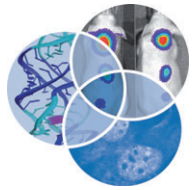
Advanced Review



Advances in the characterization of RNA-binding proteins

Domenica Marchese,[1,2†] Natalia Sanchez de Groot,[1,2†] Nieves Lorenzo Gotor,[1,2] Carmen Maria Livi[1,2,3] and Gian G. Tartaglia[1,2,4*]

From transcription, to transport, storage, and translation, RNA depends on association with different RNA-binding proteins (RBPs). Methods based on next-generation sequencing and protein mass-spectrometry have started to unveil genome-wide interactions of RBPs but many aspects still remain out of sight. How many of the binding sites identified in high-throughput screenings are functional? A number of computational methods have been developed to analyze experimental data and to obtain insights into the specificity of protein–RNA interactions. How can theoretical models be exploited to identify RBPs? In addition to oligomeric complexes, protein and RNA molecules can associate into granular assemblies whose physical properties are still poorly understood. What protein features promote granule formation and what effects do these assemblies have on cell function? Here, we describe the newest *in silico*, *in vitro*, and *in vivo* advances in the field of protein–RNA interactions. We also present the challenges that experimental and computational approaches will have to face in future studies.

How to cite this article:
WIREs RNA 2016. doi: 10.1002/wrna.1378

INTRODUCTION

Since 'the central dogma' was formulated (the genetic information flows from DNA to protein passing through RNA), our knowledge on the transcriptome has progressed enormously.[1,2] We now know that genes do not code for just a single protein but produce a number of variants: the activity of nuclear RNAs (snRNAs and snoRNAs) and several splicing factors generates several messenger RNAs (mRNAs) with different lengths and exon compositions (Figure 1).[2] In addition to protein production, RNA participates in other essential processes such as mRNAs expression regulation (microRNAs, miRNA; small interfering RNAs, siRNA; long noncoding RNAs, lncRNAs)[1,2] and genome protection by transposon silencing (PIWI-interacting RNA, piRNA).[3] RNAs also perform a number of structural and functional tasks, as in the case of rRNAs, which constitute 60% of the ribosome,[4] and tRNAs, which carry the amino acids to the ribosome during the translation process.[1,2]

Every aspect of RNA life, from birth (polymerases) to degradation (nucleases), involves protein binding (Figure 1). A correct interplay between RNAs and RNA-binding proteins (RBPs) is crucial for the development of cellular processes: miRNAs, for instance, require Argonaute proteins to reach target mRNAs as well as piRNAs associate with PIWI proteins to form silencing complexes that protect the germline genome from transposons.[1,2]

[†]These authors contributed equally to this article.

*Correspondence to: gian.tartaglia@crg.es

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain

[2]Universitat Pompeu Fabra (UPF), Barcelona, Spain

[3]IFOM Foundation, FIRC Institute of Molecular Oncology Foundation, Milan, Italy

[4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Conflict of interest: The authors have declared no conflicts of interest for this article.

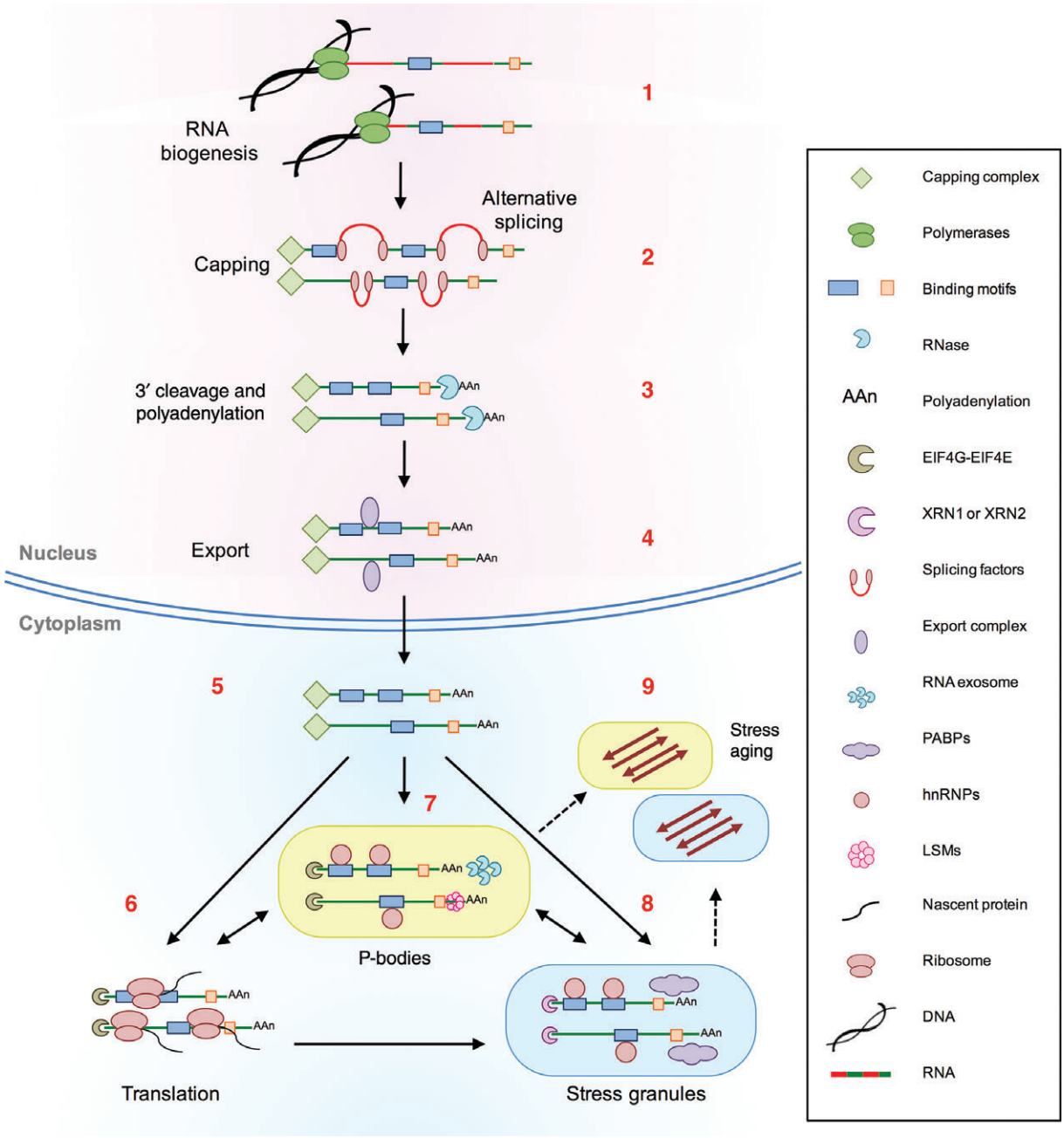


FIGURE 1 | RNA-binding proteins (RBPs) and RNA life. RNA birth is regulated by RBPs (1) that are responsible for maturation (2) and modification (3). RBPs protect (4) and transport (5) RNA around the cell to specific sites (6). Interactions are regulated through a diverse set of binding sites that allows the formation of dynamic complexes sustained by reversible contacts and involving multiple partners (7, 8). When RNA is not required, it can be stored for future needs (7) or degraded (8). The last process of the RNA life cycle is the release of nucleotides that will be employed to build new RNA molecules. When RBPs are impaired (e.g., protein mutation, concentration deregulation, etc.), half-life, arrangement and location of RNA are affected (9).

Moreover, the interplay between RBPs and RNAs can promote formation of membrane-less organelles (e.g., P-bodies, stress granules),[5–7] whose composition can be rapidly adapted to the cell state and the environment conditions (e.g., stress conditions).[8] Mutations in RBPs, aberrant interactions or altered RNA processing have been related to a number of human diseases, from neurological disorders to cancer.[9–12]

Protein–RNA interactions have been studied using a range of quantitative approaches such as electrophoretic mobility shift assay (EMSA),[13] fluorescent anisotropy/polarization,[14] Förster resonance

energy transfer (FRET),[15] surface plasmon resonance (SPR)[16] or, more recently, microscale thermophoresis (MST)[17,18] and stochastic optical reconstruction microscopy (STORM).[19] Although these approaches have proved to be powerful to assess the specificity and affinity of protein–RNA interactions, they only allow the study of single or few molecular interactions at a time.

The recent development of large-scale quantitative methods, exploiting next-generation sequencing[20,21] and protein mass spectrometry,[22] contributes to the genome-wide identification of RBPs, RNA targets and cofactors.[2] Deep-sequencing approaches combined with RBPs immunoprecipitation as well as *in vitro* methods, such as Systematic Evolution of Ligands by Exponential Enrichment SELEX,[23,24] revealed the binding ability of a number of RBPs and showed that many RBPs bind to thousands of transcripts.

In parallel to experimental advances, a number of *in silico* methods have been developed to predict protein–RNA interactions and to detect binding-sites. Computational tools are particularly useful to predict potential ribonucleoprotein associations and to narrow down a list of interaction partners for experimental validation. For instance, some RBPs recognize particular nucleotide sequences, whereas others bind to the backbone or double-stranded RNA.[25] When modeling protein–RNA interactions, many factors should be considered, including secondary structures, folding and characteristics of binding interfaces.[26] Especially physicochemical properties of amino acids, such as structural disorder and polarity, are relevant to characterize the RNA-binding ability of proteins.[27,28] Indeed, recent studies reported that in addition to classical RNA-binding domains other regions found in ribosomal proteins, translation elongation factors, zinc fingers as well as structurally disordered parts participate in contacting transcripts.[29]

Here we review the most recent experimental and computational advances for the detection of protein–RNA interactions and introduce new challenges for future developments in the field.

EXPERIMENTAL METHODS FOR DETECTION OF PROTEIN–RNA INTERACTIONS

Transcripts are never naked and form complexes with partner proteins in ribonucleoprotein particles from their birth to their degradation (Figure 1).[30] The assembly of functional complexes and the delivery to final destination involves progression through a series of intermediate complexes and subcellular localizations. For instance, Cajal bodies are sites of noncoding ribonucleoprotein particles maturation, where assembly factors gather to accelerate complicated biochemical reactions. Complexes containing mRNAs often undergo remodeling as they travel from the site of transcription to the cytoplasm where they are translated.

The advent of sequencing technologies, together with the introduction of various cross-linking chemistries, has enabled the development of new high-throughput methods for the simultaneous detection of hundreds to thousands interactions in a single experiment. The methods can be classified into 'protein-centric' approaches, which reveal RNAs bound to a known protein, and 'RNA-centric' approaches, which characterize proteins interacting with an RNA of interest.

Protein-Centric Approaches

The predominant *protein-centric* methods are based on protein immunoprecipitation from cell lines or tissues, and detection of the co-purified RNAs (Table 1 and Figure 2).

RNA immunoprecipitation (RIP) implies the purification of RNA–protein interactions in native conditions by using a protein-specific antibody, and detection of interacting RNAs by either microarray (RIP-chip) or sequencing (RIP-seq).[31,51,52] Despite the genome-wide potential, the method has practical limitations. Indeed, it is prone to detect nonspecific interactions due to the nonphysiological formation of protein–RNA complexes in solution. As a consequence, interactions identified using native purification methods often require additional validation.

RIP limitations have been overcome with the introduction of cross-linking and denaturing methods, namely cross-linking and immunoprecipitation (CLIP). CLIP combines UV cross-linking of RBPs to their cognate RNA molecules with stringent purification of protein–RNA complexes that are resolved and size-selected on an SDS-PAGE before proceeding to high-throughput sequencing of cDNA library (HITS-CLIP).[32] Photoactivable ribonucleoside enhanced CLIP (PAR-CLIP) and individual-nucleotide resolution CLIP (iCLIP) are approaches derived from modified cross-linking or library-preparation protocols that allow the identification of cross-linking sites at a single-nucleotide resolution.[32,33,53,54]

CLIP has enabled the characterization of RNA binding profiles of several RBPs involved in neurologic disorders and cancer, thus helping in understanding their role in disease. The first CLIP experiment,

TABLE 1 | List of Experimental Methods for the Identification of Protein–RNA Interactions

Method			Advantages	Challenges	Disease-Related RBP/RNA	References
Protein-centric	***In vivo***	RIP	Genome-wide Applicable to tissues	High background noise Possible artifacts Low resolution	ELAVL1 (epilepsy, cancer) CELF1 (myotonic dystrophy) IGF2BP1 (cancer) ADAR (Dyschromatosis symmetrica, spastic paraplegia)	Tenenbaum et al.[31]
		HiTS-CLIP	Genome-wide	False negative due to low cross-linking efficiency Time consuming	CELF4 (epilepsy, hyperactivity)	Licatalosi et al.[32]
		PAR-CLIP	High specificity		ELAVL1 (epilepsy, cancer)	Hafner et al.[33]
		iCLIP	High resolution (binding sites)	Challenging set-up	FMR1 (Fragile-X mental retardation, autism spectrum disorders)	Konig et al.[20]
		eCLIP		Not applicable to tissues (PAR-CLIP)	FUS (FTLD, ALS) MBNL1/2 (myotonic dystrophy) NOVA1/2 (POMA) PARK7 (Parkinon's disease) TARDBP (FTLD, ALS) RBFOX (autism)	Van Nostrand et al.[34]
	In vitro	RNA-compete	Large-scale	Nonphysiological conditions	ELAVL4	Ray et al.[35]
		SEQRS	Measurement of interaction affinity and specificity Analysis of multiple proteins		NOVA1 (POMA)	Campbell et al.[24]
		RBNS			HNRNPA0 (cancer)	Lambert et al.[36]
		RNA-MaP			MBNL1 (myotonic dystrophy)	Buenrostro et al.[37]
		HiTS-RAP			CELF1 (myotonic dystrophy)	Tome et al.[38]
		MITOMI				Martin et al.[39]
RNA-centric	***In vitro***	TRAP/RAT	Relatively easy and flexible	High background noise		Hogg and Collins[40]
		RaPID	Easy purification protocols	Possible perturbation of RNA folding		Slobodin and Gerst[41]
		RiboTrap		Large amounts of starting material		Michlewski and Cáceres[42]
		RNA-assisted chromatography				
		Protein microarray	Large-scale Relatively easy and fast	Possible artifacts due to nonphysiological conditions (protein folding, accessibility, post-translational modifications, etc.,)0	*TP53* *HRAS* *MYC* (*cancer*) *BCL2* PWRN1 (Prader Willi syndrome)	Scherrer et al.[43] Siprashvili et al.[44]

TABLE 1 | Continued

Method		Advantages	Challenges	Disease-Related RBP/RNA	References
In vivo	MS2-BioTRAP	Fast and easy set-up Large-scale	Challenges associated with cell transfection		Tsai et al.[45]
		Low amounts of starting material	Possible artifacts due to tag		
	ChIRP CHART RAP-MS	Study of protein–RNA interactions under physiological conditions	Time and cost consuming Large amounts of starting material required Challenging set-up	MALAT1 (cancer) NEAT1 (cancer)	Chu et al.[46] Simon[47]
	Interactome capture	Study of protein–RNA interactions under physiological conditions	Possible artifacts (positive and false negative) due to cross-linking		Castello et al.[29]
		Identification of unknown RBPs	Time consuming Challenging set-up		

RIP, RNA Immunoprecipitation, HiTS-CLIP, High-throughput sequencing of RNA isolated by cross-linking immunoprecipitation, PAR-CLIP, Photoactivable ribonucleoside enhanced CLIP, iCLIP, individual-nucleotide resolution Cross-Linking and ImmunoPrecipitation, SEQRS, *in vitro* selection high-throughput sequencing of RNA and sequence specificity landscape, RBNS, RNA Bind-n-Seq, RNA-Map, RNA on a massively parallel array, HiTS-RAP, High-throughput sequencing RNA affinity profiling, RNA-MITOMI, RNA mechanically induced trapping of molecular interactions, TRAP, Tandem RNA affinity purification, RAT, RNA affinity in tandem, RaPID, RNA-binding protein purification and identification, MS2-BioTRAP, MS2 *in vivo* biotin tagged RNA affinity purification, ChIRP, Chromatin isolation by RNA purification, CHART, Capture hybridization analysis of RNA targets, RAP-MS, RNA antisense purification-mass spectrometry.

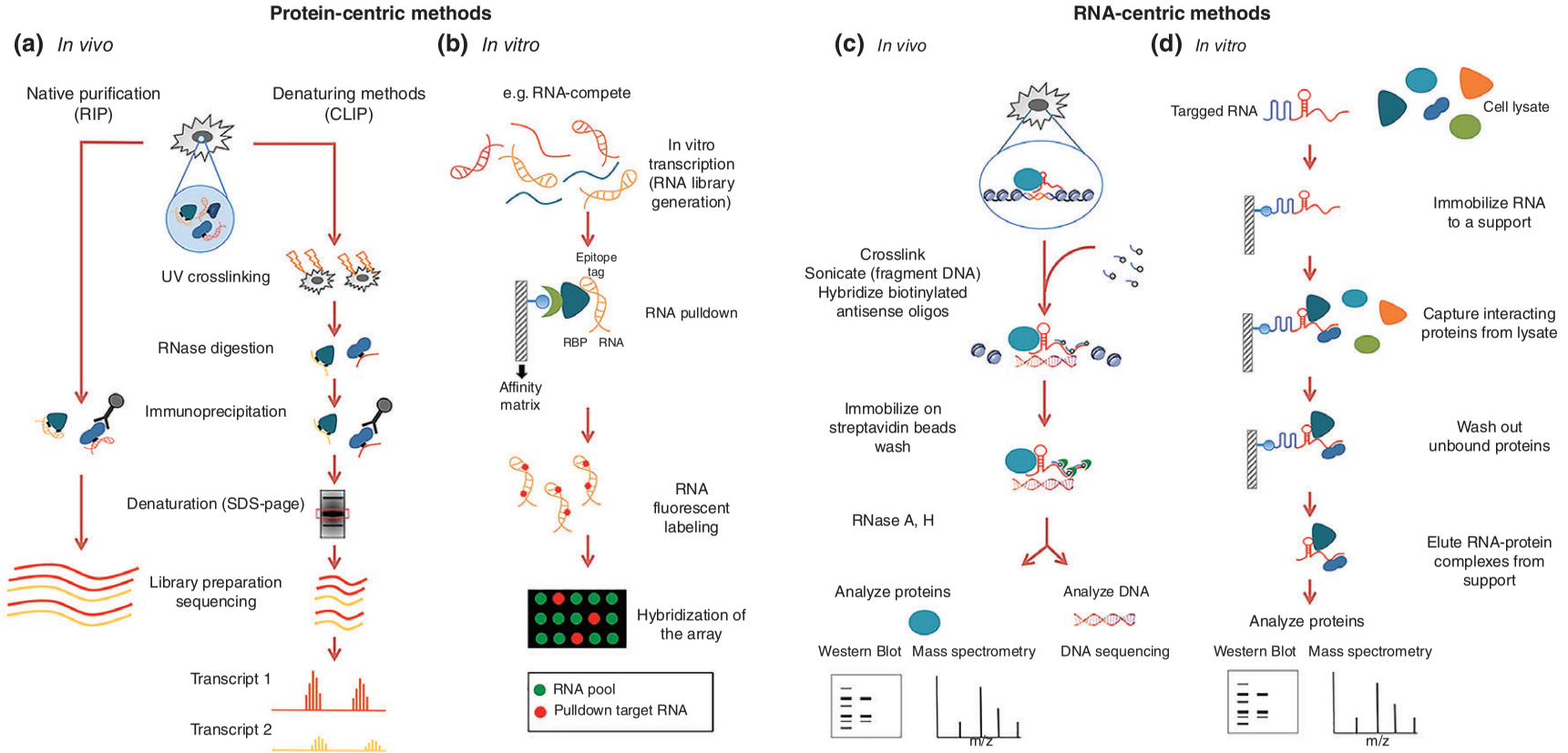


FIGURE 2 | Schematic representation of experimental methods for the identification of protein–RNA interactions. **Protein-centric methods**. (a) *In vivo* approaches include native purification protocols (RNA Immunoprecipitation, RIP) and denaturing protocols (Cross-linking and Immunoprecipitation, CLIP). In the fist case, RNAs bound to a specific protein are immunoprecipitated from cell lysate in native conditions by using a protein-specific antibody and, after wash and protein removal by proteinase K treatment, RNAs are reverse transcribed and identified through RNA sequencing.[48] In the second case, cells are UV cross-linked to 'freeze' protein–RNA complexes. RNA is digested to obtain fragments of a defined size and the obtained complexes are immunoprecipitated and resolved on an SDS-PAGE. After isolation from membrane and proteinase K digestion, the RNA fragments are reverse transcribed and sequenced.[49] In red and yellow are represented simulated enrichment values of a target and nontarget RNA respectively (lower part of the panel). (b) As an example of *in vitro* approaches, the schematic workflow of **RNA compete** is represented.[35] RNA libraries are generated by *in vitro* transcription. Transcripts are incubated with the protein of interest immobilized on an affinity matrix (e.g., streptavidin-biotin tag system) and the bound fragments are then fluorescently labeled and detected by hybridization on a microarray platform. **RNA-centric methods**. (c) *In vivo* approaches for the identification of proteins bound to an RNA of interest often derive from methods used for the identification of genomic DNA *loci* targeted by noncoding RNAs. Cells are cross-linked and lysed. Chromatin is sheared and protein–RNA–DNA complexes are pull-down by using biotinylated oligos complementary to the sequence of the RNA. After RNA digestion, proteins can be identified by western blot or mass spectrometry analysis.[46] (d) *In vitro* approaches commonly exploit the use of RNA tags to immobilize the RNA of interest onto an affinity matrix. Upon incubation with cell lysate, proteins bind to the immobilized RNA. After washes, the protein–RNA complexes are eluted from the matrix and proteins are characterized by western blot or mass spectrometry.[50]

published by Ule et al., shed light on the role of NOVA 1 and 2 proteins in paraneoplastic opsoclonus-myoclonus-ataxia (POMA), an autoimmune neurologic disease characterized by abnormal motor inhibition.[55] NOVA proteins were found to regulate alternative splicing of RNAs encoding multiple components of inhibitory synapses, such as GABAβ 2 receptor and GIRK2, which mediate slow inhibitory postsynaptic potentials, the K+ voltage-gated channel KCNQ3 or the nicotinic acetylcholine receptors β2 and α2, highly represented in the GABAergic interneurons.[55]

The combination of RIP-chip and CLIP approaches has allowed the identification of targets of the mammalian ELAVl family whose four members, ELAVL 1,2,3 and 4, (also known as HuR, HuB, HuC, and HuD) are implicated in different cancers and neurological diseases.[56,57] In particular, ELAVL1, the only member of the family expressed in both neuronal and non neuronal cells, has been found particularly abundant in breast, ovary, colon, and brain cancers[58] associated with poor prognosis.[59] RIP-chip and CLIP have identified a transcriptome-wide list of targets (including Cyclin D1, E1, A2 and B1, EGF, c-Myc, p27, COX-2, and BRCA1) and indicated that ELAVL1 functions as a major hub for regulating RNA metabolism in the cell at different levels, from pre-mRNA alternative splicing to mature mRNA stability and microRNA biogenesis.[57,60] The other three members, ELAVL2–4 (nELAVl), are highly enriched in neurons and show a unique hierarchical expression during cortical development.[61] HITS-CLIP on mice cortical tissue has revealed that nELAVl regulated transcripts are mainly involved in synaptic cytoskeleton assembly and disassembly, amino acid and sugar biosynthetic pathways. In particular, nELAVl have been found to regulate the alternative splicing of the gene coding for the glutaminase enzyme, the major responsible for the synthesis of the excitatory neurotransmitter glutamate.[56]

The complex protocol and bioinformatics required for data analysis represent the main disadvantage of CLIP approaches (Figure 2). As a matter of fact, high experimental failure rates are reported. Recently, an enhanced CLIP (eCLIP) protocol has been developed.[34] In addition to reducing hands-on time to as few as 4 days, eCLIP dramatically decreases required library amplification by 1000 fold and enhances the rate of success at generating libraries with high usable reads percentages across diverse RBPs, maintaining the single-nucleotide resolution of previous methods. In addition, paired-matched input controls improve the signal-to-noise ratio for the discovery of authentic binding sites. 102 eCLIP experiments for 73 diverse RBPs have been generated (available at https://www.encodeproject.org), providing an unprecedented source of data for the study of protein–RNA functional networks. Despite its successful applications, a major concern when using CLIP is that UV-induced cross-linking is still poorly understood at the biophysical level. Only a very small percentage (1–5%) of protein–RNA complexes present in the cells can be efficiently cross-linked, and it is not clear which types of interactions might be unseen.[62] For instance, several RBP families do not directly associate with nucleic acid bases but interact with other elements (i.e., the sugar phosphate backbone) showing low cross-linking efficiency. Moreover, UV only cross-links direct protein–RNA interactions but it does not capture interactions occurring with protein complexes, thus providing just part of the information.

The study of *in vivo* interactions is limited to contacts formed in a certain cell type and at a specific time point. To better understand the physicochemical properties controlling protein–RNA interactions, a number of new methods allow the *in vitro* screening of interactions between proteins and libraries of randomly generated RNA sequences by combining the use of microarray and microfluidic platforms with molecule fluorescent labeling and RNA sequencing technologies. While *in vitro* evolution **SELEX** has bias towards highest-affinity targets,[36] other methods enable the characterization of lower-specificity and medium-range affinities with single proteins or multiprotein complexes.[63] Incubation of RNA libraries with a protein of interest immobilized on an affinity matrix is followed by fluorescent labeling of selected RNAs and hybridization to a microarray, in the case of **RNA-compete** (Figure 2), or deep sequencing in the case of *in vitro* selection high-throughput sequencing of RNA and sequence specificity landscape (**SEQRS**) and RNA Bind-n-Seq (**RBNS**).[25,36] These methods are often applied to identify *consensus* elements of RBPs (see also *Computational Methods for Detection of Protein–RNA Interactions*).

Quantitative analysis of RNA on a massively parallel array (RNA-MaP) and high-throughput sequencing RNA affinity profiling (HiTS-RAP), consist in the *in situ* synthesis of RNA libraries inside an *Illumina* sequencing flow cell followed by incubation with a fluorescently labeled protein and quantification of molecular interactions.[37,38] These approaches, together with RBNS, provide quantitative measurements of dissociation constants (Kd) through the use of multiple protein concentrations. Moreover, methods relying on the use of fluorescent proteins, such as

the previously mentioned RNA-MaP and HiTS-RAP, as well as the RNA mechanically induced trapping of molecular interactions (RNA-MITOMI),[39] provide effective visualization of multiple proteins simultaneously, thus revealing the effects of protein partners. These approaches have great range of applicability, especially considering that RNA metabolism is regulated by multiple RBPs to form functional particles (e.g., small nuclear ribonucleoproteins, telomerases, ribosomal subunits, UTR-regulatory complexes, etc).

RNA-Centric Approaches

RNA-centric methods aim to identify RBPs targeting a single RNA of interest. The majority of the existing methods exploit tagged RNAs as a bait to capture and characterize all proteins bound to it by mass spectrometry analysis (Table 1 and Figure 2). *In vitro* synthesized RNA can be chemically tagged through the incorporation of modified ribonucleotides that contain biotin, fluorescent dyes, or, alternatively, natural or artificial aptamers (e.g., S1, D8, MS2 hairpin loop, etc.,) can be incorporated.[40–42] In addition to their *in vitro* application (e.g., TRAP, RAT, RaPID, RiboTrap, RNase-assisted RNA chromatography), RNA tagging system can be applied *in vivo*, as in the example of MS2-BioTRAP method, where a MS2 hairpin loop tagged RNA and the bacteriophage MS2 coat protein fused to a protein tag (e.g., HB or streptavidin), are co-expressed to capture the RNA of interest by exploiting the high affinity interaction between the MS2 protein and the MS2 hairpin loop.[45] All these methods are relatively flexible, and their *in vivo* applicability allows the study of protein–RNA interactions in physiological conditions. Nevertheless, the incorporation of a tag into the RNA bait may alter its secondary structures and possibly the formation of ribonucleoprotein complexes. In addition, MS2-BioTRAP is only applicable in easy-to-transfect cells and the overexpression of at least one of the two interacting molecules might lead to experimental artifacts.

Capture hybridization analysis of RNA targets (CHART), Chromatin isolation by RNA purification (ChIRP) and RNA antisense purification (RAP) are methods designed to identify DNA and proteins targeted by noncoding-RNAs *in vivo* (Figure 2). These approaches involve cell cross-linking and pull-down of the RNA of interest using short biotinylated oligodeoxyribonucleotides that are complementary to the endogenous RNA. After reversion of cross-linking, the RNA-associated DNA and proteins are identified by sequencing and mass spectrometry (MS) analysis, respectively.[46,47,64,65] The design of antisense oligonucleotides with high affinity to accessible single-stranded regions of RNAs is often a challenging step in these approaches. Moreover, these protocols are time consuming and usually require large amounts of starting material in order to generate sufficient product for MS.

CHART was first applied for the characterization of protein interactors of the human long noncoding RNAs NEAT1 (nuclear-enriched abundant transcript 1) and MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) that localize to nuclear speckles and paraspeckles, respectively.[66] Speckles and paraspeckles are nuclear bodies in close proximity to each other, comprising many RBPs, including splicing factors. Interestingly, paraspeckles have been implicated in sequestering RNAs that respond to cellular stress. Although NEAT1 and MALAT1 have specific binders, the overlap of their protein interactome suggested potential redundancy or cooperation in regulating nuclear organization around nuclear bodies.

ChiRP and RAP-MS have been applied to the discovery of the protein interactome of *Xist* (X-inactive specific transcript), a lncRNA required for X chromosome inactivation (XCI) of one of the two X chromosomes in female cells, enabling dosage compensation between XX females and XY males.[66] Through an RAP-MS approach, McHugh and colleagues recently characterized ten *Xist*-specific interactors in mouse embryonic stem cells (mESC).[67] One of them, SAF-A (scaffold attachment factor-A, also known as HNRNPU) was previously shown to be required for tethering *Xist* to the inactive X chromosome in differentiated cells, while five of these proteins are implicated in transcriptional repression, chromatin regulation, and nuclear organization. These results partially overlap with a ChiRP-MS based study conducted on the same year by Chu and colleagues, in which 81 *Xist* interactors were described.[68] These proteins are mainly involved in chromatin modification, nuclear matrix and RNA remodeling pathways. Notably, this analysis reveals two sets of proteins that interact with *Xist* in a developmentally regulated manner, shedding light on a potential step-wise assembly of *Xist* binding proteins from the pluripotent state to cell differentiation.

A number of other methods based on the *in vitro* or *in vivo* screening of protein libraries have been developed. As an example, protein microarrays have been recently used to test the binding of proteins with a specific RNA *in vitro*.[43,44] The RNA is fluorescently labeled and hybridized on a protein chip. This method allows the screening of thousands of proteins in less than a day, and relatively small

amounts of RNA are required. On the other hand, the quality of the data is less robust because it strongly depends on the proteins conditions on the array (i.e., folding, RBDs accessibility, post-translational modifications, etc).

Finally, the interactome capture, a method developed in 2012 by Castello et al., allows the simultaneous recovery and characterization of the whole proteome associated with mRNAs in living cells.[29] This approach, originally applied to HeLa and Hek293 cells,[69] and lately to mouse embryonic stem cells[70] and yeast,[71] led to the discovery of new potential RBPs lacking the canonical RNA-binding domains but enriched in features such as protein disorder and repetitive low complexity amino acid regions. The majority of these candidate RBPs have also enzymatic activity, and, while the RNA binding capacity of some of them has been validated, further evidence should be provided for the large majority of these proteins.

Future Experimental Challenges

Despite the enormous progress done in the last few decades, it is clear that each of the experimental methods offers a partial, and in some cases potentially erroneous, view on protein–RNA interactions. In this scenario, two main questions arise: to what extent *in vitro* interactions are relevant *in vivo*? Does *in vivo* binding imply functionality?

As regulatory regions interact with competing RBPs and RNAs, only a small percentage of *in vitro* determined binding sites is actually occupied *in vivo*. Very often, binding regions become inaccessible due to mRNA localization into granular assemblies (see section *Beyond the Protein–RNA Complex: Membrane-Less Organelles*).[72] It should be also mentioned that RNAs lacking annotated biding sites are found to interact *in vivo* because of indirect interactions or as a result of molecule sequestration in subcellular compartments, where high concentrations favor low affinity or unspecific interactions.[71]

Most importantly, *in vivo* occupancy not always indicates functionality. As a matter of fact, to unravel the complexity of the RNA biology, it will be essential to develop new tools for the integration of interaction data with global functional assays, such as for instance Ribosome profiling.[73]

COMPUTATIONAL METHODS FOR DETECTION OF PROTEIN–RNA INTERACTIONS

A large number of computational methods address the problem of characterizing RNA partners of specific RBPs. **MEME**,[74] **RBPmap**,[75] **SeAMotE**,[76] and **RNAcontext**[77] identify motifs enriched in RNA targets.[26] The use of sequence motifs allows discovery of targets in RNA datasets, but *ab initio* predictions of RNA interactions are not possible without previous experimental knowledge on the RBPs of interest.

A different class of computational approaches aims to identify RNA-binding residues[78–80] and RNA-binding regions using primary-, secondary- or tertiary-structure information.[81–83] Methods such as **Struct-NB**,[84] **PRIP**,[85] **SPOT-Seq**,[86] and **OPRA**[87] predict the RNA-binding ability by identifying regions in the protein surface that accommodate nucleotide chains. As three-dimensional structures are needed to perform the calculations, these approaches are limited by the existence of available templates. Yet, using a library of 1164 nonredundant protein–RNA complexes (95% sequence identity cutoff) and the folding recognition technique SPARKS X[88] the **SPOT-Seq-RNA** approach[89] has been used to characterize 2418 novel RBPs in the human proteome of which 291 are reported in interactome capture experiments[29] (see RNA-centric methods in section *Experimental Methods for Detection of Protein–RNA Interactions*).

Algorithms relying on primary-structure features have a clear advantage over tertiary-structure methods that require three-dimensional references to compare structures[87] and atomistic details[90] to study interactions. Methods based on primary-structure exploit evolutionary information (i.e., conservation of specific residues in sequence alignments), secondary-structure propensities and information on physicochemical properties of amino acids (e.g., hydrophobicity). The binding elements are often classified using machine-learning methods such as Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB). For further information about the algorithms, we refer the Reader to reviews describing features and techniques in detail.[81–93] In the following text we will provide a short description of the most used algorithms (Table 2) and their published performances.

BindN[68] predicts DNA- and RNA-binding residues through a SVM trained on hydrophobicity, dissociation constant (pKa), and molecular mass of amino acids. The method was validated on a dataset of 100 protein–RNA complexes characterized through X-ray crystallography and nuclear magnetic resonance (NMR), and shows a cross-validation accuracy of 0.69. Using a slightly different codification of physicochemical properties and integrating them with evolutionary information, the **BindN+** algorithm[78] reaches an accuracy to of 0.78. The evolutionary information is derived from a position-

TABLE 2 | List of Computational Methods for the Identification of Protein–RNA Interactions

Prediction	Examples	Advantages	Disadvantages	References
Binding motif (RNA)	MEME SeAMotE	*de novo* binding site discovery	High-throughput data are required as input Sequence complexity is a limitation	Bailey et al.[94] Agostini et al.[76]
Binding residue	Pprint BindN+ RNAbindR+	Evolutionary information	RNA-binding domains cannot be identified	Kumar et al.[79] Wang et al.[78] Walia et al.[80]
Domain (protein)	HMMER	Domain recognition	Annotation of RNA-binding domains are required	Finn et al.[95] Livi et al.[83]
	*cat*RAPID *signature*	Annotation of RNA-binding domains are not required	Single amino acid resolution has not been implemented	
RNA–protein interaction	*cat*RAPID RPISeq	Runs on high-throughput data High sensitivity	RNA < 1200 nt Protein < 750 aa Low specificity Max 100 sequences per run	Bellucci et al.[96] Agostini et al.[97] Muppirala et al.[98]

specific scoring matrix (PSSM) generated by PSI-BLAST,[99] that searches against a nonredundant database of protein sequences. With the same approach, **Pprint**[100] combines evolutionary information from PSSMs with SVMs, thus predicting RNA-binding sites with an accuracy of 0.81. Similarly, **RNA-ProB**[101] employs a smoothed PSSM encoding scheme to predict RNA-binding sites on proteins and shows an accuracy of 0.89. Slightly different types of physicochemical properties are combined in **RNABindR**,[27] that exploits accessible surface area, sequence specific entropy, hydrophobicity, secondary structure propensities, and electrostatics. In the testing phase, RNA-BindR identifies interface residues with an accuracy of 0.85. Adding the contribution of **HomPRIP**,[55] a sequence homology-based method for prediction of RNA-binding sites, **RNABindR+** shows an accuracy of 0.83 on a larger benchmark set of 200 proteins.[80] The working principles of these algorithms are based on different assumptions: the conservation, physicochemical properties or topological properties of the binding site. Hence, it is challenging to compare the algorithms; their applicability depends on the kind of dataset that needs to be analyzed. For more details, the Reader can refer to comparative studies that describe the tools in more detail.[92,102]

While BindN,[68] Pprint,[100] RNAProB,[101] and RNABindR[27] predict the RNA-binding ability of individual amino acids without considering the sequence context, **HMMER**[82] is designed to perform homology searches in protein sequences using Hidden Markov Models and multiple sequence alignments of domain families. HMMER functionality is limited to previously annotated domain-families, as the tool does not perform *de novo* detection of binding-domains. A recent method, ***cat*RAPID *signature***,[83] identifies RNA-binding regions by considering physicochemical properties that are present in known RBPs.[83] *cat*RAPID *signature* exploits properties such as hydrophobicity, secondary structure, disorder, and burial.[83] Each feature defines a unique *signature*, or profile, containing position-specific information arranged in sequential order from the N- to the C-terminus.[28,103] In addition to the RNA-binding score, *cat*RAPID *signature* predicts regions contacting RNA. On a test set of mouse proteins harboring noncanonical RNA-binding domains,[70] *cat*RAPID *signature* shows an accuracy of 0.71 in predicting the RNA-binding ability. When applied to newly discovered RBPs, the algorithm discriminates RBPs from nonRBPs with an area under the receiver operating curve of 0.76.[29,69] In Table 3, we show the performances of catRAPID *signature*,[83] BindN+,[78] Pprint,[100] and RNABindR+[80] to identify nonclassical RNA-binding domains.[70]

TABLE 3 | Performances on Detecting RNA-Binding Regions

Method	ACC[a]	sens[b]	spec[c]	prec[d]
***cat*RAPID**	0.67	0.76	0.60	0.65
BindN+	0.38	0.37	0.39	0.38
PPrint	0.47	0.49	0.45	0.49
RNAbindR+	0.48	0.53	0.42	0.48

We analyzed 102 proteins containing nonclassical RNA-binding domains and 102 without annotated RNA-binding domains with *cat*RAPID *signature* and three other algorithms.[27,68,83,100] The performances are measured using **a**. accuracy, **b**. sensitivity, **c**. specificity, and **d**. precision.

We also compared the prediction performances of BindN,[93] BindN+,[78] Pprint,[79] RNAproB,[101] RNABindR+,[80] and *cat*RAPID *signature*[83] using structural data from 90 folds, 126 families and 100 superfamilies retrieved from the SCOP database[102] (Figure 3). In all the classifications, *cat*RAPID *signature* and BindN+ show the best performances[102] recognizing contacts identified at a distance cutoff of 3.5 Å. We note that while BindN,[93] BindN+,[78] Pprint,[79] RNAproB,[101] and RNABindR+[80] have been designed to predict the RNA-binding ability of individual residues in protein sequences, *cat*RAPID *signature* identifies binding regions (Table 2).

Only a few methods address the question whether a protein of interest interacts with a specific RNA exploiting the properties of both interaction partners. *cat*RAPID[96,97] estimates the interaction potential through van der Waals, hydrogen bonding and secondary structure propensities of both protein and RNA sequences (total of 10 features per molecule)[81] and allows identification of binding partners with high confidence (training accuracy = 0.89; testing accuracy = 0.78; see also Cirillo et al.[91]; Table 2). *cat*RAPID identified the interaction of the N-terminus of Fragile Mental Retardation Protein FMRP[104] with dendritic nontranslatable brain cytoplasmic RNA BC1[105] before crystallographic studies

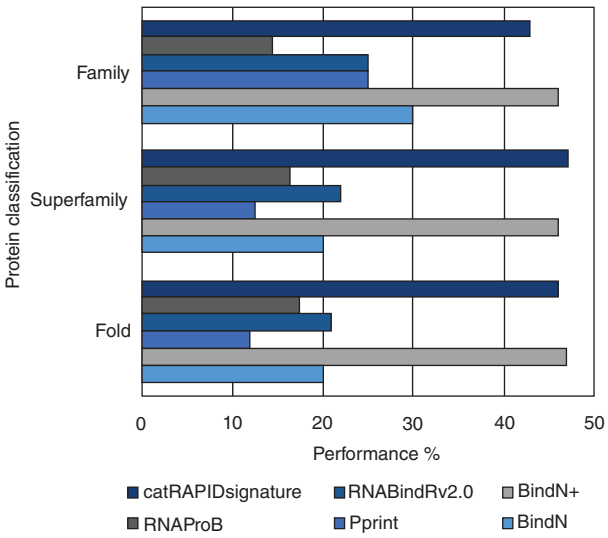


FIGURE 3 | Methods to predict RNA binding sites. We calculated performances of BindN,[93] BindN+,[78] Pprint,[79] RNAproB,[101] RNABindR+,[80] and *cat*RAPID *signature*[83] on a set of proteins whose RNA-binding sites have been validated through X-ray and NMR techniques. As in a recent work,[102] three protein classes 'fold,' 'family,' and 'superfamily' were retrieved from SCOP.[102] Performances were estimated using the formula (sensitivity + specificity)/2 on 90 folds, 126 families, and 100 superfamilies (details at http://service.tartaglialab.com/static_files/shared/documentation_signature.html).

determined the presence of tandem Agenet and KH domains.[106] Similarly, *cat*RAPID predictions for the physical binding of heterogeneous nuclear ribonucleoprotein U SAF-A with the long noncoding *Xist*[107] have been recently confirmed by ChiRP-MS experiments.[68] In addition, *cat*RAPID can be used to perform proteome- and transcriptome-wide calculations,[97] which is useful to complement experimental approaches, such as interactome capture[29] (see also *Experimental Methods for Detection of Protein–RNA Interactions*) and to identify target partners.[108] Owing to the complexity of the conformational space, a *fragmentation* procedure,[107,109] based on division of polypeptide and nucleotide sequences into overlapping regions, is used for RNAs longer than 1000 nucleotides and proteins longer than 750 amino acids.

Another method called **RPISeq** predicts interactions combining protein and RNA features in SVM and RF classifiers.[98] In RPISeq, RNA sequences are encoded by frequency of nucleotide tetrads (i.e., 4-mer combinations of [A,C,G,U]), while protein sequences are represented using 3-mer of 7 amino acid types ([A,G,V], [I,L,F,P], [Y,M,T,S], [H,N,Q,W], [R,K], [D,E], and [C]). More specifically, the RNA is represented by a $4 \times 4 \times 4$ or 256-dimensional vector, in which each feature represents the 4-mer normalized frequency appearing in a RNA sequence (e.g., CCAU, AUUG, and GACA). The protein instead is encoded by a $7 \times 7 \times 7$ or 343-dimensional vector, where each element of the vector corresponds to the normalized frequency of the corresponding triple of amino acids in the sequence. The protein and RNA vectors serve as input for the SVM and RF to predict whether the protein–RNA pair interacts. RPISeq shows significantly high performances in the training phase (accuracy = 0.89) and high sensitivity / low specificity when applied to other sets.[81,92,106]

Future Computational Challenges

*cat*RAPID and other computational tools[96,98] predict the interaction propensity of a protein–RNA pair to interact, dismissing other proteins involved in the physical binding. As a matter of fact, RBPs often form complexes and bind together to their target RNAs (see also 'RIP' section in *Experimental Methods for Detection of Protein–RNA Interactions*).[110] A recent computational approach developed to predict transcription factor binding sites suggests that additional information retrieved from protein–protein interactions improves the performances of current predictions.[111] Thus, as done for DNA-

binding proteins, integration of protein–protein network layers[112] could boost considerably interaction predictions. We envisage that network-based methods will shed light on the complex life of newly discovered RBPs, their dynamics and role in human diseases.[113]

A number of reports indicate that RBPs with many partners are enriched in structural disorder.[114,115] The lack of stable tertiary structure increases the ability of RBPs to interact with multiple partners[116] and promotes the formation of transient ribonucleoprotein complexes, which are discussed in section *Beyond the Protein–RNA Complex: Membrane-Less Organelles*. Moreover, some RBPs have been found to be enriched in glycine, arginine, lysine, and tyrosine motifs[29] as well as a number of other patterns[71] that could promote transient interactions. Thus, to reach a complete and biologically relevant understanding of protein–RNA interactions, complex formation should be studied as a dynamic process. Data based on techniques such as surface plasmon resonance could allow measurement of interactions in real time providing both equilibrium and kinetic information for the development of new methods.[16]

BEYOND THE PROTEIN–RNA COMPLEX: MEMBRANE-LESS ORGANELLES

RBPs and RNAs interact through different types of contacts that facilitate not only the formation of oligomeric complexes but also other types of dynamic assemblies.[31] Interestingly, a number of RBPs associate through weak contacts present in structurally disordered regions[117] to trigger a process known as liquid-to-liquid phase separation.[118] Through liquid-liquid demixing, ribonucleoprotein assemblies form distinct compartments in the cytoplasm and nucleoplasm[119] that *in vivo* can act as membrane-less organelles.[5–7] The liquid nature of these assemblies allows rapid diffusion favoring chemical reactions on biological timescales.

Two of the most common macromolecular assemblies are processing-bodies (P-bodies) and stress granules.[7,120] These assemblies are associated with mRNA degradation and storage, respectively, and are built in response to different environmental *stimuli*.[8] P-bodies and stress granules are conserved throughout evolution and appear in yeast, plant, nematode, fly, and mammalian cells.[121]

Ribonucleoprotein assemblies have special physicochemical properties that can be investigated using computational approaches such as the *multiclever*Machine.[122,123] With respect to the yeast proteome (globular proteins with sequence redundancy < 40%), the *multiclever*Machine predicts that proteins forming P-bodies and stress granules are depleted in hydrophobicity and enriched in structural disorder and RNA-binding ability (p-values < 10^{-5}; Fisher's exact test; Figure 4), which is in agreement with experimental evidence.[72,126] Hydrophobicity and structural disorder highly discriminate the two sets (areas under the ROC curve AUC in the range of 0.70–0.80; calculations available at http://www.tartaglialab.com/cs_multi/confirm/1207/c51ef6cff3/).

An important property of the membrane-less organelles is the wide *spectrum* of structural states,[126] which is promoted by the different type of interactions between protein and RNA components. The RNA content influences viscoelasticity of assemblies, as well as exchange rate with the surrounding *milieu*, fusion and fission kinetics.[127] As a matter of fact, the protein–RNA contacts can be quite diverse: strong to stabilize RNAs from synthesis to degradation or weak to bind RNAs at a precise moment and place.[128] A better understanding of the physical characteristics of ribonucleoprotein components is crucial to investigate their function and dysfunction.

RBP Assemblies and Disease

Current literature indicates that the structure of a ribonucleoprotein assembly depends on a delicate equilibrium between protein and RNA components that is regulated by the quality control machinery (e.g., chaperones).[129] Incorrect assembly/disassembly of such macrostructures can jeopardize molecular cell homeostasis. Intriguingly, RBPs found in liquid phase-separated assemblies are enriched in structural disorder that is also associated with a high risk of misfolding and the formation of toxic protein aggregates.[126,129] Neurons appear to be especially susceptible to failures in proteostasis, a major source of neurodegenerative diseases. Indeed, several dementias and motor-neuron diseases are associated with accumulation of disordered RBPs, such as TAR-DNA-binding protein 43 in Amyotrophic Lateral Sclerosis (ALS) or Ataxin 1 in Ataxia.[130]

In this context, chaperones perform the important task to keep protein homeostasis and prevent disease. They are upregulated upon cellular stress (e.g., diabetes-induced [glyc]oxidation stresses) to limit the accumulation of damage[131] via folding assistance, blocking aberrant interactions, disaggregating proteins, and facilitating protein degradation.[132–136] We speculate that the ability to form phase-separated

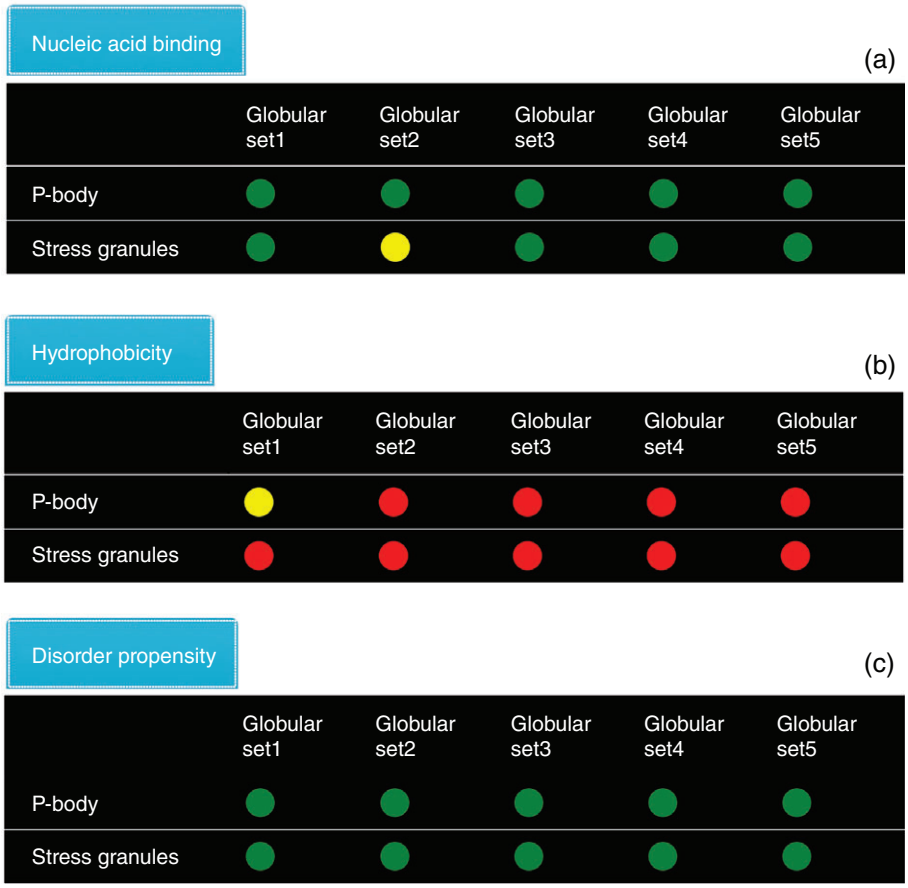


FIGURE 4 | Features of yeast proteins forming ribonucleoprotein assemblies. Using physicochemical properties, the *multiclever*Machine approach[122,123] discriminates P-bodies and stress granules from other globular proteins. The datasets employed in this analysis comprise 52 P-body and 62 stress granule proteins, previously reported in experimental works,[124] as well as five random sets of 100 globular proteins from Astral SCOPe 2.05 (<40% sequence identity).[125] Three specific properties are reported: (a) nucleic acid binding, (b) hydrophobicity, and (c) disorder propensity. For each feature, enrichment or depletion of a set is indicated with a specific color: green indicates that proteins contained in P-bodies or stress granules are enriched with respect to proteome subsamples; red means depletion with respect to the proteome subsamples; yellow indicates no significant differences between the sets. As observed in previous experimental works,[63,104] proteins found in P-bodies and stress granules are more structurally disordered and prone to bind nucleic acids as well as less hydrophobic (p-values $<10^{-5}$; Fisher's exact test). Sets and results can be accessed at http://www.tartaglialab.com/cs_multi/confirm/1207/c51ef6cff3/ where additional statistical analyses are also reported.

ribonucleoprotein assemblies requires a tightly regulated control machinery, that in stress conditions and aging can be compromised and drive to formation of toxic aggregates. Thus, chaperones can be applied to develop therapeutic interventions, from controlling the initial aggregative events in the ER, to blocking the assembly of cytotoxic aggregates.

Future Challenges in the Characterization of RBP Assemblies

A major challenge in the future will be to identify the components of each ribonucleoprotein assembly and pinpoint the defects associated with RBP dysfunction.[128,137] Currently the characterisation of solid phase separated aggregates is possible.[138] However, the study of the liquid phase assemblies is more complex due to their fast exchange with the cell *milieu*, i.e., their rapid dissolution impeding its isolation.

Solid aggregates appear in many diseases, including Huntington's, Creutzfeldt-Jakob, and Alzheimer's disease. This suggests a common hallmark[139] and the existence of common pathways for pharmacotherapeutic targeting.[137,140] In this context, it is crucial to understand the cellular mechanisms controlling health conditions as well as the factors triggering toxicity. Interestingly, the metastable nature of the liquid assemblies offers the possibility to use endogenous biochemical pathways to reverse pathological states.[137]

As a matter of fact, prion protein neurodegeneration can be reduced by inhibition of RNA granules pathway and stimulation of protein synthesis.[140]

CONCLUSION

RBPs functions include protection, modification, and transport of transcripts to their translation or degradation sites (Figure 1). The scenario is particularly complex considering the fact that not only RBPs regulate RNAs but also RNAs regulate RBPs.[141,142] Experimental techniques aiming to reveal RNA–protein contacts will have to uncover more details of such interactions (see 'protein-centric' and 'RNA-centric' experimental methods). The study of RBPs will benefit from new computational approaches taking into account information of protein–protein networks (see section *Computational Methods for Detection of Protein–RNA Interactions*).

An important aspect to consider while building computational models is that many proteins exhibit RNA-binding activity without having canonical RNA-binding domains.[30,71,117] As observed for protein–protein interactions, structurally disordered regions bind with high specificity and low affinity.[143] Hence, noncanonical domains, which are enriched in structural disorder, could promote low affinity interactions with RNA molecules.[108] Accordingly, computational methods predict that the RNA-binding ability of structural disorder is more pronounced in noncanonical RBPs, indicating that unfolded regions promote RNA interactions.[108] With respect to the RNA-binding ability of full-length proteins, the contribution of disorder is high at low-interaction propensities, which suggests that a large number of transcripts can be targeted by noncanonical domains.[108] How many other RBPs lacking canonical binding domains exist in nature?[69,83] The *repertoire* of RBPs needs to be expanded in order to link transcriptomic with metabolomic properties,[144] which will involve the development of new computational strategies.[80,83]

Most importantly, future computational and experimental investigations will have to focus on noncanonical protein–RNA complexes.[145] Indeed, transient interactions, promoted by structurally disordered regions in RBPs, often induce phase-separation in the nucleus or cytoplasm (see section *Beyond the Protein–RNA Complex: Membrane-Less Organelles*).[6] The functionality of phase-separated protein–RNA assemblies is still under investigation but appearance of solid protein–RNA aggregates has been linked with devastating diseases, such as for instance ALS (see section *RBP Assemblies and Disease*).[137,140] Thus, studying protein–RNA assemblies will be essential to better understand the etiopathogenesis of specific diseases and to design new therapeutic strategies.

ACKNOWLEDGMENTS

The research leading to this work has received funding from the European Union Seventh Framework Programme (FP7/2007-2013), through the European Research Council, under grant agreement RIBOMYLOME_309545 (Gian Gaetano Tartaglia), and from the Spanish Ministry of Economy and Competitiveness (BFU2014-55054-P). We also acknowledge support from AGAUR (2014 SGR 00685) and the Spanish Ministry of Economy and Competitiveness 'Centro de Excelencia Severo Ochoa 2013–2017' (SEV-2012-0208).

REFERENCES

1. Morris KV, Mattick JS. The rise of regulatory RNA. *Nat Rev Genet* 2014, 15:423–437.
2. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet* 2014, 15:829–845.
3. Iwasaki YW, Siomi MC, Siomi H. PIWI-interacting RNA: its biogenesis and functions. *Annu Rev Biochem* 2015, 84:405–433.
4. Verschoor A, Srivastava S, Grassucci R, Frank J. Native 3D structure of eukaryotic 80s ribosome: morphological homology with *E. coli* 70S ribosome. *J Cell Biol* 1996, 133:495–505.
5. Courchaine E, Neugebauer KM. Paraspeckles: paragons of functional aggregation. *J Cell Biol* 2015, 210:527–528.
6. Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Jülicher F, Hyman AA. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 2009, 324:1729–1732.
7. Chen D, Huang S. Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells. *J Cell Biol* 2001, 153:169–176.

8. Giménez-Barcons M, Díez J. Yeast processing bodies and stress granules: self-assembly ribonucleoprotein particles. *Microb Cell Fact* 2011, 10:73.
9. Ward AJ, Cooper TA. The pathobiology of splicing. *J Pathol* 2010, 220:152–163.
10. Nussbacher JK, Batra R, Lagier-Tourenne C, Yeo GW. RNA-binding proteins in neurodegeneration: Seq and you shall receive. *Trends Neurosci* 2015, 38:226–236.
11. Castello A, Fischer B, Hentze MW, Preiss T. RNA-binding proteins in Mendelian disease. *Trends Genet* 2013, 29:318–327.
12. Lukong KE, Chang K, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. *Trends Genet* 2008, 24:416–425.
13. Setzer DR. Measuring equilibrium and kinetic constants using gel retardation assays. *Methods Mol Biol* 1999, 118:115–128.
14. Wilson, G.M., Sutphen, K., Chuang Ky, null, and Brewer, G. (2001) Folding of A + U-rich RNA elements modulates AUF1 binding. Potential roles in regulation of mRNA turnover. *J Biol Chem*, 276:8695–8704.
15. Walter NG, Hampel KJ, Brown KM, Burke JM. Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer. *EMBO J* 1998, 17:2378–2391.
16. Katsamba PS, Park S, Laird-Offringa IA. Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods* 2002, 26:95–104.
17. Martin D, Charpilienne A, Parent A, Boussac A, D'Autreaux B, Poupon J, Poncet D. The rotavirus nonstructural protein NSP5 coordinates a [2Fe-2S] iron-sulfur cluster that modulates interaction to RNA. *FASEB J* 2013, 27:1074–1083.
18. Schulz S, Doller A, Pendini NR, Wilce JA, Pfeilschifter J, Eberhardt W. Domain-specific phosphomimetic mutation allows dissection of different protein kinase C (PKC) isotype-triggered activities of the RNA binding protein HuR. *Cell Signal* 2013, 25:2485–2495.
19. Sunwoo H, Wu JY, Lee JT. The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. *Proc Natl Acad Sci USA* 2015, 112: E4216–4225.
20. Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. iCLIP—transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp* 2011, 50.
21. Ascano M, Hafner M, Cekan P, Gerstberger S, Tuschl T. Identification of RNA–protein interaction networks using PAR-CLIP. *WIREs RNA* 2012, 3:159–177.
22. Mann M. Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* 2006, 7:952–958.
23. Marshall KA, Ellington AD. In vitro selection of RNA aptamers. *Methods Enzymol* 2000, 318:193–214.
24. Campbell ZT, Bhimsaria D, Valley CT, Rodriguez-Martinez JA, Menichelli E, Williamson JR, Ansari AZ, Wickens M. Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep* 2012, 1:570–581.
25. Cook KB, Hughes TR, Morris QD. High-throughput characterization of protein–RNA interactions. *Brief Funct Genomics* 2015, 14:74–89.
26. Livi CM, Blanzieri E. Protein-specific prediction of mRNA binding using RNA sequences, binding motifs and predicted secondary structures. *BMC Bioinformatics* 2014, 15:123.
27. Terribilini M, Sander JD, Lee J-H, Zaback P, Jernigan RL, Honavar V, Dobbs D. RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res* 2007, 35(Web Server issue):W578–584.
28. Klus P, Bolognesi B, Agostini F, Marchese D, Zanzoni A, Tartaglia GG. The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities. *Bioinformatics* 2014, 30:1601–1608.
29. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 2012, 149:1393–1406.
30. Tartaglia GG. The grand challenge of characterizing ribonucleoprotein networks. *Front Mol Biosci* 2016, 3:24.
31. Tenenbaum SA, Carson CC, Lager PJ, Keene JD. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci USA* 2000, 97:14085–14090.
32. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 2008, 456:464–469.
33. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp A-C, Munschauer M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010, 141:129–141.
34. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, et al. Robust transcriptome-wide discovery of RNA-binding protein

binding sites with enhanced CLIP (eCLIP). *Nat Methods* 2016, 6:508–514.

35. Ray D, Kazan H, Chan ET, Castillo LP, Chaudhry S, Talukder S, Blencowe BJ, Morris Q, Hughes TR. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol* 2009, 27:667–670.
36. Lambert N, Robertson A, Jangi M, McGeary S, Sharp PA, Burge CB. RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* 2014, 54:887–900.
37. Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, Snyder MP, Greenleaf WJ. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat Biotechnol* 2014, 32:562–568.
38. Tome JM, Ozer A, Pagano JM, Gheba D, Schroth GP, Lis JT. Comprehensive analysis of RNA-protein interactions by high-throughput sequencing-RNA affinity profiling. *Nat Methods* 2014, 11:683–688.
39. Martin L, Meier M, Lyons SM, Sit RV, Marzluff WF, Quake SR, Chang HY. Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nat Methods* 2012, 9:1192–1194.
40. Hogg JR, Collins K. RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA* 2007, 13:868–880.
41. Slobodin B, Gerst JE. RaPID: an aptamer-based mRNA affinity purification technique for the identification of RNA and protein factors present in ribonucleoprotein complexes. *Methods Mol Biol* 2011, 714:387–406.
42. Michlewski G, Cáceres JF. RNase-assisted RNA chromatography. *RNA* 2010, 16:1673–1678.
43. Scherrer T, Mittal N, Janga SC, Gerber AP. A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS One* 2010, 5: e15499.
44. Siprashvili Z, Webster DE, Kretz M, Johnston D, Rinn JL, Chang HY, Khavari PA. Identification of proteins binding coding and non-coding human RNAs using protein microarrays. *BMC Genomics* 2012, 13:633.
45. Tsai BP, Wang X, Huang L, Waterman ML. Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach. *Mol Cell Proteomics* 2011, 10: M110.007385.
46. Chu C, Quinn J, Chang HY. Chromatin isolation by RNA purification (ChIRP). *J Vis Exp* 2012, 61:3912.
47. Simon MD. Capture hybridization analysis of RNA targets (CHART). *Curr Protoc Mol Biol*, Wiley Online Library, NY UUEE, **Chapter 21**, Unit 21.25, 2013.
48. Keene JD, Komisarow JM, Friedersdorf MB. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* 2006, 1:302–307.
49. Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* 2005, 37:376–386.
50. Faoro C, Ataide SF. Ribonomic approaches to study the RNA-binding proteome. *FEBS Lett* 2014, 588:3649–3664.
51. Trifillis P, Day N, Kiledjian M. Finding the right RNA: identification of cellular mRNA substrates for RNA-binding proteins. *RNA* 1999, 5:1071–1082.
52. Brooks SA, Rigby WF. Characterization of the mRNA ligands bound by the RNA binding protein hnRNP A2 utilizing a novel in vivo technique. *Nucleic Acids Res* 2000, 28:E49.
53. König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* 2010, 17:909–915.
54. Wang T, Xiao G, Chu Y, Zhang MQ, Corey DR, Xie Y. Design and bioinformatics analysis of genome-wide CLIP experiments. *Nucleic Acids Res* 2015, 43:5263–5274.
55. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. *Science* 2003, 302:1212–1215.
56. Ince-Dunn G, Okano HJ, Jensen KB, Park W-Y, Zhong R, Ule J, Mele A, Fak JJ, Yang C, Zhang C, et al. Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability. *Neuron* 2012, 75:1067–1080.
57. Lu Y-C, Chang S-H, Hafner M, Li X, Tuschl T, Elemento O, Hla T. ELAVL1 modulates transcriptome-wide miRNA binding in murine macrophages. *Cell Rep* 2014, 9:2330–2343.
58. Denkert C, Koch I, von Keyserlingk N, Noske A, Niesporek S, Dietel M, Weichert W. Expression of the ELAV-like protein HuR in human colon cancer: association with tumor stage and cyclooxygenase-2. *Mod Pathol* 2006, 19:1261–1269.
59. Abdelmohsen K, Gorospe M. Posttranscriptional regulation of cancer traits by HuR. *WIREs RNA* 2010, 1:214–229.
60. Mukherjee N, Corcoran DL, Nusbaum JD, Reid DW, Georgiev S, Hafner M, Ascano M, Tuschl T, Ohler U, Keene JD. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell* 2011, 43:327–339.

61. Okano HJ, Darnell RB. A hierarchy of Hu RNA binding proteins in developing and adult neurons. *J Neurosci* 1997, 17:3024–3037.
62. Darnell RB. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *WIREs RNA* 2010, 1:266–286.
63. Jankowsky E, Harris ME. Specificity and nonspecificity in RNA-protein interactions. *Nat Rev Mol Cell Biol* 2015, 16:533–544.
64. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell* 2013, 154:26–46.
65. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 2013, 341:1237973.
66. West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, Tolstorukov MY, Kingston RE. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell* 2014, 55:791–802.
67. McHugh CA, Chen C-K, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 2015, 521:232–236.
68. Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY. Systematic discovery of Xist RNA binding proteins. *Cell* 2015, 161:404–416.
69. Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* 2012, 46:674–690.
70. Kwon SC, Yi H, Eichelbaum K, Föhr S, Fischer B, You KT, Castello A, Krijgsveld J, Hentze MW, Kim VN. The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol* 2013, 20:1122–1130.
71. Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume A-M, Schwarzl T, Curk T, Foehr S, Huber W, et al. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat Commun* 2015, 6:10127.
72. Mitchell SF, Jain S, She M, Parker R. Global analysis of yeast mRNPs. *Nat Struct Mol Biol* 2013, 20:127–133.
73. Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet* 2014, 15:205–213.
74. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic Acids Res* 2015, 43(W1): W39–49.
75. Paz I, Kosti I, Ares M, Cline M, Mandel-Gutfreund Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res* 2014, 42(Web Server issue):W361–367.
76. Agostini F, Cirillo D, Ponti RD, Tartaglia GG. SeA-MotE: a method for high-throughput motif discovery in nucleic acid sequences. *BMC Genomics* 2014, 15:925.
77. Kazan H, Ray D, Chan ET, Hughes TR, Morris Q. RNAcontext: a new method for learning the sequence and structure binding preferences of RNA-binding proteins. *PLoS Comput Biol* 2010, 6:e1000832.
78. Wang L, Huang C, Yang MQ, Yang JY. BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features. *BMC Syst Biol* 2010, 4:1–9.
79. Kumar M, Gromiha MM, Raghava GPS. Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins* 2008, 71:189–194.
80. Walia RR, Xue LC, Wilkins K, El-Manzalawy Y, Dobbs D, Honavar V. RNABindRPlus: a predictor that combines machine learning and sequence homology-based methods to improve the reliability of predicted RNA-binding residues in proteins. *PLoS One* 2014, 9:e97725.
81. Cirillo D, Agostini F, Tartaglia GG. Predictions of protein-RNA interactions. *WIREs Comput Mol Sci* 2012, 3:161–175.
82. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. HMMER web server: 2015 update. *Nucleic Acids Res* 2015, 43 (W1):W30–38.
83. Livi CM, Klus P, Delli Ponti R, Tartaglia GG. catRAPID signature: identification of ribonucleoproteins and RNA-binding regions. *Bioinformatics* 2015, 32:773–775.
84. Towfic F, Caragea C, Gemperline DC, Dobbs D, Honavar V. Struct-NB: predicting protein-RNA binding sites using structural features. *Int J Data Min Bioinform* 2010, 4:21–43.
85. Maetschke SR, Yuan Z. Exploiting structural and topological information to improve prediction of RNA-protein binding sites. *BMC Bioinformatics* 2009, 10:341.
86. Zhao H, Yang Y, Zhou Y. Highly accurate and high-resolution function prediction of RNA binding proteins by fold recognition and binding affinity prediction. *RNA Biol* 2011, 8:988–996.
87. Pérez-Cano L, Fernández-Recio J. Optimal protein-RNA area, OPRA: a propensity-based method to identify RNA-binding sites on proteins. *Proteins* 2010, 78:25–35.
88. Yang Y, Faraggi E, Zhao H, Zhou Y. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between

predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics* 2011, 27:2076–2082.

89. Zhao H, Yang Y, Janga SC, Kao CC, Zhou Y. Prediction and validation of the unexplored RNA-binding protein atlas of the human proteome. *Proteins* 2014, 82:640–647.
90. Ivani I, Dans PD, Noy A, Pérez A, Faustino I, Hospital A, Walther J, Andrio P, Goñi R, Balaceanu A, et al. Parmbsc1: a refined force field for DNA simulations. *Nat Methods* 2016, 13:55–58.
91. Cirillo D, Livi CM, Agostini F, Tartaglia GG. Discovery of protein–RNA networks. *Mol BioSyst* 2014, 10:1632–1642.
92. Si J, Cui J, Cheng J, Wu R. Computational prediction of RNA-binding proteins and binding sites. *Int J Mol Sci* 2015, 16:26303–26317.
93. Wang L, Brown SJ. BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucl Acids Res* 2006, 34(suppl 2):W243–248.
94. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME Suite: tools for motif discovery and searching. *Nucl Acids Res* 2009, 37(suppl 2):W202–208.
95. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 2011, 39(Web Server issue):W29–37.
96. Bellucci M, Agostini F, Masin M, Tartaglia GG. Predicting protein associations with long noncoding RNAs. *Nat Methods* 2011, 8:444–445.
97. Agostini F, Zanzoni A, Klus P, Marchese D, Cirillo D, Tartaglia GG. catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. *Bioinformatics* 2013, 29:2928–2930.
98. Muppirala UK, Honavar VG, Dobbs D. Predicting RNA-protein interactions using only sequence information. *BMC Bioinformatics* 2011, 12:489.
99. Liu T, Geng X, Zheng X, Li R, Wang J. Accurate prediction of protein structural class using auto covariance transformation of PSI-BLAST profiles. *Amino Acids* 2012, 42:2243–2249.
100. Kumar M, Gromiha MM, Raghava GPS. SVM based prediction of RNA-binding proteins using binding residues and evolutionary information. *J Mol Recognit* 2011, 24:303–313.
101. Cheng C-W, Su EC-Y, Hwang J-K, Sung T-Y, Hsu W-L. Predicting RNA-binding sites of proteins using support vector machines and evolutionary information. *BMC Bioinformatics* 2008, 9(Suppl 12):S6.
102. Nagarajan R, Gromiha MM. Prediction of RNA binding residues: an extensive analysis based on structure and function to select the best predictor. *PLoS One* 2014, 9:e91140.
103. Agostini F, Cirillo D, Livi CM, Ponti RD, Tartaglia GG. ccSOL omics: a webserver for large-scale prediction of endogenous and heterologous solubility in *E. coli*. *Bioinformatics* 2014, 30:2975–2977.
104. Johnson R, Noble W, Tartaglia GG, Buckley NJ. Neurodegeneration as an RNA disorder. *Prog Neurobiol* 2012, 99:293–315.
105. Zalfa F, Giorgi M, Primerano B, Moro A, Di Penta A, Reis S, Oostra B, Bagni C. The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* 2003, 112:317–327.
106. Myrick LK, Hashimoto H, Cheng X, Warren ST. Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. *Hum Mol Genet* 2015, 24:1733–1740.
107. Agostini F, Cirillo D, Bolognesi B, Tartaglia GG. X-inactivation: quantitative predictions of protein interactions in the Xist network. *Nucleic Acids Res* 2013, 41:e31.
108. Cirillo D, Marchese D, Agostini F, Livi CM, Botta-Orfila T, Tartaglia GG. Constitutive patterns of gene expression regulated by RNA-binding proteins. *Genome Biol* 2014, 15:R13.
109. Cirillo D, Agostini F, Klus P, Marchese D, Rodriguez S, Bolognesi B, Tartaglia GG. Neurodegenerative diseases: quantitative predictions of protein-RNA interactions. *RNA* 2013, 19:129–140.
110. Will CL, Lührmann R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 2011, 3: a003707.
111. Cirillo D, Botta-Orfila T, Tartaglia GG. By the company they keep: interaction networks define the binding ability of transcription factors. *Nucleic Acids Res* 2015, 43:e125.
112. Levy ED, Erba EB, Robinson CV, Teichmann SA. Assembly reflects evolution of protein complexes. *Nature* 2008, 453:1262–1265.
113. Matia-González AM, Laing EE, Gerber AP. Conserved mRNA-binding proteomes in eukaryotic organisms. *Nat Struct Mol Biol* 2015, 22:1027–1033.
114. Neelamraju Y, Hashemikhabir S, Janga SC. The human RBPome: from genes and proteins to human disease. *J Proteomics* 2015, 127(Pt A):61–70.
115. Castello A, Horos R, Strein C, Fischer B, Eichelbaum K, Steinmetz LM, Krijgsveld J, Hentze MW. Comprehensive identification of RNA-binding proteins by RNA interactome capture. *Methods Mol Biol* 2016, 1358:131–139.
116. Babu MM, Kriwacki RW, Pappu RV. Structural biology. Versatility from protein disorder. *Science* 2012, 337:1460–1461.
117. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. Intrinsic disorder and protein function. *Biochemistry* 2002, 41:6573–6582.

118. Bolognesi B, Lorenzo N, Dhar R, Cirillo D, Baldrighi M, Tartaglia G, Lehner B. A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Rep* 2016, 16:222–231.
119. Li P, Banjade S, Cheng H-C, Kim S, Chen B, Guo L, Llaguno M, Hollingsworth JV, King DS, Banani SF, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* 2012, 483:336–340.
120. Martin JE, Mather K, Swash M, Dodd SM, Dale GE, Garofalo O, Leigh PN. Stress protein inclusions in cerebral vessels in dialysis encephalopathy. *Neuropathol Appl Neurobiol* 1991, 17:105–111.
121. Shah KH, Zhang B, Ramachandran V, Herman PK. Processing body and stress granule assembly occur by independent and differentially regulated pathways in *Saccharomyces cerevisiae*. *Genetics* 2013, 193:109–123.
122. Klus P, Cirillo D, Botta Orfila T, Gaetano Tartaglia G. Neurodegeneration and cancer: where the disorder prevails. *Sci Rep* 2015, 5:15390.
123. Klus P, Ponti RD, Livi CM, Tartaglia GG. Protein aggregation, structural disorder and RNA-binding ability: a new approach for physico-chemical and gene ontology classification of multiple datasets. *BMC Genomics* 2015, 16:1071.
124. Jain S, Wheeler JR, Walters RW, Agrawal A, Barsic A, Parker R. ATPase-modulated stress granules contain a diverse proteome and substructure. *Cell* 2016, 164:487–498.
125. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012, 28:3150–3152.
126. Becker LA, Gitler AD. It's all starting to come together. *Elife* 2015, 4:e09853.
127. Mitrea DM, Kriwacki RW. Phase separation in biology; functional organization of a higher order. *Cell Commun Signal* 2016, 14:1.
128. Masuda K, Marasa B, Martindale JL, Halushka MK, Gorospe M. Tissue- and age-dependent expression of RNA-binding proteins that influence mRNA turnover and translation. *Aging (Albany NY)* 2009, 1:681–698.
129. Gsponer J, Babu MM. Cellular strategies for regulating functional and nonfunctional protein aggregation. *Cell Rep* 2012, 2:1425–1437.
130. Bossy-Wetzel E, Schwarzenbacher R, Lipton SA. Molecular pathways to neurodegeneration. Published online: 01 July 2004 10: S2–S9. doi:10.1038/nm1067.
131. Jaisson S, Gillery P. Impaired proteostasis: role in the pathogenesis of diabetes mellitus. *Diabetologia* 2014, 57:1517–1527.
132. Lindberg I, Shorter J, Wiseman RL, Chiti F, Dickey CA, McLean PJ. Chaperones in neurodegeneration. *J Neurosci* 2015, 35:13853–13859.
133. Ojha J, Masilamoni G, Dunlap D, Udoff RA, Cashikar AG. Sequestration of toxic oligomers by HspB1 as a cytoprotective mechanism. *Mol Cell Biol* 2011, 31:3146–3157.
134. Binger KJ, Ecroyd H, Yang S, Carver JA, Howlett GJ, Griffin MDW. Avoiding the oligomeric state: B-crystallin inhibits fragmentation and induces dissociation of apolipoprotein C-II amyloid fibrils. *FASEB J* 2013, 27:1214–1222.
135. Cascella R, Conti S, Tatini F, Evangelisti E, Scartabelli T, Casamenti F, Wilson MR, Chiti F, Cecchi C. Extracellular chaperones prevent Aβ42-induced toxicity in rat brains. *Biochim Biophys Acta Mol Basis Dis* 2013, 1832:1217–1226.
136. Mannini B, Mulvihill E, Sgromo C, Cascella R, Khodarahmi R, Ramazzotti M, Dobson CM, Cecchi C, Chiti F. Toxicity of protein oligomers is rationalized by a function combining size and surface hydrophobicity. *ACS Chem Biol* 2014, 9:2309–2317.
137. Wolozin B. Regulated protein aggregation: stress granules and neurodegeneration. *Mol Neurodegener* 2012, 7:56.
138. de Groot NS, Sabate R, Ventura S. Amyloids in bacterial inclusion bodies. *Trends Biochem Sci* 2009, 34:408–416.
139. King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res* 2012, 1462:61–80.
140. Vanderweyde T, Youmans K, Liu-Yesucevitz L, Wolozin B. Role of stress granules and RNA-binding proteins in neurodegeneration: a mini-review. *Gerontology* 2013, 59:524–533.
141. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature* 2012, 482:339–346.
142. Lee S, Kopp F, Chang T-C, Sataluri A, Chen B, Sivakumar S, Yu H, Xie Y, Mendell JT. Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. *Cell* 2016, 164:69–80.
143. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, et al. Classification of intrinsically disordered regions and proteins. *Chem Rev* 2014, 114:6589–6631.
144. Castello A, Hentze MW, Preiss T. Metabolic enzymes enjoying new partnerships as RNA-binding proteins. *Trends Endocrinol Metab* 2015, 26:746–757.
145. Sanchez de Groot N, Gomes RA, Villar-Pique A, Babu MM, Coelho AV, Ventura S. Proteome response at the edge of protein aggregation. *Open Biol* 2015, 5:140221–140221.