breaking the emulsions, streptavidin-coated magnetic beads bound with biotin-labeled RNA were used to capture the RBPs and corresponding encoding DNA. The selected gene expression cassettes are subsequently amplified and identified. (f) Phage display. Libraries of proteins are displayed on phage particles as fusions to coat proteins. In phage library selections, RBP-specific clones can be selected by binding to immobilized target followed by washing to remove non-binding phage. Retained phage can be amplified and individual binding clones can be subjected to DNA sequencing to decode the sequences of the displayed proteins. (g) mRNA display. A DNA library is transcribed into RNA, which is subsequently translated into mRNA-displayed proteins. The mRNA/DNA-polypeptide fusions can be selected over immobilized selection targets for several rounds. Following binding selection, those library members that stay bound to the immobilized target are amplified and sequenced for identification. (h) Protein microarray. Specific RNA of interest is synthesized *in vitro*, labelled with Cy-dyes and bound to a protein microarray. In subsequent analyses, proteins interacting preferentially with individual RNA are identified.

the cytoplasm [120]. The advantage of this method is that the RNA–protein interactions are studied *in vivo* in a mammalian cell, where any possible required accessory factors, chaperons and post-translational modifying enzymes are available to help in the complex reassembly. For this reason this screening method is particularly appropriate for studying mammalian complexes that may require post-translational modifications or multiple cellular components for binding. On the other hand, large domains fused to Tat may sterically hinder the formation of transcription elongation complexes and it is also possible that endogenous nuclear proteins or RNAs may compete for RNA binding, thereby preventing activation.

#### 7.4. Northwestern technology

Northwestern technology is a powerful *in vitro* tool to obtain cDNA clones of proteins that specifically interact with defined RNA sequences [121]. The screen utilizes short  $^{32}\text{P}$  labelled RNA probes to isolate cDNA library clones expressing proteins that engage in sequence-specific interactions with the probe (Fig. 2d). The rationale of this screening approach is based on the plaque lift technique originally used for screening cDNA libraries constructed in bacteriophages [122]. Sägesser and colleagues improved this screening protocol taking advantage of the well-characterized human U1A/U1-RNA model system for RNA–protein interaction [123]. Afterwards, they successfully applied this screening protocol for the characterization of RBPs that specifically interact with the potato spindle tuber viroid (PSTVd) RNA [124].

This method is rapid and inexpensive. It requires basic information about the specific RNA of interest and does not require knowledge of the amino acid sequences of the putative RBPs. However, multiple factors can interfere with the assay sensitivity, especially plaque size and density, the type of probe used, the number and complexity of the target sequence and the properties of the solid phase.

#### 7.5. *In vitro* compartmentalization (IVC)

*In vitro* compartmentalization (IVC) is a cell-free screening system that generates “artificial cells” for the directed evolution of proteins [125]. IVC utilizes micro-compartments for genotype-phenotype linkage, unlike any other techniques used in conventional *in vitro* display. These cell-like compartments are aqueous droplets of water-in-oil (w/o) emulsions, where on average each droplet contains a single gene and resembles an artificial cell in enabling transcription and translation of the resulting proteins to occur (Fig. 2e). Chen and colleagues demonstrated for the first time the application of the IVC method to identify RBPs by exploiting the well-known interaction between zinc-finger proteins and their encoding DNA sequences [126].

One major advantage of IVC is that the expressed protein does not need to be directly bound to the nucleic acid. This precludes any type of modification in the protein preventing the alteration of its activity. Additionally, the high capacity of IVC, the ease of preparing emulsions, and stability over a broad range of temperatures, makes IVC emulsions an attractive reaction vessel for screening very large gene libraries without cloning and transforming steps. However, some technical limitations must be considered [127]. The first is the stringency of the genotype-phenotype link since it is possible to form double emulsion droplets containing multiple compartments encapsulating unrelated genes [128]. The second limitation is a technical issue when using membrane proteins or any other protein that requires post-translation modification [129].

#### 7.6. Phage and mRNA display

Phage display and mRNA display have been used to identify variant domains of RBPs that bind to RNAs with altered specificities.

Phage display involves the insertion of a gene encoding a protein of interest into a phage coat protein gene, causing the phage to display the protein on its outside while containing the gene for the protein on its inside, resulting in a connection between genotype and phenotype (Fig. 2f). In this way, by immobilizing an RNA target to beads or to the surface of a microtiter plate well, large protein libraries can be screened against the RNA target and amplified in a process called *in vitro* selection. Phage display has been successfully used to investigate the RNA-binding specificity of the mammalian spliceosomal protein U1A, which binds to hairpin II of the U1 small nuclear RNA (U1hpII) [130], as well as to isolate single zinc fingers that bind complex RNA structures with high affinity and specificity [131]. A number of phage libraries displaying cyclic peptides and linear peptides are commercially available. However, the size of the libraries is limited by the transformation efficiency of bacterial cells and only a limited number of individual clones can be examined easily making this method time-consuming, difficult to scale-up and labour-intensive. mRNA display, also called mRNA–protein fusion [132] or *in vitro* virus [133], relies on the covalent coupling of mRNA to the nascent polypeptide (Fig. 2g). Numerous mRNA-display selections have isolated more than one hundred chemically distinct RNA-binding peptides [134]. Most of the experiments have used the RNA-binding domain from phage  $\lambda$  N protein owing to its small size and high affinity for its cognate RNA. Despite its advantages, mRNA display has some limitations. The major concerns are the possibilities that the covalently attached mRNA interferes with the function of the protein, or that the target RNA interacts with the displayed protein. Other weaknesses of mRNA display are limitations in the display of membrane-bound proteins owing to their low expression level in *in vitro* translation systems [135] and of proteins whose biological functions relies on complex formation.

#### 7.7. Protein microarrays

Protein microarrays have been used to screen for known proteins that unexpectedly interact with a specific RNA *in vitro*. This approach involves the labelling of the RNA target with fluorescent dyes, such as the cyanine Cy3 or Cy5, and their subsequent screening with a fluorescence scanner (Fig. 2h).

Two studies using protein microarrays and RNA probes to search for new RNA-binding proteins in yeast identified approximately 200 unexpected RBPs [136,137]. Interestingly, more than half of the identified proteins are well known enzymes, and many of them involved in RNA metabolism. This study indicated direct connections between metabolic status and post-transcriptional gene regulation. Moreover, most of the RBPs identified lack recognizable RNA-binding domains, which underlies the limitations of current bioinformatics methods that rely upon homology to known RNA binding domains. Recently, Siperashvili and colleagues used human protein microarrays to identify 137 RNA–protein interactions specific for 10 coding and non-coding RNAs [138]. In particular, Stau1, a well-known RNA-binding protein involved in RNA stability and localization [139], was found to interact with a 256 bp sequence of the TP53 3' UTR, regulating TP53 mRNA in a previously uncharacterized way.

The use of protein arrays to monitor RNA–protein interactions has several advantages. As hundreds and even thousands of proteins can be analysed simultaneously in a single experiment, it is a very effective high-throughput method taking less than a day to complete. The reagents required are readily available, such as many commercial protein arrays. Moreover, the assay requires a

minimal amount of RNA and no large-scale cell culture for protein isolation. Although protein arrays encompass a broad range of applications on a proteome scale, the greatest limitation to their utility is that the quality of the data obtained from screening a protein array depends on a complete and functional representation of proteins on the chip. Misfolded proteins or inaccessible protein-binding domains may lead to false-negative or false-positive results. As the expression and analysis of RNA–protein interactions are carried out in an artificial environment relative to the cell there is not always a direct correlation between protein abundance and activity. Moreover, this artificial environment may not represent the physiological conditions of the biological system under study.

Screening for RNA sequence-specific proteins using large libraries has a higher sensitivity in comparison to the MS-based affinity strategies and allows for the identification of poorly expressed RNA-binding proteins. However, except for the mammalian Tat-hybrid system, these methods are limited to binary interactions, and interactions that require specific post-translational modifications are likely to be missed. For this reason detection of cDNA clones from large libraries encoding any possible RNA-binding proteins has not yet become a consolidated ribonomic strategy. Moreover, these methods present some restrictions regarding the size of the RNA target, and almost all of them are not appropriate to investigate full-length transcripts or long RNAs.

Regardless of the RNA-centric approaches used, relevant RBP candidates should be identified from several independent biological replicates followed by functional validation. Indeed, due to the high affinity RNA binding from some proteins [103,140], the candidate interacting proteins must undergo further biological validations. For instance, computational approaches can be an excellent complement to experiments by either suggesting new targets of investigation or providing independent validation and ranking of experimental results.

## 8. Computational methods

In parallel to the progress made in methods for experimentally mapping RNA–protein interactions, the development of bioinformatic tools has improved the *in silico* prediction of such interactions. However, whereas the methodology for prediction and modelling of proteins and protein–protein complexes is well established [141,142], there are fewer methods capable of predicting and modelling RNA and RNA–protein complexes [143,144].

Ideally, the prediction of RNA–protein complex structure should be based on the knowledge of atomic structures of the components, as determined by X-ray crystallography, Cryo-electron microscopy (Cryo-EM) or Nuclear Magnetic Resonance (NMR) spectrometry. Single molecule studies using FRET pairs can determine either the dynamics or the force-versus-extension traces within the RNA. In addition, different structure-specific probe methods, such as dimethyl sulfate (DMS) [145], hydroxyl radical footprinting [146], or the 2'-hydroxyl acylation RNA (SHAPE) [147,148], can be used for predicting RNA secondary structures. As experimentally determined structures of components of the complex are not available in many cases, computationally-modelled structures can be used instead. To this end, a great number of softwares have been developed that allow for reasonably accurate and practically useful prediction of RNA 3D structures and interactions.

Over the past three decades several computational programs have been developed for the prediction of RNA secondary structures. Multiple approaches have been used including: free energy minimization using thermodynamic parameters, such as Mfold [149] and RNAfold [150]; knowledge-based predictions based on known RNA structures, such as Contrafold [151] and RNASHapes

[152]; comparative sequence alignment algorithms, such as Dyn-align [153] and RNAforester [154]; and combinations of these. RNA tertiary-structure predictions programs have also emerged in order to overcome some of the limitations of the previously cited programs. Some of the most recently available programs for RNA tertiary-structure predictions include FARNA [155], NAST [46], iFoldRNA [156] and RNA2D3D [157].

Several software tools can be used to predict RNA-binding residues from protein sequence alone or from protein structure. Some indicative examples include RNA BindR Plus [158], OPRA [159], DRNA [160] and KYG [161]. In addition, various methods have been established for the identification of potential protein–RNA docking sites. These are modifications of methods used to identify protein–protein docking sites that view nucleic acid molecules as receptors and/or ligands, instead of proteins. Such methods include HADDOCK [162], GRAMM [163], HEX [164], PatchDock [165], and FTDock [166]. Only recently, the docking method called 3dRPC [167] that takes special features of RNA surfaces into account was developed with a specific purpose of identifying protein–RNA docking sites.

Finally, Tartaglia and colleagues developed catRAPID, the first computational method capable of predicting RNA–protein interactions in a large scale [168]. The algorithm evaluates the interaction propensities of polypeptide and nucleotide chains based on physicochemical properties such as secondary structure information, hydrogen bonding and Van der Waals forces. Moreover, the authors validated their algorithm on a large collection of protein associated with lncRNAs, the NPinter dataset [169]. However, prediction of lncRNAs function is generally hampered by poor sequence homology and lack of interaction data.

Computational approaches are highly needed for predicting the RNP interactome and complement experimental methods. These bioinformatic tools should be useful to uncover and extend the network of RBPs, thus achieving better understanding of RNA biology. Over the last decade, great advances have been made in ribonomic approaches, computational softwares and technologies for RNA investigations. However, both RNA-based computational approaches as well as ribonomic RNA-centric methods require further technical developments and refinements.

Development of computational methods able to predict RNP structures, either from the individual components or directly from their sequences has been highly sought owing to the difficulty in the structure determination of RNA–protein complexes. The existing approaches for RNP modelling can provide useful predictions despite their various limitations. One problem specific to RNA modelling is the relative paucity of experimentally-determined RNA and RNP structures that can be used as templates. Another limitation is related with the fact that docking programs seldom take into account conformational changes that may occur upon binding, both in protein and RNA components.

## 9. Conclusions and future directions

Exciting new discoveries have revealed the complex versatility of RNPs and their importance in a large number of functions within the cells. The broad characterization of RNP components, interactions, dynamics and function are fundamental steps in required to uncover the role of post-transcriptional regulation in gene expression. Understanding the effects of linked gene regulatory networks on cellular function, organism development and the evolution of biological complexity will be achieved with more comprehensive analysis.

The combination of ribonomic approaches with cross-linking techniques and high-throughput sequencing has provided a systematic mapping of RBP-binding sites [40,170,171]. Next



generation sequencing methods has allowed a broader and quantitative detection of RNAs, including rare and previously unknown RNA molecules [172]. Novel large-scale proteomic approaches now enable the quantitative measurement of hundreds of proteins in parallel [173,174]. Concomitant analysis of the changes in the RNA and protein levels upon RBP depletion or overexpression could therefore provide global information of the different status of mRNAs and of their downstream effects [175,176].

However, little information is available regarding the full complement of RBPs that are associated with a specific RNA. Such information will be important to unravel the combinatorial regulation of mRNAs by multiple RBPs and understand how the assorted RBPs change with environmental cues. This becomes even more important in light of the rapidly increasing number of ncRNAs – in particular the lncRNAs – that play important roles for gene expression control in cell-differentiation and development.

Elucidating the composition of RNPs requires technological advances in RNA-centric high-throughput methods [177,178], as well as refinements in bioinformatics tools. Future steps are needed to obtain a complete picture of the ribonome. As knowledge of post-transcriptional RNA operons increases, surely new functions for multifunctional proteins and connections within regulatory pathways will be uncovered, but this is a difficult task to achieve with traditional methods.

A first barrier to understanding RNPs is the identification and characterization of their total RNA and RBP components [179]. As the number of known RNA–protein interactions increases, the binding and localization of RBPs over the genome and transcriptome will begin to be better evaluated. Fast identification of all proteins bound to an individual RNA of interest will provide a deeper characterization of the molecular mechanisms and the functional role of virtually any specific coding and non-coding RNAs involved in human diseases.

Furthermore, a comprehensive understanding of RNP biology will require specific knowledge of the interplay between different inputs and each component of the system and how they play their role in controlling and regulation. RNP biology needs to focus on the dynamics of post-transcriptional processes and their interactions on a large scale rather than the old static analysis of individual component interactions. The RNA–protein network appears to be very dynamic and responds to environmental or developmental signals by altering the RNA and protein content of RNPs. Thereby, post-translational modifications can alter subcellular localization or RNA-binding activity of the RBP [180,181]. The RNA–protein network is further complicated by the possibility that multiple states of a specific mRNA population may exist and therefore not every RNA molecule may have the same fate. RNP systems are normally studied within cell populations and therefore reflect the cell average [182]. Single cell approaches may provide a better definition of RNP interactions and compositions within individual cells. Such studies are necessary to provide a deeper understanding of the RNPs function from individual cells to tissues, and multicellular organisms. Ribonomic analysis can identify the emerging of new RNA-binding domains during evolution upon studies of samples from multiple species.

Finally, a major challenge ahead will be the analysis, visualization and integration of a large scale data from different levels of gene expression, which in turn will provide new insights into their cellular and physiological function. A robust understanding of RNP system biology will rise from a large-scale data integration on RNP, protein–protein and RNA–protein interactions, protein and transcript levels, protein and RNA localization and RNP dynamics [183,184].

## Acknowledgements

The Australian Research Council (ARC) and the University of Sydney supported this work. We would like to thank Janine Flores and James Walshe for critical reading of the manuscript.

## References

- [1] Staley, J.P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315–326.
- [2] Mansfield, K.D. and Keene, J.D. (2009) The ribonome: a dominant force in coordinating gene expression. *Biol. Cell* 101, 169–181.
- [3] Glisovic, T., Bachorik, J.L., Yong, J. and Dreyfuss, G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 582, 1977–1986.
- [4] Orphanides, G. and Reinberg, D. (2002) A unified theory of gene expression. *Cell* 108, 439–451.
- [5] Lasko, P. (2003) Cup–ling oskar RNA localization and translational control. *J. Cell Biol.* 163, 1189–1191.
- [6] Buckanovich, R.J., Posner, J.B. and Darnell, R.B. (1993) Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron* 11, 657–672.
- [7] Valente, L. and Nishikura, K. (2005) ADAR gene family and A-to-I RNA editing: diverse roles in posttranscriptional gene regulation. *Prog. Nucleic Acid Res. Mol. Biol.* 79, 299–338.
- [8] Ryan, K., Calvo, O. and Manley, J.L. (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* 10, 565–573.
- [9] Vinciguerra, P. and Stutz, F. (2004) MRNA export: an assembly line from genes to nuclear pores. *Curr. Opin. Cell Biol.* 16, 285–292.
- [10] Bohl, F., Kruse, C., Frank, A., Ferring, D. and Jansen, R.P. (2000) She2p, a novel RNA-binding protein tethers ASH1 mRNA to the Myo4p myosin motor via She3p. *EMBO J.* 19, 5514–5524.
- [11] St Johnston, D. (2005) Moving messages: the intracellular localization of mRNAs. *Nat. Rev. Mol. Cell Biol.* 6, 363–375.
- [12] Gebauer, F. and Hentze, M.W. (2004) Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell Biol.* 5, 827–835.
- [13] Huang, Y.S. and Richter, J.D. (2004) Regulation of local mRNA translation. *Curr. Opin. Cell Biol.* 16, 308–313.
- [14] Ānkō, M.-L. and Neugebauer, K.M. (2012) RNA-protein interactions in vivo: global gets specific. *Trends Biochem. Sci.* 37, 255–262.
- [15] Janga, S.C. (2012) From specific to global analysis of posttranscriptional regulation in eukaryotes: posttranscriptional regulatory networks. *Brief Funct. Genomic* 11 (6), 505–521.
- [16] McHugh, C.A., Russell, P. and Guttman, M. (2014) Methods for comprehensive experimental identification of RNA–protein interactions. *Genome Biol.* 15 (203), <http://dx.doi.org/10.1186/gb4152>.
- [17] Attar, N. (2014) The RBPome: where the brains meet the brawn. *Genome Biol.* 15 (402), <http://dx.doi.org/10.1186/gb4153>.
- [18] Lukong, K.E., Chang, K.W., Khandjian, E.W. and Richard, S. (2008) RNA-binding proteins in human genetic disease. *Trends Genet.* 24, 416–425.
- [19] Kechavarzi, B. and Janga, S.C. (2014) Dissecting the expression landscape of RNA-binding proteins in human cancer. *Genome Biol.* 15 (R14), 1–16.
- [20] Kerner, P., Degnan, S.M., Marchand, L., Degnan, B.M. and Vervoort, M. (2011) Evolution of RNA-binding proteins in animals: insights from genome-wide analysis in the sponge *Amphimedon queenslandica*. *Mol. Biol. Evol.* 28, 2289–2303.
- [21] Anantharaman, V., Koonin, E.V. and Aravind, L. (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* 30 (7), 1427–1464.
- [22] Keene, J.D. and Tenenbaum, S.A. (2002) Eukaryotic mRNPs may represent posttranscriptional operons. *Mol. Cell* 9, 1161–1167.
- [23] Tamburino, A.M., Ryder, S.P. and Walthout, A.J.M. (2013) A compendium of *Caenorhabditis elegans* RNA binding proteins predicts extensive regulation at multiple levels. *G3 – Genes Genom. Genet.* 3 (2), 297–304.
- [24] Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D. and Brown, P.O. (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol.* 6 (10), e255, <http://dx.doi.org/10.1371/journal.pbio.0060255>.
- [25] Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* 46 (5), 674–690.
- [26] Lunde, B.M., Moore, C. and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* 8, 479–490.
- [27] Gao, F.B., Carson, C.C., Levine, T. and Keene, J.D. (1994) Selection of a subset of mRNAs from combinatorial 3' untranslated region libraries using neuronal RNA-binding protein Hel-N1. *Proc. Natl. Acad. Sci. U.S.A.* 91, 11207–11211.
- [28] Tenenbaum, S.A., Carson, C.C., Lager, P.J. and Keene, J.D. (2000) Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14085–14090.
- [29] Hieronymus, H. and Silver, P.A. (2004) A systems view of mRNP biology. *Genes Dev.* 18, 2845–2860.
- [30] Keene, J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Gen.* 8, 533–543.

- [31] McKee, A.E. and Silver, P.A. (2007) Systems perspectives on mRNA processing. *Cell Res.* 17, 581–590.
- [32] Gerber, A.P., Herschlag, D. and Brown, P.O. (2004) Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol.* 2, E79.
- [33] Gerber, A.P., Luschnig, S., Krasnow, M.A., Brown, P.O. and Herschlag, D. (2006) Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4487–4492.
- [34] Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P. and Krause, H.M. (2007) Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* 131, 174–187.
- [35] Czaplinski, K. and Singer, R.H. (2006) Pathways for mRNA localization in the cytoplasm. *Trends Biochem. Sci.* 31, 687–693.
- [36] Ule, J., Jensen, K., Mele, A. and Darnell, R.B. (2005) CLIP: a method for identifying protein–RNA interaction sites in living cells. *Methods Mol. Biol.* 37, 376–386.
- [37] Keene, J.D., Komisarow, J.M. and Friedersdorf, M.B. (2006) RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat. Protoc.* 1, 302–307.
- [38] Ray, D., Kazan, H., Chan, E.T., Pena Castillo, L., Chaudhry, S., Talukder, S., Blencowe, B.J., Morris, Q. and Hughes, T.R. (2009) Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat. Biotechnol.* 27, 667–670.
- [39] Scheibe, M., Butter, F., Hafner, M., Tuschl, T. and Mann, M. (2012) Quantitative mass spectrometry and PAR-CLIP to identify RNA-protein interactions. *Nucleic Acids Res.* 40, 9897–9902.
- [40] Zhang, C. and Darnell, R.B. (2011) Mapping in vivo protein–RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* 29, 607–614.
- [41] Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M.J., Jungkamp, A.C., Munschauer, M., et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141.
- [42] Paz, I., Kosti, I., Ares Jr., M., Cline, M. and Mandel-Gutfreund, Y. (2014) RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.*
- [43] Kazan, H., Ray, D., Chan, E.T., Hughes, T.R. and Morris, Q. (2010) RNAcontext: a new method for learning the sequence and structure binding preferences of RNA-binding proteins. *PLoS Comput. Biol.* 6 (7), e1000832, <http://dx.doi.org/10.1371/journal.pcbi.1000832>.
- [44] Auweter, S.D., Oberstrass, F.C. and Allain, F.H. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* 34, 4943–4959.
- [45] Wang, X., Juan, L., Lv, J., Wang, K., Sanford, J.R. and Liu, Y. (2011) Predicting sequence and structural specificities of RNA binding regions recognized by splicing factor SRSF1. *BMC Genomics* 12 (Suppl. 5), S8.
- [46] Jonikas, M.A., Radmer, R.J., Laederach, A., Das, R., Pearlman, S., Herschlag, D. and Altman, R.B. (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* 15, 189–199.
- [47] Lee, M.H. and Schedl, T. (2001) Identification of in vivo mRNA targets of GLD-1, a maxi-KH motif containing protein required for *C. elegans* germ cell development. *Genes Dev.* 15, 2408–2420.
- [48] Ánkó, M.L., Morales, L., Henry, I., Beyer, A. and Neugebauer, K.M. (2010) Global analysis reveals SRp20- and SRp75-specific mRNPs in cycling and neural cells. *Nat. Struct. Mol. Biol.* 17, 962–970.
- [49] Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.
- [50] Komili, S., Farny, N.G., Roth, F.P. and Silver, P.A. (2007) Functional specificity among ribosomal proteins regulates gene expression. *Cell* 131, 557–571.
- [51] Halbeisen, R.E., Galgano, A., Scherrer, T. and Gerber, A.P. (2008) Post-transcriptional gene regulation: from genome-wide studies to principles. *Cell. Mol. Life Sci.* 65, 798–813.
- [52] Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P.Y., Soll, S.J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M. and Tuschl, T. (2008) Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA* 14, 2580–2596.
- [53] Rouault, T.A., Hentze, M.W., Haile, D.J., Harford, J.B. and Klausner, R.D. (1989) The iron-responsive element binding protein: a method for the affinity purification of a regulatory RNA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.* 86, 5768–5772.
- [54] Sharma, S. (2008) Isolation of a sequence-specific RNA binding protein, polypyrimidine tract binding protein, using RNA affinity chromatography. *Methods Mol. Biol.* 488, 1–8.
- [55] Tombelli, S., Minunni, M. and Mascini, M. (2005) Analytical applications of aptamers. *Biosens. Bioelectron.* 20, 2424–2434.
- [56] Srisawat, C., Goldstein, I.J. and Engelke, D.R. (2001) Sephadex-binding RNA ligands: rapid affinity purification of RNA from complex RNA mixtures. *Nucleic Acids Res.* 29 (2), e4.
- [57] Srisawat, C. and Engelke, D.R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA* 7, 632–641.
- [58] Leonov, A.A., Sergiev, P.V., Bogdanov, A.A., Brimacombe, R. and Dontsova, O.A. (2003) Affinity purification of ribosomes with a lethal G2655C mutation in 23 S rRNA that affects the translocation. *J. Biol. Chem.* 278, 25664–25670.
- [59] Li, Y. and Altman, S. (2002) Partial reconstitution of human RNase P in HeLa cells between its RNA subunit with an affinity tag and the intact protein components. *Nucleic Acids Res.* 30, 3706–3711.
- [60] Carey, J., Cameron, V., de Haset, P.L. and Uhlenbeck, O.C. (1983) Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. *Biochemistry* 22 (11), 2601–2610.
- [61] Bardwell, V.J. and Wickens, M. (1990) Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res.* 18, 6587–6594.
- [62] Das, R., Zhou, Z. and Reed, R. (2000) Functional association of U2 snRNP with the ATP-independent spliceosomal complex. *E. Mol. Cell* 5, 779–787.
- [63] Lioka, H., Loisel, D., Haystead, T.A. and Macara, I.G. (2011) Efficient detection of RNA–protein interactions using tethered RNAs. *Nucleic Acids Res.* 39 (8), e53, <http://dx.doi.org/10.1093/nar/gkq1316>.
- [64] Blencowe, B.J. and Lamond, A.I. (1999) Purification and depletion of RNP particles by antisense affinity chromatography. *Methods Mol. Biol.* 118, 275–287.
- [65] Lingner, J. and Cech, T.R. (1996) Purification of telomerase from *Euplotes aedicularis*: requirement of a primer 3' overhang. *Proc. Natl. Acad. Sci. U.S.A.* 93, 10712–10717.
- [66] Soeno, Y., Taya, Y., Stasyk, T., Huber, L.A., Aoba, T. and Huttenhofer, A. (2010) Identification of novel ribonucleo-protein complexes from the brain-specific snoRNA MBII-52. *RNA* 16, 1293–1300.
- [67] Walton, S.P., Stephanopoulos, G.N., Yarmush, M.L. and Roth, C.M. (2002) Thermodynamic and kinetic characterization of antisense oligodeoxynucleotide binding to a structured mRNA. *Biophys. J.* 82, 366–377.
- [68] Krause, H. and Simmonds, A. (2006) Trap-tagging: a novel method for the identification and purification of RNA–protein complexes, in: Google Patents.
- [69] Hogg, J.R. and Collins, K. (2007) RNA-based affinity purification reveals 75K RNPs with distinct composition and regulation. *RNA* 13, 868–880.
- [70] Carey, J. and Uhlenbeck, O.C. (1983) Kinetic and thermodynamic characterization of the R17 coat protein–ribonucleic acid interaction. *Biochemistry* 22, 2610–2615.
- [71] Lim, F., Downey, T.P. and Peabody, D.S. (2001) Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7. *J. Biol. Chem.* 276, 22507–22513.
- [72] Leppek, K. and Stoecklin, G. (2014) An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.* 42, e13.
- [73] Leppek, K., Schott, J., Reitter, S., Poetz, F., Hammond, M.C. and Stoecklin, G. (2013) Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell* 153, 869–881.
- [74] Tsai, B.P., Wang, X., Huang, L. and Waterman, M.L. (2011) Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach. *Mol. Cell. Proteomics* 10 (M110), 007385.
- [75] Tagwerker, C., Flick, K., Cui, M., Guerrero, C., Dou, Y., Auer, B., Baldi, P., Huang, L. and Kaiser, P. (2006) A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivo cross-linking. *Mol. Cell. Proteomics* 5, 737–748.
- [76] von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P. (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 417, 399–403.
- [77] McBride, K.M., Gazumyan, A., Woo, E.M., Schwickert, T.A., Chait, B.T. and Nussenzweig, M.C. (2008) Regulation of class switch recombination and somatic mutation by AID phosphorylation. *J. Exp. Med.* 205, 2585–2594.
- [78] Cheng, H.L., Vuong, B.Q., Basu, U., Franklin, A., Schwer, B., Astarita, J., Phan, R.T., Datta, A., Manis, J., Alt, F.W., et al. (2009) Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2717–2722.
- [79] Swaffield, J.C., Melcher, K. and Johnston, S.A. (1995) A highly conserved ATPase protein as a mediator between acidic activation domains and the TATA-binding protein. *Nature* 374, 88–91.
- [80] Slobodin, B. and Gerst, J.E. (2011) rapID: an aptamer-based mRNA affinity purification technique for the identification of RNA and protein factors present in ribonucleoprotein complexes. *Methods Mol. Biol.* 714, 387–406.
- [81] Slobodin, B. and Gerst, J.E. (2010) A novel mRNA affinity purification technique for the identification of interacting proteins and transcripts in ribonucleoprotein complexes. *RNA* 16, 2277–2290.
- [82] Lee, H.Y., Haurwitz, R.E., Apffel, A., Zhou, K., Smart, B., Wenger, C.D., Laderman, S., Bruhn, L. and Doudna, J.A. (2013) RNA–protein analysis using a conditional CRISPR nuclease. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5416–5421.
- [83] Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L.M., Krijgsvelde, J. and Hentze, M.W. (2013) System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.* 8 (3), 491–500.
- [84] Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149 (6), 1393–1406.
- [85] Zielinski, J., Kilk, K., Peritz, T., Kannanayakal, T., Miyashiro, K.Y., Eiriksdottir, E., Jochems, J., Langel, U. and Eberwine, J. (2006) In vivo identification of ribonucleoprotein–RNA interactions. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1557–1562.
- [86] Bell, T.J., Eiriksdottir, E., Langel, U. and Eberwine, J. (2011) PAIR technology: exon-specific RNA-binding protein isolation in live cells. *Methods Mol. Biol.* 683, 473–486.

- [87] Liu, D.G. and Sun, L. (2005) Direct isolation of specific RNA-interacting proteins using a novel affinity medium. *Nucleic Acids Res.* 33, e132.
- [88] Beach, D.L. and Keene, J.D. (2008) Ribotrap: targeted purification of RNA-specific RNPs from cell lysates through immunofluorescence precipitation to identify regulatory proteins and RNAs. *Methods Mol. Biol.* 419, 69–91.
- [89] Michlewski, G. and Caceres, J.F. (2010) RNase-assisted RNA chromatography. *RNA* 16, 1673–1678.
- [90] Gong, C., Popp, M.W. and Maquat, L.E. (2012) Biochemical analysis of long non-coding RNA-containing ribonucleoprotein complexes. *Methods* 58, 88–93.
- [91] Gumireddy, K., Li, A., Yan, J., Setoyama, T., Johannes, G.J., Orom, U.A., Tchou, J., Liu, Q., Zhang, L., Speicher, D.W., et al. (2013) Identification of a long non-coding RNA-associated RNP complex regulating metastasis at the translational step. *EMBO J.* 32, 2672–2684.
- [92] Yoon, J.H., Srikantan, S. and Gorospe, M. (2012) MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs. *Methods* 58, 81–87.
- [93] Chu, C., Quinn, J. and Chang, H.Y. (2012) Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.*
- [94] Ulitsky, I. and Bartel, D.P. (2013) LincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46.
- [95] Simon, M.D. (2013) Capture hybridization analysis of RNA targets (CHART). *Current protocols in molecular biology/edited by Frederick M Ausubel [et al.]. Chapter 21, Unit 21 25.*
- [96] Simon, M.D., Wang, C.L., Kharchenko, P.V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I. and Kingston, R.E. (2011) The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20497–20502.
- [97] Blackwell, B.J., Lopez, M.F., Wang, J., Krastins, B., Sarracino, D., Tollervey, J.R., Dobke, M., Jordan, I.K. and Lunyak, V.V. (2012) Protein interactions with piALU RNA indicates putative participation of retroRNA in the cell cycle, DNA repair and chromatin assembly. *Mobile Genet. Elem.* 2, 26–35.
- [98] Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S. and Wu, J.Y. (2011) RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol. Cell. Biol.* 31, 1812–1821.
- [99] Patel, G.P., Ma, S. and Bag, J. (2005) The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex. *Nucleic Acids Res.* 33, 7074–7089.
- [100] Mehta, A. and Driscoll, D.M. (1998) A sequence-specific RNA-binding protein complements apobec-1 To edit apolipoprotein B mRNA. *Mol. Cell. Biol.* 18, 4426–4432.
- [101] Ward, A.M., Bidet, K., Yinglin, A., Ler, S.G., Hogue, K., Blackstock, W., Gunaratne, J. and Garcia-Blanco, M.A. (2011) Quantitative mass spectrometry of DENV-2 RNA-interacting proteins reveals that the DEAD-box RNA helicase DDX6 binds the DB1 and DB2 3' UTR structures. *RNA Biol.* 8, 1173–1186.
- [102] Talukdar, I., Sen, S., Urbano, R., Thompson, J., Yates 3rd, J.R. and Webster, N.J. (2011) HnRNP A1 and hnRNP F modulate the alternative splicing of exon 11 of the insulin receptor gene. *PLoS ONE* 6, e27869.
- [103] Harris, D., Zhang, Z., Chaubey, B. and Pandey, V.N. (2006) Identification of cellular factors associated with the 3'-nontranslated region of the hepatitis C virus genome. *Mol. Cell. Proteomics* 5, 1006–1018.
- [104] Zielinski, J., Kilk, K., Peritz, T., Kannanayakal, T., Miyashiro, K.Y., Eiríksson, E., Jochems, J., Langel, U. and Eberwine, J. (2006) In vivo identification of ribonucleoprotein-RNA interactions. *PNAS* 103, 1557–1562.
- [105] Tacheny, A., Dieu, M., Arnould, T. and Renard, P. (2013) Mass spectrometry-based identification of proteins interacting with nucleic acids. *J. Proteomics* 94, 89–109.
- [106] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999.
- [107] Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386.
- [108] Ly, L. and Wasinger, V.C. (2011) Protein and peptide fractionation, enrichment and depletion: tools for the complex proteome. *Proteomics* 11, 513–534.
- [109] Schirmer, E.C., Yates 3rd, J.R. and Gerace, L. (2003) MudPIT: a powerful proteomics tool for discovery. *Discov. Med.* 3, 38–39.
- [110] Mili, S. and Steitz, J.A. (2004) Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* 10, 1692–1694.
- [111] Greenberg, J.R. (1979) Ultraviolet light-induced crosslinking of mRNA to proteins. *Nucleic Acids Res.* 6, 715–732.
- [112] SenGupta, D.J., Zhang, B., Kraemer, B., Pochart, P., Fields, S. and Wickens, M. (1996) A three-hybrid system to detect RNA-protein interactions in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 93, 8496–8501.
- [113] Bernstein, D.S., Buter, N., Stumpf, C. and Wickens, M. (2002) Analyzing mRNA-protein complexes using a yeast three-hybrid system. *Methods* 26, 123–141.
- [114] Harada, K., Martin, S.S. and Frankel, A.D. (1996) Selection of RNA-binding peptides in vivo. *Nature* 380, 175–179.
- [115] Greenblatt, J., Nodwell, J.R. and Mason, S.W. (1993) Transcriptional antitermination. *Nature* 364, 401–406.
- [116] Greenblatt, J., Mah, T.F., Legault, P., Mogridge, J., Li, J. and Kay, L.E. (1998) Structure and mechanism in transcriptional antitermination by the bacteriophage lambda N protein. *Cold Spring Harb. Symp. Quant. Biol.* 63, 327–336.
- [117] Roy, S., Dellling, U., Chen, C.H., Rosen, C.A. and Sonenberg, N. (1990) A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev.* 4, 1365–1373.
- [118] Southgate, C., Zapp, M.L. and Green, M.R. (1990) Activation of transcription by HIV-1 Tat protein tethered to nascent RNA through another protein. *Nature* 345, 640–642.
- [119] Tan, R. and Frankel, A.D. (1998) A novel glutamine-RNA interaction identified by screening libraries in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 95, 4247–4252.
- [120] Nakamura, R.L., Landt, S.G., Mai, E., Nejim, J., Chen, L. and Frankel, A.D. (2012) A cell-based method for screening RNA-protein interactions: identification of constitutive transport element-interacting proteins. *PLoS ONE* 7, e48194.
- [121] Qian, Z. and Wilusz, J. (1993) Cloning of a cDNA encoding an RNA binding protein by screening expression libraries using a northwestern strategy. *Anal. Biochem.* 212, 547–554.
- [122] Benton, W.D. and Davis, R.W. (1977) Screening lambda gt10 recombinant clones by hybridization to single plaques in situ. *Science* 196, 180–182.
- [123] Sagesser, R., Martinez, E., Tsagris, M. and Tabler, M. (1997) Detection and isolation of RNA-binding proteins by RNA-ligand screening of a cDNA expression library. *Nucleic Acids Res.* 25, 3816–3822.
- [124] Martinez de Alba, A.E., Sagesser, R., Tabler, M. and Tsagris, M. (2003) A bromodomain-containing protein from tomato specifically binds potato spindle tuber viroid RNA in vitro and in vivo. *J. Virol.* 77, 9685–9694.
- [125] Tawfik, D.S. and Griffiths, A.D. (1998) Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* 16, 652–656.
- [126] Chen, Y., Mandic, J. and Varani, G. (2008) Cell-free selection of RNA-binding proteins using in vitro compartmentalization. *Nucleic Acids Res.* 36, e128.
- [127] Nishikawa, T., Sunami, T., Matsuura, T. and Yomo, T. (2012) Directed evolution of proteins through in vitro protein synthesis in liposomes. *J. Nucleic Acids* 2012, <http://dx.doi.org/10.1155/2012/923214>.
- [128] Fallah-Araghi, A., Baret, J.C., Ryckelynck, M. and Griffiths, A.D. (2012) A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution. *Lab Chip* 12, 882–891.
- [129] Girard, P., Pecreaux, J., Lenoir, G., Falson, P., Rigaud, J.L. and Bassereau, P. (2004) A new method for the reconstitution of membrane proteins into giant unilamellar vesicles. *Biophys. J.* 87, 419–429.
- [130] Laird-Offringa, I.A. and Belasco, J.G. (1995) Analysis of RNA-binding proteins by in vitro genetic selection: Identification of an amino acid residue important for locking U1A onto its RNA target. *Proc. Natl. Acad. Sci.* 92, 11859–11863.
- [131] Friesen, W.J. and Darby, M.K. (1997) Phage display of RNA binding zinc fingers from transcription factor IIIA. *J. Biol. Chem.* 272, 10994–10997.
- [132] Roberts, R.W. and Szostak, J.W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12297–12302.
- [133] Nemoto, N., Miyamoto-Sato, E., Husimi, Y. and Yanagawa, H. (1997) In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett.* 414, 405–408.
- [134] Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T. and Roberts, R.W. (2001) Large libraries reveal diverse solutions to an RNA recognition problem. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12374–12378.
- [135] Lin, H. and Cornish, V.W. (2002) Screening and selection methods for large-scale analysis of protein function. *Angew. Chem.* 41, 4402–4425.
- [136] Scherrer, T., Mittal, N., Janga, S.C. and Gerber, A.P. (2010) A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS ONE* 5, e15499.
- [137] Tsvetanova, N.G., Klass, D.M., Salzman, J. and Brown, P.O. (2010) Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*. *PLoS ONE* 5.
- [138] Siprashvili, Z., Webster, D.E., Kretz, M., Johnston, D., Rinn, J.L., Chang, H.Y. and Khavari, P.A. (2012) Identification of proteins binding coding and non-coding human RNAs using protein microarrays. *BMC Genomics* 13, 633.
- [139] Kim, Y.K., Furic, L., Desgroseillers, L. and Maquat, L.E. (2005) Mammalian Staufen1 recruits Upf1 to specific mRNA 3' UTRs so as to elicit mRNA decay. *Cell* 120, 195–208.
- [140] Goracznik, R. and Gunderson, S.I. (2008) The regulatory element in the 3'-untranslated region of human papillomavirus 16 inhibits expression by binding CUG-binding protein 1. *J. Biol. Chem.* 283, 2286–2296.
- [141] Wichadakul, D., McDermott, J. and Samudrala, R. (2009) Prediction and integration of regulatory and protein-protein interactions. *Methods Mol. Biol.* 541, 101–143.
- [142] Moreira, I.S., Fernandes, P.A. and Ramos, M.J. (2010) Protein-protein docking dealing with the unknown. *J. Comput. Chem.* 31, 317–342.
- [143] Laing, C. and Schlick, T. (2010) Computational approaches to 3D modeling of RNA. *J. Phys. Condens. Matter* 22, 283101.
- [144] Rother, K., Rother, M., Boniecki, M., Puton, T. and Bujnicki, J.M. (2011) RNA and protein 3D structure modeling: similarities and differences. *J. Mol. Model.* 17, 2325–2336.
- [145] Wells, S.E., Hughes, J.M., Igel, A.H. and Ares Jr., M. (2000) Use of dimethyl sulfate to probe RNA structure in vivo. *Methods Enzymol.* 318, 479–493.



- [146] Sclavi, B., Woodson, S., Sullivan, M., Chance, M. and Brenowitz, M. (1998) Following the folding of RNA with time-resolved synchrotron X-ray footprinting. *Methods Enzymol.* 295, 379–402.
- [147] Merino, E.J., Wilkinson, K.A., Coughlan, J.L. and Weeks, K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* 127, 4223–4231.
- [148] Mortimer, S.A. and Weeks, K.M. (2009) Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat. Protoc.* 4, 1413–1421.
- [149] Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* 9, 133–148.
- [150] Hofacker, I.L. and Stadler, P.F. (2006) Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics* 22, 1172–1176.
- [151] Do, C.B., Woods, D.A. and Batzoglou, S. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* 22, e90–e98.
- [152] Giegerich, R., Voss, B. and Rehmsmeier, M. (2004) Abstract shapes of RNA. *Nucleic Acids Res.* 32, 4843–4851.
- [153] Mathews, D. (2004) Predicting the secondary structure common to two RNA sequences with Dynalign, current protocols in bioinformatics/editorial board, Andreas D Baxeavanis [et al.], Chapter 12, Unit 12 4.
- [154] Hochsmann, M., Voss, B. and Giegerich, R. (2004) Pure multiple RNA secondary structure alignments: a progressive profile approach. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 1, 53–62.
- [155] Das, R. and Baker, D. (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14664–14669.
- [156] Sharma, S., Ding, F. and Dokholyan, N.V. (2008) IFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics* 24, 1951–1952.
- [157] Martinez, H.M., Maizel Jr., J.V. and Shapiro, B.A. (2008) RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. *J. Biomol. Struct. Dyn.* 25, 669–683.
- [158] Terribilini, M., Sander, J.D., Lee, J.H., Zaback, P., Jernigan, R.L., Honavar, V. and Dobbs, D. (2007) RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res.* 35, W578–W584.
- [159] Perez-Cano, L. and Fernandez-Recio, J. (2010) Optimal protein-RNA area, OPRA: a propensity-based method to identify RNA-binding sites on proteins. *Proteins* 78, 25–35.
- [160] Zhao, H., Yang, Y. and Zhou, Y. (2011) Structure-based prediction of RNA-binding domains and RNA-binding sites and application to structural genomics targets. *Nucleic Acids Res.* 39, 3017–3025.
- [161] Kim, O.T., Yura, K. and Go, N. (2006) Amino acid residue doublet propensity in the protein-RNA interface and its application to RNA interface prediction. *Nucleic Acids Res.* 34, 6450–6460.
- [162] Dominguez, C., Boelens, R. and Bonvin, A.M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125, 1731–1737.
- [163] Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C. and Vakser, I.A. (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U.S.A.* 89, 2195–2199.
- [164] Ritchie, D.W. and Kemp, G.J. (2000) Protein docking using spherical polar Fourier correlations. *Proteins* 39, 178–194.
- [165] Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. and Wolfson, H.J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33, W363–W367.
- [166] Gabb, H.A., Jackson, R.M. and Sternberg, M.J. (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* 272, 106–120.
- [167] Huang, Y., Liu, S., Guo, D., Li, L. and Xiao, Y. (2013) A novel protocol for three-dimensional structure prediction of RNA-protein complexes. *Sci. Rep.* 3, 1887.
- [168] Agostini, F., Zanzoni, A., Klus, P., Marchese, D., Cirillo, D. and Tartaglia, G.G. (2013) CatRAPID omics: a web server for large-scale prediction of protein-RNA interactions. *Bioinformatics* 29, 2928–2930.
- [169] Bellucci, M., Agostini, F., Masin, M. and Tartaglia, G.G. (2011) Predicting protein associations with long noncoding RNAs. *Nat. Methods* 8, 444–445.
- [170] Grozdanov, P.N. and Macdonald, C.C. (2014) High-throughput sequencing of RNA isolated by cross-linking and immunoprecipitation (HITS-CLIP) to determine sites of binding of CstF-64 on nascent RNAs. *Methods Mol. Biol.* 1125, 187–208.
- [171] Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A. and Darnell, R.B. (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302, 1212–1215.
- [172] Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63.
- [173] Ong, S.E. and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* 1, 252–262.
- [174] Schiess, R., Mueller, L.N., Schmidt, A., Mueller, M., Wollscheid, B. and Aebersold, R. (2009) Analysis of cell surface proteome changes via label-free, quantitative mass spectrometry. *Mol. Cell. Proteomics* 8, 624–638.
- [175] Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P. and Bartel, D.P. (2008) The impact of microRNAs on protein output. *Nature* 455, 64–71.
- [176] Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R. and Rajewsky, N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63.
- [177] Brockmann, R., Beyer, A., Heinisch, J.J. and Wilhelm, T. (2007) Posttranscriptional expression regulation: what determines translation rates? *PLoS Comput. Biol.* 3, e57.
- [178] Lackner, D.H., Beilharz, T.H., Marguerat, S., Mata, J., Watt, S., Schubert, F., Preiss, T. and Bahler, J. (2007) A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol. Cell* 26, 145–155.
- [179] Issel-Tarver, L., Christie, K.R., Dolinski, K., Andrada, R., Balakrishnan, R., Ball, C.A., Binkley, G., Dong, S., Dwight, S.S., Fisk, D.G., et al. (2002) Saccharomyces genome database. *Methods Enzymol.* 350, 329–346.
- [180] Huttelmaier, S., Zenklusen, D., Lederer, M., Dichtenberg, J., Lorenz, M., Meng, X., Bassell, G.J., Condeelis, J. and Singer, R.H. (2005) Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1. *Nature* 438, 512–515.
- [181] Paquin, N., Menade, M., Poirier, G., Donato, D., Drouet, E. and Chartrand, P. (2007) Local activation of yeast ASH1 mRNA translation through phosphorylation of Khd1p by the casein kinase Yck1p. *Mol. Cell* 26, 795–809.
- [182] Levisky, J.M. and Singer, R.H. (2003) Gene expression and the myth of the average cell. *Trends Cell Biol.* 13, 4–6.
- [183] Ge, H., Walhout, A.J. and Vidal, M. (2003) Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Genetics* 19, 551–560.
- [184] Hegde, P.S., White, I.R. and Debouck, C. (2003) Interplay of transcriptomics and proteomics. *Curr. Opin. Biotechnol.* 14, 647–651.