

Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression

Shivendra Kishore, Sandra Luber and Mihaela Zavolan

Advance Access publication date | December 2010

Abstract

Eukaryotic cells express a large variety of ribonucleic acid-(RNA)-binding proteins (RBPs) with diverse affinity and specificity towards target RNAs that play a crucial role in almost every aspect of RNA metabolism. In addition, specific domains in RBPs impart catalytic activity or mediate protein–protein interactions, making RBPs versatile regulators of gene expression. In this review, we elaborate on recent experimental and computational approaches that have increased our understanding of RNA–protein interactions and their role in cellular function. We review aspects of gene expression that are modulated post-transcriptionally by RBPs, namely the stability of polymerase II-derived mRNA transcripts and their rate of translation into proteins. We further highlight the extensive regulatory networks of RBPs that implement a combinatorial control of gene expression. Taking cues from the recent development in the field, we argue that understanding spatio-temporal RNA–protein association on a transcriptome level will provide invaluable and unexpected insights into the regulatory codes that define growth, differentiation and disease.

Keywords: RNA-binding proteins; RNA-binding domains; RBP–RNA interaction; RBP regulatory networks; RBP target identification

INTRODUCTION

As they emerge from the RNA polymerase, nascent RNA transcripts are immediately covered with RNA-binding proteins (RBPs) that assist in various aspects of RNA metabolism (biogenesis, maturation, transport, cellular localization, turnover) and function [1]. The composition of ribonucleoprotein (RNP) complexes further undergoes dynamic remodeling not only as the RNA matures but also in response to changes in intra- and extra-cellular conditions. This is due to both changes in the protein composition of the different sub-cellular compartments to which an RNA molecule is targeted, as well as to post-translational modifications that

modulate the ability of RBPs to bind RNA and to interact with other proteins. It is generally thought that individual RBPs have hundreds to thousands of targets [2] and that the combinatorial binding of multiple RBPs to mRNAs contributes to the specification of cell identity or state [3]. Interestingly, increasingly many interactions of RBPs with their own mRNAs or the mRNAs encoding other RBPs are found [2, 4–6] and the function of these regulatory feedback loops is insufficiently understood. A few emerging paradigms are that RBPs coordinately regulate the localization, stability and translation of many targets [2, 7, 8] and that post-transcriptional regulation by miRNA-guided RNP complexes is a

Corresponding author. Mihaela Zavolan, Biozentrum, University of Basel and Swiss Institute of Bioinformatics, Klingelbergstrasse 50–70, CH-4056 Basel, Switzerland. E-mail: mihaela.zavolan@unibas.ch

Shivendra Kishore is an experimental postdoctoral fellow.

Sandra Luber is a computational postdoctoral researcher in the group of Mihaela Zavolan at the Biozentrum, University of Basel.

Mihaela Zavolan is an associate professor at the Biozentrum, University of Basel. She is an MD and holds a PhD degree in Computer Science. Her group combines computational and experimental techniques to study post-transcriptional gene regulation.

subtle mechanism to stabilize gene expression levels, perhaps in contrast with the binary mode (on/off) of transcriptional regulation implemented by transcription factors (TFs) [9]. Post-transcriptional miRNA-dependent regulation also appears to provide a means for rapidly responding to changes in environmental conditions [10] or for highly sub-compartment-specific protein expression [11]. Newly developed high-throughput approaches to globally identify the targets of RBPs in a specific cell type under specific conditions have greatly improved our understanding of how RBPs can systematically and coherently affect the levels of proteins in individual pathways [12]. In this review, we present the current understanding of this continuously evolving field and discuss the future impacts that the recent developments in experimental technologies may bring in our understanding of cellular function.

ARCHITECTURE OF RBPs

RBPs are defined through their ability to bind RNA targets, which they do through one or more RNA-binding domains (RBDs). In the current release (24.0) of the Pfam database [13] over 1100 human genes are annotated with a function in RNA metabolism. One may expect that the variety of functions is reflected in a variety of structures and RBDs, but this is not the case. About 40 different types of domains that are involved in RNA recognition are known to date though the numbers are constantly increasing. Some RBDs are only found in single species or in proteins with specific functions like in cap-binding or viral proteins [14], but a few RBDs implement the RNA-binding activity of a large number of proteins (for an overview about common RBDs, see [15]). The presence of multiple domains enables a more specific and higher affinity interaction of the RBP with its target. Surprisingly, only a few RBDs, among which are the RNA-recognition motif (RRM) [16], the heterogeneous nuclear (hn) RNP K homology (KH) [17], the double-stranded RBD (dsRBD) [18], Pumilio (PUF repeats) homology (PUM-H) [19], zinc finger (ZnF) [20] and Piwi/Argonaute/Zwilli (PAZ) [21] domain, are well studied (see Table 1 for details about several common RBDs).

Apart from RBDs, RBPs frequently contain domains that impart catalytic or protein-protein interaction activities. For instance, all adenosine

deaminases (ADARs) contain two or three dsRBDs coupled with a catalytic deaminase domain that is involved in editing their double-stranded nucleic acid substrates [22]. On the other hand, the dsRBDs in Drosha and Dicer are coupled with two RNase III domains that bring about endonucleolytic cleavage of primary- and precursor-miRNA, respectively [23, 24]. Several RBPs contain domains that are involved in protein-protein interaction and participate in large RNP complexes that function in maturation, localization and stability of RNA molecules. Prominent examples are splicing factors, heterogeneous nuclear ribonucleoproteins (hnRNPs) and serine/arginine-rich (SR) proteins that assist spliceosomes in splice site selection. Finally, atypical RRM domains that can bind proteins rather than RNA have been described [25, 26]. For instance, the third RRM of *Drosophila* ELAV protein has been suggested to serve as a bifunctional domain that can interact with both protein and RNA [27].

Other sequence elements within RBPs specify the sub-cellular localization that is essential for their function. For instance, the nucleolar localization of H/ACA box small nucleolar RNP (snoRNP) is attributed to the core protein Dyskerin that possesses a nuclear localization signal (NLS), as well as an acidic/lysine-rich domain (KKE/D motif) that is found in many other nucleolar proteins [28].

Finally, the specificity of interaction between RBP and RNA targets can lie in guide RNAs as opposed to RBDs. Increasingly many classes of guide RNA are being described, among these being the miRNAs [29] and endo-siRNAs [30–32] that associated with Argonaute proteins, piRNAs that associate with Piwi proteins [33–35] and small nuclear and small nucleolar RNAs (sn- and sno-RNAs) [36, 37] that associate with multiple proteins in a complex.

Thus, the combinatorial arrangement of various domains in RBPs imparts a tremendous structural and functional diversity to RBPs that enables them to modify and regulate virtually every step of target RNA metabolism and function.

ROLE OF RBPs IN mRNA STABILITY AND TRANSLATION

Recent discoveries particularly in the field of miRNAs revealed novel mechanisms by which the levels of mRNA available for translation are controlled post-transcriptionally, either by modulating

Table 1: Characteristics of several common RNA-binding motif/domains

Domain	Topology	Consensus sequence	Usual recognition mode (see also refs [14, 121, 125])	Representative proteins	Pathways involved
RRM	β $\alpha\beta$ β $\alpha\beta$ [126]	RNPI (K/R-G-F/Y-G/A-F/Y/V/L-X-F/Y) and RNP2 stretches (V/I/L-F/Y-V/I/L-X-N/L) [14]	2–8 nt. Two nucleotides bound by three conserved amino acids of β sheet; other nucleotides often bound by β strands or loops [16]	Most SR proteins and hnRNPs (A1, A2/B1, C1/C2, D, E, G, F/H), U1A, PTB, Hu family proteins, U2AF65, RBMY, FOX1, Hrp1, Prp24, FIR	Most post-transcriptional gene expression processes like RNA processing export, stability and translation
KH (hnRNP KH domain)	type I: β $\alpha\alpha\beta$ β α ; type II: $\alpha\beta$ β $\alpha\alpha\beta$ [127]	(I/L/V)-I-G-X-X-G-X-X-(I/L/V) [128]	4 nt of ssRNA bound by cleft of loops, α helices and β strands [129] more extensive recognition in some cases	hnRNP K, PCBP, KSRP, Nova, SFI, FMRP	Transcriptional regulation, splicing and translational control
RGG box	Helical β -spiral structure [130]	Repeats of R-G-G tripeptide with interspersed aromatic (FY) residues [131]	Binds to G-quartet motif [132]	hnRNPs (U, A1, G), FMRP, Nucleolin, EWS	mRNA and rRNA biogenesis
PUM-H (Pumilio-homology)	PUF repeats with three α helices each [14]	–	Binds to UGUR tetranucleotide binding pocket formed by two repeats; nucleotide of ssRNA recognized by amino acids of $\alpha 2$ of each repeat, usually 8 PUF motifs in each domain, each PUF repeat recognizes 1 nt [133]	Pum1, Pum2	mRNA stability and translation
ZF-C2H2 (ZnF C2H2 type)	β β α [134–137]	(F/Y)-X-C-X ₂₋₅ -C-X ₃ -(F/Y)-X ₅₋₉ -X ₂ -H-X ₃₋₅ -H [134, 138, 139] (α : hydrophobic residue; X: amino acid)	Binds to DNA, protein or RNA targets nucleotides of ssRNA recognized by protein side chains of α helices	TFIIIA	rRNA biogenesis
ZF-CCCH (ZnF CCCH type)	–	C-X ₈ – C-X ₅ – C-X ₃ – H [140]	Each CCCH ZnF module binds to UAUU [140] ssRNA nucleotides recognized by aromatic side chains and protein main-chain amino acids	Ts1ID, TTP	mRNA stability
DEAD Box	–	–	–	RNA helicases, eIF4A, Ded1, p68/p72	Pre-mRNA splicing, translation initiation, miRNA biogenesis
dsRBD (dsRBD)	$\alpha\beta$ β β α [141–143]	–	Nucleotides of dsRNA recognized by sugar-phosphodiester backbone of $\alpha 1$, $\alpha 2$, loops between $\beta 1/\beta 2$ and $\beta 2/\beta 3$	ADARs, TRBP, Staufen, DAI kinase, Vaccinia E3L	post-transcriptional gene silencing, development, translational regulation
PAZ	–	–	–	PIWI and dicer family of proteins	Post-transcriptional gene silencing

mRNA stability or by relocalizing mature mRNAs into translationally silent compartments such as the processing (P)-bodies. In the following section we review the expanding role of RBPs in regulating mRNA stability and translation.

Nuclear factors regulating stability, coding potential and translation rate of mRNA

The regulatory roles of nuclear factors extend beyond the processing of precursors into mature RNA forms that can be exported to the cytoplasm. Editing enzymes with nuclear localization such as ADARs that act on RNAs modulate the levels of adenosine to inosine modification in their target RNAs, which in turn may change the coding potential of exported mRNAs. A prominent example is the serotonin receptor 2C mRNA which through its five ADAR-edited sites may generate 32 mRNA isoforms encoding 24 different proteins with varied efficacy of G-protein coupling [38]. Similarly, transfer RNA (tRNA) editing by ADARs acting on tRNAs (ADATs) at or near the anticodon position modulates codon recognition during mRNA translation [39]. Non-sense mediated decay (NMD) is a mechanism by which transcripts that are incorrectly or incompletely spliced thereby carrying pre-mature termination codons are degraded by a specific cellular machinery [40]. Though this is primarily a post-transcriptional surveillance mechanism that prevents synthesis of truncated proteins that may otherwise have toxic effects in the cell, NMD has been recruited to tightly regulate bursts of protein synthesis at distinct sub-cytoplasmic locations. For instance in neurons, translationally quiescent Arc/Arg3.1 mRNAs with introns in the 3'-UTR migrate to dendrites and upon activating stimuli at the synapse, generate a single polypeptide before being degraded by NMD [41]. There is increasing evidence that alternative processing of the mRNA by 3'-end processing factors leads to differential susceptibility to post-transcriptional regulation in different cell types or phases of the cell cycle [42]. Shorter 3'-UTRs are preferentially generated in cancer cells and are associated with increased mRNA stability and translation, whereas longer 3'-UTRs are predominantly generated in non-transformed cells in which they offer more possibilities for regulation.

Cytoplasmic regulation of mRNA stability and translation

The composition of mRNPs changes as they are exported from nucleus to cytoplasm. SR proteins like ASF/SF2, that accompanied the transcript through the splicing process, along with newly assembled translation initiation factors, aid in the recruitment of the translation machinery [43, 44]. In contrast, binding of the miRNA-induced silencing complex (miRISC) brings about decapping, deadenylation and translational repression of the mRNA [45]. A few examples of miRNA-mediated translational activation have also been reported [46–48]. For example, it has been reported that miR369-3 stimulates translation of TNF- α mRNA by recruiting an Argonaute2-FXR1 complex in quiescent cells arrested at G0/G1 phase [46, 49, 50]. Interestingly, miR-369-3 repressed translation in actively dividing cells. AU-rich elements (AREs) located in 3'-UTRs serve as docking sites for various proteins that modulate mRNA stability [51]. ARE-binding proteins (ARE-BPs) of the Hu family (HuB, HuC, HuD and HuR) typically stabilize transcripts, while other ARE-BPs like AUF1 and TIA-1 primarily destabilize their target mRNAs [52]. The distinction between stabilizing and destabilizing factors is starting to become blurred because various atypical examples have been uncovered. For instance, binding of the usually stabilizing RBP HuR to an ARE located in the c-myc transcript recruits the let-7 miRNA to repress translation and promote subsequent degradation [53]. Likewise, a number of studies suggest that RBPs can assist in the recruitment of the RISC complex to its site of action either directly or indirectly [54, 55]. Recently, it has been shown that binding of Pumilio to the 3'-UTR of tumor suppressor p27 transcript induces a local change in the 3'-UTR secondary structure, exposing the miR-221/miR-222 target site which is ultimately responsible for p27 silencing in proliferating cells. In this study, no direct interaction between Argonaute and Pumilio was detected [55]. On the contrary, the RBP Dnd1 has been found to counteract miR-221-mediated translation repression of p27 by binding to an ARE in the 3'-UTR of p27 mRNA and prohibiting RISC assembly [56]. The effect of RBPs on mRNA stability can be coupled though not always coherently, with their effect on translation rate. An interesting example is the Cox-2 mRNA, whose stability is enhanced by both HuR and CUGBP2 proteins, though HuR enhances and

CUGBP2 inhibits its translation [57]. It is noteworthy that many RBPs are involved in multiple, quite distinct processes. For example, as mentioned earlier, the splicing factor ASF/SF2 also assists in mRNA export to the cytoplasm where it promotes mRNA translation [43, 44]. Another splicing factor, TIA-1, functions in the assembly of stress granules (SGs) in the cytoplasm. SGs are formed when translation initiation is impaired due to external stress or lack of translation initiation factors. The repressed mRNAs are bound by TIA-1 through its RBD and the self-aggregates of the protein through QN-rich prion-like domains promote SG assembly [58]. Likewise, cold inducible RBP (CIRP) migrates from nucleus to cytoplasmic SGs upon oxidative stress and acts as a translational repressor, independent of TIA-1 mediated SG formation [59].

REGULATORY AND AUTOREGULATORY NETWORKS OF RBPs

Since RBPs play crucial roles in the metabolism of a multitude of RNAs, it is important to regulate precisely their quantity and localization within cellular compartments. Hence, diverse mechanisms have evolved to ensure precise turnover and localization of RBPs. Furthermore, because RBPs orchestrate specific cellular responses under conditions such as stress, the levels of RBPs have to be rapidly adjusted under such conditions. Mechanisms for rapidly turning on or off the expression or activity of RBPs have also been evolved, starting at the level of transcription. Some RBPs such as Nova and SNRPN are expressed in a tissue-specific manner (in this case in neurons). Others are regulated at the level of splicing. The *Staufen* gene, for e.g. has multiple isoforms which have been shown to exhibit different RNA-binding activity in rat [60]. The phosphorylation state of a splicing factor is a major determinant that regulates its shuttling from nucleus to cytoplasm [61]. SR proteins contain one or more stretches of serine/arginine residues. Phosphorylation at serine residues influences the localization of SR proteins within different sub-cellular compartments and releases them from speckles where they are relatively inactive, into the nucleus where they can actively bind to pre-mRNAs [62, 63].

Complex-regulatory networks in which RBPs regulate their own and each other's expression have been already uncovered. For example, RNA

editing enzyme ADAR2 can edit its own pre-mRNA to activate a cryptic 3'-acceptor splice site, which changes the predicted reading frame of the mature ADAR2 transcript and generates a non-functional protein variant [64]. Similarly, the splicing factor Transformer2- β 1 (Tra2- β 1) autoregulates its expression through alternative splicing. High levels of Tra2- β 1 promote binding of Tra2- β 1 to its own pre-mRNA leading to the generation of a functionally inactive splice variant [4]. Quite generally, the genes encoding SR proteins contain highly conserved so-called 'poison exons' that harbor premature stop codons and are used to tightly regulate their protein level through a feedback mechanism that involves NMD [65]. The alternatively spliced introns in the 3'-UTR of the SR protein-encoding mRNAs serve a similar purpose [65]. In the case of the splicing factors SC-35 and Tra-2beta1 mRNA destabilization is induced in an autoregulatory manner [4, 5]. Among the many potential targets of miRNAs, splicing factors can of course trigger broad downstream effects. For example, upregulation of miR-23a/b in the post-natal development of cardiac tissue leads to tissue-specific repression of the CUGBP1 and CUGBP2 splicing factors, which in turn shifts the splicing of a conserved set of genes into an adult pattern [66]. Similarly, the neuron-specific miR-124 plays a crucial role in setting up the neuronal-splicing program by repressing the expression of polypyrimidine tract-binding protein (PTB), a global repressor of alternative splicing in non-neuronal cells. PTB and its neuron-specific paralog nPTB have largely overlapping target sets, though PTB is a stronger repressor of alternatively spliced exons compared to nPTB. More importantly, PTB promotes an exon exclusion event in both its own transcript as well as that of nPTB [6]. The exon exclusion results in enhanced transcript degradation via non-sense-mediated RNA decay [67, 68]. Repression of PTB by miR-124 leads to the expression of functionally active nPTB which further defines the pattern of splicing in the neuronal tissue [69]. On the other hand, miR-133 has been shown to directly downregulate nPTB during muscle development by binding to its 3'-UTR [70]. MicroRNAs regulate many other RBPs with central cellular functions. A very interesting example is the auto-regulatory loop involving the let-7 miRNA that down-regulates the expression of the endoribonuclease Dicer, which is involved in pre-miRNA processing. Since let-7 is itself processed

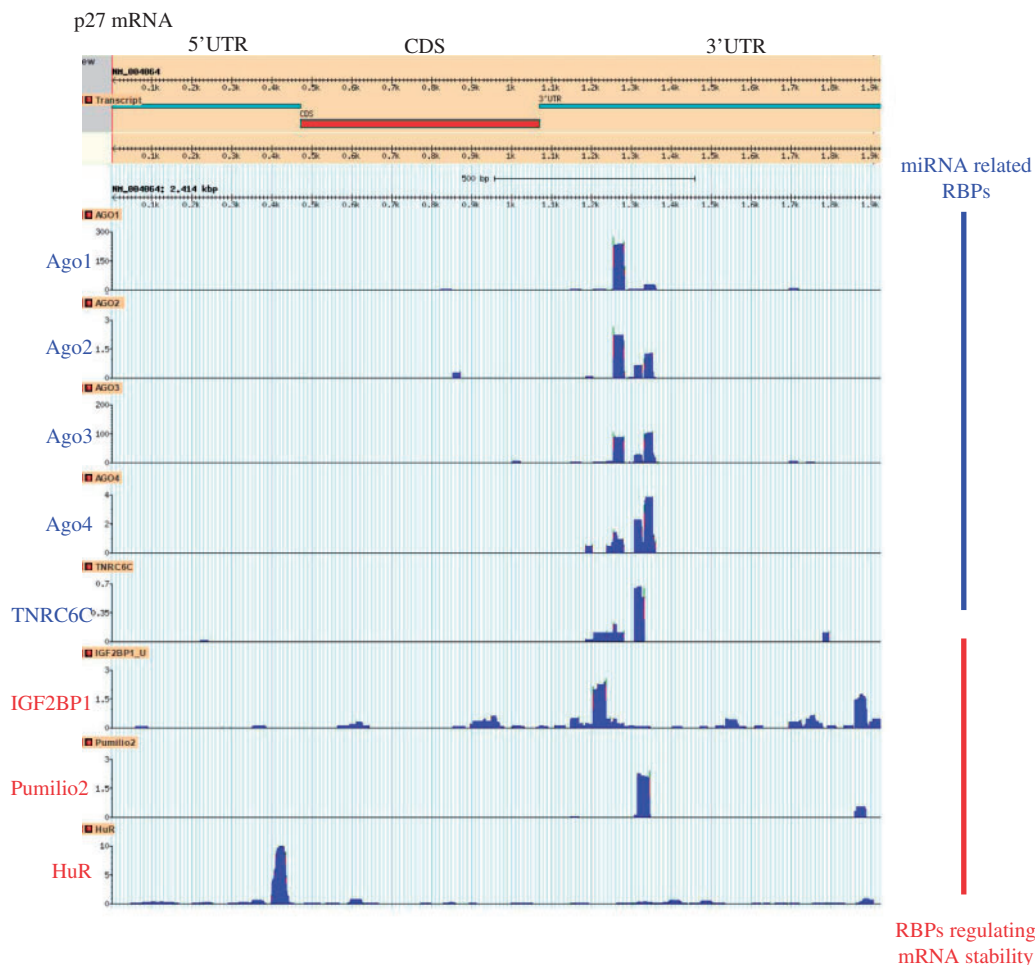


Figure 1: Footprints of several RBPs (CDS: coding sequence) on the p27 (Kip1) transcript are shown. Information on RBP-binding sites were obtained from 4-thiouridine CLIP performed on individual RBPs (see reference [96] and unpublished data from M.Z. lab). Overlapping RBP clusters suggest cross-talks between various RBPs that ultimately determine the fate of the p27 transcript.

by Dicer, it is likely that this negative feedback loop is involved in setting the miRNA expression at appropriate levels within cells [71].

A few recent studies suggest a strong interplay between the ARE-BPs and miRNPs in modulating gene expression. Computational studies have shown that effective miRNA target sites reside within A/U-rich regions inside 3'-UTRs [72, 73]. Several proteins have been described to bind AREs and regulate mRNA stability, and one therefore expects that ARE-BPs modulate the interaction of miRNA with their targets. An example that we already mentioned is that of the ARE-BP HuR that regulates the expression of *c-myc* by recruiting the let-7 miRNA [53]. In contrast, binding of HuR to the ARE of cationic amino acid transporter (CAT-1) mRNA upon stress relieves miR-122-mediated translational

repression of CAT-1 mRNA in liver tissue [74]. Similarly, the Deadend ARE-BP (Dnd1) has been shown to inhibit the access of a miRNA to its target site [56]. Interestingly, results from our lab show that binding sites for several RBPs on their target RNAs occur in close vicinity, suggesting extensive crosstalks between RBP complexes that probably result in precise regulation of gene expression levels in a variety of conditions. Binding sites that were experimentally identified in an mRNA transcript of the cell cycle related p27 (Kip1) gene for a number of proteins are shown in Figure 1. Similar to splicing factors, RBPs that regulate mRNA stability are also frequently found to regulate their own and each other's expression. For example, the AU-rich element-binding protein (ARE-BP) Tristetrapoline (TTP) negatively regulates its own

expression by binding to an ARE in the 3'-UTR of its own mRNA [75]. Similarly, in *Drosophila*, another neuronal ARE-BP Elav has been suggested to regulate its own mRNA by binding to ARE elements in its 3'-UTR [76]. Cross-regulatory RNP interactions have also been identified. For example, HuR stabilizes TIA-1 mRNA and upregulates TIA-1 expression while TIAR suppresses TIA-1 expression [77].

Experiments in which RBP-bound mRNAs were identified after RBP immunoprecipitation (RIP-Chip) indicated that RBPs typically bind a large number of mRNAs. Moreover, they revealed that the targets of an RBP are not a random subset of the transcripts that are expressed in the cell, but rather functionally related transcripts, this being the essence of the 'regulon theory' in eukaryotes (similar to operon model in prokaryotes) [78]. A recent example consistent with this model is that of the RBP Musashi-1, whose knockdown results in the upregulation of a significant number of apoptosis-related genes, encoding components of SGs [79].

EXPERIMENTAL METHODS TO IDENTIFY TARGETS OF RBPs

Given the multitude of targets and functions of RBPs, various methods have been developed over the years to investigate their binding specificity. The electrophoretic mobility shift assay is a technique that is used currently to detect interactions between proteins and nucleic acids *in vitro* [80, 81]. It is based on the principle that an oligonucleotide migrates slower in an electric field when bound by the protein and it is of course suited only to document that a given RBP binds a given oligonucleotide sequence. Another *in vitro* procedure called Systematic Evolution of Ligands by Exponential Enrichment has been widely used to identify aptamers (ssDNA or RNA) that specifically bind to a protein of interest starting from a large library of randomly generated oligonucleotides [82]. Through multiple rounds of selection, purification under stringent conditions and amplification, high-affinity interactors can be selectively enriched. It has been found however that the high-affinity targets obtained *in vitro* are frequently not the most specific. The recently developed RNAcompete method uses a single binding reaction to determine *in vitro* the affinities an RBP for a complete set of *k*-mers (oligonucleotides) that are

presented in structured and unstructured RNA contexts. This method has been applied to various RBPs and the results have been shown to be consistent with previous *in vivo* studies [83]. One drawback of these *in vitro* studies is that they employ recombinant proteins which limits their use when the proteins of interest are unstable, difficult to express or purify to homogeneity.

Genome-wide microarray analyses upon RBP knockdown or overexpression may identify *in vivo* targets of RBPs. However, such analyses profile global steady state levels of mRNAs which poorly correlate with the cellular protein levels [84], and it is therefore unclear how informative such analyses are. RNA immunoprecipitation followed by microarray-based identification of protein-bound RNAs (RIP-Chip) [85–89] has been widely used to identify *in vivo* targets of RBPs. Although this method has the advantage of identifying interactions occurring *in vivo*, the RNA–protein complex cannot be washed stringently, thereby having the potential of isolating many false positive targets. Also, because with this approach one isolates large RNA molecules, uncovering the sequence or structure specificity of the protein from RIP-Chip data requires complex computational analyses. To avoid artifacts arising from interactions that occur after cell lysis, methods that rely on cross-linking RNA–protein complexes *in vivo* by ultraviolet (UV) irradiation (at 254 nm) have been proposed [90]. Although such methods have been in use for decades, advances in technology such as high-throughput sequencing and higher computational power have dramatically increased the power of the procedure.

The cross-linking and immunoprecipitation assay (CLIP) exploits covalent protein–nucleic acid cross-linking to stringently purify complexes of RNAs with a specific RBP. Moreover, nuclease digestion of RNA sequence stretches that are not protected by the RBP enables the isolation of binding sites at very high resolution [91, 92]. The protected RNA fragments are then ligated to adapters at both 5'- and 3'-end, converted into cDNA and PCR-amplified. Initial CLIP experiments employed sequencing of cloned PCR-amplified fragments in a suitable vector. Based on the location of YCAAY sequence motifs with respect to intron and exon boundaries, this method generated a genome-wide RNA map predicting the pattern of NOVA-dependent splicing in the brain [91]. A more

recent version of the protocol called HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking and immunoprecipitation) has provided a more comprehensive picture of the function of this nervous system-specific RBP, revealing the role of NOVA in alternative polyadenylation of transcripts in the brain [93]. HITS-CLIP has further been used to identify functional miRNA–mRNA interaction sites [94] and, in combination with custom microarray-based analyses, to identify specific pathways targeted by RBPs [95]. Though highly successful, this method has a few limitations. UV irradiation at shorter wavelengths (254 nm) may introduce nucleic acids breaks as well as modifications in the nucleotides that may hamper accurate mapping of the cDNAs to the corresponding genomic loci. Given that the binding sites of RBPs are usually very short, <10 nt [14], the nuclease digestion has to be very precisely monitored and optimized. Overdigestion will lead to RNA fragments that are too short to be unambiguously mapped to the genome, while underdigestion to fragments that are too long to be sequenced in their entirety yielding reads that do not contain interaction sites.

Recently, another promising variant of this method called photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) has been proposed. PAR-CLIP uses photoreactive analogs of ribonucleosides to crosslink RNA–protein complexes *in vivo* [96]. When 4-thiouridine is used as a uridine analog, crosslinked 4-thiouridines are subsequently reverse transcribed and PCR amplified as cytosines. Loci to which many reads with thymidine to cytosine mutations are mapped, represent in all likelihood binding sites to which the protein has been crosslinked. The crosslink–diagnostic mutations enable one to pinpoint RBP-binding sites at site resolution. An even more specific mutational pattern has been observed when PAR-CLIP was applied to Argonaute proteins: the location of thymidine-to-cytosine mutation was very frequently found immediately upstream of the region of complementarity between the mRNA and the 5′-end (known as the seed region) of an abundantly expressed miRNA [96]. Using this information, one can identify with accuracy the miRNA that is involved in regulating a particular transcript at a specific site. Another advantage of PAR-CLIP is that the photoreactive analogs can be crosslinked at higher wavelengths (365 nm) which

are less damaging to the nucleic acids. While use of 4-thiouridine is suitable for RBPs that have uridines in their binding motif or very close to it, it is not suitable for RBPs that bind to uridine depleted stretches. For such cases, 6-thioguanosine has been proposed as an alternative [96]. Preliminary studies suggest that 6-thioguanosine crosslinking also yields diagnostic guanosine to adenosine mutations that could be used to accurately pin-point the location of binding sites. However, this change appears to occur less frequently than the thymidine-to-cytosine transition which is observed when 4-thiouridine is used. In addition, earlier reports indicated that 6-thioguanosine is toxic to cells and prevents RNA and protein synthesis beyond certain concentrations [97], hence the incorporation frequency of 6-thioguanosine in the RNA cannot be as high as that of 4-thiouridine. Also, the crosslinking efficiency of 6-thioguanosine is several folds lower than that of 4-thiouridine [96] making its use more limited compared to that of 4-thiouridine. Finally, incorporation of photoreactive analogs in animal tissues remains a daunting task. Therefore, this technique has so far been used in analyses of cell lines.

In yet another promising development in this field, a CLIP variant known as iCLIP exploits the propensity of reverse transcriptase to terminate at a nucleotide remains modified after the protein that was crosslinked to it was digested away [98]. Sequence reads obtained from iCLIP are thus expected to start exactly at the site of crosslink enabling identification of RBP-binding sites at nucleotide resolution. It has been argued that with previous methods, many bona fide binding sites would have caused the reverse transcriptase to stop before reaching the 5′-adaptor, yielding aborted cDNAs that would be selectively lost during the PCR cycles due to the lack of the annealing site for the sequencing primer. Further studies will be necessary to establish to what extent this occurs, because at least in PAR-CLIP, large numbers of sequence reads with crosslink diagnostic mutations inside the reads have been obtained.

Taken together, all the CLIP variants have been employed with demonstrable success in the transcriptome-wide identification of binding sites for numerous RBPs. The improvements in the technique are enabling increasingly many groups to apply it to a variety of proteins [91, 93, 96, 98–100], and we expect that in the not-too-distant future, catalogs of RBP-binding sites will be available, similar to

resources that have been constructed for TFs. However, additional work will be needed to establish protocols that would allow one to obtain quantitative data on RBP–RNA interactions *in vivo*. It is clear that the choices made in the many steps of the CLIP protocol can lead to the identification of specific subsets of binding sites. For example, in order to identify the position of the crosslink with methods other than iCLIP, the sequence read has to cover the RBP-binding site. This is only ensured when the length of the fragments that are subjected to deep sequencing is relatively short, and thus the nuclease digestion has to be carried out to the extent that little beyond the RBP-protected fragment is left intact. The complete digestion products will likely be too short to be unambiguously mapped, and almost complete digestion is very difficult to control. As alternatives, one may use incomplete digestion followed by computational analyses to identify the short-binding sites presumably located at some distance from the sequenced reads, or digestion by a nuclease that does not cleave between all dinucleotide pairs. In fact, of the nucleases that have been employed in CLIP, only RNase I does not have a preferred nucleotide for cleavage, while the others have specific preferences. RNase T1 cleaves 3' of guanosine [101] and RNase A cleaves only after cytosine and uridine [102]. Another commonly used nuclease, micrococcal nuclease cleaves preferentially 5' of an adenosine or thymidine [103]. Thus, the choice of nuclease may influence the subset of sites that is isolated through CLIP, enriching for those that do not contain optimal nuclease cleavage sites in the immediate vicinity of the RBP binding site. It is also known that 254 nm UV irradiation preferentially crosslinks pyrimidines [104] with crosslinks normally not occurring between base paired regions [105]. Thus, one may expect that binding sites that are embedded in pyrimidine-rich regions will be preferentially enriched in the 254 nm CLIP data. Overall, a large number of techniques to decipher RNA–protein interactions is available today, each having its own set of advantages and limitations, which need to be weighted in relationship to the problem at hand.

Rapid advancements in the experimental methods have facilitated identification of RBP-binding sites at a transcriptome-wide level and lead to fascinating insights into the intricate regulatory networks involving RBPs. Although additional work will be needed to increase the accuracy of these methods, it

should soon be possible to obtain comprehensive maps of RNA–protein interactome in a variety of cell types and in a variety of conditions.

COMPUTATIONAL INFERENCE OF RBP SPECIFICITY

Availability of large-scale data sets of binding sites for individual RBPs opens the door for detailed computational studies aiming to uncover the factors that govern the specificity of RBP–RNA interactions. Numerous methods are available for inferring binding motifs in nucleic acid sequences. Though most of the methods have been in fact developed in the context of inferring the specificity and binding sites of TFs these methods can readily be applied to RBPs. A relatively recent description of the general framework of sequence motif finding can be found in ref. [106]. Depending on the available input data, the problem of inferring sequence motifs that are bound by RBPs comes in various flavors, a typical setting being the following. Through a high-throughput experimental method (for e.g. RIP-Chip) one has isolated a large number of putative targets of the RBP of interest and the question is to find the location of the RBP-binding sites in these targets as well as the sequence motif that is recognized by the RBP. In this case, one can employ one of the available motif finding tools [107–110] to simultaneously infer the binding specificity of the protein and the location of the binding sites in the targets. The Phylogibbs algorithm [108, 109] can take into account not only individual sequences, but also multiple alignments of orthologous regions from a set of species, in order to identify sequence motifs that are both over-represented in the input sequences, and evolutionarily conserved. Motif finders have been successfully used to identify binding motifs of RBPs. For example, Phylogibbs was used to recover the known binding motifs of the Pumilio and Quaking proteins and to discover the specificity of Insulin growth factor 2-binding proteins from PAR-CLIP data [96]. It has been argued, however, that a further increase in specificity of binding-site identification can be obtained by taking into account clustering of sequence elements [96, 111]. A rationale could be the multi-domain structure of RBPs, each domain binding a specific sequence element in the RNA. Methods that would take advantage of this property remain to be developed, though for the determination of binding motifs in which

deletions and insertions are allowed in the presumed binding site, one can employ for example the GLAM2 [112] motif finder.

A further complication for the prediction of RBP-binding sites in contrast to binding sites of TFs is that the RNAs are not double helical, but form complex structures. Many RBPs bind only to RNAs that are in a single-stranded conformation while others recognize specific structural elements. For these reasons, one expects that methods that consider both the sequence and structure of the putative targets should have increased accuracy. A simple way of taking into account RNA secondary structure is to extract from sequence-based predictions of RBP-binding sites those that are also predicted to be located in relatively accessible regions. Tafer *et al.* [113] and Hausser *et al.* [73] studied the effect of structural accessibility of target sites in the context of miRNA-binding sites and Li *et al.* [114] in the context of RBPs. Indeed, they found that functional miRNA- and RBP-binding sites are located in regions that form less stable secondary structures. A more sophisticated approach is to consider the secondary structure in the prediction of RBP-binding sites. For instance, in the MEMERIS program [115] the estimated probability of a given region assuming a single-stranded conformation features as a prior probability that a sequence-specified binding site is located in that region of the sequence.

Another set of tools attempts to identify structural elements that are important for the interaction with the RBP. For instance, Lopez de Silanes *et al.* [116] used a stochastic context-free grammar model and inferred that the TIA-1 protein appears to recognize U-rich elements in the context of bent stem-loop structures. Rabani *et al.* [117] developed the RNAPROMO tool to identify over-represented structural motifs among the predicted secondary structures of a set of input RNA sequences assumed to have a motif in common. To what extent specific secondary structure motifs are important for the recognition of binding sites by RBPs remains to be determined. Structural motifs that are required for the recognition by specific proteins are known [118], though Ray *et al.* [83] found that most of the nine proteins for which binding affinities to a large number of oligonucleotides were measured do not appear to require that their target sites are presented in a particular structural configuration. As we learn more about the dynamic associations between RBPs and RNAs and as the accuracy of

secondary-structure prediction algorithms improve, predictive models of the accessibility and occupancy of individual binding sites in the RNA in specific cellular conditions will become more common. As a few examples, we mention the attempts to predict the effect of RNA secondary structures on RNA-ligand binding [119] or the impact of protein binding on the secondary structure of ssRNAs [120]. Finally, computational prediction of RBP-RNA interaction has also been approached from a computational structural biology point of view. Zheng *et al.* [121] designed a distance-dependent statistical potential function to predict RNA sequences that would be recognized by various RRM and KH homology domains, as well as to predict the affinities of interaction. It would be extremely interesting to study the relationship between these predictions, the affinities of RBPs for oligonucleotides that were estimated by RNAcompete [83] and *in vivo* binding data obtained through CLIP. The challenge now lies in integrating these diverse datasets to derive regulatory codes that can predict tissue-specific regulation of gene expression. Barash *et al.* [122] have illustrated the importance of combining multiple features like the occurrence of motifs for various splicing factors and other regulatory sequences, to unravel the splicing code which ultimately explains tissue specific alternative splicing events [122]. Such studies are within reach and would greatly improve our understanding of the combinatorial regulation by RBPs.

PERSPECTIVE

In spite of the fact that RBPs have been extensively studied for several decades, it is clear that there is much left to learn about their roles and interactions in the regulation of gene expression. Several novel RBDs have been recently discovered that remain to be characterized (see for e.g. [123, 124]). Technological advances enable characterization of RBP-binding sites at nucleotide resolution opening numerous avenues of investigation. Novel functions are being identified for even well-studied RBPs. Dynamical changes in the sets of targets of individual RBPs in response to the cellular state and external stimuli are beginning to be characterized and dynamic interplays of multiple RBPs on the same target are being identified. The aim in the long run is to construct a detailed, temporally resolved map of RBP-RNA interactions that we expected will reveal novel pathogenic mechanisms and will aid in the

development of more specifically tailored and effective therapeutic agents.

Key Points

- RBPs play a crucial role in post-transcriptional control of gene expression.
- Combinatorial arrangement of various RBDs along with other functional domains impart structural and functional diversity to RBPs that enables them to modify and regulate virtually every step of target RNA metabolism and function.
- The expression level and sub-cellular localization of RBPs is precisely controlled by complex regulatory and autoregulatory networks.
- Individual RBPs have hundreds of targets and precise control of gene expression occurs due to specific spatio-temporal association of multiple RBPs with their target RNA.
- Several computational and experimental tools that are now available to study global protein–RNA interactions will provide us with invaluable insights into principles of gene regulation.

Acknowledgements

We thank Lukasz Jaskiewicz for critical reading of the article and for useful suggestions. S.K. is supported by an EMBO LTF-fellowship and S.L. by a Swiss National Foundation grant (number CRSII3_127454/1) to M.Z.

References

1. Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* 2002;**3**:195–205.
2. Keene JD, Tenenbaum SA. Eukaryotic mRNPs may represent posttranscriptional operons. *Mol Cell* 2002;**9**:1161–7.
3. Smith CW, Valcarcel J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* 2000;**25**:381–8.
4. Stoilov P, Daoud R, Nayler O, *et al.* Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. *Hum Mol Genet* 2004;**13**:509–24.
5. Sureau A, Gattoni R, Dooghe Y, *et al.* SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J* 2001;**20**:1785–96.
6. Boutz PL, Stoilov P, Li Q, *et al.* A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev* 2007;**21**:1636–52.
7. Selbach M, Schwanhauser B, Thierfelder N, *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* 2008;**455**:58–63.
8. Baek D, Villen J, Shin C, *et al.* The impact of microRNAs on protein output. *Nature* 2008;**455**:64–71.
9. Cai L, Dalal CK, Elowitz MB. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* 2008;**455**:485–90.
10. Krol J, Busskamp V, Markiewicz I, *et al.* Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. *Cell* 2010;**141**:618–31.
11. Schrott G. microRNAs at the synapse. *Nat Rev Neurosci* 2009;**10**:842–9.
12. Mukherjee N, Lager PJ, Friedersdorf MB, *et al.* Coordinated posttranscriptional mRNA population dynamics during T-cell activation. *Mol Syst Biol* 2009;**5**:288.
13. Finn RD, Mistry J, Tate J, *et al.* The Pfam protein families database. *Nucleic Acids Res* 2010;**38**:D211–22.
14. Auweter SD, Oberstrass FC, Allain FH. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res* 2006;**34**:4943–59.
15. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* 2007;**8**:479–90.
16. Clery A, Blatter M, Allain FH. RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* 2008;**18**:290–8.
17. Valverde R, Edwards L, Regan L. Structure and function of KH domains. *FEBS J* 2008;**275**:2712–26.
18. Doyle M, Jantsch MF. New and old roles of the double-stranded RNA-binding domain. *J Struct Biol* 2002;**140**:147–53.
19. Jenkins HT, Baker-Wilding R, Edwards TA. Structure and RNA binding of the mouse Pumilio-2 Puf domain. *J Struct Biol* 2009;**167**:271–6.
20. Klug A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* 2010;**43**:1–21.
21. Wang Y, Sheng G, Juraneck S, *et al.* Structure of the guide-strand-containing argonaute silencing complex. *Nature* 2008;**456**:209–13.
22. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* 2010;**79**:321–49.
23. Han J, Lee Y, Yeom KH, *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 2006;**125**:887–901.
24. Bernstein E, Caudy AA, Hammond SM, *et al.* Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 2001;**409**:363–6.
25. Kielkopf CL, Rodionova NA, Green MR, *et al.* A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* 2001;**106**:595–605.
26. Selenko P, Gregorovic G, Sprangers R, *et al.* Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell* 2003;**11**:965–76.
27. Toba G, White K. The third RNA recognition motif of Drosophila ELAV protein has a role in multimerization. *Nucleic Acids Res* 2008;**36**:1390–9.
28. Heiss NS, Girod A, Salowsky R, *et al.* Dyskerin localizes to the nucleolus and its mislocalization is unlikely to play a role in the pathogenesis of dyskeratosis congenita. *Hum Mol Genet* 1999;**8**:2515–24.
29. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;**116**:281–97.
30. Kawamura Y, Saito K, Kin T, *et al.* Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 2008;**453**:793–7.

31. Ghildiyal M, Seitz H, Horwich MD, *et al.* Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 2008;**320**:1077–81.
32. Tam OH, Aravin AA, Stein P, *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 2008;**453**:534–8.
33. Lau NC, Seto AG, Kim J, *et al.* Characterization of the piRNA complex from rat testes. *Science* 2006;**313**:363–7.
34. Aravin A, Gaidatzis D, Pfeffer S, *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 2006;**442**:203–7.
35. Girard A, Sachidanandam R, Hannon GJ, *et al.* A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 2006;**442**:199–202.
36. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell* 2009;**136**:701–18.
37. Decatur WA, Fournier MJ. RNA-guided nucleotide modification of ribosomal and other RNAs. *J Biol Chem* 2003;**278**:695–8.
38. Werry TD, Loiacono R, Sexton PM, *et al.* RNA editing of the serotonin 5HT_{2C} receptor and its effects on cell signalling, pharmacology and brain function. *Pharmacol Ther* 2008;**119**:7–23.
39. Grosjean H, Auxilien S, Constantinesco F, *et al.* Enzymatic conversion of adenosine to inosine and to N1-methylinosine in transfer RNAs: a review. *Biochimie* 1996;**78**:488–501.
40. Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 2007;**76**:51–74.
41. Giorgi C, Yeo GW, Stone ME, *et al.* The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. *Cell* 2007;**130**:179–91.
42. Sandberg R, Neilson JR, Sarma A, *et al.* Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 2008;**320**:1643–7.
43. Michlewski G, Sanford JR, Caceres JF. The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Mol Cell* 2008;**30**:179–89.
44. Sanford JR, Gray NK, Beckmann K, *et al.* A novel role for shuttling SR proteins in mRNA translation. *Genes Dev* 2004;**18**:755–68.
45. Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* 2010;**79**:351–79.
46. Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: microRNAs can up-regulate translation. *Science* 2007;**318**:1931–4.
47. Orom UA, Nielsen FC, Lund AH. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* 2008;**30**:460–71.
48. Henke JI, Goergen D, Zheng J, *et al.* microRNA-122 stimulates translation of hepatitis C virus RNA. *Embo J* 2008;**27**:3300–10.
49. Vasudevan S, Steitz JA. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* 2007;**128**:1105–18.
50. Vasudevan S, Tong Y, Steitz JA. Cell-cycle control of microRNA-mediated translation regulation. *Cell Cycle* 2008;**7**:1545–9.
51. Chen CY, Shyu AB. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci* 1995;**20**:465–70.
52. Barreau C, Paillard L, Osborne HB. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* 2005;**33**:7138–50.
53. Kim HH, Kuwano Y, Srikantan S, *et al.* HuR recruits let-7/RISC to repress c-Myc expression. *Genes Dev* 2009;**23**:1743–8.
54. Jing Q, Huang S, Guth S, *et al.* Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell* 2005;**120**:623–34.
55. 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.
56. Kedde M, Strasser MJ, Boldajipour B, *et al.* RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* 2007;**131**:1273–86.
57. Sureban SM, Murmu N, Rodriguez P, *et al.* Functional antagonism between RNA binding proteins HuR and CUGBP2 determines the fate of COX-2 mRNA translation. *Gastroenterology* 2007;**132**:1055–65.
58. Buchan JR, Parker R. Eukaryotic stress granules: the ins and outs of translation. *Mol Cell* 2009;**36**:932–41.
59. De Leeuw F, Zhang T, Wauquier C, *et al.* The cold-inducible RNA-binding protein migrates from the nucleus to cytoplasmic stress granules by a methylation-dependent mechanism and acts as a translational repressor. *Exp Cell Res* 2007;**313**:4130–44.
60. Monshausen M, Putz U, Rehbein M, *et al.* Two rat brain staufen isoforms differentially bind RNA. *J Neurochem* 2001;**76**:155–65.
61. Caceres JF, Sreaton GR, Krainer AR. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev* 1998;**12**:55–66.
62. Misteli T, Caceres JF, Spector DL. The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 1997;**387**:523–7.
63. Misteli T, Caceres JF, Clement JQ, *et al.* Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J Cell Biol* 1998;**143**:297–307.
64. Rueter SM, Dawson TR, Emeson RB. Regulation of alternative splicing by RNA editing. *Nature* 1999;**399**:75–80.
65. Lareau LF, Inada M, Green RE, *et al.* Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 2007;**446**:926–9.
66. Kalsotra A, Wang K, Li PF, *et al.* MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. *Genes Dev* 2010;**24**:653–8.
67. Wollerton MC, Gooding C, Wagner EJ, *et al.* Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol Cell* 2004;**13**:91–100.
68. Ashiya M, Grabowski PJ. A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. *RNA* 1997;**3**:996–1015.
69. Makeyev EV, Zhang J, Carrasco MA, *et al.* The MicroRNA miR-124 promotes neuronal differentiation by triggering

- brain-specific alternative pre-mRNA splicing. *Mol Cell* 2007;**27**:435–48.
70. Boutz PL, Chawla G, Stoilov P, *et al.* MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes Dev* 2007;**21**:71–84.
71. Tokumaru S, Suzuki M, Yamada H, *et al.* let-7 regulates Dicer expression and constitutes a negative feedback loop. *Carcinogenesis* 2008;**29**:2073–7.
72. Grimson A, Farh KK, Johnston WK, *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007;**27**:91–105.
73. Hausser J, Landthaler M, Jaskiewicz L, *et al.* Relative contribution of sequence and structure features to the mRNA binding of Argonaute/EIF2C-miRNA complexes and the degradation of miRNA targets. *Genome Res* 2009;**19**:2009–20.
74. 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.
75. Tchen CR, Brook M, Saklatvala J, *et al.* The stability of tristetraprolin mRNA is regulated by mitogen-activated protein kinase p38 and by tristetraprolin itself. *J Biol Chem* 2004;**279**:32393–400.
76. Samson ML. Evidence for 3' untranslated region-dependent autoregulation of the *Drosophila* gene encoding the neuronal nuclear RNA-binding protein ELAV. *Genetics* 1998;**150**:723–33.
77. 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.
78. Keene JD. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* 2007;**8**:533–43.
79. Nikpour P, Emadi-Baygi M, Steinhoff C, *et al.* The RNA binding protein Musashi1 regulates apoptosis, gene expression and stress granule formation in urothelial carcinoma cells. *J Cell Mol Med* 2010, May 14 [Epub ahead of print].
80. Garner MM, Revzin A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res* 1981;**9**:3047–60.
81. Fried M, Crothers DM. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res* 1981;**9**:6505–25.
82. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 1990;**249**:505–10.
83. 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.
84. Gygi SP, Rochon Y, Franza BR, *et al.* Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 1999;**19**:1720–30.
85. Lopez de Silanes I, Zhan M, Lal A, *et al.* Identification of a target RNA motif for RNA-binding protein HuR. *Proc Natl Acad Sci USA* 2004;**101**:2987–92.
86. 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.
87. Baroni TE, Chittur SV, George AD, *et al.* Advances in RIP-chip analysis: RNA-binding protein immunoprecipitation-microarray profiling. *Methods Mol Biol* 2008;**419**:93–108.
88. Calaluce R, Gubin MM, Davis JW, *et al.* The RNA binding protein HuR differentially regulates unique subsets of mRNAs in estrogen receptor negative and estrogen receptor positive breast cancer. *BMC Cancer* 2010;**10**:126.
89. Mazan-Mamczarz K, Kuwano Y, Zhan M, *et al.* Identification of a signature motif in target mRNAs of RNA-binding protein AUF1. *Nucleic Acids Res* 2009;**37**:204–14.
90. Wagenmakers AJ, Reinders RJ, van Venrooij WJ. Cross-linking of mRNA to proteins by irradiation of intact cells with ultraviolet light. *Eur J Biochem* 1980;**112**:323–30.
91. Ule J, Jensen KB, Ruggiu M, *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* 2003;**302**:1212–5.
92. Ule J, Jensen K, Mele A, *et al.* CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* 2005;**37**:376–86.
93. Licatalosi DD, Mele A, Fak JJ, *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 2008;**456**:464–9.
94. Chi SW, Zang JB, Mele A, *et al.* Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 2009;**460**:479–86.
95. Ule J, Ule A, Spencer J, *et al.* Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* 2005;**37**:844–52.
96. Hafner M, Landthaler M, Burger L, *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010;**141**:129–41.
97. Melvin WT, Milne HB, Slater AA, *et al.* Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. *Eur J Biochem* 1978;**92**:373–9.
98. 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.
99. Sanford JR, Coutinho P, Hackett JA, *et al.* Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS One* 2008;**3**:e3369.
100. Bohnsack MT, Martin R, Granneman S, *et al.* Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol Cell* 2009;**36**:583–92.
101. Donis-Keller H, Maxam AM, Gilbert W. Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res* 1977;**4**:2527–38.
102. Raines RT. Ribonuclease A. *Chem Rev* 1998;**98**:1045–66.
103. Cuatrecasas P, Fuchs S, Anfinsen CB. Catalytic properties and specificity of the extracellular nuclease of *Staphylococcus aureus*. *J Biol Chem* 1967;**242**:1541–7.
104. Noah JW, Shapkina T, Wollenzien P. UV-induced cross-links in the 16S rRNAs of *Escherichia coli*, *Bacillus subtilis* and *Thermus aquaticus* and their implications for ribosome structure and photochemistry. *Nucleic Acids Res* 2000;**28**:3785–92.

105. Wilms C, Noah JW, Zhong D, *et al.* Exact determination of UV-induced crosslinks in 16S ribosomal RNA in 30S ribosomal subunits. *RNA* 1997;**3**:602–12.
106. van Nimwegen E. Finding regulatory elements and regulatory motifs: a general probabilistic framework. *BMC Bioinformatics* 2007;**8**:S4.
107. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994;**2**:28–36.
108. Siddharthan R, Siggia ED, van Nimwegen E. PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny. *PLoS Comput Biol* 2005;**1**:e67.
109. Siddharthan R, van Nimwegen E. Detecting regulatory sites using PhyloGibbs. *Methods Mol Biol* 2007;**395**: 381–402.
110. Bailey T, Elkan C. Unsupervised learning of multiple motifs in biopolymers using expectation maximization. *Mach Learn* 1995;**21**:51–80.
111. Ule J, Stefani G, Mele A, *et al.* An RNA map predicting Nova-dependent splicing regulation. *Nature* 2006;**444**: 580–6.
112. Frith MC, Saunders NF, Kobe B, *et al.* Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol* 2008;**4**:e1000071.
113. Tafer H, Ameres SL, Obernosterer G, *et al.* The impact of target site accessibility on the design of effective siRNAs. *Nat Biotechnol* 2008;**26**:578–83.
114. Li X, Quon G, Lipshitz HD, *et al.* Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA* 2010;**16**:1096–107.
115. Hiller M, Pudimat R, Busch A, *et al.* Using RNA secondary structures to guide sequence motif finding towards single-stranded regions. *Nucleic Acids Res* 2006;**34**:e117.
116. Lopez de Silanes I, Galban S, Martindale JL, *et al.* Identification and functional outcome of mRNAs associated with RNA-binding protein TIA-1. *Mol Cell Biol* 2005;**25**:9520–31.
117. 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.
118. Leontis NB, Westhof E. Analysis of RNA motifs. *Curr Opin Struct Biol* 2003;**13**:300–8.
119. Hackermuller J, Meisner NC, Auer M, *et al.* The effect of RNA secondary structures on RNA-ligand binding and the modifier RNA mechanism: a quantitative model. *Gene* 2005;**345**:3–12.
120. Forties RA, Bundschuh R. Modeling the interplay of single-stranded binding proteins and nucleic acid secondary structure. *Bioinformatics* 2010;**26**:61–7.
121. Zheng S, Robertson TA, Varani G. A knowledge-based potential function predicts the specificity and relative binding energy of RNA-binding proteins. *FEBS J* 2007;**274**: 6378–91.
122. Barash Y, Calarco JA, Gao W, *et al.* Deciphering the splicing code. *Nature* 2010;**465**:53–9.
123. Anantharaman V, Aravind L. Novel conserved domains in proteins with predicted roles in eukaryotic cell-cycle regulation, decapping and RNA stability. *BMC Genomics* 2004;**5**:45.
124. Zhang Z, Theler D, Kaminska KH, *et al.* The YTH domain is a novel RNA binding domain. *J Biol Chem* 2010;**285**: 14701–10.
125. Tian B, Bevilacqua PC, Diegelman-Parente A, *et al.* The double-stranded-RNA-binding motif: interference and much more. *Nat Rev Mol Cell Biol* 2004;**5**:1013–23.
126. Nagai K, Oubridge C, Jessen TH, *et al.* Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* 1990;**348**:515–20.
127. Grishin NV. KH domain: one motif, two folds. *Nucleic Acids Res* 2001;**29**:638–43.
128. Siomi H, Matunis MJ, Michael WM, *et al.* The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res* 1993;**21**:1193–8.
129. Du Z, Lee JK, Tjhen R, *et al.* Crystal structure of the first KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.7 Å. *J Biol Chem* 2005;**280**:38823–30.
130. Ghisolfi L, Joseph G, Amalric F, *et al.* The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties. *J Biol Chem* 1992;**267**:2955–9.
131. Kiledjian M, Dreyfuss G. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J* 1992;**11**:2655–64.
132. Darnell JC, Jensen KB, Jin P, *et al.* Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* 2001;**107**:489–99.
133. Wang X, Zamore PD, Hall TM. Crystal structure of a Pumilio homology domain. *Mol Cell* 2001;**7**:855–65.
134. Pavletich NP, Pabo CO. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 1991;**252**:809–17.
135. Lee MS, Gippert GP, Soman KV, *et al.* Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 1989;**245**:635–7.
136. Parraga G, Horvath SJ, Eisen A, *et al.* Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* 1988;**241**:1489–92.
137. Brayer KJ, Kulshreshtha S, Segal DJ. The protein-binding potential of C2H2 zinc finger domains. *Cell Biochem Biophys* 2008;**51**:9–19.
138. Miller J, McLachlan AD, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J* 1985;**4**:1609–14.
139. Lu D, Searles MA, Klug A. Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* 2003;**426**:96–100.
140. Hall TM. Multiple modes of RNA recognition by zinc finger proteins. *Curr Opin Struct Biol* 2005;**15**:367–73.
141. Bycroft M, Grunert S, Murzin AG, *et al.* NMR solution structure of a dsRNA binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J* 1995;**14**: 3563–71.
142. Kharrat A, Macias MJ, Gibson TJ, *et al.* Structure of the dsRNA binding domain of *E. coli* RNase III. *EMBO J* 1995;**14**:3572–84.
143. Chang KY, Ramos A. The double-stranded RNA-binding motif, a versatile macromolecular docking platform. *FEBS J* 2005;**272**:2109–17.