From specific to global analysis of posttranscriptional regulation in eukaryotes: posttranscriptional regulatory networks

Sarath Chandra Janga

Abstract

Regulation of gene expression occurs at several levels in eukaryotic organisms and is a highly controlled process. Although RNAs have been traditionally viewed as passive molecules in the pathway from transcription to translation, there is mounting evidence that their metabolism is controlled by a class of proteins called RNA-binding proteins (RBPs), as well as a number of small RNAs. In this review, I provide an overview of the recent developments in our understanding of the repertoire of RBPs across diverse model systems, and discuss the computational and experimental approaches currently available for the construction of posttranscriptional networks governed by them. I also present an overview of the different roles played by RBPs in the cellular context, based on their *cis*-regulatory modules identified in the literature and discuss how their interplay can result in the dynamic, spatial and tissue-specific expression maps of RNAs. I finally present the concept of posttranscriptional network of RBPs and their cognate RNA targets and discuss their cross-talk with other important posttranscriptional regulatory molecules such as microRNAs s, resulting in diverse functional network motifs. I argue that with rapid developments in the genome-wide elucidation of posttranscriptional networks it would not only be possible to gain a deeper understanding of regulation at a level that has been under-appreciated in the past, but would also allow us to use the newly developed high-throughput approaches to interrogate the prevalence of these phenomena in different states, and thereby study their relevance to physiology and disease across organisms.

Keywords: gene regulation; expression; posttranscriptional control; RNA; RNA-binding proteins; eukaryotes

INTRODUCTION

Regulation of gene expression is a complex process known to be controlled at several levels. In all organisms from the prokaryote *Escherichia coli* to the higher eukaryotes, gene expression is first regulated at the transcriptional level where transcription factors facilitate the RNA synthesis in response to internal or external stimuli [1–4]. On the other hand, at the protein level, several posttranslational modifications, such as phosphorylation by kinases, sumoylation by ubiquitin ligases and acetylation by acetyl–transferases, are known to spatially and temporally control the availability of functional protein products within the cell. However, a much less understood level of regulation of gene expression, which occurs between these two layers, is the posttranscriptional control of RNAs. In contrast to prokaryotes where transcription and translation are coupled, in eukaryotes transcription takes place in the nucleus and translation in the cytoplasm. This uncoupling of

Corresponding author. Sarath Chandra Janga, School of Informatics, Indiana University Purdue University, Indianapolis, Indiana, Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 719 Indiana Ave Ste 319, Walker Plaza Building, IN 46202, USA. Tel: 317-278-4147; Fax: 317-278-9201; E-mail: scjanga@iupui.edu

Sarath Chandra Janga obtained his PhD from the MRC Laboratory of Molecular Biology & University of Cambridge in 2010. He is currently an Assistant Professor of Informatics at the School of Informatics, Indiana University Purdue University at Indianapolis and a faculty member of the Center for Computational Biology and Bioinformatics at the Indiana University School of Medicine. Sarath's research interests include understanding the design principles and constraints imposed on gene regulatory systems within the broader field of computational and systems biology that his lab works on.

transcription and translation provides extensive opportunities for an additional layer of gene expression control at posttranscriptional level. The presence of this posttranscriptional control is also evidenced by a number of studies, which showed that in general, there is a poor correlation between the mRNA and protein pools in eukaryotic cells [5–10]. It is now increasingly accepted that this level is controlled by a complex interplay of numerous RNA-associated factors with the major protein players being the RNA-binding proteins (RBPs) [11–14]. Therefore, RBPs provide an additional layer of plasticity in controlling gene expression. They have been shown to be involved in the regulation of several biological processes such as embryo development and stem cell differentiation [15–17], T-cell activation [18], angiogenesis [19], etc. To understand the mechanism of how these processes are regulated and affected by RBPs, several large-scale studies have been performed to identify the genome-wide RNA targets of RBPs using numerous recently developed methods [20–24]. Due to their central role in controlling gene expression at the posttranscriptional level, alteration in expression or mutations in either RBPs or their binding sites in target transcripts have also been reported to be the cause of several human diseases such as muscular atrophies, neurological disorder and cancer (reviewed extensively in [25–27]).

In this review, I provide a comprehensive overview of this rapidly developing area of posttranscriptional regulatory networks, formed by RBPs and their cognate RNA targets, due to the avalanche of data from several high-throughput technologies. I organize it into three major sections namely, (i) computational and experimental methods for identifying RBPs and their RNA targets, (ii) *cis*-regulatory elements and the global regulation by RBPs and finally, (iii) discuss the structure, dynamics and cross-talk of these posttranscriptional networks with other posttranscriptional players, namely microRNAs (miRNAs), based on recent studies.

COMPUTATIONAL AND EXPERIMENTAL IDENTIFICATION OF RBPS AND THEIR RNA TARGETS

A fundamental area of exploration in elucidating posttranscriptional networks is the identification of the repertoire of RBPs across organisms, and several approaches both computational and experimental have been developed in recent years. Computational approaches typically involve either the sequence-based identification of the set of protein-coding genes containing the *bonafide* RNA-binding domains or employing both sequence and structural protein–RNA templates to predict novel proteins that have the ability to bind to RNA. Sequence-based methods rely on the assumption that homologous sequences have the same broad biological function i.e. RNA-binding ability and use a set of known RBPs and/or their RNA-binding domains to identify homologs across a genome, which is usually followed by manual curation of the collected set to identify a high-confidence set of RBPs [20, 28]. However, because lack of sequence similarity does not always reflect a lack of structural similarity, several structure-based methods are increasingly being used for predicting the repertoire of DNA/RNA-binding proteins in genomic sequences to complement the sequence-based methods [29–33]. Structure-based methods commonly use machine learning techniques to collect and train the sequence information on protein–RNA contacts for all available complexes from protein databank, which is then employed for predicting new members of the structural family. As such, structure-based methods encompass both template-based methods which, employ known structures as template to predict unknown structures, as well as purely structure-based methods which work only for proteins with known structures [34]. Structure-based methods, in addition to complementing sequence-based methods, were shown to exhibit about 10% improvement either in precision or sensitivity than Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) and other sensitive sequence search techniques [29, 31]. The experimental techniques comprise of employing the protein chip of an organism of interest to probe for the potential binding of the cellular RNA molecules, and are analogous to the attempts at characterizing the repertoire of DNA-binding proteins [35–38]. Another strategy developed to identify the RBPs attached to known RNA molecules is the Peptide nucleic acid (PNA)-assisted identification of RBPs (PAIR) [39]. This assay utilizes a specific mRNA-binding probe (PNA) that has the ability to cross the cell membrane and bind to the RNA of interest. This probe also contains the photoactivable amino acid adduct *p*-benzophenylalaline (Bpa), which can covalently cross-link with the RBP, associated with the RNA, on photoactivation. After delivery of PNA,

cells are exposed to ultra violet light for cross-linking of PNA to RBPs associated with RNA of interest. Cells are then lysed, treated with RNase, and PNA–RBP adducts are isolated by using sense oligo (bound to PNA)-coupled magnetic beads. Following which RBPs are identified by mass spectrometry. This method has been used to identify the RBPs associated with ankylosis (ank) RNA, a pan-neuronal dendritically localized RNA [40]. The ank RNA encodes for an inorganic pyrophosphate transporter and its mutation is known to cause a generalized, progressive form of arthritis accompanied by mineral deposition, formation of bony outgrowths and joint destruction [41]. The dendritic localization of RNAs is a rare event occurring for only ∼5% of the cellular RNAs [42]. Table 1 summarizes a list of commonly used computational and experimental methods for identifying RBPs and their targets.

RBPs bind to their RNA targets with the help of several domains having different specificity and affinity. Some of the most common domains are RRM (RNA recognition motif), KH (K homology domain), SR (serine arginine domain), Zn-finger, Pumilio/FBF (PUF domain) and Sm [11]. Table 2 shows the most frequently occurring RNA-binding domains in the yeast, *Saccharomyces cerevisiae*, along with the commonly appearing partner domains in the conventional list of 560 RBPs reported by Hogan *et al.* [20]. A large number of RBPs have been predicted in several model organisms, including humans on the basis of the presence of these commonly occurring domains. For instance, in *Caenorhabditis elegans*, approximately 500 proteins are annotated as RBPs on the basis of the presence of one or more RNA-binding domains. In the yeast, *S. cerevisiae*, about 560 proteins have been reported as conventional RBPs using sequence-based searches and literature mining [20], with an additional approximately 200 reported from proteomic screens as unconventional RBPs [60, 61]. In humans, more than 1000 proteins are considered as *bona fide* RBPs, with 497 containing at least one RRM domain [62]. Other than these putative conventional RBPs (on the basis of previously known RNA-binding domains), several metabolic enzymes have also been shown to bind to RNA molecules [63]. For example Aco1, a TCA cycle enzyme, in yeast *S. cerevisiae* binds to several RNAs encoded by the mitochondrial genome [20] with several other enzymes identified as novel RBPs from proteome-wide screens in yeast [60, 61]. Likewise, recent studies have also shown the ability of RBPs to bind to DNA, suggesting that some of the known RBPs might act as unconventional DNA-binding proteins [35]. These examples indicate the potential for the existence of novel classes of RBPs in eukaryotes with yet to be discovered functional roles. In fact, in a recent study by Castello *et al.* [56], the authors described a comprehensive approach by integrating the two different versions of the CLIP protocols followed by detailed proteomic analysis, to define a high-confidence set of 860 mRNA-binding proteins in the HeLa cells, which suggested that RBPs are common in intermediary metabolism, frequently function as kinases and report the presence of novel domain architectures that have the ability to bind to RNA but the RNA-binding domain is yet to be discovered. The study also strengthened the prevailing notion that RBPs are significantly unstructured and enriched in short repetitive motifs [56, 64].

Although several RBPs have been identified on the basis of conservation of domains in different organisms, identifying the targets of these RBPs using computational means has been a very challenging task with relatively limited progress in recent years [43]. However, several experimental methods have been employed to identify the targets of RBPs, both *in vitro* and *in vivo*. Traditionally, RNA targets for known RBPs have been identified *in vitro* by using cross-linking immunoprecipitation followed by electromobility shift assays [65, 66]. More recently, one hybrid [67] and three hybrid assays [44] have been used to identify the interaction of an RBP and RNA molecule *in vivo*. In addition to these methods, there is increasing use of SELEX (systematic evolution of ligands by exponential enrichment)-based approaches [68] for high-throughput elucidation of the binding sites of RBPs [54, 55]. For instance, Riordan *et al.* [54] took advantage of the yeast genome-wide collection of Tandem Affinity Purification (TAP)-tagged strains to conduct SELEX binding reactions by adding *in vitro* transcribed RNA pools, consisting of 30 randomized bases flanked by two 20 base constant regions, to a cell lysate containing the TAP-tagged RBP of interest. This allowed them to represent in each reaction, ∼600-fold coverage of all 20-mers in their randomized pool. Four cycles of selection were performed for each of the 12 different RBPs studied, which allowed them to confirm or discover novel binding sites for 10 of the RBPs. In an attempt to speed up the discovery of consensus binding sites of RBPs

Table I: Different computational and experimental methods to identify novel RBPs, their targets or RBP–RNA interactions

Method	Description	Reference
Sequence-based	These methods to identify RBPs rely on the assumption that homologous proteins have the same broad biological function, i.e RNA-binding ability and use a set of known RBPs and/or their RNA-binding domains to identify homologs across a genome.	[20, 28]
Structure-based	These methods integrate the currently available RNA–protein structural templates and sequence properties of the binding interfaces to predict proteins that fall into the same structural scaffold. Can be subdivided into template-based ones, which use known structures as template to predict unknown structures or purely structure-based methods, which can only predict proteins whose structures are known. Similar methods are used to predict the binding sites of RBPs although most have been attributed with limited success in predicting them [49].	[29–34]
Three hybrid	*In vivo* yeast genetic method to detect and analyze the RNA–RBP interaction of known RNA and RBPs. This method is based on the binding of bifunctional RNA to both of the two hybrid proteins which activates the expression of reporter gene.	[53]
RNA compete	*In vitro* identification of RNA-binding specificity of a RBP. High concentration of RNA is used and incubated with tagged RBP. This high concentration provides competition for binding and hence this technique gets its name. RBP–RNA complexes are purified and a microarray is used to identify the specific binding sites of RBP.	[57]
RIP-chip	*In vivo* identification of RNA targets of a RBP of interest. Cells are lysed and RBP–RNA complexes are immunoprecipitated in native state. Target RNA is extracted from the RBP–RNA complexes and identified by the microarray method where total RNA of the cell is used as a control.	[59]
CLIP	*In vivo* identification of RNA targets for RBP of interest. Cells are treated with ultraviolet light to covalently cross-link RBP–RNA complexes. Cells are then lysed and RBP–RNA complexes are immunoprecipitated, and the RNA identified by RT–PCR. Modifications of this method employ sequencing technologies to identify the RNA targets of an RBP (referred to as High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) or CLIP-seq).	[22, 63]
iCLIP	iCLIP is a modification of the CLIP protocol that captures the truncated cDNAs by replacing one of the inefficient intermolecular RNA ligation steps with a more efficient intramolecular cDNA circularization. This allows the sequencing of the truncated cDNAs, which are typically lost in CLIP, thereby provide insights into the position of the cross-link site at nucleotide resolution.	[24, 65]
PAR-CLIP	This method relies on the incorporation of photoreactive ribonucleoside analogs, such as 4-thiouridine (4-SU) and 6-thioguanosine (6-SG) into nascent RNA transcripts by living cells. Irradiation of the cells by UV light induces efficient cross-linking of photoreactive nucleoside-labeled cellular RNAs to interacting RBPs. Immunoprecipitation of the RBP of interest is followed by isolation of the cross-linked and coimmunoprecipitated RNA. The isolated RNA is converted into a cDNA library and deep sequenced using Solexa technology.	[21, 64]
PAIR	*In vivo* identification of novel RBPs. A mRNA-binding PNA probe is delivered into the cells. Cells are then exposed to UV light that enables PNA to bind with RBP. Cells are lysed and PNA–RNA–RBP complexes are immunoprecipitated and RBPs are identified by mass spectrometry.	[40]
SERF	*In vitro* selection of RNA fragments that bind to RBP. A random pool of fragmented RNA is generated and incubated with RBP in a test tube. The RBP–RNA complex is extracted by filtration on nitrocellulose membrane. Selection cycle is repeated several times and selected RNA fragments are cloned and identified by the consensus sequences binding to RBP.	[118]
TRAP	*In vivo* system for identification of RNA–RBP interactions in yeast. This involves the transformation of reporter mRNA encoding GFP protein and expression of RBP of interest. Fluorescence intensity of the GFP is measured to know the binding of the RBP of interest. High affinity interactions lead to low expression and low fluorescence intensity.	[119]
SNAAP	*In vitro* method used to identify mRNAs bound to specific RBPs. Purified tagged RBP is treated with cell lysate. This is followed by immunoprecipitation of the mRNP using antibody against tag. Target mRNA is identified by the differential display method	[120]
Ribotrap	Expression of a reporter mRNA containing a 3′-UTR recognition site for a known RBP is followed by RBP immunoprecipitation and analysis of associated RNP components by mass spectrometry.	[121]
SELEX	Immunoprecipitation of RBPs bound to artificial RNAs *in vitro*, followed by cDNA sequencing to identify sequence motifs. In genomic SELEX genome-based RNA pool is used generated by random priming and *in vitro* transcription to reduce complexity and increase sensitivity.	[55, 56]
Interactome capture	In this approach two different versions of the CLIP protocols, i.e conventional and PAR, are applied to the living cells and the resulting fractions are analyzed by high-resolution LC-MS/MS followed by detailed proteomic analysis.	[47]
RaPID	Identification of RNP components associated to RNA-aptamer tagged mRNA *in vivo* by mass spectrometry, which allows the detection of different RNA species captured in the same RNP by quantitative real-time PCR	[122]
Quantitative proteomics	*In vitro* method to identify RBPs bound to specific RNA sequences. An RNA aptamer is tagged with an RNA sequence and then incubated with cell lysate. The RNA aptamer–RNA–RBPs complex is purified and RBPs are identified by using mass spectrometry.	[123]

Table 2: Common RNA-binding domains in the putative list of RBPs from yeast, *S. cerevisiae* [20] along with their frequency of occurrence

Domain	Pfam accession	Description	Protein frequency	Frequently occurring partner domains
RRM.I	PF00076	RRM. Many eukaryotic proteins containing one or more copies of this putative RNA-binding domain of about 90 amino acids. They are known to bind single-stranded RNAs.	0.105	RRM.I, Lsm.interact
DEAD	PF00270	DEAD/DEAH box helicase. Members of this family include the DEAD and DEAH box helicases.	0.042	Helicase C,
KH.I	PF00013	KH is a domain of 70 amino acids and is present in diverse RBPs.	0.015	KH.I
PUF	PF00806	Pumilio-family RNA-binding repeat. Puf domain usually occurs as a tandem repeat of eight domains.	0.013	PUF, RRM.I
WD40	PF00400	WD-40 repeats (also known as WD or beta-transducin repeats) are short, approximately 40 amino acid motifs, often terminating in a Trp-Asp (W-D) dipeptide.	0.013	WD40

Note: Also shown are the domains most often associated with these RNA-binding domains according to the Pfam [124] domain database.

belonging to different families, in both structured and unstructured RNA contexts, the Hughes group [45] developed RNA-compete method, which systematically analyzes binding specificities of a complete range of k-mers in a single binding reaction using the microarray technology. The authors identified the binding sites of nine different RBPs using this approach and concluded that it is sufficient to represent the binding preferences of these RBPs with a 7-mer. However, the above methods have limitations in their ability to identify new targets on a genomic scale in *in vivo* conditions. Therefore, other *in vivo* assays have been developed to identify the novel targets of a RBP such as RNP (Ribonucleoprotein) immunoprecipitation-microarray (RIP-chip), ultraviolet (UV) cross-linking and immunoprecipitation (CLIP) and photo-activatable-ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP). These high-throughput immunoprecipitation assays usually work on a similar concept wherein (i) the RBP complex together with its target RNAs is first purified and (ii) the target RNA identified. However, they differ in the specific procedure used for purifying RBP–RNA complexes and identification of target RNAs. In the RNP RIP-chip method, instead of treating cells with UV light to cross-link RBP–RNA complexes—which is the case in the other two methods, cells are lysed directly and native RBP–RNA complexes for the RBP of interest is purified from the cell lysate using immunoprecipitation. Following this, RNA is isolated from the complexes and identified by using high-density

microarrays [46, 69]. The targets of the Puf family of RBPs and other RBPs in *S. cerevisiae* have been identified by using a modified RIP-chip method, where TAP RBPs were used to facilitate the immunoprecipitation step [20, 70]. These studies showed that the RNA targets vary widely from 1 to 1300 for the studied RBPs in yeast. For instance, Nop13, responsible for pre-18 s rRNA processing, has two RNA targets, whereas Npl3 and Mex67, both involved in mRNA export, have 1266 and 1150 RNA targets respectively [20, 71]. One of the main drawbacks of RIP-chip is that indirect targets of an RBP, i.e. often those RNAs bound by other members of an RNP complex, are also identified as targets of an RBP under study, thereby detecting nonspecific interactions. In addition, in RIP-chip protocol, RNA–protein complexes can be rearranged after cell lysis, due to the absence of cross-linking, thereby producing artifactual results [72]. Moreover, because the resulting microarray data is of low resolution, the binding site in the co-purified RNA remains unresolved. To address some of these challenges, UV CLIP method was developed [22]. In this method, cells are exposed to UV C light of 254 nm to cross-link RBP–RNA molecules in living cells. Then the cells are lysed and cross-linked RBP–RNA complexes are immunoprecipitated using antibody against the RBP of interest. Further, RNA is isolated from the complexes and identified by RT–PCR or through one of the sequencing protocols (which are referred to as HITS-CLIP or CLIP-seq methods [23, 47]). For instance, in a study to discover the targets of the

splicing factor Nova, 34 transcripts were identified by using the CLIP method [22]. PAR–CLIP method developed by Tuschl *et al.* [21, 49] relies on the incorporation of photoreactive ribonucleoside analogs, such as 4-thiouridine (4-SU) and 6-thioguanosine (6-SG) into nascent RNA transcripts of living cells. Irradiation of the cells by UV A light of 365 nm induces efficient cross-linking of photoreactive nucleoside-labeled cellular RNAs to interacting RBPs. Immunoprecipitation of the RBP of interest is followed by isolation of the cross-linked and co-immunoprecipitated RNA. The isolated RNA is converted into a cDNA library and deep sequenced, using Solexa technology. One characteristic feature of cDNA libraries prepared by PAR-CLIP is that the precise position of cross-linking can be identified by mutations residing in the sequenced cDNA, thereby making it possible to separate them from the background of sequences derived from abundant cellular RNAs. This is in contrast to the most first generation UV-CLIP methods in which location of the cross-link is not readily identifiable within the sequenced cross-linked fragments, making it difficult to separate UV-cross-linked target RNA segments from background noncross-linked RNA fragments also present in the sample. However, recently, the iCLIP (individual-nucleotide resolution CLIP) method was introduced which has the ability to capture many cDNAs that truncate prematurely at the cross-linked nucleotide and are lost during the standard CLIP library preparation protocol but are otherwise, very informative in identifying the position of the cross-link site at base pair resolution [48, 73]. For instance, this method has been applied to show that hnRNP particles, which are widely believed to control pre-mRNA processing in the nucleus, recognize uridine tracts with a defined long-range spacing consistent with the tetramer organization of the hnRNP C proteins [24]. The authors also found that hnRNP particles predominantly assemble on introns and exons but remain generally excluded from splice sites, whereas integration of transcriptome-wide iCLIP data and alternative splicing profiles provided insights into how the positioning of hnRNP particles on the transcripts determines the inclusion of alternative exons.

Although all the CLIP variants discussed above produce large amounts of sequencing data of high quality, which can help in the identification of target RNAs, as well as in the inference of binding sites,

they are not sufficient to quantitatively compare the extent of binding nor are they a true reflection of the binding affinities of different sites. This is because CLIP read counts from all these techniques are not necessarily a direct measure of RBP affinity, as they can be affected by factors, such as half-life of the bound RNA region or the cross-linking efficiency of the given sequence. An additional factor to be considered in this context is that, owing to biases in the PCR amplification step, libraries can result in thousands of sequences that originated from a single cDNA. This can lead to data redundancy with limited information content causing a bias in interpreting the affinity of binding and hence current CLIP protocols do not provide data at a resolution which can be used for quantitative modeling of posttranscriptional networks [73]. In addition, because the library preparation protocols for these techniques require a large number of enzymatic steps, they can also potentially affect binding site detection. For instance, it is important to optimize the conditions of partial RNase digestion, as over-digestion can decrease the number of identified sites [74]. In fact, the key to high resolution in CLIP methods is the RNase treatment that removes unprotected RNA fragments. Also the cross-linking efficiency with UV C light (used for HITS-CLIP and iCLIP techniques) or UV A light (used for PAR-CLIP) varies significantly for different proteins and hence, the optimal conditions need to be experimentally determined individually for a protein of interest [74]. Furthermore, because the number of steps involved in all the CLIP protocols is large, significant amount of material can be lost by the time cDNA library is generated from cross–linked RNA. As a consequence, the resulting cDNA libraries rarely contain the full range of RNA-binding sites [73]. In addition to these common limitations of CLIP methods, there are also technique-specific drawbacks. For instance, whereas PAR-CLIP is limited to cultured cells that can efficiently incorporate nucleoside analogs, HITS-CLIP or iCLIP can provide a challenge to data analysis for mining the binding site, especially when the sequencing depth is low. One frequent approach employed by recent studies to identify functional interactions from a vast number of sites identified in these screens and to remove the false positives, is to integrate diverse layers of complementary information obtained from genome-wide studies using technologies such as microarray and RNA-seq [23, 24, 75]. Such data

integration can not only provide new biological insights but also provide an integrated view of the posttranscriptional regulatory landscape. In the future, it might be more common to integrate CLIP data with other high-throughput technologies like ribosome profiling [76] and native elongating transcript sequencing (NET-seq) [77], developed recently by Churchman and Weissman to gain a deeper understanding of the interplay between diverse layers of regulatory control.

CIS-REGULATORY ELEMENTS AND GLOBAL REGULATION BY RBPS

RBPs are key regulators of different steps in the metabolism of RNA in eukaryotes. They participate in the processing of pre-mRNA, which includes splicing, poly-adenylation and capping to produce mature mRNA. Following which, they are responsible for mediating the transport of mRNA from nucleus to cytoplasm. RBPs are also found to facilitate and control the localization, translation, stability and degradation of mRNA (Figure 1). To regulate the different steps of RNA metabolism, RBPs bind to RNA and form RNP complexes. Depending upon whether RBPs are bound to pre-mRNA or mRNA, RNPs are classified as heteronuclear RNPs (hnRNPs) or messenger RNPs (mRNPs), respectively. RNPs are inherently highly dynamic complexes due to their ability to associate and dissociate with various RBPs to mediate different steps of RNA metabolism. Some RBPs associated with RNP complexes are known to remain bound to their target RNA during all the steps of the RNA processing, from splicing to translation. For instance, SF2/ASF, a member of the SR class of RBPs in mammals, is found to facilitate splicing, export and translation initiation of its target RNA [78, 79] with recent iCLIP studies on two of the SR family members indicating that family members map to large nonoverlapping target genes with distinct *in vivo* consensus binding motifs [80]. The latter study also reported the evidence for cross-regulation between SR family members by alternative splicing. Similarly, Npl3, a yeast SR protein, has also been shown to interact with pre-mRNA and regulate the events from splicing to translational elongation [81]. Yet another example is that of the neuronal Embryonic Lethal Abnormal Vision (ELAV) protein, which regulates the fate of its target RNA by mediating

the events from poly-adenylation to translation [82]. On the other hand, several RBPs are also responsible for participating in specific steps of RNA metabolism such as the Nova protein, which is associated with splicing in neuronal cells [22, 75]. Tap protein, like its yeast homolog Mex67, was reported to be a *bona fide* mRNA nuclear export factor [83]. All these examples highlight (i) the role of RBPs in regulating the expression of genes in multiple steps at posttranscriptional level and (ii) the complex combinatorial interplay of different RBPs to integrate various posttranscriptional events to fine-tune the availability of transcripts both spatially and temporally. In addition to the key roles of RBPs in controlling expression by binding to RNA targets, there is mounting genome-wide evidence for their ability to alter the RNA sequence [84, 85]. Cellular RNAs have been documented to have more than a 100 structurally distinct posttranscriptional modifications at thousands of sites (http://rna-mdb.cas.albany.edu/RNAmods/). Although many of these modifications have been known for decades, the enzymes that posttranscriptionally modify RNA and the roles of covalent changes of RNA have been less investigated. A prevailing view of RNA modification is that it has an adaptive role that can fine-tune the structures and functions of mature RNAs to influence gene expression, but recent studies have also hinted that RNA modification may have other regulatory functions. For instance, some posttranscriptional RNA modifications can be dynamic and might have roles analogous to those of posttranslational protein modifications [84, 86, 87].

Unlike the DNA-binding transcription factors, whose sequence specificity can be predominantly inferred from the DNA sequence they are bound to [88], functional motifs of RBPs are difficult to predict from sequences alone. This has been attributed to the fact that although the RNA recognition elements of RBPs are highly conserved, unlike their DNA counterparts, it is most often the structure and not the sequence that is conserved [89–92]. It is due to this observation that the prediction of binding sites of RBPs or computationally associating a significantly enriched site in a group of transcripts to a specific RBP continues to be a challenging problem [54, 93, 94]. In a study by Riordan *et al.* [54], the authors used an integrated computational and experimental approach to refine and discover the RNA recognition elements for 29 RBPs that have been previously studied by the authors using
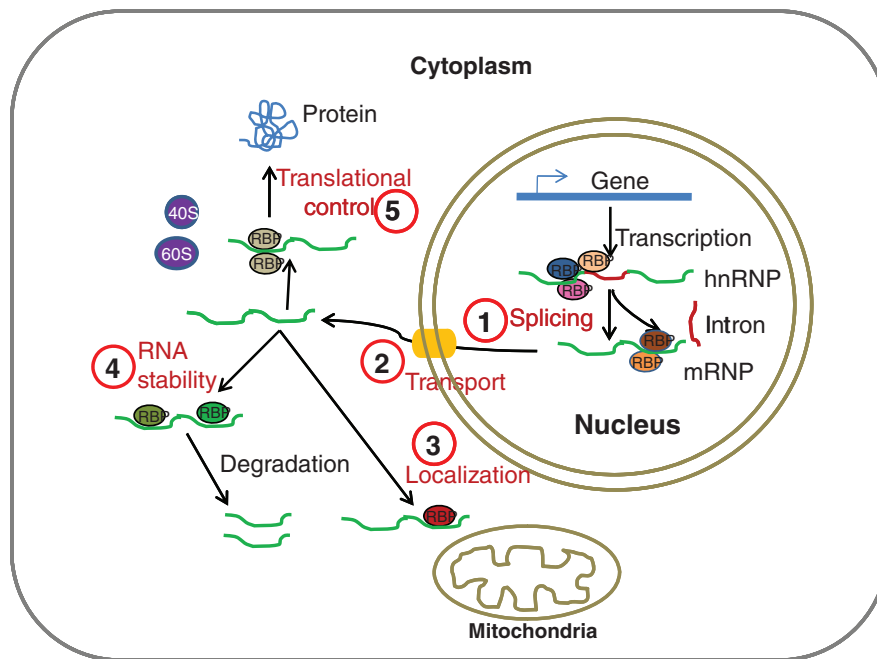
Figure 1: Schematic diagram showing the extensive role of RBPs in various posttranscriptional processes at different locations in eukaryotic cells. The circled number indicates the process in which RBPs are involved. RBPs are major players in splicing pre-mRNAs into mature mRNAs in the nucleus, which are then exported into the cytoplasm by various other RBPs. Depending upon whether RBPs are bound to pre-mRNA or mRNA to form a RNP complex, RNPs are classified as hnRNPs or mRNPs, respectively. In addition, RBPs are responsible for the localization of mRNAs to distinct subcellular compartments such as the mitochondria. In the cytoplasm, RBPs are also involved in governing the stability of transcripts by binding the substrate RNAs, and in controlling the translation of mRNAs into corresponding protein products. Often, multiple RBPs can bind to a single RNA at one or more locations giving rise to a plethora of combinatorial possibilities at every step of posttranscriptional control. For this reason, RBPs have been found to be playing a major role in the cause of several disorders due to changes in regulation they bring about at posttranscriptional level.

RIP-chip screens. Inspired by the FIRE algorithm developed by Elemento *et al.* [95] which is capable of enumerating the *N*-mers that are specifically enriched in the target set of genes/transcripts compared with a genomic background by calculating their frequency distributions, the authors described the relative filtering by nucleotide enrichment (REFINE) approach. REFINE first searches the specific RNAs, identified as targets of a RBP, for segments that contain sequence patterns over-represented in the target set relative to the whole transcriptome, then uses existing tools to identify motifs in these segments. Briefly, the approach involves the identification of all possible hexamers that are enriched in the set of mRNAs bound by a RBP, as the first step. Segments comprising these enriched hexamers, along with three flanking residues on each side and intervening sequences of up to 12 bases that connect two adjacent hexamers are then selected. The resulting filtered target segments are used as input sequences for the Multiple Em for Motif Elicitation (MEME) [96] motif-finding algorithm. The motifs identified by MEME are then evaluated for specific enrichment in target RNAs and statistical significance is evaluated based on random simulations. The authors also perform SELEX experiments for 12 of the RBPs resulting in the identification of a total of eight novel candidate motifs and confirming six earlier reported binding sites. In a complementary approach, Goodarzi *et al.* [93] proposed the TEISER (Tool for Eliciting Informative Structural Elements in RNA) framework for identifying the structural motifs that are informative of the stability of the transcripts measured on a genome-wide scale. In this framework, structural motifs are defined in terms of context-free grammars that represent hairpin structures, as well as primary sequence information. TEISER then employs mutual information to measure the regulatory consequences of the presence or absence of each of approximately 100 million

different seed context-free grammars available in the system. Applying this framework to the RNA stability data from a breast cancer cell line, the authors report eight high-confidence putative structural motifs likely controlling the stability of a vast number of transcripts. Employing mass spectrometry and cross-linking methods, they show that the strongest of their structural motif predictions is bound by HNRPA2B1 (heterogenous nuclear ribonucleoprotein A2/B1) and is responsible for controlling the stability of approximately 4000 transcripts in this cell line. In an attempt to identify the motifs responsible for controlling the decay rates of the transcripts in early *Drosophila* development, Thomsen *et al* [94] grouped the transcripts with common degradation patterns and analyzed the 3'-untranslated region (3'-UTR) sequences using SYLAMER software [97] to identify over-represented motifs of word lengths six or eight. This analysis allowed the identification of 27 significant motifs associated with severe decay patterns, all of which the authors found to be complementary to miRNAs identified in *Drosophila* or other metazoans suggesting that most of the identified motifs might contribute to the degradation of unstable mRNAs. Gene Ontology (GO) analysis of the motif encoding genes further supported their role in the degradation of these transcripts [94].

Most of the studied RBPs in the literature recognize short and degenerate RNA motifs that are typically located at several sites in the genome. Thus, it is not sufficient to determine whether a protein interacts with a particular RNA, but it is important to define the full landscape of interactions of the protein with the RNA on a genome-wide scale. UV Cross-linking and RIP-chip approaches discussed in the previous section have played an instrumental role in improving our global understanding of the binding potential of RBPs. Although the computational methods for the identification of binding sites using the data generated from these technologies are still in their infancy, I discuss the general framework involved in the mapping of binding sites using CLIP-based techniques, all of which eventually generate large amounts of data due to the increasing use of sequencing technologies. Downstream analysis of the sequencing reads obtained from CLIP techniques can be grouped into two major steps: (i) mapping the reads onto the genome and (ii) identification of the binding sites. Read mapping involves the use of algorithms like

Bowtie [98] and TopHat [99] to map high-throughput sequencing reads onto the genome. The later approach which can handle spliced versions of a transcript would be more appropriate when the RNA targets of a RBP being analyzed are in their mature form and hence would map to spaced locations on the genome. To take into account the sequencing errors and cross-link-induced point mutations (from PAR-CLIP [21] or HITS-CLIP [100] data), it is common to allow one or more mismatches in the alignment. This is often achieved by employing short-read nucleotide alignment programs such as GSNAP [101] or Segemehl [102], which allow gapped alignments. In the second step for discovering the binding sites, the main assumption is that the occupancy of the RBP at the authentic sites would clearly be outnumbered (after read normalization) compared with the locations where it is unlikely to have a true functional role. Whereas this may not always be the case, most of the original studies employed this approach borrowed from our understanding of the CHIP-seq datasets for transcription factors and chromatin modifiers. This approach not only filters a vast number of low occupancy sites but also provides a high-confidence list of target sites where the RBP is likely to bind with higher affinity. In fact, when such read clusters with high-RBP occupancy are identified from replicate experiments ensuring that binding at a given site is reproducible and/or the calculation of the significant enrichment over the background signal occurs in the surrounding areas of the same gene then the binding site can be confidently called a *bona fide* site [23, 24, 74, 103]. Once high-affinity sites are identified based on the mapping of clusters of CLIP reads, cross-link nucleotide can be identified through U-to-C transitions in PAR-CLIP data, through deletion sites in the HITS-CLIP reads and in iCLIP the cross-link nucleotide is located one nucleotide upstream of truncation sites. By aligning the high-affinity site regions around the cross-link nucleotide across the genome and employing motif detection algorithms like MEME [96] it should be possible to identify the binding site of an RBP.

In prokaryotes, it has been long known that the genes involved in similar processes tend to cluster on chromosomes and are transcribed together using the same promoter, thus forming operons such as the well studied, Gal, Lac operons. On the other hand, in eukaryotes, chromosomal operons are rare. However, with the availability of wealth of

information on RBP–RNA interactions on a genome-wide scale in model organisms, recently, the concept of posttranscriptional operons has been proposed in eukaryotes [104]. According to this concept, diverse RNAs related to a common biological process are regulated by similar RBPs. For instance, in yeast, a study of the RBP–RNA interactions by modified RIP-chip method hasrevealed that each member of Puf family RBPs bind with functionally and cytotopically related RNAs [70]. Puf1 and Puf2 have been shown to bind to mRNAs of membrane-associated proteins. Similarly, Puf3 binds to cytoplasmic mRNAs of mitochondrial proteins. Likewise, the Nova protein was found to regulate splicing of pre-mRNA encoding components of inhibitory synapses and a stem–loop binding protein (SLBP) was involved solely in splicing and translation of replication-dependent histone RNAs [105]. Further examples in support of posttranscriptional operons have been reviewed extensively elsewhere [13, 106]. These examples demonstrate the role of RBPs in view of posttranscriptional operons for coordinating the expression of functionally related genes in eukaryotes. Given the advantages involved in spatially and temporally controlling the posttranscriptional events by selectively binding to compartment or location or process-specific transcripts, it is possible to speculate that in eukaryotes, RBPs act as mediators in facilitating operon-like organization, which by the virtue of polycistronic transcripts and coupled transcription/translation is inherent in prokaryotes.

GENOME-SCALE POSTTRANSCRIPTIONAL REGULATORY NETWORK DIRECTED BY RBPS

The development of several high-throughput approaches such as CLIP and RIP-chip has increased the amount of data for targets of RBPs in diverse organisms. This data on the interaction between RBPs and their targets can be utilized to construct RBP–RNA interaction network, which is also typically referred to as posttranscriptional regulatory network. A posttranscriptional network is represented in the form of a directional network with each edge corresponding to a regulatory link between the nodes as shown in Figure 2A. In this directed network, one set of nodes are RBPs forming the regulatory proteins, whereas the other set of nodes are RNAs encoded by either protein-coding or nonprotein-coding genes referred to as the target nodes. These two nodes (regulator node and target node) are joined by an arrow starting at the regulator node and directing towards the target node. The target RNA can be encoded by any functional gene in the genome including genes encoded for RBPs. This network can also contain loops as a link starting from RBP and targeting itself, typically referred to as auto-regulation of an RBP (Figure 2B). This loop structure suggests that RBP can bind to its own RNA and control its metabolism at transcript level. There are several examples suggesting the auto-regulation of RBPs at posttranscriptional level. For instance, in humans, RBPs such as AUF1, HuR, KSRP, NF90, TIA-1 and TIAR were reported to associate with their own mRNA and with that of other RBPs [107]. In fact, in yeast more than one-third of the currently studied RBPs have been found to auto-regulate their expression at posttranscriptional level with cross-regulation between RBPs being a common theme [108].

The availability of the network of post-transcriptional interactions for a considerable fraction of RBPs in model systems such as *S. cerevisiae* [20], makes it possible to address several questions concerning the structure and organization of post-transcriptional networks directed by RBPs. Table 3 summarizes some of the properties which govern the structure of this network. It is evident from this table that the majority of the mRNA transcriptomes encoded by ~70% of the genes in yeast, has significant associations with at least one of the RBPs screened for RNA interactions. In fact, on average, each distinct yeast mRNA was found to interact with three of the RBPs, suggesting the potential for a combinatorial and multidimensional network of regulation. Indeed, I found that the average connectivity of a node in this network was approximately seven, indicating that most nodes in this network have more number of targets and/or more number of RBPs controlling them. Other measures of centrality like betweenness and closeness, which provide a measure of the importance of a node in a network, shown in this table, also reflect this trend (see [109] and references therein for definitions). For instance, the average length of the shortest path between two nodes in this network, which gives an indication of the distance between nodes suggests that most nodes are separated by no more than three edges—a measure reflecting the dense networking in this network
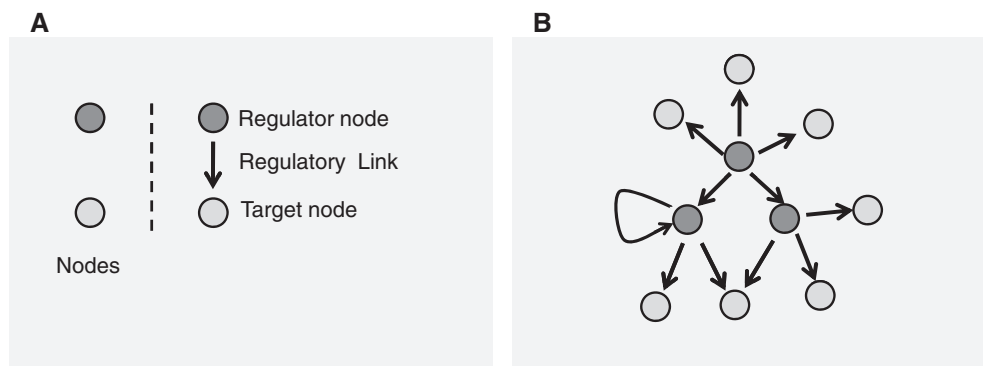
Figure 2: Concept figure showing the RBP-mediated posttranscriptional regulatory network. (**A**) Dark (Regulator) and light (Target) gray circles denote nodes in the network. These nodes are linked to each other via a directional arrow starting from regulator (which is an RBP in the network) and pointing toward target (which may be an RNA or miRNA) in the directional network. These linked nodes simply indicate that RBP (Dark gray circle) binds to RNA/miRNA of target gene (Light gray circle) and regulates its metabolism. (**B**) Shows an illustration of a network representing a dense set of RBP–RNA interactions with different RBPs having diverse targets. The targets of one RBP in the network may be RNA of other genes or miRNA (dark and light circle linked by arrow), the RNA of the RBP itself (loop from dark circle) and the RNA of other RBPs (two dark circles linked by an arrow).

Table 3: Properties defining the structure of the posttranscriptional network of RBPs and their target RNAs in the yeast, *S. cerevisiae*

Property	Definition	Value[a]
No. of edges	Each edge corresponds to a single RBP–RNA interaction. Hence, total edges represent all the interactions in the posttranscriptional network	19 396
No. of vertices/nodes	Total number of nodes, which comprise of both the RBPs, as well as the RNAs, encoding for both protein coding and noncoding genes. This network comprises of 41 RBPs that are screened for their RNA targets.	5398
Degree or connectivity	Degree or connectivity refers to the number of interactions a protein or RNA has in this network—the higher the connectivity (i.e. hub nodes) the more the number of targets and/or more the number of RBPs controlling it.	7.18
Clustering coefficient	Clustering coefficient of a node reflects the extent to which the neighbors of a given node are interconnected among themselves to what is expected theoretically, and indicates the cohesiveness or local modularity of the network. Average value taken over all nodes reflects the modularity of the network.	0.37
Betweenness	Betweenness centrality of a node measures the number of shortest paths between all pairs of nodes in the network that pass through a node of interest—the higher the number of paths that pass through a node, the more important it is.	43.11
Average path length	Average length of the shortest paths between all pairs of nodes in the network.	2.65
Closeness	Closeness centrality is defined as the inverse of the average length of all the shortest paths from a node of interest to all other nodes in the network—note that closeness centrality defined this way implies that higher the closeness value, the higher the importance (centrality) of a node.	0.38
Diameter	The diameter of a network is the length of the longest path among all the shortest paths defined between two nodes. It gives an estimation of the farthest distance between nodes in the network.	6
Graph density	The density of a network is the ratio of the number of edges to the number of total possible edges.	1.33×10^{-3}
Power law fit (exponent-alpha)	Fitting a power–law distribution function to the degree distribution of the network to study whether the network is likely to exhibit a scale-free network structure.	1.77

Source: The dataset employed for characterizing the network structure was obtained from Hogan *et al.* [20].
Note: All the network properties are calculated using *igraph*, a publicly available R package for analyzing graphs [http://cneurocvs.rmki.kfki.hu/igraph/ & http://www.r-project.org]. [a]The average values for the entire network are reported for properties that are defined for specific node or edge.

[110]. Similarly, the diameter of a network that refers to the longest of all the shortest paths between pairs of nodes is about six, indicating that two nodes in this network are separated by no more than six edges. The clustering coefficient which is a proxy for the modularity of the network shows that neighbors of most nodes tend to be highly interconnected among themselves, forming a dense and cohesive network of regulatory linkages at this level of regulation. Finally, although incomplete in size, the scaling exponent of this network is about 1.8, which suggests that the network obeys a scale-free topology with a power–law degree distribution [110]. In simple terms, scale-free topology of a network refers to a degree distribution of nodes, where most nodes in the network have low degrees, whereas few nodes are very highly connected and are referred to as hubs in the network.

Due to their central role in controlling gene expression at the posttranscriptional level, alteration in expression or mutations in either RBPs or their RNA targets have been reported to be the cause of several human diseases such as muscular atrophies, neurological disorders and cancer [25–27, 111]. In particular, disorders such as myotonic dystrophy (DM) and oculopharyngeal muscular dystrophy (OPMD) have been attributed with RNA's gain-of-function. For instance, CUG repeat expansion in the case of DM protein kinase (DMPK) [26] and GCG repeat expansion in exon 1 of the RBP, PABPN1 in the case of OPMD [25] are illustrated examples of RNA's gain-of-function. On the other hand, diseases like opsoclonus-myoclonus ataxia (POMA) and spinal muscular atrophy (SMA) have been reported to be due to the RBPs' loss of function [25], suggesting that mutations in either RBP or any of its interacting RNA target sequences can lead to extensive variations in their expression patterns and result in a number of diseases. In line with these studies discussing the impact of changes in the expression levels of RBPs or their targets being associated with diseases and fitness defects, a study of the posttranscriptional regulatory network in yeast showed that RBPs, as a functional class, are rapidly turned over (i.e. less stable) at the transcript level (quickly degraded) and are tightly controlled at the protein level with very little cell-to-cell variation in their expression, with highly connected RBPs being more abundant and more tightly regulated than less connected ones [112]. Tight regulation of RBPs indicates that variations in the expression levels of these key posttranscriptional regulators can have significant impacts on the functioning of the cell, thereby leading to a disease phenotype, with the effect of the perturbation being more profound for global posttranscriptional controllers.

CROSS-TALK AND NETWORK MOTIFS IN POSTTRANSCRIPTIONAL REGULATORY NETWORK

miRNAs are small, noncoding RNAs that are predominantly known to repress gene expression through interaction with 3′-UTRs of mRNAs and are predicted to target >50% of all human protein-coding genes, enabling them to have numerous regulatory roles in many physiological and developmental processes [113, 114]. Therefore, miRNAs on a genomic scale are known to form a dense network of posttranscriptional interactions controlling a wide range of processes, with recent studies pointing to the downregulation of this class of molecules as a common feature in a number of cancers [115]. It is also becoming increasingly clear that mRNAs do not exist in a truly naked state but instead, are likely bound by multiple molecules, existing as mRNPs. Many of these molecules, including microRNAs and RBPs can regulate gene expression most notably through their interactions with the UTRs of mRNA. Such regulatory mechanisms rely on miRNA and RBP binding activity to common target RNAs and are probably under tight spatio-temporal control [116]. Emerging evidence suggests that *cis*-regulatory code targeted by microRNAs might be the same as that read by RBPs, providing a mechanism to ensure that the appropriate regulatory elements are utilized for the correct expression of a multifunctional mRNA transcript. In particular, several examples wherein microRNAs that would bind the mRNA in a manner that would compete with essential secondary structures such as stems or loops, making a microRNA–mRNA complex and RBP–mRNA complex mutually exclusive, have been found [116–119]. One early example of the interplay between RBP and miRNAs is that of HuR, whose translocation from the nucleus to the cytoplasm following exposure to stress was shown to relieve cationic amino acid transporter 1 (CAT1) mRNA from miR-122-mediated repression in the cytoplasm of liver cells [120]. Recent studies have revealed that

the influence of HuR on many bound transcripts depends on HuRs interplay with miRNAs, which associate with the same mRNAs [118], resulting in the interactions of HuR and miRNAs being both competitive and cooperative that govern expression of shared target mRNAs [119]. Competition between HuR and miRNAs typically results in enhanced gene expression if the HuR–mRNA interaction prevails and in repression if the miRNA remains associated. It was also found that cooperation between HuR and miRNAs lead to lower expression of the shared mRNA [119]. Kedde *et al.* [117] showed that PUM1 binding results in a conformation change that allows miRNA-binding sites in p27 mRNA to be accessible for regulation. In a global analysis of the targets of the human PUF family proteins PUM1 and PUM2 in human cancer cells, PUM-binding motifs were enriched in the vicinity of the predicted miRNA-binding sites. Likewise, high-confidence miRNA-binding sites were generally enriched in the 3′-UTRs of the experimentally determined PUM targets, predicting strong cross-talk between human PUF proteins and miRNA targeting [121]. Indeed, as discussed above, PUM1 induces a structural change within the 3′-UTR of the tumor suppressor p27 transcript, thus permitting access to miR-221/ 222 with consequences for cell cycle progression [117]. There is also increasing evidence in support of the cross-talk between miRNAs, as well as between RBPs and miRNAs. For instance, a tripartite motif protein, TRIM71 with ubiquitin ligase activity is a target of let-7 miRNA. TRIM71 drives AGO degradation through ubiquitylation, thereby interfering with miRNA function [122]. Also pluripotency factor lin-28 binds the pre-let-7 RNA and inhibits its processing by the Dicer ribonuclease in embryonic stem cells and embryocarcinoma cells. In embryonic neural stem cells, lin-28 is downregulated by mir-125 and let-7, allowing processing of pre-let-7 to proceed, thereby forming a feedback loop controlling pre-let-7 maturation [123]. All these studies imply that microRNAs and RBPs may be targeting, at least in part, overlapping regions of the RNA transcript. More specifically, this hypothesis presents a mechanism, whereby miRNAs may modulate the posttranscriptional regulatory code in a manner that masks or reveals the regulatory targets of RBPs. As such, interplay between miRNAs and RBPs on target 3′-UTRs can rapidly modulate target expression under specific conditions. Also binding of RBPs near miRNA target sites can potentially regulate miRNA function, either directly by affecting miRNA-binding or indirectly through a switch in RNA secondary structure [116]. Indeed, several mechanisms have been proposed for potential miRNA interactions with mRNA that would influence RBP binding sites [124]. These studies not only support the interplay between miRNAs and RBPs as a means of spatially and temporally controlling the expression of the target genes in a combinatorial fashion, but also suggest a cross-talk between these classes of regulatory molecules. Asmore data on genome-wide protein–RNA maps become available, it should be possible to elucidate the different network motifs and the frequency with which they are found in posttranscriptional regulatory networks. However, it is possible to speculate at this point that miRNAs and RBPs might interact with each other to control their mutual expression while they cooperate or compete for the binding sites on the target genes' transcript forming at least three node network motifs in posttranscriptional regulatory networks.

CONCLUSION

Although the postgenomic era has introduced the genomic complement of thousands of genomes, it has also left us with several unanswered questions regarding the functional relevance of the genes harbored by an organism and of the principles that govern the regulation of such genes. It is noteworthy to mention that even in a model organism like *S. cerevisiae*, regulation of gene expression at the posttranscriptional level is rather poorly understood. Nevertheless with recent improvements and availability of high-throughput approaches to the study of RBPs, such as RNA-sequencing and immunoprecipitation protocols, we can expect to see a wealth of data detailing the dynamic, spatial and tissue-specific nature of the interactions governed by these exciting class of regulatory molecules. Such advances would undoubtedly allow us to gain a deeper understanding of regulation at a level that has been under-appreciated over the past decades. Given the unprecedented detail at which these high-throughput technologies can reveal the link between the regulatory elements on the target RNAs and the RBPs specific to environmental conditions, it is possible to use these approaches to interrogate the prevalence of these phenomena in different states and thereby study their relevance to physiology and disease in diverse model systems.

Key Points

- Posttranscriptional control of RNA by RBPs plays an important role in diverse cellular processes such as splicing, transcript stability, RNA editing, RNA localization and translation rate.
- About 5–10% of the protein coding genes in eukaryotes are predicted to encode for RBPs, to control diverse cellular targets in a temporal and spatial context by forming dynamic RNP complexes.
- RBPs can be grouped into conventional and unconventional classes with the unconventional RBPs typically exhibiting moonlighting/secondary functions in addition to their ability to bind RNA.
- *Cis*-regulatory binding sites of an RBP are difficult to predict and to build consensus, because of sequence heterogeneity among bound sites that might otherwise be identical in their structure.
- RBPs auto-regulate their expression, cross-regulate other RBPs and control their target RNAs to form a dense posttranscriptional regulatory network on a genomic scale. RBPs also interact with miRNA and mutually control each other's expression to form dynamic network motifs that are perturbed in disease conditions.

References

1. Martinez-Antonio A, Janga SC, Salgado H, *et al*. Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*. *Trends Microbiol* 2006;**14**:22–7.
2. Babu MM, Janga SC, de Santiago I, *et al*. Eukaryotic gene regulation in three dimensions and its impact on genome evolution. *Curr Opin Genet Dev* 2008;**18**:571–82.
3. Janga SC, Collado-Vides J. Structure and evolution of gene regulatory networks in microbial genomes. *Res Microbiol* 2007;**158**:787–94.
4. Babu MM, Luscombe NM, Aravind L, *et al*. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* 2004;**14**:283–291.
5. Ideker T, Thorsson V, Ranish JA, *et al*. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 2001;**292**:929–34.
6. Greenbaum D, Colangelo C, Williams K, *et al*. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* 2003;**4**:117.
7. Gygi SP, Rochon Y, Franza BR, *et al*. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 1999;**19**:1720–1730.
8. Schwanhausser B, Busse D, Li N, *et al*. Global quantification of mammalian gene expression control. *Nature* 2011;**473**:337–342.
9. Ghazalpour A, Bennett B, Petyuk VA, *et al*. Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet* 2011;**7**:e1001393.
10. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 2012;**13**:227–32.
11. Glisovic T, Bachorik JL, Yong J, *et al*. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* 2008;**582**:1977–86.
12. Mata J, Marguerat S, Bahler J. Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci* 2005;**30**:506–14.
13. Keene JD. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* 2007;**8**:533–43.
14. Collins LJ, Penny D. The RNA infrastructure: dark matter of the eukaryotic cell? *Trends Genet* 2009;**25**:120–8.
15. Lee MH, Schedl T. *RNA-binding Proteins*. St. Louis, USA: WormBook 2006,1–13. http://genetics.wustl.edu/tslab/files/2010/03/RNAbindingproteins.pdf.
16. Ratti A, Fallini C, Cova L, *et al*. A role for the ELAV RNA-binding proteins in neural stem cells: stabilization of Msi1 mRNA. *J Cell Sci* 2006;**119**:1442–52.
17. Cassar PA, Stanford WL. Integrating post-transcriptional regulation into the embryonic stem cell gene regulatory network. *J Cell Physiol* 2012;**227**:439–49.
18. Mukherjee N, Lager PJ, Friedersdorf MB, *et al*. Coordinated posttranscriptional mRNA population dynamics during T-cell activation. *Mol Syst Biol* 2009;**5**:288.
19. Chang SH, Hla T. Gene regulation by RNA binding proteins and microRNAs in angiogenesis. *Trends Mol Med* 2011;**17**:650–8.
20. Hogan DJ, Riordan DP, Gerber AP, *et al*. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol* 2008;**6**:e255.
21. Hafner M, Landthaler M, Burger L, *etal*. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010;**141**:129–41.
22. Ule J, Jensen KB, Ruggiu M, *et al*. CLIP identifies Nova-regulated RNA networks in the brain. *Science* 2003;**302**:1212–5.
23. Licatalosi DD, Mele A, Fak JJ, *et al*. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 2008;**456**:464–9.
24. Konig J, Zarnack K, Rot G, *et al*. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* 2010;**17**:909–15.
25. Lukong KE, Chang KW, Khandjian EW, *et al*. RNA-binding proteins in human genetic disease. *Trends Genet* 2008;**24**:416–25.
26. Musunuru K. Cell-specific RNA-binding proteins in human disease. *Trends Cardiovasc Med* 2003;**13**:188–95.
27. Kim MY, Hur J, Jeong S. Emerging roles of RNA and RNA-binding protein network in cancer cells. *BMB Rep* 2009;**42**:125–130.
28. Galante PA, Sandhu D, de Sousa Abreu R, *et al*. A comprehensive in silico expression analysis of RNA binding proteins in normal and tumor tissue: Identification of potential players in tumor formation. *RNA Biol* 2009;**6**:426–33.
29. Zhao H, Yang Y, Zhou Y. Highly accurate and high-resolution function prediction of RNA binding proteins

by fold recognition and binding affinity prediction. *RNA Biol* 2011;**8**:988–96.

30. Kumar M, Gromiha MM, Raghava GP. SVM based prediction of RNA-binding proteins using binding residues and evolutionary information. *J Mol Recognit* 2011;**24**: 303–13.

31. Shao X, Tian Y, Wu L, *et al*. Predicting DNA- and RNA-binding proteins from sequences with kernel methods. *J Theor Biol* 2009;**258**:289–93.

32. Spriggs RV, Murakami Y, Nakamura H, *et al*. Protein function annotation from sequence: prediction of residues interacting with RNA. *Bioinformatics* 2009;**25**: 1492–7.

33. Yu X, Cao J, Cai Y, *et al*. Predicting rRNA-, RNA-, and DNA-binding proteins from primary structure with support vector machines. *J Theor Biol* 2006;**240**:175–84.

34. Zhao H, Yang Y, Zhou Y. Structure-based prediction of RNA-binding domains and RNA-binding sites and application to structural genomics targets. *Nucleic Acids Res* 2011; **39**:3017–25.

35. Hu S, Xie Z, Onishi A, *et al*. Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* 2009;**139**:610–22.

36. Hall DA, Zhu H, Zhu X, *et al*. Regulation of gene expression by a metabolic enzyme. *Science* 2004;**306**: 482–84.

37. Fasolo J, Snyder M. Protein microarrays. *Methods Mol Biol* 2009;**548**:209–22.

38. Zhu H, Bilgin M, Bangham R, *et al*. Global analysis of protein activities using proteome chips. *Science* 2001;**293**: 2101–5.

39. Zeng F, Peritz T, Kannanayakal TJ, *et al*. A protocol for PAIR: PNA-assisted identification of RNA binding proteins in living cells. *Nat Protoc* 2006;**1**:920–7.

40. Zielinski J, Kilk K, Peritz T, *et al*. In vivo identification of ribonucleoprotein-RNA interactions. *Proc Natl Acad Sci USA* 2006;**103**:1557–62.

41. Ho AM, Johnson MD, Kingsley DM. Role of the mouse ank gene in control of tissue calcification and arthritis. *Science* 2000;**289**:265–70.

42. Eberwine J, Belt B, Kacharmina JE, *et al*. Analysis of subcellularly localized mRNAs using in situ hybridization, mRNA amplification, and expression profiling. *Neurochem Res* 2002;**27**:1065–77.

43. Walia RR, Caragea C, Lewis BA, *et al*. Protein-RNA interface residue prediction using machine learning: an assessment of the state of the art. *BMC Bioinformatics* 2012; **13**:89.

44. SenGupta DJ, Zhang B, Kraemer B, *et al*. A three-hybrid system to detect RNA-protein interactions in vivo. *Proc Natl Acad Sci USA* 1996;**93**:8496–501.

45. Ray D, Kazan H, Chan ET, *et al*. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol* 2009;**27**:667–70.

46. Tenenbaum SA, Carson CC, Lager PJ, *et al*. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci USA* 2000;**97**: 14085–90.

47. Darnell RB. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA* 2010;**1**:266–86.

48. Konig J, Zarnack K, Rot G, *et al*. iCLIP–transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp* 2011;**50**:2638.

49. Ascano M, Hafner M, Cekan P, *et al*. Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip Rev RNA* 2012;**3**:159–77.

50. Stelzl U, Nierhaus KH. SERF: in vitro election of random RNA fragments to identify protein binding sites within large RNAs. *Methods* 2001;**25**:351–7.

51. Paraskeva E, Atzberger A, Hentze MW. A translational repression assay procedure (TRAP) for RNA-protein interactions in vivo. *Proc Natl Acad Sci USA* 1998;**95**: 951–6.

52. Rodgers ND, Jiao X, Kiledjian M. Identifying mRNAs bound by RNA-binding proteins using affinity purification and differential display. *Methods* 2002;**26**: 115–22.

53. Beach DL, Keene JD. Ribotrap: targeted purification of RNA-specific RNPs from cell lysates through immunoaffinity precipitation to identify regulatory proteins and RNAs. *Methods Mol Biol* 2008;**419**:69–91.

54. Riordan DP, Herschlag D, Brown PO. Identification of RNA recognition elements in the *Saccharomyces cerevisiae* transcriptome. *Nucleic Acids Res* 2011;**39**:1501–9.

55. Galarneau A, Richard S. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat Struct Mol Biol* 2005;**12**:691–8.

56. Castello A, Fischer B, Eichelbaum K, *et al*. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 2012;**149**:1393–406.

57. Slobodin B, Gerst JE. RaPID: an aptamer-based mRNA affinity purification technique for the identification of RNA and protein factors present in ribonucleoprotein complexes. *Methods Mol Biol* 2011;**714**:387–406.

58. Butter F, Scheibe M, Morl M, *et al*. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci USA* 2009;**106**:10626–31.

59. Finn RD, Mistry J, Tate J, *et al*. The Pfam protein families database. *Nucleic Acids Res*;**38**:D211–22.

60. Scherrer T, Mittal N, Janga SC, *et al*. A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS One* 2010;**5**:e15499.

61. Tsvetanova NG, Klass DM, Salzman J, *et al*. Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*. *PLoS One* 2010;**5**:e12671.

62. Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 2005;**272**: 2118–31.

63. Ciesla J. Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? *Acta Biochim Pol* 2006; **53**:11–32.

64. Tompa P, Csermely P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J* 2004;**18**: 1169–75.

65. Pinero DJ, Hu J, Connor JR. Alterations in the interaction between iron regulatory proteins and their iron responsive element in normal and Alzheimer's diseased brains. *Cell Mol Biol* 2000;**46**:761–76.

66. Thomson AM, Rogers JT, Walker CE, *et al*. Optimized RNA gel-shift and UV cross-linking assays for

characterization of cytoplasmic RNA-protein interactions. *Biotechniques* 1999;**27**:1032–39, 1042.

67. Wilhelm JE, Vale RD. A one-hybrid system for detecting RNA-protein interactions. *Genes Cells* 1996;**1**:317–23.

68. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 1990;**249**:505–10.

69. Keene JD, Komisarow JM, Friedersdorf MB. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* 2006;**1**:302–7.

70. Gerber AP, Herschlag D, Brown PO. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol* 2004;**2**: E79.

71. Hieronymus H, Silver PA. Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. *Nat Genet* 2003;**33**:155–61.

72. Mili S, Steitz JA. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* 2004;**10**: 1692–4.

73. Konig J, Zarnack K, Luscombe NM, *et al*. Protein-RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet* 2011;**13**:77–83.

74. Kishore S, Jaskiewicz L, Burger L, *et al*. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods* 2011;**8**:559–64.

75. Ule J, Stefani G, Mele A, *et al*. An RNA map predicting Nova-dependent splicing regulation. *Nature* 2006;**444**: 580–6.

76. Ingolia NT, Ghaemmaghami S, Newman JR, *et al*. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009;**324**: 218–23.

77. Churchman LS, Weissman JS. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 2011;**469**:368–73.

78. Sanford JR, Gray NK, Beckmann K, *et al*. A novel role for shuttling SR proteins in mRNA translation. *Genes Dev* 2004;**18**:755–68.

79. Zhong XY, Wang P, Han J, *et al*. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell* 2009;**35**:1–10.

80. Anko ML, Muller-McNicoll M, Brandl H, *et al*. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol* 2012;**13**:R17.

81. Gross T, Richert K, Mierke C, *et al*. Identification and characterization of srp1, a gene of fission yeast encoding a RNA binding domain and a RS domain typical of SR splicing factors. *Nucleic Acids Res* 1998;**26**:505–11.

82. Pascale A, Amadio M, Quattrone A. Defining a neuron: neuronal ELAV proteins. *Cell Mol Life Sci* 2008;**65**:128–40.

83. Gruter P, Tabernero C, von Kobbe C, *et al*. TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol Cell* 1998;**1**:649–59.

84. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, *et al*. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 2012;**485**:201–6.

85. Cantara WA, Crain PF, Rozenski J, *et al*. The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* 2011;**39**:D195–201.

86. He C. Grand challenge commentary: RNA epigenetics? *Nat Chem Biol* 2010;**6**:863–5.

87. Shikanai T. RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol Life Sci* 2006;**63**:698–708.

88. Zhao Y, Stormo GD. Quantitative analysis demonstrates most transcription factors require only simple models of specificity. *Nat Biotechnol* 2011;**29**:480–3.

89. Parker BJ, Moltke I, Roth A, *et al*. New families of human regulatory RNA structures identified by comparative analysis of vertebrate genomes. *Genome Res* 2011;**21**:1929–43.

90. George AD, Tenenbaum SA. Web-based tools for studying RNA structure and function. *Methods Mol Biol* 2011; **703**:67–86.

91. Foat BC, Stormo GD. Discovering structural cis-regulatory elements by modeling the behaviors of mRNAs. *Mol Syst Biol* 2009;**5**:268.

92. Rabani M, Kertesz M, Segal E. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. *Proc Natl Acad Sci USA* 2008;**105**: 14885–90.

93. Goodarzi H, Najafabadi HS, Oikonomou P, *et al*. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature* 2012;**485**: 264–68.

94. Thomsen S, Anders S, Janga SC, *et al*. Genome-wide analysis of mRNA decay patterns during early Drosophila development. *Genome Biol* 2010;**11**:R93.

95. Elemento O, Slonim N, Tavazoie S. A universal framework for regulatory element discovery across all genomes and data types. *Mol Cell* 2007;**28**:337–50.

96. Bailey TL, Boden M, Buske FA, *et al*. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 2009;**37**:W202–8.

97. van Dongen S, Abreu-Goodger C, Enright AJ. Detecting microRNA binding and siRNA off-target effects from expression data. *Nat Methods* 2008;**5**:1023–5.

98. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–9.

99. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;**25**: 1105–11.

100. Zhang C, Darnell RB. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* 2011;**29**:607–14.

101. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 2010; **26**:873–81.

102. Hoffmann S, Otto C, Kurtz S, *et al*. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput Biol* 2009;**5**:e1000502.

103. Yeo GW, Coufal NG, Liang TY, *et al*. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* 2009; **16**:130–7.

104. Keene JD, Tenenbaum SA. Eukaryotic mRNPs may represent posttranscriptional operons. *Mol Cell* 2002;**9**: 1161–7.

105. Townley-Tilson WH, Pendergrass SA, Marzluff WF, *et al*. Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein. *RNA* 2006;**12**:1853–67.

106. Keene JD, Lager PJ. Post-transcriptional operons and regulons co-ordinating gene expression. *Chromosome Res* 2005;**13**:327–37.

107. Pullmann R, Jr, Kim HH, Abdelmohsen K, *et al*. Analysis of turnover and translation regulatory RNA-binding protein expression through binding to cognate mRNAs. *Mol Cell Biol* 2007;**27**:6265–78.

108. Mittal N, Scherrer T, Gerber AP, *et al*. Interplay between posttranscriptional and posttranslational interactions of RNA-binding proteins. *J Mol Biol* 2011;**409**:466–79.

109. Junker BH, Koschutzki D, Schreiber F. Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics* 2006;**7**:219.

110. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;**5**:101–13.

111. Cooper TA, Wan L, Dreyfuss G. RNA and disease. *Cell* 2009;**136**:777–93.

112. Mittal N, Roy N, Babu MM, *et al*. Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. *Proc Natl Acad Sci USA* 2009;**106**:20300–5.

113. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009;**136**:215–33.

114. Friedman RC, Farh KK, Burge CB, *et al*. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009;**19**:92–105.

115. Lu J, Getz G, Miska EA, *et al*. MicroRNA expression profiles classify human cancers. *Nature* 2005;**435**:834–8.

116. van Kouwenhove M, Kedde M, Agami R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat Rev Cancer* 2011;**11**:644–56.

117. Kedde M, van Kouwenhove M, Zwart W, *et al*. A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nat Cell Biol* 2010;**12**:1014–20.

118. Uren PJ, Burns SC, Ruan J, *et al*. Genomic analyses of the RNA-binding protein Hu antigen R (HuR) identify a complex network of target genes and novel characteristics of its binding sites. *J Biol Chem* 2011;**286**:37063–6.

119. Srikantan S, Tominaga K, Gorospe M. Functional interplay between RNA-binding protein HuR and microRNAs. *Curr Protein Pept Sci* 2012;**13**:372–9.

120. Bhattacharyya SN, Habermacher R, Martine U, *et al*. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* 2006;**125**:1111–24.

121. Galgano A, Forrer M, Jaskiewicz L, *et al*. Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLoS One* 2008;**3**:e3164.

122. Rybak A, Fuchs H, Hadian K, *et al*. The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2. *Nat Cell Biol* 2009;**11**:1411–20.

123. Rybak A, Fuchs H, Smirnova L, *et al*. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat Cell Biol* 2008;**10**:987–93.

124. George AD, Tenenbaum SA. MicroRNA modulation of RNA-binding protein regulatory elements. *RNA Biol* 2006;**3**:57–59.