

## SURVEY AND SUMMARY

# Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases

Magdalena Jazurek<sup>†</sup>, Adam Ciesiolka<sup>†</sup>, Julia Starega-Roslan, Katarzyna Bilinska and Włodzimierz J. Krzyzosiak<sup>\*</sup>

Department of Molecular Biomedicine, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

Received June 24, 2016; Revised August 31, 2016; Accepted September 01, 2016

### ABSTRACT

**RNA–protein complexes play a central role in the regulation of fundamental cellular processes, such as mRNA splicing, localization, translation and degradation. The misregulation of these interactions can cause a variety of human diseases, including cancer and neurodegenerative disorders. Recently, many strategies have been developed to comprehensively analyze these complex and highly dynamic RNA–protein networks. Extensive efforts have been made to purify *in vivo*-assembled RNA–protein complexes. In this review, we focused on commonly used RNA-centric approaches that involve mass spectrometry, which are powerful tools for identifying proteins bound to a given RNA. We present various RNA capture strategies that primarily depend on whether the RNA of interest is modified. Moreover, we briefly discuss the advantages and limitations of *in vitro* and *in vivo* approaches. Furthermore, we describe recent advances in quantitative proteomics as well as the methods that are most commonly used to validate robust mass spectrometry data. Finally, we present approaches that have successfully identified expanded repeat-binding proteins, which present abnormal RNA–protein interactions that result in the development of many neurological diseases.**

### INTRODUCTION

RNA–protein interactions play a key role in the regulation and coordination of gene expression in cells. Every stage of the RNA life cycle, including RNA synthesis, maturation, modification, transport, and degradation, is tightly

controlled by a multitude of RNA-binding proteins (RBPs). RNA molecules and their interacting protein partners form distinct, highly dynamic ribonucleoprotein (RNP) particles, which comprise the basic unit underlying these posttranscriptional events (1–3). Any defects in RBP expression and function as well as mutations in target RNA molecules can disrupt protein–RNA networks and cause human diseases, such as cancer, autoimmune pathologies, metabolic and neurological diseases (4–7). Moreover, interactions between viral RNA and host cell proteins mediate various aspects of viral replication, leading to infectious disease development (8–10). Therefore, intensive efforts are being undertaken to explore protein–RNA interactions, not only to better understand the complex interplay between RNAs and their associated RBPs in the regulation of fundamental cellular processes but also to gain more insight into the pathogenesis of numerous diseases.

To date, various genetic, biochemical or microscopic *in vitro* and *in vivo* methods have been developed to study the RNA-binding proteome (11–16). Several of these methods allow the comprehensive identification of new RNA-associated factors, whereas others characterize known or suspected RNA–protein interactions in detail. Undoubtedly, recent developments in high-throughput technologies, such as RNA affinity purification combined with mass spectrometry (MS), protein microarrays and next-generation sequencing, have significantly contributed to deciphering the repertoire of RNP complexes (17–20). Among the widely used methods that explore the RNA–protein interactome, RNA-centric approaches employ MS to identify the protein partners associated with a specific RNA, which is used as bait. Using this RNA-based strategy, RNP complexes are formed *in vitro* or *in vivo*, and a given RNA obtained from *in vitro* synthesis or cell lysates is immobilized to a chromatographic matrix either covalently or non-covalently. Non-

<sup>\*</sup>To whom correspondence should be addressed. Tel: +48 61 8528503; Fax: +48 61 8520532; Email: wlojdkrzy@ibch.poznan.pl

<sup>†</sup>These authors contributed equally to this work as first authors.

specifically binding proteins are removed by several extensive washing steps, and RNP complexes are eluted from the solid support for MS analysis. Despite the availability of various RNA affinity purification methods, the identification of relevant RNA–protein complexes remains technologically challenging. This difficulty is predominantly associated with the isolation of low-abundance proteins that specifically interact with an RNA of interest from complex protein mixtures containing highly abundant proteins that bind non-specifically to RNA. Another important issue that can impede RNA affinity purification is the fact that RNA–protein interactions in cells are highly dynamic and can undergo extensive remodeling. This transient nature of RNP complexes principally results from the fact that RNA is structurally very flexible and can adopt a large variety of tertiary structures (21,22). With regard to the above difficulties in identifying specific RBPs, many investigators are continuing to optimize existing strategies by stabilizing the aptamer structures that are used to tag RNA molecules of interest, using quantitative proteomics, or developing new approaches that ensure the specific elution or identification of RBPs that associate with RNA in living cells.

In this review, we first present RNA affinity purification approaches that are currently used for the identification of proteins that bind to a given RNA. We briefly discuss the advantages and limitations of purifying of *in vitro*- and *in vivo*-assembled RNP complexes, including the increasing demand for quantitative MS. In the second part, we focus on the methods used to determine proteins that bind to expanded RNA repeats, which are involved in aberrant RNA–protein interactions that lead to the development of many neurological diseases. This particular group of disorders is associated with the expansion of simple repetitive elements within specific single genes. Depending on the location of the mutation, various pathomechanisms might be involved, one of which is mutant RNA *gain-of-function*. According to this mechanism, the expression of transcripts harboring repeats of abnormal length leads to the formation of RNA foci that capture specific RNA-binding proteins, resulting in their altered function. It is likely that the proteins that have been identified thus far represent only a small fraction of mutant repeat-binding proteins and that most such proteins await discovery. Thus, there is a need to use more advanced protein capture approaches and to develop new, more efficient methods for identifying proteins that are associated with these unusual types of transcripts to gain better insight into the RNA-triggered mechanisms that contribute to these disorders.

## DESIGN AND IMMOBILIZATION OF RNA BAIT

### Covalent linking

Different protocols to immobilize the RNA bait on a solid support have been developed. This is achieved by chemically modifying bait RNAs by introducing RNA tags or by using antisense strategies (Supplementary Table S1). In the first case, *in vitro* transcribed RNA baits are covalently linked to the solid support, as is the case for oxidized RNAs linked to cyanogen-activated Sepharose beads (23) or for adipic acid dihydrazide agarose beads (24–27). Unfortunately, covalent

attachment of bait RNAs does not permit elution of RNA-binding proteins in a highly specific manner. Therefore, elution of RBPs is performed primarily using highly denaturing buffers (24,28) or by digesting RNA with robust RNases (29). Moreover, covalent attachment of bait RNA has been used almost exclusively with shorter transcripts (<100 nt), such as short mRNA regulatory motifs (25,26), specific pri-miRNAs (27), pre-miRNAs (30) and miRNAs (31).

### Biotin tagging

In an attempt to circumvent some of previously mentioned limitations, non-covalent methods to immobilize bait RNAs to a solid support have been used. Among the most common and efficient approaches is chemical modification of bait RNAs by incorporating specifically modified rNTPs during *in vitro* transcription (32–34). These modifications include biotin (vitamin H), desthiobiotin, and digoxigenin. Among these, biotinylation of the bait RNAs is the most widely used, as biotin shows unprecedented, high affinity to streptavidin/avidin - proteins isolated from *Streptomyces avidinii* ( $K_d \sim 10^{-13}$  M) (Supplementary Table S1). This binding was shown to be very rapid, highly specific, and resistant to high salt concentration, extreme heat, pH and proteolysis. The caveat of this strategy is that incorporation of biotin internally into bait RNAs may cause changes in the RNA structure and formation of nonphysiological RNA–protein complexes. Therefore, biotin tagging of *in vitro* transcribed bait RNAs is carried out at their 5' or 3' ends using T4 polynucleotide kinase or T4 RNA ligase (35). This single modification is unlikely to perturb the natural function of the molecule due to the small size of biotin (MW = 244.31 g/mol). However, 3' and 5' end biotinylation reactions can be inefficient or require long reaction times, especially for large (e.g. telomerase RNA, 451 nt) (36) or highly structured bait RNAs (e.g. let-7 pre-miRNA) (37). Furthermore, the high affinity of biotin to streptavidin/avidin makes native elution, which is performed by adding excess biotin, inefficient; therefore, elution of biotinylated RNAs is generally performed with highly denaturing buffers or with RNases. To circumvent these problems, desthiobiotin, which has a much lower affinity for streptavidin ( $K_d \sim 10^{-5}$  M) can be used instead of biotin (38). Tagging of RNA bait with biotin is principally restricted to the analysis of *in vitro* assembled RNP complexes due to the limited transfection efficiency of biotinylated RNAs and the high cost of obtaining a sufficient quantity of biotinylated transcripts for transfection.

### MS2 aptamer tagging

The MS2 aptamer (23 nt long) is a commonly used, naturally occurring RNA stem–loop aptamer that is utilized for the *in vitro* or *in vivo* isolation of RNA-binding proteins (Supplementary Table S1). The MS2 RNA stem–loop structure can bind specifically and with a high affinity to a coat protein from an *Escherichia coli* bacteriophage, MS2cp (39–41). In this approach, repeats of the MS2-binding RNA stem–loop are incorporated (during *in vitro* or *in vivo* transcription) into an RNA of interest, and the tagged RNA complex is purified by coupling the MS2 protein to a solid

support or resin. Generally, the 3' end of the RNA is tagged with MS2, but occasionally the 5' end is used. Two variants of MS2 aptamers are used: the wild-type MS2 and high-affinity C-loop MS2 with  $K_d$  values of 1–3 and 0.2–0.6 nM, respectively. Due to this high binding affinity between MS2cp and MS2 aptamer, the limiting step in the affinity purification of RNA-binding proteins is the elution of the purified complex under native conditions. MS2cp is often fused with another protein, such as maltose-binding peptide (MBP) or streptavidin-binding protein (SBP), to ensure that the elution is specific and to reduce the identification of background proteins. The MBP domain of the fusion protein is captured on amylose beads, and elution of intact protein-RNA complexes from the beads is facilitated by adding excess maltose (42,43). In the second case, SBP fusion allows for the purification of RNA-protein complexes using streptavidin-conjugated beads and a specific, native elution using biotin as a competitor (44). Alternatively protease cleavage site (TEV) can be inserted between MS2cp and another fused protein attached to the resin (45,46). The MS2 approach was successfully used to identify the RBPs associated with long non-coding RNAs, highly stable and abundant RNPs, such as the U1 small nuclear ribonucleoprotein particles (snRNP) (39), and less stable mRNA-binding proteins (44).

#### PP7 aptamer tagging

The PP7 stem-loop aptamer is a different, naturally occurring RNA aptamer that is similar to the MS2 system and is also used for *in vitro* and *in vivo* purifications of RNA-binding proteins (Supplementary Table S1) (41,47). This 25 nt long stem-loop aptamer can be fused to the 5' or 3' end of an RNA of interest and was shown to bind with high specificity to the *Pseudomonas aeruginosa* bacteriophage 7 coat protein PP7 (47). Although the PP7 coat protein maintains a high affinity to aptamer-tagged RNA molecules across a wide range of ionic strength and pH, the PP7 system is used less frequently than the MS2 system, possibly due to a higher dissociation constant ( $K_d \sim 1$  nM). Furthermore, the PP7 system was shown to efficiently isolate both stable RNA-protein complexes, i.e. 7SK RNPs (48), as well as more transient RBPs associated with nascent mRNAs (49).

#### S1 and D8 aptamer tagging

Another approach to purifying specific RNA-protein complexes is to use artificial RNA aptamers with high binding affinity to known proteins or to small molecule ligands. In contrast to the described bacteriophage systems, these aptamers do not require the synthesis of recombinant proteins. Using SELEX approaches, the artificial S1 aptamer (44-nt long) that binds to streptavidin and the D8 aptamer (33-nt long) that binds to Sephadex (polysaccharide dextran B512) were identified (Supplementary Table S1) (50,51). Because the binding affinity of the S1 aptamer to streptavidin ( $K_d \sim 70$  nM) is higher than that of D8-Sephadex, this artificial aptamer is preferentially used. Historically, the S1 aptamer has been used extensively for the *in vitro* and *in vivo* purification of RNA-protein complexes, such as those associated with RNase P ribonuclease (51,52), 28S rRNA (45,53)

and mRNAs (54–56). Furthermore, the weaker affinity of the S1-streptavidin interaction compared with the biotin-streptavidin interaction is commonly used to enable the specific elution of the S1-tagged RNAs from the streptavidin resin using excess biotin (57). Moreover, by optimizing the S1 aptamer structure and repeat conformation (60 nt), a 15-fold increase in the purification of a specific RNA-protein complex was observed when compared to that obtained using the MS2 and PP7 systems (58). This modified 4xS1m aptamer was used to capture and identify (from cellular extracts; an *in vitro* strategy) both known and novel RBPs that bind to AU-rich element of tumor necrosis factor  $\alpha$  (58).

#### Tobramycin and streptomycin aptamer tagging

Tobramycin aptamer (40 nt) and streptomycin aptamer (46 nt) are also artificial, stem-loop oligonucleotides that were identified using the SELEX approach and were shown to bind with high affinity to tobramycin ( $K_d \sim 5$  nM) and streptomycin ( $K_d \sim 1$   $\mu$ M) matrices, respectively. Purification of RNA-protein complexes using these two aptamers is performed mainly *in vitro* (Supplementary Table S1). Using tobramycin aptamer-tagged pre-mRNA, structurally intact and catalytically active pre-spliceosomal complexes were purified under physiological conditions for the first time (59,60). The elution of native spliceosomes was performed using a competitive approach in the presence of excess tobramycin (61). One copy of this aptamer tag was generally attached to an RNA of interest at either the 5' or 3' end of the RNA or internally. In turn, streptomycin aptamer-tagged (StreptoTag) U1 snRNA was used to identify the complement of proteins that are associated with 5' splice site selection during pre-mRNA splicing (62). The streptomycin-tagged RNA-protein complexes were bound with high affinity ( $K_d \sim 1$   $\mu$ M) to an affinity column containing Sepharose-immobilized streptomycin. This association is highly dependent on the presence of  $Mg^{2+}$  ions (63). Following binding, the complexes were recovered from the affinity matrix by elution with excess streptomycin. One copy of the aptamer tag was attached to an RNA of interest, usually (but not always) at the 3' end of the RNA. Furthermore, a StreptoTag was successfully used to purify yeast and phage RNA-binding proteins, as well as proteins that are associated with group II intron, viral and bacterial ncRNAs (noncoding RNAs). Furthermore, certain bacterial mRNPs that bind ncRNAs have also been identified (62,64,65).

#### The CRISPR/Csy4 system

Recently, a novel RNA affinity tag has been developed that utilizes a *Pseudomonas aeruginosa* CRISPR/Csy4 system for the highly efficient purification of RNA-binding proteins associated with specific RNAs (Supplementary Table S1) (66). In this system, *in vitro* generated RNA transcripts engineered with a short 16 nt hairpin (5 bp stem and 5 nt loop) were shown to irreversibly bind, with exceptionally high affinity ( $K_d = 50$  pM), to an inactive, biotinylated form of Csy4 endoribonuclease. Upon immobilization of biotinylated Csy4 on a solid support, e.g. streptavidin-conjugated beads, Csy4 catalytic activity can be rescued by the addition of imidazole, to cleave the hairpin-tagged RNA, removing the tag and releasing the remaining RNA together

with its associated proteins to the solution. This highly specific elution step was shown to generate fewer false positives compared to other elution methods. Although this system was used to *in vitro* purify RNA-binding proteins associated with relatively short RNAs (pre-miRNAs) tagged with this 16 nt hairpin at their 5' ends, longer transcripts, different positions of the hairpin within studied RNAs, and utilization of this system in an *in vivo* context should be considered (66).

### Antisense oligomers for RNA capture

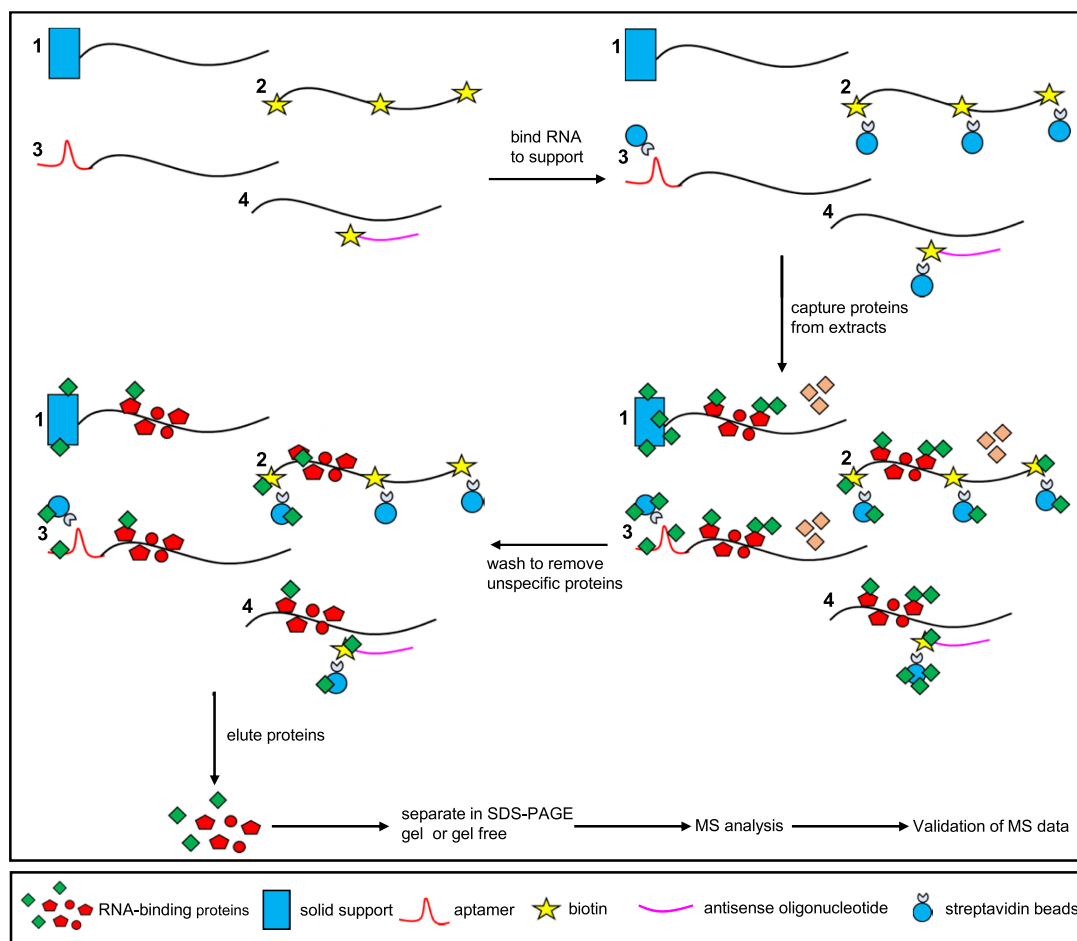
Antisense affinity capture approaches are another powerful tool for analyzing protein complexes bound to specific RNA. These approaches allow isolation of *in vitro* transcribed or, more importantly, endogenously expressed specific RNAs from cells together with associated proteins (Supplementary Table S1). In principle, antisense oligonucleotides, modified or unmodified, are first attached to a chromatographic support, e.g. by using the streptavidin–biotin interaction, and are used to isolate the RNA bait of interest by nucleic acid complementarity. These RNP complexes can then be eluted in their native form via excess competitor oligonucleotides or simply by using denaturing buffers or RNases. The DNA modifications that are commonly used to enhance the specificity and affinity of antisense oligonucleotides to target RNAs are 2'-O methylation and 2'-O-alkylation (67–69). Moreover, in a recently developed PAIR strategy (PNA-assisted identification of RBPs), described in detail later in the text, PNA oligomers were used to isolate a specific endogenously expressed ankyloses mRNP complex with satisfactory efficiency. PNAs are peptide nucleic acid analogues that bind RNA with high sequence specificity, forming highly stable RNA–PNA hybrids and increasing the stability and affinity of these duplexes by significantly increasing their melting temperatures (13). In contrast to previously described methods, antisense affinity capture does not require chemical or sequence modification of the RNA of interest. Moreover, these methods are suitable for cells that are difficult to transfect with exogenous expression constructs, theoretically enabling any endogenous RNA of interest to be isolated. The major caveat of these antisense methods is that due to RNA secondary structure, some mRNA sequences may not be accessible to antisense oligonucleotides, and therefore, the design of a working fishing probe is a challenging and time-consuming step. Moreover, these methods have been thus far applied to isolate RNP complexes that are extremely abundant in cells, and therefore, these strategies may not be suitable for purification of all RNPs. Several of the most prominent examples of antisense affinity capture methods are isolation of U4/U6 snRNPs from HeLa nuclear extracts (70), purification of telomerase complexes (71), and more recently, identification of protein components constituting the small nucleolar ribonucleoprotein (snoRNP) complex MBII-52 from mouse brains (72). Furthermore, a comprehensive analysis of an mRNA-bound proteome referred to as interactome capture, which used immobilized oligo(dT) probes, was performed to capture and analyze poly(A)-tailed RNAs and their interacting proteins (73,74).

### PURIFICATION OF *IN VITRO*- AND *IN VIVO*-ASSEMBLED RNA–PROTEIN COMPLEXES

The methods used to isolate novel RNA-binding protein complexes can be grouped into two main classes: *in vitro* and *in vivo* purification strategies (14,15,75). To date, most of the known RNA–protein interactions have been identified using *in vitro* RNA pull-down assays (Figure 1). As shown in Supplementary Table S1, the *in vitro* reconstitution of RNP complexes employs various RNA-tagging strategies, such as covalent linking, biotin labeling, or the introduction of artificial or natural aptamers, which ensure that the RNA of interest is immobilized to a solid support. To increase the specificity of the identification of *in vitro* interactions, intensive efforts have been undertaken to improve existing purification strategies, which predominantly rely on aptamer structure stabilization and the use of specific elution strategies or quantitative proteomics (54,57,58,76). Although *in vitro* methods are the most widely used, controllable and efficient strategies, they depend on the formation of RNA–protein complexes from synthetic target RNAs and cell extracts. These complexes may not represent genuine RNA–protein complexes because immobilized RNAs may not fold properly and nonspecific RNA protein interactions can form during the purification process. These methods also do not allow the analysis of RNA–protein interactions that are formed in response to environmental conditions.

To overcome the limitations that are associated with *in vitro* RNA purification methods, *in vivo* purification approaches that capture RNA–protein complexes that are present in living cells can be performed (Supplementary Table S1). These *in vivo* methods permit the RNA bait to form in a native, intracellular environment; therefore, protein complexes that are more physiologically relevant are formed (Figure 2). However, these experiments are more technically challenging, especially if the target RNA is of low abundance in the cell, and result in a relatively low recovery of the RNP complexes (usually not exceeding 20%) (46,49,51,72,76). Another advantage of *in vivo* strategies is the application of intracellular RNA–protein crosslinking, which can be performed to 'freeze' physiological RNA–protein complexes and to preserve transient protein–RNA interactions within the cell. Furthermore, crosslinking approaches enable specific RNA-binding protein complexes to be purified under fully denaturing conditions, thereby limiting the identification of false-positive contaminant interactions.

Two main strategies are extensively used to purify *in vivo*-formed RNP complexes; these strategies differ in the presence or absence of target RNA sequence modification. Aptamer-tagged RNA affinity purification uses the first strategy, and antisense RNA capture uses the second. In methods using various RNA affinity tags, cells are transfected with plasmids expressing aptamer-tagged RNAs of interest, and depending on the strategy used, additional plasmid-expressed proteins that allow for the specific isolation of RNP complexes. Several improvements in aptamer strategies have been developed to significantly increase the purification of biologically relevant endogenous RNA–protein interactions (44,46,48,49).



**Figure 1.** *In vitro* RNA affinity capture approaches. Four strategies can be used to immobilize an *in vitro* transcribed RNA of interest on a solid support. (1) RNA can be covalently linked to a solid support. (2) RNA can be chemically tagged through incorporation during *in vitro* synthesis of biotin-containing ribonucleotides or after transcription by attachment of a biotin tag by T4 RNA ligase. In this case, immobilization of target RNA is possible due to interactions between biotin and streptavidin beads. (3) Various natural or artificial aptamers can be attached co-transcriptionally to the RNA of interest. Using this tagging strategy, the RNA of interest is bound to chromatographic support through aptamer-ligand interactions. (4) RNA baits can be also isolated with antisense oligonucleotides, which are coupled with various beads. RNAs (1, 2, 3, 4) are then used to assemble ribonucleoprotein complexes using cell lysates. After incubation with various cellular extracts, RNAs with associated proteins are pulled down and washed to remove non-specifically bound proteins. Then, RNA-binding proteins are released from RNA using various elution strategies depending on the used methods. Eluted proteins are usually separated by SDS-PAGE, and the protein composition is analyzed by mass spectrometry (MS). The authenticity of the MS data is subsequently confirmed using different validation methods.

### The RAT system

One of them was the RNA affinity in tandem method (RAT), in which efficient isolation of *in vivo* assembled RNP complexes was achieved by two steps of purification due to the presence in the target non-coding RNAs of two affinity tags: PP7 and tobramycin (48,49). Using this method, endogenously formed RNP complexes were first recovered by recombinant PP7cp tagged with a TEV protease cleavage site. Following elution by TEV protease, tagged RNP complexes are bound to tobramycin resin and eluted by elevation of the buffer pH and denaturation. The high RNP yield and enrichment predominantly results from the strong association between recombinant PP7cp and its cognate binding site. Additionally, this optimized form of PP7cp, in addition to higher RNA binding affinity, had also negligible dimer aggregation properties. It appears that despite exogenous expression of RAT-tagged target RNAs, this method allows

purification of physiologically relevant RNP complexes because of the similar expression levels of tagged and endogenous RNAs of interest (48,49).

### The RaPID system

Another optimized *in vivo* system, RBP purification and identification (RaPID), uses a novel fusion protein, MS2cp-GFP-SBP, which enables both affinity purification of RNA-protein complexes using streptavidin-conjugated beads due to interaction with SBP and visualization of intracellular localization of target mRNAs bearing the MS2 aptamer using microscopy due to a fluorescence reporter (MS2cp-GFP) (44,77). Specific and strong binding of MS2cp to the MS2 loops and of SBP to streptavidin as well as formaldehyde crosslinking allows for stringent washing conditions, which prevent isolation of nonspecific RNA-protein interactions. Another advantage of RaPID analysis

is simple and specific elution of RNP complexes by competition with biotin. Additionally, to avoid MS2cp aggregation following high expression levels, genes encoding fusion proteins were placed under the control of an inducible promoter, which allow tightly regulated stable expression in both yeast and mammalian cells (44,77).

### The MS2-BioTRAP system

An *in vivo* biotin-tagged RNA affinity purification (MS2-BioTRAP) approach that integrates an RNA-tagging strategy with UV crosslinking and SILAC-based quantitative MS has also been developed for the isolation of *in vivo* assembled RNA–protein complexes under native or fully denaturing conditions (Figure 2A) (46). In this strategy, the target RNA is tagged with 4 copies of MS2 aptamers that are recognized by optimized MS2cp that is linked to a specific signal sequence for *in vivo* biotinylation (HB tag). The presence of this specific biotinylation tag allows for the rapid and efficient one-step purification of target RNA together with its associated protein complexes by endogenously biotinylated MS2cp-HB proteins (serving as a ‘fishing rod’) and streptavidin-coated beads. MS2-BioTRAP was successfully used to identify proteins that are associated with cellular IRES elements (46). Recently, this method was also applied to confirm *in vivo* MS data that were obtained from an *in vitro* analysis of RNA-binding proteins that interact with mouse Nanog mRNA (78).

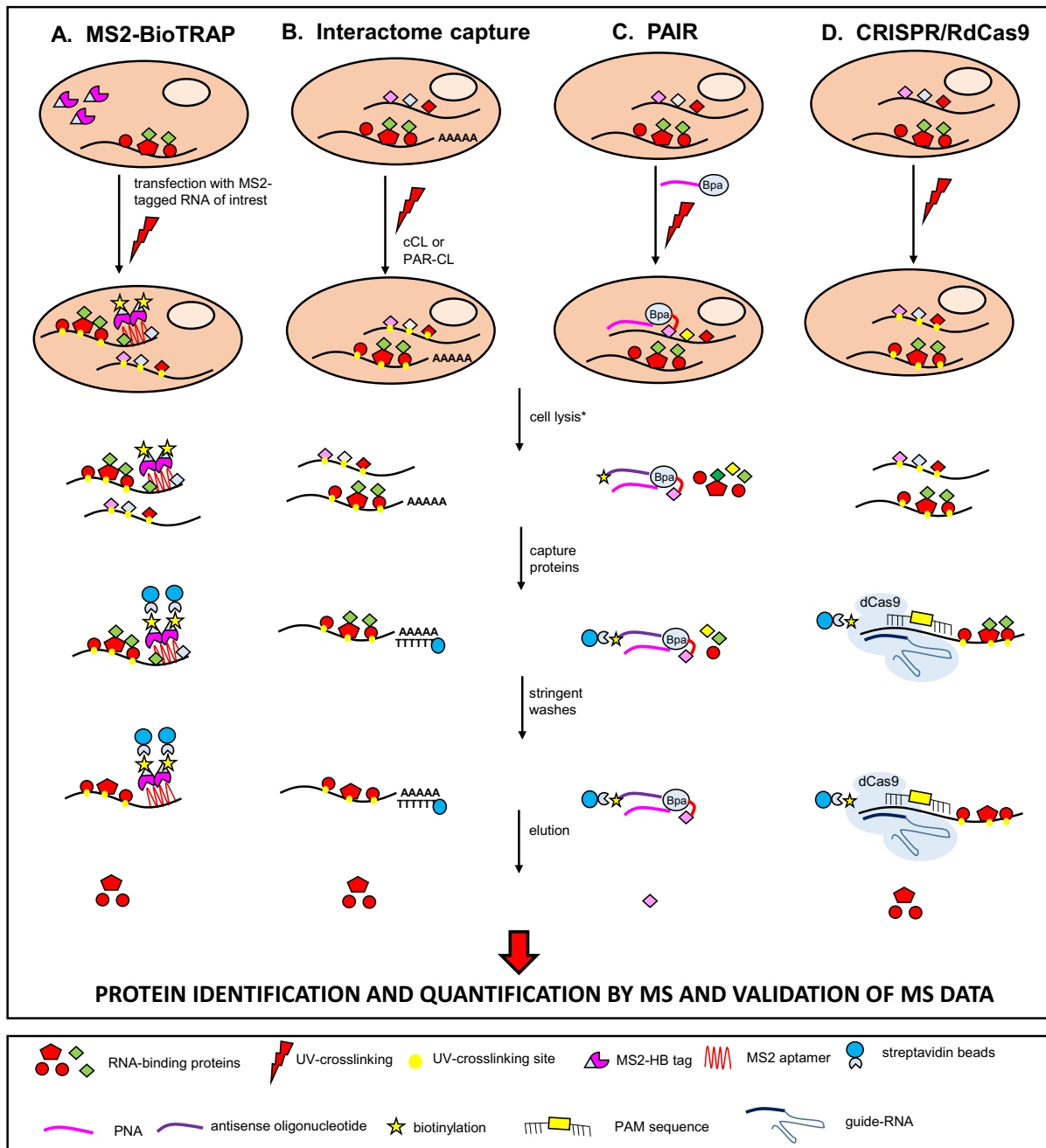
### Antisense affinity capture methods

In general, one of the major limitations of using the *in vivo* aptamer tagging of target RNAs is the requirement for cell transfection, and thus, analysis of exogenously assembled RNA–protein complexes. Furthermore, gene overexpression may result in the aggregation or mislocalization of interacting molecules, thus leading to the formation of non-physiological RNA–protein complexes. Additionally, aptamer-tagged RNAs cannot be used in cells that are difficult to transfect. Therefore, methods utilizing nucleic acid hybridization, in which there is no need for the modification of RNAs of interest, are also used for the purification of *in vivo*-assembled RNP complexes (Supplementary Table S1). Interactome capture has allowed the complete repertoire of predominantly mRNA interactome proteins in cultured cells (HeLa and HEK293 cells, embryonic stem cells, yeast) to be identified using antisense oligonucleotides that are complementary to polyA tails (Figure 2B) (73,74,79,80). An important feature of the interactome capture technique is the use of two conditions that promote only direct RNA–protein UV crosslinking: (i) conventional UV crosslinking (cCL), in which cells are irradiated with UV light at 254 nm to crosslink naturally photoreactive nucleotide bases with amino acids such as Phe, Trp, Tyr, Cys and Lys and (ii) photoactivatable-ribonucleotide-enhanced crosslinking (PAR-CL), which utilizes photoactivatable nucleotide 4-thiouridine incorporation into RNA in living cells and subsequent UV irradiation at 365 nm. This method can be applied to study RNP composition under different biological cues and environmental stimuli. More recently, modification of the interactome capture method

has been developed, which allows identification of mRNA–protein interactions with subcellular resolution due to a multiple purification procedure (81). Using this strategy, called serial RNA interactome capture, the first human nuclear RNA interactome in myeloid leukemia cells was obtained (81). Despite the notable advantages of interactome capture over the above-mentioned purification techniques, this approach is limited to identification of only mRNA-binding proteins. It appears that the PAIR technology has the potential to isolate RBP complexes for any RNAs expressed in living cells (Figure 2C) (13,82,83). This method uses a cell membrane-penetrating peptide to efficiently deliver a linked PNA oligomer, which is complementary to target endogenous RNAs, into living cells. The presence of a photoactivatable amino acid, *p*-benzoylphenylalanine, in the PNA sequence promotes capture of the adjacent RBP by UV crosslinking. Endogenously formed RNA–protein complexes can be isolated by using a biotinylated PNA oligomer coupled to streptavidin magnetic beads that is complementary to a specific RNA of interest. In general, PAIR technology, such as interactome capture, can be applied to analyze RBP dynamics in varied biological settings and can be compatible with quantitative proteomics. Using PAIR, it was possible to identify RBPs associated with ankyloses RNA (13).

### The CRISPR/RdCas9 system

As has been shown recently, the CRISPR/Cas9 system that has been designed for genome editing can also be used for RNA-guided binding and/or for the cleavage of specific ssRNA sequences. The developed system (CRISPR/RdCas9) has great potential because it can be used to selectively and effectively purify endogenously expressed, untagged RNA–protein complexes from cells (Figure 2D) (84). This system comprises three crucial elements: nuclease-inactive Cas9 protein (dCas9), a dCas9-associated guide RNA (sgRNA) that matches the target ssRNA, and a short PAM motif-presenting DNA oligonucleotide (PAMmer). This PAMmer is presented in *trans* and sits upstream of the target ssRNA sequence. The system needed some modification to capture specific ssRNA sequences. First, the original Cas9 protein was mutated (D10A;H840A) to abolish its catalytic activity, thereby preventing cleavage of the target ssRNA. Second, a 5′ overhang covering part of the target sequence was added to PAMmer beyond the PAM motif to generate specific RNA recognition that is programmed by sgRNA. Third, the PAMmer sequence was mismatched at the PAM motif to achieve a specificity of the CRISPR/RdCas9 system for RNA rather than the corresponding DNA locus. Finally, to prevent the cellular RNase-H-mediated cleavage of target ssRNA during the pull-down step, the PAMmer DNA oligonucleotide was chemically stabilized using LNA, 2′ OMe, or 2′-F ribose modifications (84,85). Thus far, this CRISPR/RdCas9 approach has only been applied to the selective capture of an endogenous, untagged, approximately 1500-nt long GAPDH transcript from HeLa cell extracts under physiological salt conditions. As a proof-of-concept, dCas9 was bound to solid resin by the site-specific biotin labeling of dCas9. However, the complement of proteins associated with GAPDH mRNA



**Figure 2.** Selected *in vivo* RNA-centric approaches to identify novel RNA-binding proteins. (A) The MS2 *in vivo* biotin-tagged RNA affinity purification (MS2-BioTRAP) strategy relies on the co-expression in living cells of a MS2-tagged RNA of interest and MS2 coat proteins fused to an HB tag, which contains a signal sequence for *in vivo* biotinylation. Generally, these recombinant MS2 coat proteins are stably expressed. After UV crosslinking, cells are lysed, and the associated proteins are captured by streptavidin-coupled beads. (B) Interactome capture allows identification of RBPs that specifically associate with mRNAs in living cells. This approach employs two strategies that differ in the type of *in vivo* UV crosslinking that covalently links RNAs with interacting RBPs: conventional crosslinking (cCL-254 nm) and photoactivatable-ribonucleoside enhanced crosslinking (PAR-CL-365 nm). After cell lysis, covalently bound RBPs are isolated using oligo(dT) magnetic beads. (C) Peptide nucleic acid (PNA)-assisted identification of RBP (PAIR) technology uses a specific mRNA-binding probe, PNA, containing the photoactivatable amino acid adduct *p*-benzophenylalanine (Bpa). PNA can cross the cell membrane of living cells due to coupling with a cell-penetrating peptide and hybridizes to complementary sequences of the endogenous RNA of interest. UV light induces covalent crosslinks between Bpa and the nearest RBP. After cell lysis and RNase treatment, PNA-RBP complexes are captured by hybridization of a biotinylated oligonucleotide antisense to PNA, coupled to streptavidin. (D) The CRISPR/RdCas9 system may represent a future RNA-based approach. Using this system, proteins bound to endogenous unmodified RNAs of interest could be captured using catalytically inactive biotinylated dCas9 tethered to streptavidin beads. Specific recruitment of dCas9-guide-RNA to a given RNA is possible using a protospacer adjacent motif (PAM) *in trans* as a separate DNA oligonucleotide. UV crosslinking of living cells before pull-down experiments might additionally increase the specificity of the identified proteins. (A, B, C, D) After stringent washing conditions to remove non-specific RNA-protein interactions, bound proteins are eluted from RNA and subjected to proteomic analyses. Obtained MS data are subsequently validated to confirm biologically relevant RNA-protein interactions. (\*) in case of PAIR technology cell lysis is followed by RNase treatment.

was not determined (84). More recently, the application of this CRISPR/RdCas9 system for *in vivo* RNA studies was demonstrated, not for RNA–protein complex purification, but for recognizing and visualizing specific endogenous mRNAs by live confocal microscopy imaging (85). To track the journey of mRNA from the nucleus to the cytoplasm, researchers have used a nuclear localization signal-tagged dCas9 fused to GFP, and this did not alter target mRNA abundance and localization or the amount of translated protein.

### PROTEOMIC ANALYSIS OF ISOLATED RBPs

To comprehensively identify RNA-binding proteins by MS, two different approaches are being used: non-quantitative and quantitative proteomics. The non-quantitative analysis requires one or two dimensional gel-based separation of eluted proteins from the RNA bait and the control purifications. The gel is stained for total protein, and protein bands that are exclusively present in the RNA bait proteome and not in the control purification are then processed and identified by MS analysis. The major caveats of this non-quantitative proteomic approach are that it impedes identification of proteins that due to a low abundance are not visible on a gel, and it results in difficulties in confirming the binding specificity of the abundant cellular proteins that also bind, to a certain extent non-specifically, to the solid support (beads) (86). To overcome these issues, quantitative proteomics can be used. There in a single MS analysis quantification and comparison of the protein interactomes from control and RNA bait purifications can be simultaneously performed. Several different approaches have been developed to perform this type of analysis (Figure 3) (87). These strategies rely on the chemical (ICAT, iTRAQ, dimethyl, <sup>18</sup>O labeling) or metabolic labeling (e.g. SILAC, <sup>15</sup>N, <sup>13</sup>C labeling) of either already isolated protein complexes or whole cells/organisms (Figure 3). Following isotopic labeling, differentially tagged protein pools are simultaneously analyzed by MS and compared to provide direct quantification. Among the quantitative proteomics strategies, the simplest, most reproducible and most frequently used approach to identify RNA-binding proteins is stable isotope labeling by amino acids in cell culture (SILAC) (Figure 3B) (88,89). In this method, by growing two cell populations in medium containing different amino acid isotopes (e.g. ‘light’ lysine or ‘heavy’ lysine), cells are metabolically labeled to generate differentially tagged protein pools for MS analysis. During these analyses, the isotopically labeled proteins from different cells can be compared to provide direct relative quantification. The main advantage of the SILAC approach is that it allows identification of true binding partners from non-specific interactors by comparing the ratios of peptides from the experimental and control samples. Moreover, the SILAC approach can monitor the changes between different interactomes in terms of protein composition (control versus RNA bait) and also enables quantification of various post-translational modifications (e.g. phosphorylation, acetylation, ubiquitination) (90). Furthermore, this technique was already successfully used to identify many novel RNA-binding proteins that interact with various classes of RNAs, such as pri- or pre- miRNAs (30), telomeric repeat-

containing RNA (91), spliceosomal snRNAs (92,93), viral RNAs (61,76) and expanded GGGCC hexanucleotide repeats (94).

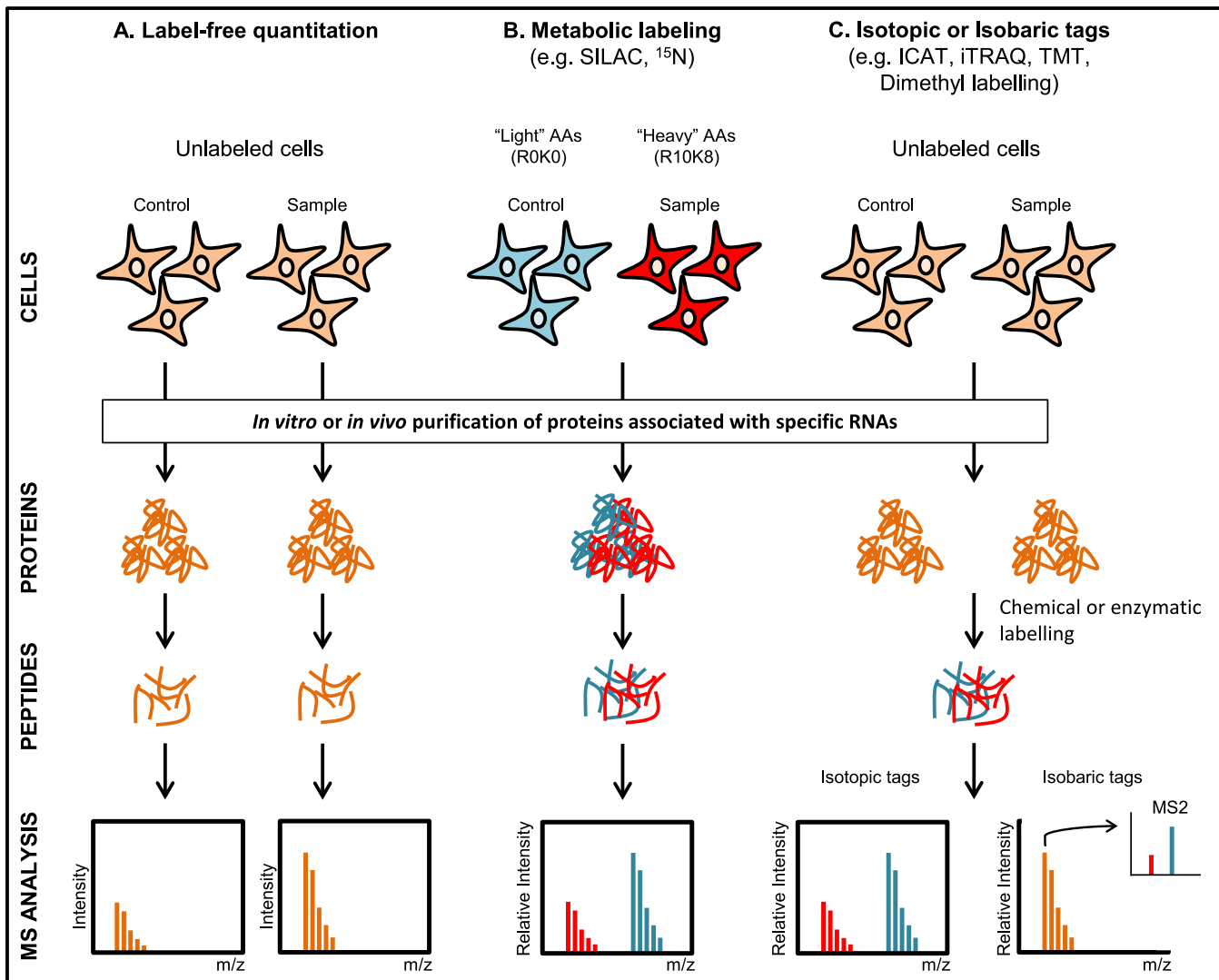
### VALIDATION OF MS DATA

Generally, MS analysis of RNA pull-down assays identifies tens of proteins, among which only some directly bind to a given RNA. Moreover, many of the identified proteins can interact with RNA nonspecifically/non-physiologically or indirectly through other proteins. Therefore, one of the required steps in RNA affinity purification is validation of MS-identified RNA–protein interactions. A variety of *in vitro* and *in vivo* methods allow not only verification of the authenticity of MS data but also provide a detailed analysis of RNA–protein interactions. Using *in vivo* methods is particularly important because they indicate which RNA–protein interactions are truly biologically relevant. Interestingly, these methods have also been successfully applied to identify putative RNA–protein interactions. In this case, proteins that potentially bind to the RNA of interest are selected based on the sequence, structure or functional similarities.

Among *in vitro* methods that are commonly used in validation of MS data are filter binding assay (FBA) and electrophoretic mobility shift assay (EMSA). These methods are also mainstays for determination of direct RNA–protein interactions. FBA is based on the premise that proteins, but not RNA molecules, can bind to nitrocellulose membrane filters; therefore, only RNAs associated with proteins can be retained on a membrane filter and assayed, while unbound RNA will pass through (95,96). EMSA relies on the fact that the electrophoretic mobility of RNA–protein complexes is usually lower than that of the free RNA (11,97–99). The typical FBA or EMSA experiment is performed with a constant trace amount of <sup>32</sup>P-labeled RNA that is titrated with increasing concentrations of purified protein, generally recombinant protein. Because radiolabeled RNA is used, FBA and EMSA are extremely sensitive assays. Additionally, EMSA, in contrast to FBA, allows not only the determination of binding affinities but also the post-electrophoretic visualization of the RNP complexes. Despite the many advantages of FBA and EMSA, these methods suffer from the limitation that the binding reaction is not measured in free, aqueous solution, and therefore, they do not provide real-time kinetic data. When true equilibrium conditions are necessary, particularly with highly dynamic RNA–protein interactions, fluorescence anisotropy (FPA) can be employed. This solution-based technique measures the rate of depolarization of a fluorophore during its lifetime, which is often in the low nanosecond range and depends on the molecular volume and/or flexibility of the labeled RNA molecules (100–102). In general, RNA–protein complexes, due to their larger size, rotationally diffuse more slowly and retain more emission polarization than RNA alone, which rapidly rotates and more effectively depolarizes the emission. Thus, changes in FPA reflect the dynamics of protein binding to a fluorescent RNA substrate.

UV crosslinking assays and footprinting are other *in vitro* methods that can be used to confirm RNA–protein in-





**Figure 3.** Major quantitative proteomic approaches for the identification and analysis of proteins associated with specific RNAs. The stage in each workflow when samples are isotopically labeled for quantitative MS analysis is indicated by blue (light – control sample) and red (heavy – RBPs of interest). (A) The exception is label-free quantitation, where the samples are separately collected, prepared, and analyzed by MS, after which the data from the control and studied sample are compared using multiple approaches (peak intensities and spectral counting). To account for any experimental variations, label-free quantification experiments should be more carefully controlled than stable isotope approaches. (B) In the case of metabolic labeling of cells in culture, e.g. SILAC, labeling of proteins is performed *in vivo* by growing cells in medium containing different isotope-labeled amino acids, with arginine (R) and lysine (K) being the most commonly used. The cells used for control purification (RNA aptamer tag only) and cells that express the aptamer-tagged RNA of interest are grown in light medium (R0K0) and heavy medium (R10K8), respectively. Then, the cells or protein extracts used as a source for RBP isolation are combined and processed together for the quantitative analysis. By combining these samples early in the labeling workflow, this strategy has the lowest risk of experimental bias. (C) If metabolic labeling is not possible (e.g. human tissues) or cost prohibitive (e.g. mouse model organisms), alternative approaches, such as chemical or enzymatic labeling of isolated RNA-binding proteins, are applied. This labeling can be performed by adding isotopic (ICAT, dimethyl labeling) or isobaric mass tags (iTRAQ, TMT) to already purified proteins or to peptides generated after proteolytic cleavage. The resulting differentially labeled peptides, from control and the sample of interest, are then pooled together to be analyzed by MS. Unlike isotopic labeling methods that use MS1 precursor ion spectrum for relative quantification, when isobaric mass tags of identical masses and chemical properties are chosen, relative quantification is obtained from MS2 spectra representing peptide fragment ions generated after collision-induced dissociation.

teractions identified by MS. The first method utilizes UV irradiation to covalently attach any proteins that are directly bound to the RNA of interest (103,104). Briefly, radiolabeled RNA probes and purified recombinant proteins or protein extracts are incubated to form RNP complexes spontaneously. The binding reaction is then exposed to UV light, followed by RNase digestion to remove RNA not covalently bound to the protein. The RNP complexes are analyzed by SDS-PAGE, and the signals visualized by phos-

phoimaging (103,104). The second assay, also known as a protection assay, is based on the ability of ligands (e.g. proteins) to protect RNA from cleavage at its binding site. Therefore, in addition to identification of an interaction between RNA and protein, footprinting also allows a detailed analysis of the sequence recognized by a given protein (105,106). In a classical footprinting assay, the RNA fragment, which is usually radiolabeled at one end, is cut by a chemical or enzymatic cleavage agent (sequence- and

structure-specific RNases) in the presence or absence of the protein of interest. After cleavage, the resulting ladders are resolved on a denaturing polyacrylamide gel and visualized by autoradiography. The gaps in the array indicate RNA binding sites protected by the protein (105,106).

The yeast three-hybrid (Y3H) system is a useful tool to validate RNA–protein interactions in an *in vivo* context. This powerful genetic method involves the expression in yeast cells of three chimerical molecules, which assemble to activate the reporter genes, HIS3 and LacZ (16,107,108). The first hybrid protein consists of a DNA-binding protein linked to an RNA-binding protein. The second chimeric protein is a fusion protein of a transcription-activating domain and the RNA-binding domain of interest. The third element of Y3H is a hybrid RNA molecule that promotes the interaction of the two hybrid proteins by providing the two specific RNA targets for the RNA-binding proteins. However, this approach requires a careful interpretation because the most common problem of Y3H analysis is the detection of a large number of false positive clones that may prohibit isolation of genuine RNA–protein partners. Therefore, it is necessary to perform additional control steps to confirm the biological relevance of the identified interaction (16,107,108).

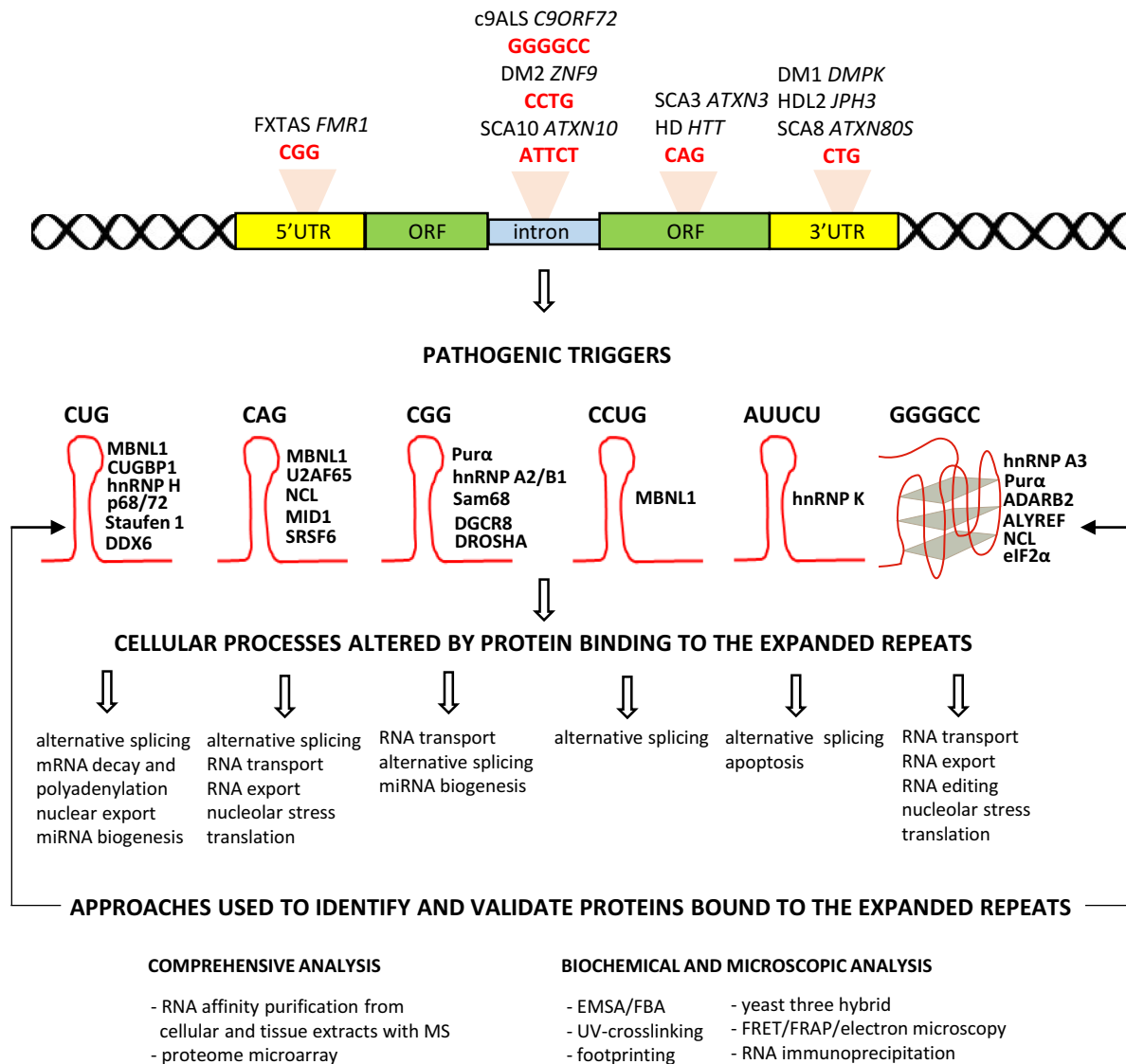
Antibody-based RNA-immunoprecipitation (RIP) is a very common method for studying *in vivo*-formed RNA–protein complexes. Using a high-quality antibody, an RNA-binding protein is immunoprecipitated together with its associated RNA, which can then be analyzed using RT-PCR, qPCR or next-generation sequencing (109–111). Furthermore, using different lysis buffers, it is possible to isolate either weakly or strongly associated RBPs. However, due to the post-lysis reorganization of co-immunoprecipitated complexes (112) and the inability to distinguish direct from indirect RNA–protein interactions, several improvements in RIP analysis, such as CLIP (113,114), HiTS CLIP (115,116), PAR-CLIP (18,117), iCLIP (118,119) and eCLIP (120,121), have been developed. These approaches not only confirm the direct binding of proteins to the RNA of interest but also determine the exact binding site of the RNA.

A powerful tool for validation of MS data *in vivo* is various microscopic analyses. RNA fluorescence *in situ* hybridization combined with immunofluorescence (FISH-IF) allows detection and localization of RNA–protein complexes at the cellular level in fixed cells or tissues (12,122,123). FISH-IF utilizes fluorescently labeled nucleic acid probes complementary to the desired RNA and antibodies detecting the protein of interest. Recent advances in RNA FISH, such as single-molecule FISH (smFISH), allow for the analysis of proteins co-localizing with individual RNA molecules in single cells (123). However, because of the limitations of optical resolution, the FISH-IF technique does not provide evidence for the direct interaction between proteins and RNA. Recently, various methods based on adaptation of proximity ligation assay (PLA) have been successfully used for more precise determination of RNA–protein interactions *in situ* (124–126). PLA uses proximity probes—oligonucleotides attached to antibodies against two epitopes—to guide the formation of circular DNA strands when bound in close proximity (<40 nm). This newly created DNA molecule can serve as template for

localized rolling-circle amplification (RCA), which results in coiled-single stranded DNA. These PLA products can be easily detected by hybridizing complementary fluorescently labeled oligonucleotides (127–129). Among PLA-based microscopic analyses that allow visualization and quantification of RNA–protein interactions *in situ* in single cells with single-interaction sensitivity is a method that combines peptide-modified, multiply-labeled tetravalent RNA imaging probes (MTRIPs) with proximity ligation and RCA (124,130). Fluorescence resonance energy transfer (FRET) analysis enables direct visualization of RNA–protein complexes in live and fixed cells. FRET is a photophysical phenomenon in which energy transfer between two adjacent fluorophores, donor and acceptor, are strongly dependent on the distance between these molecules (102,131). Therefore, FRET is observed only when the donor- and acceptor-labeled molecules are in close proximity (typically, 2–10 nm) and can directly indicate the interaction. Typically, RNA is stained with SytoxOrange or labeled with MS2, while the RNA-binding protein is tagged with a fluorescence protein (130,132,133). Using FRET, it is possible to analyze temporal and spatial association of proteins with RNA inside cells. Recent advances in developing the CRISPR/RdCas9 system have resulted in the ability to track the dynamics of RNA–protein interactions in live cells (85). An indisputable power of this method is the recognition of unmodified, endogenous RNAs. Most techniques that are used for live-cell RNA imaging require the incorporation of exogenous tags that might affect RNA folding, localization or stability (134). The CRISPR/RdCas9 system has the potential to be highly selective towards its target transcripts, as suggested by smFISH-IF results (135).

## METHODS FOR IDENTIFYING PROTEINS THAT BIND TO EXPANDED RNA REPEATS

Most of the previously described RNA affinity purification approaches are used to study various types of RNA–protein interactions that are essential for the normal regulation of RNA biogenesis, RNA processing and stability, or mRNA translation (Supplementary Table S1). However, in some diseases, abnormal RNA–protein interactions occur, which disrupt the physiological function of bound proteins. Such diseases include dominantly inherited disorders that are associated with the presence of long repeat expansions in the non-coding or coding regions of individual genes (Figure 4). In this case, the interaction between proteins and mutant RNA often results in the immobilization of these proteins in specific structures that are termed RNA foci, which are the pathogenic hallmarks of this type of disease (136–139). The identification of proteins that are bound to CUG repeats in myotonic dystrophy type 1 (DM1) led to the development of the RNA *gain-of-function* model, which posits that expanded repeats sequester RNA-binding proteins from their normal function (140–143). Since then, proteins that are trapped not only by CUG repeats but also by CAG, CGG, CCUG, AUUCU and GGGGCC repeats have been the subject of many studies, the results of which have broadened our understanding of the pathomechanisms of these disorders (Figure 4) (Table 1 and Supplementary Table S2). These abnormal RNA–protein interactions affect the alter-



**Figure 4.** Schematic representation of the RNA *gain-of-function* mechanism in the selected repeat expansion disorders. Upper: Localization of expanded simple repeats in disease genes. Middle: Transcripts containing repeat expansions function as pathogenic agents via the sequestration of specific RNA-binding proteins, resulting in their impaired cellular function. Lower: Methods applied to capture and identify proteins that are associated with different expanded repeats.

native splicing of specific pre-mRNAs (144), alter the use of alternative polyadenylation sites of a number of mRNAs (145), change nuclear transport and export (146,147), affect translation (148), induce nucleolar stress (94), and dysregulate miRNA processing (149). What makes these expanded repeats so attractive for cellular proteins? The structures that are formed by the expanded repeat RNAs likely trigger protein recruitment. Depending on the type of expansion involved, mutant transcripts can adopt *in vitro* stable hairpin or G-quadruplex structures (94,150,151). Despite intensive studies on RNA-mediated toxicity in these disorders, many questions regarding the molecular disease-causing mechanisms remain unanswered. Determination of the complete protein composition of these specific RNP complexes will help to elucidate these RNA *gain-of-function* mechanisms. Proteins that are recruited into mutant RNA

foci may be attractive therapeutic targets for these as-yet incurable disorders. In the following chapter, we briefly describe approaches that have been used thus far to identify proteins that bind to various types of expanded repeats.

### CUG/CCUG-binding proteins

CTG repeat expansions (50 to >3500 repeats) in the 3' UTR of the *DMPK* gene and CCTG repeat expansions (75 to ~11 000 repeats) in the first intron of the *ZNF9* gene are causative agents of myotonic dystrophy type 1 (DM1) and myotonic dystrophy type 2, respectively (152–154). Proteins that bind to CUG/CCUG repeat expansions were initially identified using *in vitro* biochemical assays, such as EMSA and the UV crosslinking assay, followed by immunoblotting. These analyses relied on the assumption that proteins that can bind dsRNA or proteins that localize to the

**Table 1.** Proteins interacting with expanded repeat RNAs identified by RNA pull-down assays combined with WB or MS

Type of repeats	RNA affinity capture	Extract source	Identified protein	Validation method	Reference
<b>CAG</b>					
(CAG) <sub>15, 128</sub> Biotin-tagged RNA	RNA pull-down and WB	Cytoplasmic extract from human brain	PKR	IHC	(172)
(CAG) <sub>27, 78</sub> S1 aptamer-tagged RNA	RNA pull-down and WB	Transgenic flies	NCL	<i>in vitro</i> RIP	(173)
(CAG) <sub>20, 51</sub> Biotin-tagged RNA	RNA pull-down and WB	Whole-cell lysate from HeLa cells	MID1 complex (MID1, S6K, PP2Ac)	Pull-down followed by WB	(148)
<b>CGG</b>					
(CGG) <sub>105</sub> Biotin-tagged RNA	RNA pull-down and MS	Cytoplasmic extract from mouse cerebellum	hnRNP A2/B1, Pur $\alpha$	Pull-down followed by WB, EMSA, RIP	(181)
(CGG) <sub>60</sub> Biotin-tagged RNA	RNA pull-down and MS	Nuclear extract from mouse brain and COS7 cells	Sam68 and other 37 proteins identified	FISH-IF	(183)
(CGG) <sub>20, 60, 100</sub> Biotin-tagged RNA	RNA pull-down and MS	Nuclear extract from mouse brain	DROSHA, DGCR8 and other 30 proteins identified	Pull-down followed by WB, FISH-IF, EMSA, UV crosslinking	(149)
<b>CUG</b>					
(CUG) <sub>85</sub> Biotin-tagged RNA	RNA pull-down and MS	Nuclear extracts from HeLa cells	hnRNP H	FISH-IF, UV crosslinking	(103)
(CUG) <sub>95</sub> Biotin-tagged RNA	RNA pull-down and MS	Nuclear extracts from HeLa cells and myoblasts and myotubes from C2C12	p68/DDX5 and other 100 proteins identified	FISH-IF, EMSA	(104)
<b>AUUCU</b>					
(AUUCU) <sub>500</sub> Biotin-tagged RNA	RNA pull-down and MS	Nuclear extract from mouse brain	hnRNP K	Pull-down followed by WB, RIP, FISH-IF	(185)
<b>GGGGCC</b>					
(GGGGCC) <sub>23</sub> Biotin-tagged RNA	RNA pull-down and MS	Nuclear extract from HEK293 cells	hnRNP A3, hnRNP A1 and other 20 proteins identified	Pull-down followed by WB	(194)
(GGGGCC) <sub>10</sub> Biotin-tagged RNA	RNA pull-down and MS	Whole-cell lysate from mouse spinal cord	Pur $\alpha$ , Pur $\beta$ , Pur $\gamma$	Pull-down followed by WB, FBA, RIP	(195)
(GGGGCC) <sub>30</sub> Biotin-tagged RNA	RNA pull-down and MS	Nuclear extract from mouse brain	hnRNP H1, hnRNP H2 and other 30 proteins identified	FISH-IF	(196)
(GGGGCC) <sub>72, 48</sub> Biotin- and S1 aptamer-tagged RNA	RNA pull-down and WB	Nuclear extract of SH-SY5Y cells and rat brain cortex	hnRNP-H	FISH-IF	(190)
(GGGGCC) <sub>4</sub> Biotin-tagged RNA	RNA pull-down and MS	HEK293T (SILAC)	NCL, hnRNP U and other 81 proteins identified	Pull-down followed by WB, FISH-IF	(94)
(GGGGCC) <sub>5</sub> Biotin-tagged RNA	RNA pull-down and MS	Whole-cell lysate and nuclear extracts from SH-SY5Y cells, total extracts from human cerebellum	ALYREF, SRSF1, SRSF2, hnRNP A1, hnRNP H1/F and other 103 proteins identified	FISH-IF, UV crosslinking	(197)
(GGGGCC) <sub>6,5</sub> 5'Cy5-labeled RNA	Proteome array	<i>in vitro</i>	ADARB2 and other 19 proteins identified	FISH-IF, RIP, EMSA	(199)
(GGGGCC) <sub>31</sub> Biotin-tagged RNA	RNA pull-down and MS	Whole-cell lysate from mouse brain and spinal cord	hnRNP H, eIF2 $\alpha$ , eIF2 $\beta$ , RAX, ILF3	Pull-down followed by WB, FISH-IF	(198)

IHC, immunohistochemistry; WB, western blotting; FISH, fluorescence *in situ* hybridization, IF, immunofluorescence; RIP, RNA immunoprecipitation; EMSA, electrophoretic mobility shift assay; FBA, filter binding assay.

nucleus potentially interact with expanded CUG/CCUG repeats. The potential interactors were subsequently assayed for their co-localization with expanded RNAs using FISH-IF (Supplementary Table S2). In the case of EMSA analysis, *in vitro*-transcribed and radiolabeled expanded CUG/CCUG RNA was incubated with whole cell protein extracts or with a set of pre-defined, purified recombinant RBPs. The first approach identified the alternative splicing regulator CUG-BP (CUG triplet repeat, RNA-binding protein 1) as well as two CCUG-interacting multiprotein complexes: the 20S catalytic core complex of the proteasome and the CUG-BP1-eIF2 complex, which is involved in translational regulation (140,141,155). Using the second approach, CUG-binding properties were determined for mETR3 (muscarinic acetylcholine receptor M3) and PKR (interferon-induced, double-stranded RNA-activated protein kinase) (156,157). The UV crosslinking of whole cell extracts with radiolabeled RNA is also commonly used to detect protein-RNA interactions. Using this method, alternative splicing regulator MBNL1 (muscleblind-like protein 1) was identified as a protein that binds to CUG repeats (142). Structural studies showed that at least 20 CUG repeats are needed for the RNA to acquire the RNA hair-

pin secondary structure that functions as a sequestration trigger and is necessary for MBNL1 binding (143,156,158). FISH-IF not only confirmed MBNL1 co-localization with expanded CUG repeats but was also the first approach to show that MBNL1 possibly interacts with expanded CCUG RNAs (159). Interestingly, FISH co-localization of MBNL1 with nuclear CUG RNA inclusions was also demonstrated for the rarely occurring CUG repeat-related disorders spinocerebellar ataxia type 8 (SCA8) (160) and Huntington's Disease-like 2 (HDL2) (161). Using additional *in vivo* and *in vitro* approaches, it was shown that sequestration of MBNL1 by CUG/CCUG repeats leads to abnormalities in alternative splicing (144,162), mRNA localization and transport (163), stability (164), and microRNA biogenesis (165); these abnormalities which affect many cellular functions, resulting in disease. RIP has also been used to identify proteins that interact with expanded CUG repeats. Using RIP, interactions between expanded CUG repeats containing transcript and Staufen 1 (which is involved in mRNA transport, stability, and translation) (166), DDX6 (ATP-dependent RNA helicase) (167), and two transcription factors, Sp1 and RAR $\gamma$  (168), were examined. Only two studies attempted to comprehen-

sively examine the complement of proteins that bind to expanded CUG repeats (Table 1). In both studies, RNA affinity chromatography with *in vitro* transcribed and biotinylated CUG RNAs combined with MS analysis was performed to identify novel factors that bind these repeats (103,104). Among the newly identified proteins were two helicases, p68 (DDX5) and p72 (DDX17), which play a role in remodeling of RNA and RNA-complexes, and a key regulator of mRNA metabolism, hnRNP H. Some of these proteins were further shown to affect nuclear retention of mutant RNA or to favor MBNL1 binding to the CUG/CCUG repeats (103,104).

### CAG-binding proteins

Expanded CAG repeats that are localized in the coding sequences of nine functionally unrelated genes are molecular triggers for a group of neurodegenerative disorders termed polyglutamine (polyQ) diseases. There are at least nine polyQ disorders, including Huntington's disease (HD), spinocerebellar ataxia (SCA) types 1, 2, 3, 6, 7 and 17, dentatorubral-pallidoluysian atrophy (DRPLA) and spinal and bulbar muscular atrophy (SBMA). These diseases take their name from mutant proteins containing long polyQ tracts. In the most common disorders (HD and SCA3), the expression of at least 36 CAG repeats that are localized in the first exon of the *HTT* gene and at least 60 CAG repeats that are localized in the 10th exon of the *ATXN3* gene is sufficient to cause the pathogenic effects, respectively (169). Expanded CAG-repeat binding proteins were identified using similar approaches to those used in studies on CUG/CCUG-interacting proteins (Table 1 and Supplementary Table S2), which test the binding of extended CAG-repeat RNA to known RNA-binding proteins (142,156,170). The first strategy that attempted to identify proteins with CAG-binding properties was the use of an EMSA and a UV-crosslinking assay with cytoplasmic protein extracts from various human tissues (171). In this study, two CAG-repeat binding proteins (63 and 49 kDa) were identified. However, no further analysis allowing the precise identification of these proteins was performed. Another frequently used *in vitro* approach to identify CAG-binding proteins is the use of RNA pull-down assays followed by immunoblotting with antibodies that are specific for known RBPs (Table 1). The resulting *in vitro* transcribed RNAs containing different-length CAG repeats (normal or mutant) were immobilized on agarose beads either covalently or using biotin or S1 aptamer, followed by incubation with protein extracts obtained from different sources. The use of this method confirmed that PKR (protein kinase R) (172), the MID1 complex (which regulates translation) (148), and nucleolin (which regulates rRNA transcription) (173) are all CAG-repeat binding proteins. Further experiments showed that the interaction between nucleolin and expanded CAG RNA triggers the perturbation of rRNA transcription and, consequently, nucleolar stress and apoptosis (173,174). Moreover, the binding of the MID1 complex to the expanded CAG RNA results in the upregulated translation of mutant *HTT* transcripts, thereby leading to the overproduction of aberrant proteins (148). CAG-binding proteins were also identified using Y3H assays. Us-

ing this method, the CAG RNA-binding properties of two proteins, MBNL1 and PKR, were shown (170). What is the role of CAG repeat protein interactions in the pathogenesis of polyQ diseases? Co-localization of MBNL1 with RNA containing expanded CAG repeats results in aberrant alternative splicing similar to CUG/CCUG repeats (175). Proteins that may interact with expanded CAG repeats were also identified by FISH-IF. This method, however, was only used to identify potential interactions of CAG with MBNL1 (176). Another approach used to detect proteins with CAG-binding properties was RIP. In this case, interactions between candidate proteins and transcripts containing expanded repeats were studied. Using this approach, U2AF65, which regulates nuclear export of RNA (146), and SRSF6, which is involved in regulation of splicing and translation (177), were identified. Interestingly, SRSF6 has been selected to the analysis on the basis of bioinformatics predictions, which revealed binding site of SRSF6 with expanded CAG repeats (177). For U2AF65 and nucleolin, immunoprecipitation using GST-tagged fusion proteins was performed to identify direct interactions with expanded CAG RNA (146,173).

### CGG-binding proteins

The pathological expansion of CGG repeats in the 5' UTR region of the *FMRI* gene causes one of two different disorders, depending on the length of the repeats: Fragile X-associated tremor/ataxia syndrome (FXTAS) is caused by a premutation in which the expansion is between 55 and 200 CGG, whereas Fragile X syndrome (FXS) occurs when the number of CGG repeats exceed 200 triplets (178,179). The first attempt to identify potential CGG-binding proteins started with the purification of ubiquitin-positive *FMR1*-mRNA inclusions (the neuropathological hallmark of FXTAS), followed by a detailed protein characterization using MS analysis (Supplementary Table S2). This indirect analysis showed that at least two RNA-binding proteins, heterogeneous nuclear ribonucleoprotein A2 (hnRNP-A2) and MBNL1, were present in these inclusions (180). To identify proteins that bind specifically to CGG repeats, *in vitro* transcribed, biotinylated, expanded CGG-repeat RNA was used as bait (181). Putative CGG repeat-binding proteins were purified by binding to streptavidin magnetic beads, and protein identification was performed using MS analysis. Among the proteins selected for further studies were the known transcription and RNA transport regulators, Pur $\alpha$  and hnRNP A2/B1 (Table 1). The MS identification of these proteins as potential CGG-repeat-binding proteins was further confirmed using other *in vitro* biochemical assays (EMSA and immunoprecipitation with (CGG)<sub>90</sub>-eGFP assays). Moreover, Pur $\alpha$  was shown to be present in the ubiquitin-positive inclusions of FXTAS patient brains and in a *Drosophila* model of FXTAS (181). Furthermore, a medium-throughput *Drosophila* genetic screen was also used to identify potential CGG-binding proteins (182). Independent of MS analysis, hnRNP A2/B1 and CUG-BP were also shown to interact with expanded CGG repeats, as their overexpression suppresses the CGG-mediated eye toxicity in *Drosophila* (182). Other reports using *in vitro* transcribed and biotinylated RNA as bait to identify CGG-

binding proteins by MS have also been published (Table 1). In 2010, Sellier *et al.* used mouse brains or COS7 cell protein extracts to identify over 20 proteins that specifically interacted with expanded CGG repeat RNA (183). In addition, FISH-IF studies revealed that only MBNL1 and hnRNP-G co-localize with mutant RNA foci in COS7 cells expressing these expanded repeats. Surprisingly, further studies demonstrated that recruitment of MBNL1 and hnRNP-G to the CGG foci required the presence of an alternative splicing regulator, Sam68, which itself does not bind directly to CGG repeats (183). Moreover, using normal (CGG)<sub>20</sub> and mutant (CGG)<sub>60</sub> RNA as baits followed by MS analysis, the double-stranded RNA-binding protein DGCR8 that exclusively binds to mutant RNA was identified (Table 1) (149). Further detailed studies showed that DGCR8 co-localizes to aggregates formed by mutant (CGG)<sub>60</sub>, resulting in the partial sequestration of DGCR8 and its partner DROSHA, which interestingly was not identified in MS analysis. As a consequence, the processing of miRNAs is reduced, resulting in decreased levels of mature miRNAs in neuronal cells expressing expanded CGG repeats and in brain tissues from patients with FXTAS (149).

#### AUUCU-binding proteins

The ATTCT pentanucleotide repeat expansion in intron 9 of the *ATX10* gene has been associated with spinocerebellar ataxia type 10 (SCA10). Up to 15 repeats occur in healthy individuals, but extended repeats of 800–4500 AUUCU are found in patients suffering from SCA10 (184). To date, only one study identified proteins that bound specifically to expanded AUUCU repeats (185). In this report, purification of proteins bound to *in vitro* transcribed and biotinylated expanded RNA followed by MS analysis identified hnRNP K as a potential interacting protein (Table 1). Further analysis using RIP confirmed the presence of intron 9 of *ATX10* after hnRNP K pull-down. Additionally, hnRNP K showed co-localization with AUUCU expanded foci in SCA10 fibroblasts and transgenic mouse brain tissues (185,186). Furthermore, it was demonstrated that sequestration of hnRNP induces alternative splicing defects in certain genes and induces caspase-mediated apoptosis in SCA10 fibroblasts (185,186).

#### GGGGCC-binding proteins

The expanded GGGGCC repeats located in the gene *C9ORF72* have been associated with amyotrophic lateral sclerosis (c9ALS) (187,188). Less than 30 repeats are found in healthy people, while expanded repeats in the range of 500 to thousands are found in people affected with c9ALS, with an increased number of repeats associated with a more severe phenotype (188–190). The first approaches focused on the identification of proteins interacting with GGGGCC repeats were based on bioinformatics prediction followed by biochemical verification methods (Supplementary Table S2). The assumption that ASF2/SF2 and hnRNP A1 interact with GGGGCC repeats was confirmed by EMSA (191,192). Additionally, FISH-IF has been used to show co-localization of SC35, SF2 and GGGGCC repeats (190,193). As RNA *gain-of-function* model has been implicated in

c9ALS quite recently (188), the well-established global approaches were predominantly used to identify GGGGCC repeats-binding proteins among which was a pull-down method with biotinylated RNA and MS analysis (Table 1) (94,190,194–198). Alternatively, instead of biotin, the S1 aptamer has been added in a template DNA prior to *in vitro* transcription (190). Because GGGGCC repeats are difficult to synthesize due to high GC content, nonpathological repeat tracts of GGGGCC were predominantly used as bait in the pull-down analyses. Shorter repeats are believed to have a similar ability and specificity to bind selective proteins as longer ones due to the existence of the same repetitive sequence motif and formation of similar tertiary structure (194). In nearly all studies analyzing the GGGGCC-interacting proteome, MS was used to identify RNA-binding proteins. A significant improvement in this technology was the introduction of the SILAC labeling technique, which allows for a comparative quantification of proteins that are bound to different RNAs (94). Proteome arrays have also been used to identify the proteins that interact with GGGGCC RNA expansions (Table 1) (199). The authors used nearly two-thirds of the annotated human proteome as yeast-expressed, 16,368 full-length ORFs with N-terminal GST-His $\times$ 6 fusion proteins on a chip to which an RNA probe was hybridized. The disadvantages of this method are the limited number of proteins that can be bound on the chip and the limitations relating to already identified, known proteins. Moreover, only direct interactions with single proteins can be identified using the proteome array method because no protein complexes exist in that experimental system. Using this approach, ADARB2 (adenosine deaminase RNA-specific B2, a known RNA-binding protein and a member of the ADAR family, members of which are involved in RNA editing) was shown to bind to GGGGCC repeats (199). The hypothesis regarding the sequestration of proteins by expanded repeats was extensively studied for sense transcripts containing GGGGCC repeats. However, c9ALS is an example of a disease in which bidirectional transcription occurs and results in the production of both expanded sense GGGGCC tracts and antisense CCCCCG tracts. For antisense transcripts, it was demonstrated that hnRNP-K and PCBP2 (poly(rC)-binding protein) preferentially bind antisense CCCCCG4 over sense GGGGCC4 transcripts (94), and the association of these proteins with c9ALS pathomechanisms was further evaluated (197). In another study, antisense RNA foci were demonstrated to co-localize with SRSF2, hnRNP A1, hnRNP H/F, ALYREF and hnRNP K (200). Moreover, the spectrum of proteins that bind to GGGGCC repeats in a structure conformation-dependent manner was also analyzed. NCL (nucleolin) and hnRNP-U preferentially bind a G-quadruplex, whereas hnRNP-F and ribosomal protein RPL7 do not distinguish between hairpin and quadruplex structures. As a result, nucleolar stress is observed in c9ALS patients (94). Global analyses of proteins that bind to GGGGCC repeats showed that most of these proteins possess an RNA recognition motif (RRM) within the RNA-binding domain (RBD) or prion-like domain (PrLD). This RRM allows for RNA–protein interactions, whereas the PrLD allows for protein aggregation. Gene ontology classifications revealed that differ-

ent classes of proteins recognize GGGGCC repeats: pre-mRNA splicing factors (FUS, EWSR1, hnRNP A1, hnRNP A2B1, SRSF1, SRSF2, SRSF3 and TAF15), processing, stability (SAFB2) (194,197,198), mRNA export adaptors (197), transcription activator proteins (Pur $\alpha$ , Pur $\beta$  and Pury) (195,198), heterogeneous nuclear ribonucleoproteins (190,194,198), helicases (DDX21, DHX15, and DHX30) (194) and interleukins (ILF2) (194) (Table 1). The sequestration of proteins by RNA reduces the amount of sequestered protein in its native pathway and results in the dysregulation of cell functioning.

## CONCLUSIONS AND FUTURE PERSPECTIVES

Although many different approaches have been developed to immobilize specific RNAs to affinity matrices and to identify, confirm and further examine RNA–protein interactions, serious challenges remain to be addressed. Although significant advances in mass spectrometry and peptide separation methods have enabled the reliable identification of these interactions, the capture and purification of proteins bound to specific RNAs remain great challenges. Therefore, the RNA affinity capture method used should be carefully chosen depending on the biological question that needs to be answered.

RNA-binding proteins are highly versatile factors that play crucial roles in a variety of cellular processes. However, the complement of proteins that associate with any given RNA is largely unknown, and this knowledge gap needs to be filled. Only recently, using the interactome capture strategy combined with a robust MS analysis, a comprehensive identification of RBPs associated with a total pool of mRNAs has been reported (73,74,79,80). However, these studies were performed using a limited number of cell types or tissues. Surprisingly, it was shown that over 50% of the identified RBPs lack typical RNA-binding domains, which causes difficulties in identifying these proteins using computational methods. Thus, it is very likely that numerous other RNA-binding proteins await identification in other biological sources. Furthermore, researchers know very little regarding the repertoire of proteins with which the more recently discovered lncRNAs and circular RNAs may interact (201–204). Some of these proteins, which have already been shown to interact with lncRNAs, do not resemble known RNA-binding proteins. Thus it is difficult to anticipate the structure and function of these RNA–protein complexes. Issues regarding the functioning of noncoding RNAs are likely to remain the focus of research in coming years.

The complement of proteins that are bound to a specific RNA that control its ‘life’ inside cells changes in a temporal and spatial manner. These changes in protein composition, stoichiometry and modification drive RNA molecules throughout different stages of the RNA cycle - transcription, various co-transcriptional maturation events, transport and nuclear export, translation, and finally, RNA degradation. Moreover, because these events are highly regulated by both intracellular and environmental signals, research in the field of RNA biology needs to focus on elucidating the spatial and temporal dynamics of RNA–protein networks. Thus far, predominantly static analyses of the physical association between RBP and specific RNAs

have been conducted. To characterize the dynamic RNA-binding proteome, recent advances in RNA-centric, high-throughput quantitative proteomics strategies can be applied (46,89,205). The application of these methods will allow researchers to ‘freeze’, purify and identify RBP binding to specific RNAs under different conditions; e.g. during different stages of the RNA cycle, during cell cycle progression, and in cells that have been treated with different extracellular stimuli (stress conditions). Moreover, because many RNA-binding proteins interact either weakly or transiently with their target RNA throughout its ‘life’ cycle, it is crucial to streamline existing methods or to develop novel more accurate, high-throughput technologies that exploit crosslinking-MS (206,207) and next-generation sequencing (18,121,208). Thus far, among the major caveats of RNA-centric methods is their inability to explore RNA-binding proteins that associate with low abundance transcripts inside living cells (*in vivo* purification strategies). Therefore, the development of novel approaches is needed to isolate endogenously expressed, unmodified RNAs with increased specificity, together with their interacting proteins. This goal would likely be achieved by further developing recently described methods that rely on CRISPR/RdCas9 technology (84,85). Unlike the most commonly used methods for RNA–protein complex isolation, CRISPR/RdCas9 technology offers an incomparable means to selectively capture and identify endogenously expressed specific RNAs together with their interacting proteins without interfering with its RNA ‘life’ cycle. With all its advantages, this technology has still some challenges that need to be addressed such as efficient delivery of the system components to the cells, improvement of PAMmer and gRNA design and most importantly the decrease of CRISPR/RdCas9-mediated off-target effects. Some disadvantages of the CRISPR/RdCas9 technology might be likely circumvented by using the recently described CRISPR/C2c2 system from the bacterium *Leptotrichia shahii*, which, targets RNA in its native form (209).

In this review, we also highlighted the importance of studying abnormal RNA–protein complexes that accompany a class of neurodegenerative diseases; e.g. DM1, DM2, c9ALS, FXTAS, HD and a number of SCAs. As current studies show, specific RNAs containing an expanded simple repeat tract can function in pathogenesis as trapping agents for various RNA-binding proteins. Once captured, these proteins cannot properly fulfill their normal cellular roles in processes such as alternative splicing, RNA transport and nuclear export, translation, and microRNA biogenesis. The described expanded repeat-binding proteins might be only the ‘tip of the iceberg’. Most studies to date have been performed using outdated *in vitro* purification approaches due to the problems associated with capture and identification. Therefore, there is a great need to apply recently developed cutting-edge strategies to identify novel proteins that are trapped by these expanded RNA repeat tracts to better understand the mechanisms of pathogenesis and to develop treatments for these disorders.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Science Centre [2012/06/A/NZ1/00094 to W.J.K., 2014/15/B/NZ1/01880 to W.J.K., 2015/19/B/NZ2/02453 to W.J.K.; 2015/19/D/NZ5/02183 to M.J.]; Polish Ministry of Science and Higher Education, under the KNOW program for years 2014–2019. Funding for open access charge: National Science Centre – Poland.

*Conflict of interest statement.* None declared.

## REFERENCES

- Glisovic, T., Bachorik, J.L., Yong, J. and Dreyfuss, G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.*, **582**, 1977–1986.
- Kishore, S., Lubner, S. and Zavolan, M. (2010) Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. *Brief. Funct. Genomics*, **9**, 391–404.
- Licatalosi, D.D. and Darnell, R.B. (2010) RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.*, **11**, 75–87.
- Lukong, K.E., Chang, K., Khandjian, E.W. and Richard, S. (2008) RNA-binding proteins in human genetic disease. *Trends Genet.*, **24**, 416–425.
- Cooper, T.A., Wan, L. and Dreyfuss, G. (2009) RNA and disease. *Cell*, **136**, 777–793.
- Darnell, R.B. (2010) RNA regulation in neurologic disease and cancer. *Cancer Res. Treat.*, **42**, 125–129.
- Kechavarzi, B. and Janga, S.C. (2014) Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biol.*, **15**, R14.
- Keller, B.A., Volkening, K., Droppelmann, C.A., Ang, L.C., Rademakers, R. and Strong, M.J. (2012) Co-aggregation of RNA binding proteins in ALS spinal motor neurons: evidence of a common pathogenic mechanism. *Acta Neuropathol.*, **124**, 733–747.
- Dong, Y., Yang, J., Ye, W., Wang, Y., Ye, C., Weng, D., Gao, H., Zhang, F., Xu, Z. and Lei, Y. (2015) Isolation of endogenously assembled RNA–protein complexes using affinity purification based on streptavidin aptamer S1. *Int. J. Mol. Sci.*, **16**, 22456–22472.
- Habjan, M., Hubel, P., Lacerda, L., Benda, C., Holze, C., Eberl, C.H., Mann, A., Kindler, E., Gil-Cruz, C., Ziebuhr, J. et al. (2013) Sequestration by IFIT1 Impairs Translation of 2′O-unmethylated Capped RNA. *PLoS Pathog.*, **9**, e1003663.
- Gagnon, K.T. and Maxwell, E.S. (2011) Electrophoretic mobility shift assay for characterizing RNA–protein interaction. *Methods Mol. Biol.*, **703**, 275–291.
- Chaumeil, J., Augui, S., Chow, J.C. and Heard, E. (2008) Combined immunofluorescence, RNA fluorescent in situ hybridization, and DNA fluorescent in situ hybridization to study chromatin changes, transcriptional activity, nuclear organization, and X-chromosome inactivation. *Methods Mol. Biol.*, **463**, 297–308.
- Zielinski, J., Kilk, K., Peritz, T., Kannanayakal, T., Miyashiro, K.Y., Eiriksdóttir, E., Jochems, J., Langel, U. and Eberwine, J. (2006) In vivo identification of ribonucleoprotein-RNA interactions. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 1557–1562.
- Faoro, C. and Ataíde, S.F. (2014) Ribonomic approaches to study the RNA-binding proteome. *FEBS Lett.*, **588**, 3649–3664.
- Tacheny, A., Dieu, M., Arnould, T. and Renard, P. (2013) Mass spectrometry-based identification of proteins interacting with nucleic acids. *J. Proteomics*, **94**, 89–109.
- Koh, Y.Y. and Wickens, M. (2014) Determining the RNA specificity and targets of RNA-binding proteins using a three-hybrid system. *Methods Enzymol.*, **539**, 163–181.
- König, J., Zarnack, K., Luscombe, N.M. and Ule, J. (2011) Protein-RNA interactions: new genomic technologies and perspectives. *Nat. Rev. Genet.*, **13**, 77–83.
- Ascano, M., Hafner, M., Cekan, P., Gerstberger, S. and Tuschl, T. (2012) Identification of RNA–protein interaction networks using PAR-CLIP. *Wiley Interdiscip. Rev. RNA*, **3**, 159–177.
- Sutandy, F.X.R., Hsiao, F.S.-H. and Chen, C.-S. (2015) High throughput platform to explore RNA–protein interactomes. *Crit. Rev. Biotechnol.*, **8551**, 1–9.
- Tome, J.M., Ozer, A., Pagano, J.M., Gheba, D., Schroth, G.P. and Lis, J.T. (2014) Comprehensive analysis of RNA–protein interactions by high-throughput sequencing-RNA affinity profiling. *Nat. Methods*, **11**, 683–688.
- Schroeder, R., Barta, A. and Semrad, K. (2004) Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.*, **5**, 908–919.
- Doetsch, M., Schroeder, R. and Fürtig, B. (2011) Transient RNA–protein interactions in RNA folding. *FEBS J.*, **278**, 1634–1642.
- Miniard, A.C., Middleton, L.M., Budiman, M.E., Gerber, C.A. and Driscoll, D.M. (2010) Nucleolin binds to a subset of selenoprotein mRNAs and regulates their expression. *Nucleic Acids Res.*, **38**, 4807–4820.
- Buratti, E., Dörk, T., Zuccato, E., Pagani, F., Romano, M. and Baralle, F.E. (2001) Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J.*, **20**, 1774–1784.
- Patel, G.P., Ma, S. and Bag, J. (2005) The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex. *Nucleic Acids Res.*, **33**, 7074–7089.
- Talukdar, I., Sen, S., Urbano, R., Thompson, J., Yates, J.R. and Webster, N.J.G. (2011) HnRNP A1 and hnRNP F modulate the alternative splicing of exon 11 of the insulin receptor gene. *PLoS One*, **6**, e27869.
- Choudhury, N.R., de Lima Alves, F., de Andrés-Aguayo, L., Graf, T., Cáceres, J.F., Rappsilber, J. and Michlewski, G. (2013) Tissue-specific control of brain-enriched miR-7 biogenesis. *Genes Dev.*, **1**, 24–38.
- Liu, D.G. and Sun, L. (2005) Direct isolation of specific RNA-interacting proteins using a novel affinity medium. *Nucleic Acids Res.*, **33**, 1–5.
- Michlewski, G. and Cáceres, J.F. (2010) RNase-assisted RNA chromatography. *RNA*, **16**, 1673–1678.
- Choudhury, N.R., Nowak, J.S., Zuo, J., Rappsilber, J., Spoel, S.H. and Michlewski, G. (2014) Trim25 Is an RNA-Specific Activator of Lin28a/TuT4-Mediated Uridylation. *Cell Rep.*, **9**, 1265–1272.
- Nowak, J.S., Choudhury, N.R., de Lima Alves, F., Rappsilber, J. and Michlewski, G. (2014) Lin28a regulates neuronal differentiation and controls miR-9 production. *Nat. Commun.*, **5**, 3687.
- Grabowski, P.J. and Sharp, P.A. (1986) Affinity chromatography of splicing complexes: U2, U5, and U4 + U6 small nuclear ribonucleoprotein particles in the spliceosome. *Science*, **233**, 1294–1299.
- Goracznik, R. and Gunderson, S.I. (2008) The regulatory element in the 3′-untranslated region of human papillomavirus 16 inhibits expression by binding CUG-binding protein 1. *J. Biol. Chem.*, **283**, 2286–96.
- Sharma, S. (2008) Isolation of a sequence-specific RNA binding protein, polypyrimidine tract binding protein, using RNA affinity chromatography. *Methods Mol. Biol.*, **488**, 1–8.
- Paingankar, M.S. and Arankalle, V.A. (2015) Identification and characterization of cellular proteins interacting with Hepatitis E virus untranslated regions. *Virus Res.*, **208**, 98–109.
- O’Connor, C.M., Lai, C.K. and Collins, K. (2005) Two purified domains of telomerase reverse transcriptase reconstitute sequence-specific interactions with RNA. *J. Biol. Chem.*, **280**, 17533–17539.
- Piskounova, E., Viswanathan, S.R., Janas, M., LaPierre, R.J., Daley, G.Q., Sliz, P. and Gregory, R.I. (2008) Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J. Biol. Chem.*, **283**, 21310–21314.
- Hirsch, J.D., Eslamizar, L., Filanoski, B.J., Malekzadeh, N., Haugland, R.P., Beechem, J.M. and Haugland, R.P. (2002) Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: Uses for protein labeling, detection, and isolation. *Anal. Biochem.*, **308**, 343–357.
- Bardwell, V.J. and Wickens, M. (1990) Purification of RNA and RNA–protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res.*, **18**, 6587–6594.
- Keryer-Bibens, C., Barreau, C. and Osborne, H.B. (2008) Tethering of proteins to RNAs by bacteriophage proteins. *Biol. Cell*, **100**, 125–138.



41. Carey, J., Cameron, V., de Haseth, P.L. and Uhlenbeck, O.C. (1983) Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. *Biochemistry*, **22**, 2601–2610.
42. Das, R., Zhou, Z. and Reed, R. (2000) Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol. Cell*, **5**, 779–787.
43. Zhou, Z., Licklider, L.J., Gygi, S.P. and Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419**, 182–185.
44. Slobodin, B. and Gerst, J.E. (2011) RaPID: an aptamer-based mRNA affinity purification technique for the identification of RNA and protein factors present in ribonucleoprotein complexes. *Methods Mol. Biol.*, **714**, 387–406.
45. Leonov, A.A., Sergiev, P. V., Bogdanov, A.A., Brimacombe, R. and Dontsova, O.A. (2003) Affinity purification of ribosomes with a lethal G2655C mutation in 23S rRNA that affects the translocation. *J. Biol. Chem.*, **278**, 25664–25670.
46. Tsai, B.P., Wang, X., Huang, L. and Waterman, M.L. (2011) Quantitative profiling of in vivo-assembled RNA–protein complexes using a novel integrated proteomic approach. *Mol. Cell. Proteomics*, **10**, 1–15.
47. Lim, F., Downey, T.P. and Peabody, D.S. (2001) Translational repression and specific RNA binding by the coat protein of the pseudomonas phage PP7. *J. Biol. Chem.*, **276**, 22507–22513.
48. Collins, K. and Hogg, J.R. (2007) RNA-based affinity purification reveals 75K RNPs with distinct composition and regulation. *RNA*, **13**, 868–880.
49. Hogg, J.R. and Collins, K. (2007) Human Y5 RNA specializes a Ro ribonucleoprotein for 5S ribosomal RNA quality control. *Genes Dev.*, **21**, 3067–3072.
50. Srisawat, C., Goldstein, I.J. and Engelke, D.R. (2001) Sephadex-binding RNA ligands: rapid affinity purification of RNA from complex RNA mixtures. *Nucleic Acids Res.*, **29**, e4.
51. Srisawat, C. and Engelke, D.R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA*, **7**, 632–641.
52. Li, Y. and Altman, S. (2002) Partial reconstitution of human RNase P in HeLa cells between its RNA subunit with an affinity tag and the intact protein components. *Nucleic Acids Res.*, **30**, 3706–3711.
53. Vasudevan, S. and Steitz, J.A. (2007) AU-rich-element-mediated upregulation of translation by FXR1 and argonaute 2. *Cell*, **128**, 1105–1118.
54. Iioka, H., Loiselle, D., Haystead, T.A. and MacAra, I.G. (2011) Efficient detection of RNA–protein interactions using tethered RNAs. *Nucleic Acids Res.*, **39**, 1–7.
55. Dix, C.I., Soundararajan, H.C., Dzhindzhev, N.S., Begum, F., Suter, B., Ohkura, H., Stephens, E. and Bullock, S.L. (2013) Lissencephaly-1 promotes the recruitment of dynein and dynactin to transported mRNAs. *J. Cell Biol.*, **202**, 479–494.
56. Dienstbier, M., Boehl, F., Li, X. and Bullock, S.L. (2009) Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. *Genes Dev.*, **23**, 1546–1558.
57. Butter, F., Scheibe, M., Morl, M. and Mann, M. (2009) Unbiased RNA–protein interaction screen by quantitative proteomics. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 10626–10631.
58. Leppek, K. and Stoecklin, G. (2014) An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.*, **42**, e13.
59. Hartmuth, K., Urlaub, H., Vornlocher, H.-P., Will, C.L., Gentzel, M., Wilm, M. and Lührmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16719–16724.
60. Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H. and Lührmann, R. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.*, **26**, 5528–5543.
61. Ward, A.M., Bidet, K., Yinglin, A., Ler, S.G., Hogue, K., Blackstock, W., Gunaratne, J. and Garcia-Blanco, M.A. (2011) Quantitative mass spectrometry of DENV-2 RNA-interacting proteins reveals that the DEAD-box RNA helicase DDX6 binds the DB1 and DB2 3′ UTR structures. *RNA Biol.*, **8**, 1173–1186.
62. Windbichler, N. and Schroeder, R. (2006) Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer. *Nat. Protoc.*, **1**, 637–641.
63. Wallace, S.T. and Schroeder, R. (1998) In vitro selection and characterization of streptomycin-binding RNAs: recognition discrimination between antibiotics. *RNA*, **4**, 112–123.
64. Dangerfield, J.A., Windbichler, N., Salmons, B., Günzburg, W.H. and Schröder, R. (2006) Enhancement of the StreptoTag method for isolation of endogenously expressed proteins with complex RNA binding targets. *Electrophoresis*, **27**, 1874–1877.
65. Böck-Taferner, P. and Wank, H. (2004) GAPDH enhances group II intron splicing in vitro. *Biol. Chem.*, **385**, 615–621.
66. Lee, H.Y., Haurwitz, R.E., Apffel, A., Zhou, K., Smart, B., Wenger, C.D., Laderman, S., Bruhn, L. and Doudna, J.A. (2013) RNA–protein analysis using a conditional CRISPR nuclease. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 5416–5421.
67. Lingner, J. and Cech, T.R. (1996) Purification of telomerase from *Euplotes aediculatus*: requirement of a primer 3′ overhang. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 10712–10717.
68. Kurth, I., Cristofari, G. and Lingner, J. (2008) An affinity oligonucleotide displacement strategy to purify ribonucleoprotein complexes applied to human telomerase. *Methods Mol. Biol.*, **488**, 9–22.
69. Lamond, A.I., Sproat, B., Ryder, U. and Hamm, J. (1989) Probing the structure and function of U2 snRNP with antisense oligonucleotides made of 2′-OMe RNA. *Cell*, **58**, 383–390.
70. Blencowe, B.J., Sproat, B.S., Ryder, U., Barabino, S. and Lamond, A.I. (1992) Antisense probing of the human U4/U6 snRNP with biotinylated 2′-O-MeRNA oligonucleotides. *Cell*, **59**, 531–539.
71. Zaug, A.J., Lingner, J. and Cech, T.R. (1996) Method for determining RNA 3[prime] ends and application to human telomerase RNA. *Nucleic Acids Res.*, **24**, 532–533.
72. Soeno, Y., Taya, Y., Stasyk, T., Huber, L.A., Aoba, T. and Hüttenhofer, A. (2010) Identification of novel ribonucleo-protein complexes from the brain-specific snoRNA MBII-52. *RNA*, **16**, 1293–1300.
73. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M. et al. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393–1406.
74. Baltz, A.G., Munschauer, M., Schwanhauser, B., Vasile, A., Murakawa, Y., Schueler, M., Young, N., Penfold-Brown, D., Drew, K., Milek, M. et al. (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell*, **46**, 674–690.
75. McHugh, C.a., Russell, P. and Guttman, M. (2014) Methods for comprehensive experimental identification of RNA–protein interactions. *Genome Biol.*, **15**, 203.
76. Lee, N., Pimienta, G. and Steitz, J.A. (2012) AUF1/hnRNP D is a novel protein partner of the EBER1 noncoding RNA of Epstein-Barr virus. *RNA*, **18**, 2073–2082.
77. Samra, N., Atir-Lande, A., Pnueli, L. and Arava, Y. (2015) The elongation factor eEF3 (Yef3) interacts with mRNA in a translation independent manner. *BMC Mol. Biol.*, **16**, 17.
78. Guo, C., Xue, Y., Yang, G., Yin, S., Shi, W., Cheng, Y., Yan, X., Fan, S., Zhang, H. and Zeng, F. (2015) Nanog RNA binding proteins YBX1 and ILF3 affect pluripotency of embryonic stem cells. *Cell Biol. Int.*, **9999**, 1–14.
79. Mitchell, S.F., Jain, S., She, M. and Parker, R. (2013) Global analysis of yeast mRNPs. *Nat. Struct. Mol. Biol.*, **20**, 127–133.
80. Kwon, S.C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K.T., Castello, A., Krijgsvelde, J., Hentze, M.W. and Kim, V.N. (2013) The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.*, **20**, 1122–1130.
81. Conrad, T., Albrecht, A.-S., de Melo Costa, V.R., Sauer, S., Meierhofer, D. and Ørom, U.A. (2016) Serial interactome capture of the human cell nucleus. *Nat. Commun.*, **7**, 11212.
82. Zeng, F., Peritz, T., Kannanayakal, T.J., Kilk, K., Eiriksdóttir, E., Langel, U. and Eberwine, J. (2006) A protocol for PAIR: PNA-assisted identification of RNA binding proteins in living cells. *Nat. Protoc.*, **1**, 920–927.
83. Bell, T.J. and Eberwine, J. (2015) Live cell genomics: RNA exon-specific RNA-binding protein isolation. *Cell-Penetrating Peptides: Methods Protoc.*, 457–468.

84. O'Connell, M.R., Oakes, B.L., Sternberg, S.H., East-Seletsky, A., Kaplan, M. and Doudna, J. (2014) Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature*, **516**, 263–266.
85. Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A. and Yeo, G.W. (2016) Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell*, **165**, 488–496.
86. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.*, **389**, 1017–1031.
87. Gouw, J.W., Krijgsveld, J. and Heck, A.J.R. (2010) Quantitative proteomics by metabolic labeling of model organisms. *Mol. Cell. Proteomics*, **9**, 11–24.
88. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics*, **1**, 376–386.
89. Chen, X., Wei, S., Ji, Y., Guo, X. and Yang, F. (2015) Quantitative proteomics using SILAC: Principles, applications, and developments. *Proteomics*, **15**, 3175–3192.
90. Altaealar, A.F.M., Munoz, J. and Heck, A.J.R. (2012) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.*, **14**, 35–48.
91. Scheibe, M., Arnoult, N., Kappei, D., Buchholz, F., Decottignies, A., Butter, F. and Mann, M. (2013) Quantitative interaction screen of telomeric repeat-containing RNA reveals novel TERRA regulators. *Genome Res.*, **23**, 2149–2157.
92. Sharma, S., Wongpalee, S.P., Vashisht, A., Wohlschlegel, J.A. and Black, D.L. (2014) stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes Dev.*, **28**, 2518–2531.
93. Schmidt, C., Grønborg, M., Deckert, J., Bessonov, S., Conrad, T., Lührmann, R. and Urlaub, H. (2014) Mass spectrometry-based relative quantification of proteins in precatalytic and catalytically active spliceosomes by metabolic labeling (SILAC), chemical labeling (iTRAQ), and label-free spectral count. *RNA*, **20**, 406–420.
94. Haesler, A.R., Donnelly, C.J., Periz, G., Simko, E.A.J., Shaw, P.G., Kim, M.-S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R. et al. (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, **507**, 195–200.
95. Rio, D.C. (2012) Filter-binding assay for analysis of RNA–protein interactions. *Cold Spring Harb. Protoc.*, **7**, 1078–1081.
96. Stockley, P.G. (2009) Filter-binding assays. *Methods Mol. Biol.*, **543**, 1–14.
97. Hellman, L.M. and Fried, M.G. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat. Protoc.*, **2**, 1849–1861.
98. Yakhnin, A. V., Yakhnin, H. and Babbitzke, P. (2012) Gel mobility shift assays to detect protein–RNA interactions. *Methods Mol. Biol.*, **905**, 201–211.
99. Rio, D.C. (2014) Electrophoretic mobility shift assays for RNA–protein complexes. *Cold Spring Harb. Protoc.*, 435–440.
100. Wilson, G. (2005) RNA folding and RNA–protein binding analyzed by fluorescence anisotropy and resonance energy transfer. *Rev. Fluoresc.*, **2**, 223–243.
101. Mao, C., Flavin, K.G., Wang, S., Dodson, R., Ross, J. and Shapiro, D.J. (2006) Analysis of RNA–protein interactions by a microplate-based fluorescence anisotropy assay. *Anal. Biochem.*, **350**, 222–232.
102. Royer, C.A. and Scarlata, S.F. (2008) Chapter 5. Fluorescence approaches to quantifying biomolecular interactions. *Methods Enzymol.*, **450**, 79–106.
103. Kim, D.H., Langlois, M.A., Lee, K.B., Riggs, A.D., Puymirat, J. and Rossi, J.J. (2005) HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. *Nucleic Acids Res.*, **33**, 3866–3874.
104. Laurent, F.X., Sureau, A., Klein, A.F., Trouslard, F., Gasnier, E., Furling, D. and Marie, J. (2012) New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats. *Nucleic Acids Res.*, **40**, 3159–3171.
105. Fourmy, D. and Yoshizawa, S. (2012) Protein–RNA footprinting: An evolving tool. *Wiley Interdiscip. Rev. RNA*, **3**, 557–566.
106. Nilsen, T.W. (2014) RNase footprinting to map sites of RNA–protein interactions. *Cold Spring Harb. Protoc.*, 677–682.
107. Stumpf, C.R., Opperman, L. and Wickens, M. (2008) Analysis of RNA–protein Interactions Using a Yeast Three-Hybrid System. *Methods Enzymol.*, **449**, 295–315.
108. Martin, F. (2012) Fifteen years of the yeast three-hybrid system: RNA–protein interactions under investigation. *Methods*, **58**, 367–375.
109. Niranjanakumari, S., Lasda, E., Brazas, R. and Garcia-Blanco, M.A. (2002) Reversible cross-linking combined with immunoprecipitation to study RNA–protein interactions in vivo. *Methods*, **26**, 182–190.
110. Selth, L.A., Gilbert, C. and Svejstrup, J.Q. (2009) RNA immunoprecipitation to determine RNA–protein associations in vivo. *Cold Spring Harb. Protoc.*, **4**, doi:10.1101/pdb.prot5234.
111. Peritz, T., Zeng, F., Kannanayakal, T.J., Kilk, K., Eiriksdóttir, E., Langel, U. and Eberwine, J. (2006) Immunoprecipitation of mRNA–protein complexes. *Nat. Protoc.*, **1**, 577–580.
112. Mili, S. and Steitz, J.A. (2004) Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunopr. *RNA*, **10**, 1692–1694.
113. Ule, J., Jensen, K., Mele, A. and Darnell, R.B. (2005) CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods*, **37**, 376–386.
114. Jensen, K.B. and Darnell, R.B. (2008) CLIP: Crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *Methods Mol. Biol.*, **488**, 85–98.
115. Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X. et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, **456**, 464–469.
116. Zhang, C. and Darnell, R.B. (2011) Mapping in vivo protein–RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.*, **29**, 607–614.
117. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.C., Munschauer, M. et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
118. König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.*, **17**, 909–915.
119. Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J. and Ule, J. (2014) iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods*, **65**, 274–287.
120. Conway, A.E., Van Nostrand, E.L., Pratt, G.A., Aigner, S., Wilbert, M.L., Sundararaman, B., Freese, P., Lambert, N.J., Sathe, S., Liang, T.Y. et al. (2016) Enhanced CLIP uncovers IMP protein–RNA targets in human pluripotent stem cells important for cell adhesion and survival. *Cell Rep.*, **15**, 666–679.
121. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K. et al. (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods*, **13**, 508–514.
122. Shih, J.D., Waks, Z., Kedersha, N. and Silver, P.A. (2011) Visualization of single mRNAs reveals temporal association of proteins with microRNA-regulated mRNA. *Nucleic Acids Res.*, **39**, 7740–7749.
123. Kochan, J., Wawro, M. and Kasza, A. (2015) Simultaneous detection of mRNA and protein in single cells using immunofluorescence-combined single-molecule RNA FISH. *Biotechniques*, **59**, 209–221.
124. Jung, J., Lifland, A.W., Zurla, C., Alonas, E.J. and Santangelo, P.J. (2013) Quantifying RNA–protein interactions in situ using modified-MTRIPs and proximity ligation. *Nucleic Acids Res.*, **41**, 1–13.
125. Lindehell, H., Kim, M. and Larsson, J. (2015) Proximity ligation assays of protein and RNA interactions in the male-specific lethal complex on Drosophila melanogaster polytene chromosomes. *Chromosoma*, **124**, 385–395.
126. Roussis, I.M., Guille, M., Myers, F.A. and Scarlett, G.P. (2016) RNA whole-mount in situ hybridisation proximity ligation assay

- (rISH-PLA), an assay for detecting RNA–protein complexes in intact cells. *PLoS One*, **11**, e0147967.
127. Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.-J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.-G. *et al.* (2006) Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods*, **3**, 995–1000.
  128. Weibrecht, I., Leuchowius, K.-J., Clausson, C.-M., Conze, T., Jarvius, M., Howell, W.M., Kamali-Moghaddam, M. and Söderberg, O. (2010) Proximity ligation assays: a recent addition to the proteomics toolbox. *Expert Rev. Proteomics*, **7**, 401–409.
  129. Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gústafsdóttir, S.M., Ostman, A. and Landegren, U. (2002) Protein detection using proximity-dependent DNA ligation assays. *Nat. Biotechnol.*, **20**, 473–477.
  130. Zurla, C., Jung, J. and Santangelo, P.J. (2016) Can we observe changes in mRNA 'state'? Overview of methods to study mRNA interactions with regulatory proteins relevant in cancer related processes. *Analyst*, **141**, 548–562.
  131. Sun, Y., Wallrabe, H., Seo, S.A. and Periasamy, A. (2011) FRET microscopy in 2010: the legacy of Theodor Förster on the 100th anniversary of his birth. *ChemPhysChem*, **12**, 462–474.
  132. Huranová, M., Jablonski, J.A., Benda, A., Hof, M., Stanek, D. and Caputi, M. (2009) In vivo detection of RNA-binding protein interactions with cognate RNA sequences by fluorescence resonance energy transfer. *RNA*, **15**, 2063–2071.
  133. Lorenz, M. (2009) Visualizing protein – RNA interactions inside cells by fluorescence resonance energy transfer. *RNA*, **15**, 97–103.
  134. Urbaneck, M.O., Galka-Marciniak, P., Olejniczak, M. and Krzyzosiak, W.J. (2014) RNA imaging in living cells - methods and applications. *RNA Biol.*, **11**, 1083–1095.
  135. Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A. and Yeo, G.W. (2016) Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell*, **165**, 488–496.
  136. Wojciechowska, M. and Krzyzosiak, W.J. (2011) Cellular toxicity of expanded RNA repeats: Focus on RNA foci. *Hum. Mol. Genet.*, **20**, 3811–3821.
  137. Urbaneck, M.O. and Krzyzosiak, W.J. (2016) RNA FISH for detecting expanded repeats in human diseases. *Methods*, **98**, 115–123.
  138. Todd, P.K. and Paulson, H.L. (2010) RNA-mediated neurodegeneration in repeat expansion disorders. *Ann. Neurol.*, **67**, 291–300.
  139. Urbaneck, M.O., Jazurek, M., Switonski, P.M., Figura, G. and Krzyzosiak, W.J. (2016) Nuclear speckles are detention centers for transcripts containing expanded CAG repeats. *Biochim. Biophys. Acta*, **1862**, 1513–1520.
  140. Timchenko, L.T., Timchenko, N.A., Caskey, C.T. and Roberts, R. (1996) Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: Implications for myotonic dystrophy. *Hum. Mol. Genet.*, **5**, 115–121.
  141. Timchenko, L.T., Miller, J.W., Timchenko, N.A., Devore, D.R., Datar, K. V., Lin, L., Roberts, R., Caskey, C.T. and Swanson, M.S. (1996) Identification of a (CUG)<sub>n</sub> triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res.*, **24**, 4407–4414.
  142. Miller, J.W., Urbinati, C.R., Teng-Umuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A. and Swanson, M.S. (2000) Recruitment of human muscleblind proteins to (CUG)<sub>n</sub> expansions associated with myotonic dystrophy. *EMBO J.*, **19**, 4439–4448.
  143. Napierala, M. and Krzyzosiak, W.J. (1997) CUG repeats present in myotonin kinase RNA form metastable 'slippery' hairpins. *J. Biol. Chem.*, **272**, 31079–31085.
  144. Osborne, R.J.R. and Thornton, C.A. (2006) RNA-dominant diseases. *Hum. Mol. Genet.*, **15**, 162–169.
  145. Batra, R., Charizanis, K., Manchanda, M., Mohan, A., Li, M., Finn, D.J., Goodwin, M., Zhang, C., Sobczak, K., Thornton, C.A. *et al.* (2014) Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol. Cell*, **56**, 311–322.
  146. Tsoi, H., Lau, C.K., Lau, K.F. and Chan, H.Y.E. (2011) Perturbation of U2AF65/NXF1-mediated RNA nuclear export enhances RNA toxicity in polyQ diseases. *Hum. Mol. Genet.*, **20**, 3787–3797.
  147. Sun, X., Li, P.P., Zhu, S., Cohen, R., Marque, L.O., Ross, C.a., Pulst, S.M., Chan, H.Y.E., Margolis, R.L. and Rudnicki, D.D. (2015) Nuclear retention of full-length HTT RNA is mediated by splicing factors MBNL1 and U2AF65. *Sci. Rep.*, **5**, 12521.
  148. Krauss, S., Griesche, N., Jastrzebska, E., Chen, C., Rutschow, D., Achmüller, C., Dorn, S., Boesch, S.M., Lalowski, M., Wanker, E. *et al.* (2013) Translation of HTT mRNA with expanded CAG repeats is regulated by the MID1-PP2A protein complex. *Nat. Commun.*, **4**, 1511.
  149. Sellier, C., Freyermuth, F., Tabet, R., Tran, T., He, F., Ruffenach, F., Alunni, V., Moine, H., Thibault, C., Page, A. *et al.* (2013) Sequestration of DROSHA and DGCR8 by expanded CGG RNA Repeats Alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Rep.*, **3**, 869–880.
  150. Fratta, P., Mizielinska, S., Nicoll, A.J., Zloh, M., Fisher, E.M.C., Parkinson, G. and Isaacs, A.M. (2012) C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci. Rep.*, **2**, 1016.
  151. Sobczak, K., Michlewski, G., De Mezer, M., Kierzek, E., Krol, J., Olejniczak, M., Kierzek, R. and Krzyzosiak, W.J. (2010) Structural diversity of triplet repeat RNAs. *J. Biol. Chem.*, **285**, 12755–12764.
  152. Ranum, L.P.W. and Cooper, T.A. (2006) RNA-mediated Neuromuscular Disorders. *Annu. Rev. Neurosci.*, **29**, 259–277.
  153. O'Rourke, J.R. and Swanson, M.S. (2009) Mechanisms of RNA-mediated disease. *J. Biol. Chem.*, **284**, 7419–7423.
  154. Thornton, C.A. (2014) Myotonic dystrophy. *Neurol. Clin.*, **32**, 705–719.
  155. Salisbury, E., Schoser, B., Schneider-Gold, C., Wang, G.-L., Huichalaf, C., Jin, B., Sirito, M., Sarkar, P., Krahe, R., Timchenko, N.A. *et al.* (2009) Expression of RNA CCUG repeats dysregulates translation and degradation of proteins in myotonic dystrophy 2 patients. *Am. J. Pathol.*, **175**, 748–762.
  156. Tian, B., White, R.J., Xia, T., Welle, S., Turner, D.H., Mathews, M.B. and Thornton, C.A. (2000) Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA*, **6**, 79–87.
  157. Lu, X., Timchenko, N.A. and Timchenko, L.T. (1999) Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy. *Hum. Mol. Genet.*, **8**, 53–60.
  158. Michalowski, S., Miller, J.W., Urbinati, C.R., Paliouras, M., Swanson, M.S. and Griffith, J. (1999) Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucleic Acids Res.*, **27**, 3534–3542.
  159. Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshire, M.G., Harper, P.S. and Brook, J.D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.*, **11**, 805–814.
  160. Daughters, R.S., Tuttle, D.L., Gao, W., Ikeda, Y., Moseley, M.L., Ebner, T.J., Swanson, M.S. and Ranum, L.P.W. (2009) RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genet.*, **5**, e1000600.
  161. Rudnicki, D.D., Holmes, S.E., Lin, M.W., Thornton, C.A., Ross, C.A. and Margolis, R.L. (2007) Huntington's disease-like 2 is associated with CUG repeat-containing RNA foci. *Ann. Neurol.*, **61**, 272–282.
  162. Konieczny, P., Stepniak-Konieczna, E. and Sobczak, K. (2014) MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic Acids Res.*, **42**, 10873–10887.
  163. Wang, E.T., Cody, N.A.L., Jog, S., Biancolella, M., Wang, T.T., Treacy, D.J., Luo, S., Schroth, G.P., Housman, D.E., Reddy, S. *et al.* (2012) Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell*, **150**, 710–724.
  164. Masuda, A., Andersen, H.S., Doktor, T.K., Okamoto, T., Ito, M., Andresen, B.S. and Ohno, K. (2012) CUGBP1 and MBNL1 preferentially bind to 3' UTRs and facilitate mRNA decay. *Sci. Rep.*, **2**, 209.
  165. Rau, F., Freyermuth, F., Fugier, C., Villemin, J.-P., Fischer, M.-C., Jost, B., Demele, D., Gourdon, G., Nicole, A., Duboc, D. *et al.* (2011) Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat. Struct. Mol. Biol.*, **18**, 840–845.
  166. Ravel-Chapuis, A., Belanger, G., Yadava, R.S., Mahadevan, M.S., DesGroseillers, L., Cote, J. and Jasmin, B.J. (2012) The RNA-binding protein Staufen1 is increased in DM1 skeletal muscle and promotes alternative pre-mRNA splicing. *J. Cell Biol.*, **196**, 699–712.

167. Pettersson,O.J., Aagaard,L., Andrejeva,D., Thomsen,R., Jensen,T.G. and Damgaard,C.K. (2014) DDX6 regulates sequestered nuclear CUG-expanded DMPK-mRNA in dystrophia myotonica type 1. *Nucleic Acids Res.*, **42**, 7186–7200.
168. Ebralidze,A., Wang,Y., Petkova,V., Ebralidze,K. and Junghans,R.P. (2004) RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science*, **303**, 383–387.
169. Orr,H.T. and Zoghbi,H.Y. (2007) Trinucleotide Repeat Disorders. *Annu. Rev. Neurosci.*, **30**, 575–621.
170. Kino,Y., Mori,D., Oma,Y., Takeshita,Y., Sasagawa,N. and Ishiura,S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.*, **13**, 495–507.
171. McLaughlin,B.A., Spencer,C. and Eberwine,J. (1996) CAG trinucleotide RNA repeats interact with RNA-binding proteins. *Am. J. Hum. Genet.*, **59**, 561–569.
172. Peel,A.L., Rao,R.V., Cottrell,B.A., Hayden,M.R., Ellerby,L.M. and Bredesen,D.E. (2001) Double-stranded RNA-dependent protein kinase, PKR, binds preferentially to Huntington's disease (HD) transcripts and is activated in HD tissue. *Hum. Mol. Genet.*, **10**, 1531–1538.
173. Tsoi,H., Lau,T.C.-K., Tsang,S.-Y., Lau,K.-F. and Chan,H.Y.E. (2012) CAG expansion induces nuclear stress in polyglutamine diseases. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 13428–13433.
174. Tsoi,H. and Chan,H.Y.E. (2013) Expression of expanded CAG transcripts triggers nuclear stress in huntington's disease. *Cerebellum*, **12**, 310–312.
175. Mykowska,A., Sobczak,K., Wojciechowska,M., Kozłowski,P. and Krzyżosiak,W.J. (2011) CAG repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Res.*, **39**, 8938–8951.
176. Ho,T.H., Savkur,R.S., Poulos,M.G., Mancini,M.A., Swanson,M.S. and Cooper,T.A. (2005) Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. *J. Cell Sci.*, **118**, 2923–2933.
177. Sathasivam,K., Neueder,A., Gipson,T.A., Landles,C., Benjamin,A.C., Bondulich,M.K., Smith,D.L., Faull,R.L.M., Roos,R.A.C., Howland,D. et al. (2013) Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 2366–2370.
178. Hagerman,R. and Hagerman,P. (2013) Advances in clinical and molecular understanding of the FMR1 premutation and fragile X-associated tremor/ataxia syndrome. *Lancet Neurol.*, **12**, 786–798.
179. Hagerman,P. (2013) Fragile X-associated tremor/ataxia syndrome (FXTAS): Pathology and mechanisms. *Acta Neuropathol.*, **126**, 1–19.
180. Iwahashi,C.K., Yasui,D.H., An,H.J., Greco,C.M., Tassone,F., Nannen,K., Babineau,B., Lebrilla,C.B., Hagerman,R.J. and Hagerman,P.J. (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain*, **129**, 256–271.
181. Jin,P., Duan,R., Qurashi,A., Qin,Y., Tian,D., Rosser,T.C., Liu,H., Feng,Y. and Warren,S.T. (2007) Pur  $\alpha$  binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. *Neuron*, **55**, 556–564.
182. Sofola,O.A., Jin,P., Qin,Y., Duan,R., Liu,H., de Haro,M., Nelson,D.L. and Botas,J. (2007) RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. *Neuron*, **55**, 565–571.
183. Sellier,C., Rau,F., Liu,Y., Tassone,F., Hukema,R.K., Gattoni,R., Schneider,A., Richard,S., Willemsen,R., Elliott,D.J. et al. (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J.*, **29**, 1248–1261.
184. Matsuura,T., Yamagata,T., Burgess,D.L., Rasmussen,a, Grewal,R.P., Watase,K., Khajavi,M., McCall,a.E., Davis,C.F., Zu,L. et al. (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat. Genet.*, **26**, 191–194.
185. White,M.C., Gao,R., Xu,W., Mandal,S.M., Lim,J.G., Hazra,T.K., Wakamiya,M., Edwards,S.F., Raskin,S., Teive,H.A.G. et al. (2010) Inactivation of hnRNP K by expanded intronic AUUCU repeat induces apoptosis via translocation of PKCdelta to mitochondria in spinocerebellar ataxia 10. *PLoS Genet.*, **6**, e1000984.
186. White,M., Xia,G., Gao,R., Wakamiya,M., Sarkar,P.S., McFarland,K. and Ashizawa,T. (2012) Transgenic mice with SCA10 pentanucleotide repeats show motor phenotype and susceptibility to seizure: A toxic RNA gain-of-function model. *J. Neurosci. Res.*, **90**, 706–714.
187. Renton,A.E., Majounie,E., Waite,A., Simón-Sánchez,J., Rollinson,S., Gibbs,J.R., Schymick,J.C., Laaksovirta,H., van Swieten,J.C., Myllykangas,L. et al. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, **72**, 257–268.
188. DeJesus-Hernandez,M., Mackenzie,I.R., Boeve,B.F., Boxer,A.L., Baker,M., Rutherford,N.J., Nicholson,A.M., Finch,N.A., Flynn,H., Adamson,J. et al. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, **72**, 245–256.
189. Smith,B.N., Newhouse,S., Shatunov,A., Vance,C., Topp,S., Johnson,L., Miller,J., Lee,Y., Troakes,C., Scott,K.M. et al. (2013) The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. *Eur. J. Hum. Genet.*, **21**, 102–108.
190. Lee,Y.B., Chen,H.J., Peres,J.N., Gomez-Deza,J., Attig,J., Štalekar,M., Troakes,C., Nishimura,A.L., Scotter,E.L., Vance,C. et al. (2013) Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep.*, **5**, 1178–1186.
191. Reddy,K., Zamiri,B., Stanley,S.Y.R., Macgregor,R.B. and Pearson,C.E. (2013) The Disease-associated r(GGGGCC)<sub>n</sub> Repeat from the C9orf72 Gene Forms Tract Length-dependent Uni- and Multimolecular RNA G-quadruplex Structures. *J. Biol. Chem.*, **288**, 9860–9866.
192. Zamiri,B., Reddy,K., Macgregor,R.B. and Pearson,C.E. (2014) TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)<sub>n</sub> repeat of the C9orf72 gene and blocks interaction of RNA binding proteins. *J. Biol. Chem.*, **289**, 4653–4659.
193. Sareen,D., O'Rourke,J.G., Meera,P., Muhammad,a.K.M.G., Grant,S., Simpkinson,M., Bell,S., Carmona,S., Ornelas,L., Sahabian,a. et al. (2013) Targeting RNA Foci in iPSC-Derived Motor Neurons from ALS Patients with a C9ORF72 Repeat Expansion. *Sci. Transl. Med.*, **5**, 208ra149.
194. Mori,K., Lammich,S., Mackenzie,I.R.A., Forne,I., Zilow,S., Kretzschmar,H., Edbauer,D., Janssens,J., Kleinberger,G., Cruts,M. et al. (2013) HnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathol.*, **125**, 413–423.
195. Xu,Z., Poidevin,M., Li,X., Li,Y., Shu,L., Nelson,D.L., Li,H., Hales,C.M., Gearing,M., Wingo,T.S. et al. (2013) Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 7778–7783.
196. Almeida,S., Gascon,E., Tran,H., Chou,H.J., Gendron,T.F., DeGroot,S., Tapper,A.R., Sellier,C., Charlet-Berguerand,N., Karydas,A. et al. (2013) Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. *Acta Neuropathol.*, **126**, 385–399.
197. Cooper-Knock,J., Walsh,M.J., Higginbottom,A., Highley,J.R., Dickman,M.J., Edbauer,D., Ince,P.G., Wharton,S.B., Wilson,S.A., Kirby,J. et al. (2014) Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain*, **137**, 2040–2051.
198. Rossi,S., Serrano,A., Gerbino,V., Giorgi,A., Di Francesco,L., Nencini,M., Bozzo,F., Schininà,M.E., Bagni,C., Cestra,G. et al. (2015) Nuclear accumulation of mRNAs underlies G4C2-repeat-induced translational repression in a cellular model of C9orf72 ALS. *J. Cell Sci.*, **128**, 1787–1799.
199. Donnelly,C.J., Zhang,P.-W., Pham,J.T., Haeusler,A.R., Mistry,N.A., Vidensky,S., Daley,E.L., Poth,E.M., Hoover,B. et al. (2013) RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron*, **80**, 415–428.
200. Cooper-Knock,J., Bury,J.J., Heath,P.R., Wyles,M., Higginbottom,A., Gelsthorpe,C., Highley,J.R., Hautbergue,G., Rattray,M., Kirby,J. et al. (2015) C9ORF72 GGGGCC expanded repeats produce splicing dysregulation which correlates with disease severity in amyotrophic lateral sclerosis. *PLoS One*, **10**, e0127376.
201. Lasda,E. and Parker,R. (2014) Circular RNAs: diversity of form and function. *RNA*, **20**, 1829–1842.

202. Chen, L. (2016) The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.*, **17**, 205–211.
203. Kung, J.T.Y., Colognori, D. and Lee, J.T. (2013) Long noncoding RNAs: Past, present, and future. *Genetics*, **193**, 651–669.
204. Huarte, M. (2015) The emerging role of lncRNAs in cancer. *Nat. Med.*, **21**, 1253–1261.
205. Klass, D.M., Scheibe, M., Butter, F., Hogan, G.J., Mann, M. and Brown, P.O. (2013) Quantitative proteomic analysis reveals concurrent RNA–protein interactions and identifies new RNA-binding proteins in *Saccharomyces cerevisiae*. *Genome Res.*, **23**, 1028–1038.
206. Schmidt, C., Kramer, K. and Urlaub, H. (2012) Investigation of protein-RNA interactions by mass spectrometry—Techniques and applications. *J. Proteomics*, **75**, 3478–3494.
207. Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.-L., Hentze, M.W., Kohlbacher, O. and Urlaub, H. (2014) Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods*, **11**, 1064–1070.
208. Scheibe, M., Butter, F., Hafner, M., Tuschl, T. and Mann, M. (2012) Quantitative mass spectrometry and PAR-CLIP to identify RNA–protein interactions. *Nucleic Acids Res.*, **40**, 9897–9902.
209. Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L. *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, **13**, 722–736.