ADVANCED REVIEW

WIREs RNA | WILEY

From parts lists to functional significance—RNA–protein interactions in gene regulation

Cornelia Kilchert[1] | Katja Sträßer[1] | Vladislav Kunetsky[1] | Minna-Liisa Änkö[2,3]

[1]Institute of Biochemistry, Justus-Liebig University Giessen, Giessen, Germany

[2]Centre for Reproductive Health and Centre for Cancer Research, Hudson Institute of Medical Research, Melbourne, Victoria, Australia

[3]Department of Molecular and Translational Science, School of Clinical Sciences, Monash University, Melbourne, Victoria, Australia

Correspondence
Minna-Liisa Änkö, Centre for Reproductive Health and Centre for Cancer Research, Hudson Institute of Medical Research, Melbourne, VIC, Australia.
Email: minni.anko@hudson.org.au

Abstract

Hundreds of canonical RNA binding proteins facilitate diverse and essential RNA processing steps in cells forming a central regulatory point in gene expression. However, recent discoveries including the identification of a large number of noncanonical proteins bound to RNA have changed our view on RNA–protein interactions merely as necessary steps in RNA biogenesis. As the list of proteins interacting with RNA has expanded, so has the scope of regulation through RNA–protein interactions. In addition to facilitating RNA metabolism, RNA binding proteins help to form subcellular structures and membraneless organelles, and provide means to recruit components of macromolecular complexes to their sites of action. Moreover, RNA–protein interactions are not static in cells but the ribonucleoprotein (RNP) complexes are highly dynamic in response to cellular cues. The identification of novel proteins in complex with RNA and ways cells use these interactions to control cellular functions continues to broaden the scope of RNA regulation in cells and the current challenge is to move from cataloguing the components of RNPs into assigning them functions. This will not only facilitate our understanding of cellular homeostasis but may bring in key insights into human disease conditions where RNP components play a central role. This review brings together the classical view of regulation accomplished through RNA–protein interactions with the novel insights gained from the identification of RNA binding interactomes. We discuss the challenges in combining molecular mechanism with cellular functions on the journey towards a comprehensive understanding of the regulatory functions of RNA–protein interactions in cells.

This article is categorized under:

> RNA Interactions with Proteins and Other Molecules > Protein–RNA Interactions: Functional Implications
>
> RNA Interactions with Proteins and Other Molecules > RNA–Protein Complexes
>
> RNA Interactions with Proteins and Other Molecules > Protein–RNA Recognition

KEYWORDS

RNA binding protein, RNA metabolism, RNA–protein interaction

1 | INTRODUCTION

Interactions between RNA and proteins are pervasive in biology. Eukaryotic cells harbor hundreds of proteins with well-defined RNA binding domains (RBDs) that form ribonucleoprotein (RNP) complexes with RNA. RNA binding proteins accompany RNA molecules from the moment they are born at the site of transcription. Essentially all cellular RNA exists in complex with proteins rather than as free RNA. The maturation of both coding messenger RNAs (mRNAs) and noncoding RNAs (ncRNAs) is marked by RBPs that facilitate each step of the diverse RNA biogenesis pathways (Figure 1; Gehring, Wahle, & Fischer, 2017; Gerstberger, Hafner, & Tuschl, 2014; Michlewski & Caceres, 2019; Muller-McNicoll & Neugebauer, 2013; Nussbacher & Yeo, 2018). The 5′ end of mRNAs is bound by the cap binding complex and termination and polyadenylation processes depend on the activity of RBPs at the 3′ end. The molecular properties of mRNAs and ncRNAs are modulated through pre-mRNA splicing, A-to-I editing, RNA methylation and various other RNA modifications. These processes alter the information content, stability and interaction capacity of the RNA. Each of these processing and modification steps is facilitated by a suite of RBPs, thus providing multiple points of regulation in the gene expression pathway. In addition to altering the chemical and structural properties of RNAs, RBPs can define the subcellular localization of RNAs including regulated nucleo-cytoplasmic export and the formation of membraneless organelles through liquid–liquid phase separation (LLPS; Courchaine, Lu, & Neugebauer, 2016; Drino & Schaefer, 2018; Hieronymus & Silver, 2003; Wickramasinghe & Laskey, 2015). Once RNAs reach the end of their life, RNA degradation is also facilitated by RBPs in a highly controlled manner, providing a further avenue to actively modulate the cellular RNA repertoire (Hug, Longman, & Cáceres, 2016).

Many characterized RBPs such as factors first implicated in pre-mRNA splicing are multifunctional proteins with the ability to interact with multiple RNA processing machineries. (Änkö, 2014; Gerstberger et al., 2014; Muller-McNicoll & Neugebauer, 2013; Sawicka, Bushell, Spriggs, & Willis, 2008; Figure 1). This enables control of diverse activities at more than one step in RNA biogenesis. Furthermore, many RNA biogenesis steps are functionally and/or mechanistically coupled, RBPs not only facilitating multiple processes but also providing means to link successive RNA
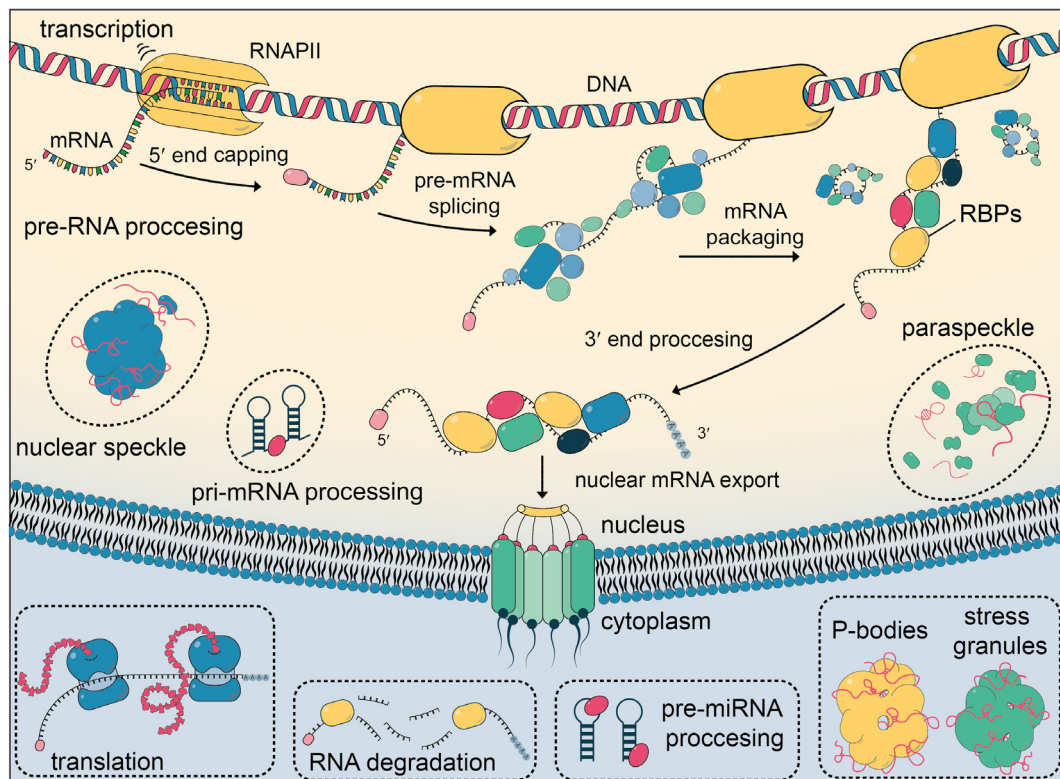

FIGURE 1 RNA binding proteins facilitate each step of RNA biogenesis of both coding and noncoding RNAs in cells. They also play roles beyond these processes by forming different types of subcellular organelles through their interactions with RNA. The RNA biogenesis steps have been drawn to occur in a step-wise manner for the purpose of visual presentation. However, they often take place simultaneously and/or co-transcriptionally in cells

regulatory steps to increase the efficiency and fidelity of gene expression (Herzel, Straube, & Neugebauer, 2018; Maniatis & Reed, 2002; Meinel & Strasser, 2015). This is exemplified by different families of splicing factors that in addition to activating or inhibiting pre-mRNA splicing have been assigned functions in other nuclear and cytoplasmic RNA regulatory processes such as transcription, mRNA export, microRNA (miRNA) processing and translation (Gehring et al., 2017; Gerstberger et al., 2014; Michlewski & Caceres, 2019; Muller-McNicoll & Neugebauer, 2013; Ratnadiwakara, Mohenska, & Anko, 2018). However, the view of RBPs solely as modulators of RNA biogenesis steps is changing. The recent identification of proteins such as metabolic enzymes that interact with RNA suggest that RNA–protein interactions are more abundant than anticipated and are also more plastic than previously thought (reviewed in Hentze, Castello, Schwarzl, & Preiss, 2018). Studies capturing the complete proteomes of polyadenylated RNAs have greatly expanded the scope of RNA–protein interactions in cellular processes by extending the list of proteins in direct contact with RNA beyond canonical RBPs (Table 1). Many of these proteins do not contain a canonical RBD but take part in RNP complexes through other protein regions. The most recent count from human cells found >1,700 proteins bound to RNA (Castello et al., 2016; Gerstberger et al., 2014; Trendel et al., 2019). The mapping of RNP constituents has drawn our attention to functions of RNPs beyond RNA metabolism. Various subcellular assemblies can be seeded around RNA–protein interactions (Drino & Schaefer, 2018; Fox, Nakagawa, Hirose, & Bond, 2018). RNA–protein interactions may also be used as recruitment mechanisms or stabilizers of macromolecular complexes (Pintacuda, Young, & Cerase, 2017). Conversely, examples of RNA controlling the activity of a protein challenge the long-held power dynamics between RNA and protein wherein the protein dominates the RNA. This is the basis of the emerging concept of protein binding RNAs (Bayraktar, Bertilaccio, & Calin, 2019; Cifuentes-Rojas, Hernandez, Sarma, & Lee, 2014).

The development of UV crosslinking and immunoprecipitation methods (CLIP-sequencing in its various forms) (Hafner et al., 2010; Huppertz et al., 2014; Licatalosi et al., 2008; Ule, Jensen, Mele, & Darnell, 2005; Van Nostrand et al., 2016), including the comprehensive survey of RBP binding sites by the ENCODE consortium (Davis et al., 2018), have resulted in a global picture of RNA targets of many canonical RBPs. Most CLIP-seq data sets have been generated from a handful of cell lines, which limits their interpretation in terms of RBP cell type specificity, cellular relevance of the interactions and dynamics of RNA–protein interactions. While these investigations are highly valuable by providing key mechanistic insights into RBP activity at a global level, further work addressing the context-dependency of the interactions is needed to fully understand the cellular and functional significance of RBP activities. Similarly, RNA interactome capture (RIC) studies have largely focused on few cell types and/or cellular conditions with a few exciting exceptions (Table 1 and references therein). At the same time, computational approaches putting together the large-scale datasets reveal RNA regulatory mechanism and allow predictions that will greatly facilitate the understanding of functional consequences of RNA–protein interaction. An example of the power of computational approaches in understanding RNA regulation was the construction of "the splicing code" that was able to predict with ~60–70% accuracy cellular splicing patterns by building on different types of large-scale datasets (Barash et al., 2010; Bretschneider, Gandhi, Deshwar, Zuberi, & Frey, 2018).

The understanding of the cellular significance of RNA–protein interactions lags behind the mapping of RBP binding sites and RNP constituents. The next step in the field is to solve cell type specific differences in RNP composition and RNP dynamics in response to various cellular cues. The comprehensive understanding of RBP activities and regulated networks in cells, tissues and organisms is an enormous undertaking but the field is moving forward from the cataloguing phase to "functional RNAomics" (Figure 2). The increasing ability to combine global methods with structural and mechanistic studies has started to reveal the molecular details of the various modes of RNA–protein interactions. From a protein centric viewpoint, a highly similar protein fold within different protein context can bind to many types of RNAs including RNA hairpin structures, individual nucleotides and linear stretches of single-stranded RNA. An RNA-centric view suggests that individual RNA molecules can interact with a wide range of protein folds, providing versatile opportunities for regulation. The integration of the comprehensive catalogues of RNA repertoires in different cells types, global RBP binding maps, proteomics profiles, in vitro assays evaluating RNA binding specificity and high-resolution RBP structures with cellular phenotypes will be the key for our functional understanding of RNP constituents in cells (Figure 2).

In this review, we discuss the central and expanding roles of RNA–protein interactions in the regulation of gene expression and cellular functions. We first describe different types of RBPs and their interactions with various types of RNAs to define an RBP and RBD, we then address how RBPs exhibit a wide range of functions beyond the classical RNA metabolic pathways. We discuss the recent advances in understanding the dynamics of RNA–protein interactions during development and in response to cellular signals that will help in closing the gap in knowledge between RNA–protein interaction networks and cellular functions.

TABLE 1 Key studies determining the RNP components using various RNA interactome capture approaches in different species

Single-condition interactomes			
Species	**Sample type**	**Enrichment method**	**Reference**
poly(A)+			
Homo sapiens	HeLa	oligo-d(T)	Castello et al. (2012)
	HEK293	oligo-d(T)	Baltz et al. (2012)
	HuH7	oligo-d(T)	Beckmann et al. (2015)
	K562 (nuclei)	oligo-d(T)	Conrad et al. (2016)
Mus musculus	Embryonic stem cells	oligo-d(T)	Kwon et al. (2013)
	HL-1 (cardiomyocytes)	oligo-d(T)	Liao et al. (2016)
	Macrophages	oligo-d(T)	Liepelt et al. (2016)
Saccharomyces cerevisiae		oligo-d(T)	Beckmann et al. (2015), Matia-Gonzalez, Laing, and Gerber (2015), Mitchell, Jain, She, and Parker (2013)
Schizosaccharomyces pombe		oligo-d(T)	Kilchert et al. (2019)
Drosophila melanogaster	Embryo	oligo-d(T)	Wessels et al. (2016), Sysoev et al. (2016)
Caenorhabditis elegans	Adult	oligo-d(T)	Matia-Gonzalez et al. (2015)
	L4-stage larvae	oligo-d(T)	Matia-Gonzalez et al. (2015)
Arabidopsis thaliana	Cultured cells	oligo-d(T)	Marondedze, Thomas, Serrano, Lilley, and Gehring (2016)
	Leaves	oligo-d(T)	Marondedze et al. (2016)
	Etiolated seedlings	oligo-d(T)	Reichel et al. (2016)
Plasmodium falciparum	Blood stage	oligo-d(T)	Bunnik et al. (2016)
Trypanosoma brucei	Blood stage	oligo-d(T)	Lueong, Merce, Fischer, Hoheisel, and Erben (2016)
Non-poly(A)+			
Homo sapiens	HEK293	Organic phase separation	Urdaneta et al. (2019)
	Huh7	Solid phase extraction	Asencio, Chatterjee, and Hentze (2018)
	MCF-7	Organic phase separation	Trendel et al. (2019)
Mus musculus	Embryonic stem cells	EU labeling	Bao et al. (2018)
Saccharomyces cerevisiae		Solid phase extraction	Shchepachev et al. (2019)
Escherichia coli		Solid phase extraction	Shchepachev et al. (2019)
Salmonella typhimurium		Organic phase separation	Urdaneta et al. (2019)
Specific RNAs (selection)			
Homo sapiens	18S/28S rRNA	LNA/DNA	Rogell et al. (2017)
	Various pre-miRNAs	Immobilized bait	Treiber et al. (2017)
	ACTB mRNA	Proximity biotinylation	Mukherjee et al. (2019)
Mus musculus	*Xist*	Antisense oligos	Minajigi et al. (2015)

(Continues)

TABLE 1 (Continued)

Comparative interactomes		
Species	**Condition**	**Reference**
Homo sapiens	SINV infection	Garcia-Moreno et al. (2019)
	Arsenite-induced stress	Trendel et al. (2019)
Mus musculus	LPS-treatment (macrophages)	Liepelt et al. (2016)
Drosophila melanogaster	Maternal-to-zygotic transition	Sysoev et al. (2016)
Danio rerio	Maternal-to-zygotic transition	Despic et al. (2017)
Caenorhabditis elegans	Induction of apoptosis (L4-stage larvae)	Matia-Gonzalez et al. (2015)
Arabidopsis thaliana	Severe drought stress	Marondedze et al. (2016)
Schizosaccharomyces pombe	Exosome mutants	Kilchert et al. (2019)

FIGURE 2 "Functional RNAomics" integrating sequencing, structural and molecular data with in vivo models will enable the comprehensive understanding of the functional significance of RNA regulation at the cellular and organism level and may reveal key insights into underlying mechanisms of human disease
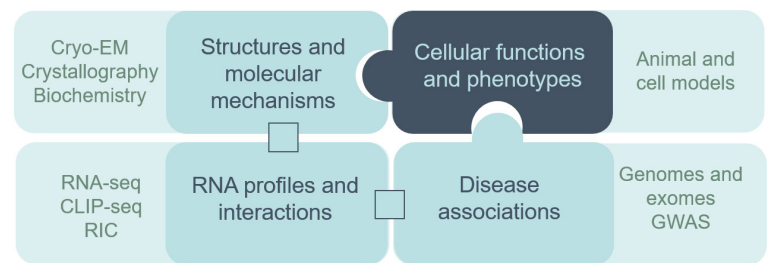
2 | DEFINING RNA BINDING PROTEINS

Canonical RBPs interact with RNA through their RBDs that have a well-defined fold and structure (Figure 3, top). Typically, an isolated RBD can interact with RNA although the surrounding protein may enhance RNA binding and is needed for the functional outcome of the interaction (Lunde, Moore, & Varani, 2007). In fact, until recently in vitro and structural studies investigating RBP activity at the amino acid level were largely conducted using isolated RBDs that possibly incorporated some additional protein regions and short target RNA sequences. The development of cryogenic electron microscopy (cryo-EM) techniques has brought a wealth of new insights into RNA–protein complexes, enabling structure determination of larger complexes without the need for crystallization (Casanal et al., 2017; Rauhut et al., 2016; Schuller, Falk, Fromm, Hurt, & Conti, 2018; Wan, Yan, Bai, Huang, & Shi, 2016). Structural studies have uncovered key insights into the molecular features of RBD–RNA interactions, including overall structure of the domains, critical amino acids within the domains and essential RNA sequence features. While genome-wide studies result in RNA consensus motifs and binding site information, structural studies when applicable can facilitate the interpretation of the binding patterns by revealing how RBDs contact RNA at the atomic resolution. A common feature that emerges is the flexibility of closely related RBD folds to interact with a wide range of RNA sequences and nucleotides such as single- and double-stranded RNA and individual nucleotides. The discovery of proteins interacting with RNA through other domains than well-defined RBDs in mammalian cells similar to phages and viruses (Bayer et al., 1995; Cai et al., 1998; Puglisi, Tan, Calnan, Frankel, & Williamson, 1992) further highlights the biochemical diversity in RNA–protein interactions and raises the question whether the novel RNA interacting domains should be classified as RBDs and perhaps more broadly how to define an RBP.

2.1 | The plasticity of canonical RNA binding domains

The diversity of domains binding to RNA demonstrates how proteins have developed a range of solutions during evolution to interact with RNA. Among the most numerous RBDs found across species are RNA recognition motifs (RRMs), KH-domains, zinc fingers, DEAD box helicase domains, and Pumilio-family (PUF) RNA binding repeats. Here we will briefly discuss these RBDs to demonstrate a common theme in RNA recognition, which is the flexibility in target
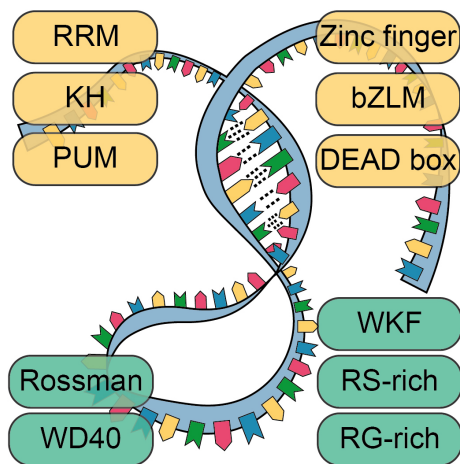
FIGURE 3 Many modes of RNA binding through RBDs and other protein domains. Common canonical (yellow) and major novel (green) RNA binding domains or protein regions discussed in this review

recognition without compromising specificity. Frequently, RBPs can recognize different types of RNA sequence elements and still distinguish targets from nontarget RNAs. We refer to databases such as SMART, RBPDB, ATtRACT, SpliceAid-F, and EuRBPDB that provide more comprehensive catalogues of RBPs and RBDs (Cook, Kazan, Zuberi, Morris, & Hughes, 2011; Giudice, Sanchez-Cabo, Torroja, & Lara-Pezzi, 2016; Giulietti et al., 2013; Letunic & Bork, 2017) (Liao et al., 2019).

The most common RBD is the RRM which consists of two short sequence motifs RNP1 and RNP2 (Burd & Dreyfuss, 1994). RRMs are found across taxonomic groups from bacteria to higher eukaryotes, the human genome encoding approximately 450 proteins with one or more RRMs (Letunic & Bork, 2017). The RRM can bind to RNA in isolation although in some RBPs such as the Sex-lethal (Sxl) multiple RRMs are required to confer specificity and high RNA binding affinity (Handa et al., 1999). The RRM is a very plastic domain forming versatile interactions with the ability to interact with both single- and double-stranded RNA, proteins and even lipids (Aubol, Serrano, Fattet, Wuthrich, & Adams, 2018; Clery, Blatter, & Allain, 2008; Clingman et al., 2014; Kuwasako et al., 2017; Scheiba et al., 2014). The RRM directly forms contacts with only a few nucleotides within the target RNA sequences (Auweter et al., 2006; Cavaloc, Bourgeois, Kister, & Stevenin, 1999; Clery et al., 2008; Petoukhov et al., 2006). RRMs are found in many spliceosomal components such as the U2 auxiliary factor 65 (U2AF65 or U2AF2) and U1 small nuclear RNP (U1A) as well as many splicing factors that are not part of the core splicing machinery such as SR proteins, RBFOX proteins and heterogeneous nuclear RNPs (hnRNPs). Although RRMs are very common in RBPs involved in pre-mRNA splicing, RRMs are not functionally limited to splicing but can convey RNA binding activity to RBPs involved in other steps of RNA metabolism. This is exemplified by the nuclear cap binding protein subunit 2 (CBP20 or NCBP2), the cytoplasmic poly(A) binding protein 1 (PABPC1) and the cleavage stimulation factor subunit 2 (CSTF2). The variety of processes regulated by RRM containing proteins further emphasize the versatility of the motif in RNA binding as the target sites for these proteins range from a poly(A) sequence to the m(7)GpppG-cap (Deo, Bonanno, Sonenberg, & Burley, 1999; Nagata et al., 2008; Wu et al., 2009).

Similar to the RRM, the K-homology RBD (KH-domain) is found in a variety of different RBPs and it can exist in isolation or in multiple copies within a single protein (Letunic & Bork, 2017; Valverde, Edwards, & Regan, 2008). The human genome encodes approximately 150 proteins with KH-domains (Letunic & Bork, 2017) and KH-domains are found across species from bacteria to higher eukaryotes exemplified by bacterial PNPases (Matus-Ortega et al., 2007), archaeal and eukaryotic exosome subunits (Oddone et al., 2007), ribosomal proteins (Wan et al., 2007) and splicing factors forming part of either the core spliceosomal machinery or acting as accessory proteins (Tadesse, Deschenes-Furry, Boisvenue, & Cote, 2008; Teplova et al., 2011). Many of the splicing factors with KH-domains are both structurally and functionally well-characterized including splicing factor 1 (SF1) binding to the branchpoint, neuronal splicing factors Nova-1 and 2, FMRP (Fragile X mental retardation protein), the STAR family RBP Sam68 (SRC associated in mitosis of 68 kDa) and hnRNP K (Lukong & Richard, 2003; Myrick, Hashimoto, Cheng, & Warren, 2015; Siomi, Choi, Siomi, Nussbaum, & Dreyfuss, 1994; Teplova et al., 2011; Wang et al., 2013, 2019; Zhang et al., 2013). Reminiscent of proteins with RRMs, KH-domain containing proteins act in diverse RNA metabolism steps and interact with different types of RNAs. Based on their structure, KH-domains can be divided into two groups—KH type 1 and 2—the two types forming different proteins folds and thus distinct RNA interaction surfaces (Grishin, 2001). The nucleotide binding sites

of KH-domains can accommodate both single-stranded RNA and DNA (Valverde et al., 2008), the domain typically interacting with just four unpaired nucleotides. Enhanced RNA specificity can be achieved by combining KH-domains in tandem within a single protein or through other protein domains that augment the interaction (Barnes et al., 2015). Mutagenesis studies on FMRP demonstrate an important commonality of KH-domains with RRMs, and perhaps all RBDs: the protein context of KH-domains within individual proteins is a critical determinant of RNA interactions and function. A single amino acid mutation that fully abolishes the function of KH-domain in FMRP had a very mild effect on a closely related KH-domains in other protein contexts (Valverde et al., 2008).

Although zinc finger proteins are generally considered DNA binding transcription factors, zinc finger containing proteins make versatile interactions with DNA, RNA, or both depending on the types and combinations of the zinc finger domains they harbor (Laity, Lee, & Wright, 2001). Thus, the zinc finger domains that are found in approximately 80 proteins in human (Letunic & Bork, 2017) represent yet another flexible RNA interacting protein domain. The main classes of RNA binding zinc fingers are C2H2 and CCHC that exhibit multiple modes of RNA binding (Hall, 2005). Zinc finger domains are often found in clusters, each domain containing multiple finger-like protrusions. Well-characterized zinc finger containing RBPs include splicing factors U2AF35 (U2 auxiliary factor 35 kDa or U2AF1), the SR protein splicing factor SRSF7 and the muscle-blind splicing regulators MBNL1-3 (Letunic & Bork, 2017). Interestingly, a single zinc finger protein can interact with both DNA and RNA through different zinc finger domains (Lu, Searles, & Klug, 2003). For instance, the transcriptional repressor CCCTC-binding factor (CTCF) contains 11 zinc fingers, two of which bind to RNA forming interactions that are essential for CTCF function in genome organization (Saldana-Meyer et al., 2019). The zinc finger nucleases harnessed as one of the first tools for genome editing are a further example of the flexibility of zinc fingers in nucleotide interactions. The modular mode of nucleotide recognition through the tandem "fingers" was exploited to engineer a desired DNA or RNA binding specificity (Klug, 2010).

Another domain that has lent itself particularly well to RBP engineering is the PUF RNA binding repeat (Filipovska, Razif, Nygard, & Rackham, 2011; Wang, Oge, Perez-Garcia, Hamama, & Sakr, 2018). Natural PUF repeats found in 14 proteins in human (Letunic & Bork, 2017) bind to RNA along a concave surface in a sequence-specific manner with each repeat recognizing a single base in an eight base sequence via interactions through three amino acids at conserved positions (Wang, McLachlan, Zamore, & Hall, 2002). PUF proteins have been engineered to recognize longer sequences, and to accommodate cytosines, which rarely occurs in nature, and are being used as protein guides for sequence-specific RNA targeting of fused proteins (Bhat et al., 2019; Campbell, Valley, & Wickens, 2014; Dong et al., 2011; Filipovska et al., 2011; Zhao et al., 2018).

The highly conserved DEAD box RNA helicases—with approximately 270 proteins encoded in the human genome (Letunic & Bork, 2017)—unwind double-stranded RNA or remodel RNA–protein interactions. They are required in nearly all nuclear and cytoplasmic gene regulatory steps involving RNA (Linder & Jankowsky, 2011). Akin to the other RBPs discussed above, DEAD box helicases are structurally very similar including their RNA interaction domains, but can interact with a vast range of RNAs and molecular machineries including the spliceosome, translating ribosomes, complexes involved in the biogenesis and assembly of small nuclear RNPs (snRNPs), as well as RNA-induced silencing complex (Linder & Jankowsky, 2011). The RNA binding site of DEAD box helicases lies within the helicase core that is structurally related to the bacterial recombinase RecA. The surrounding domains provide the structural context that allows the proteins to regulate such versatile processes (Hamann, Enders, & Ficner, 2019; Jarmoskaite & Russell, 2014). The structural studies on DEAD box helicases have given insights into how the interactions of the RBD with target RNAs are critical in defining the RBD activity. Since DEAD box helicases act as parts of large macromolecular assemblies, the overall complex likely contributes to their activity. However, as many RBPs so far, the DEAD box helicases have been structurally characterized only in isolation or with just a few co-factors.

## 2.2 | Noncanonical RNA binding proteins with novel RNA interacting domains

The development of RIC methods has enabled the unbiased cataloguing of proteins in complex with poly(A) RNA (Table 2; Baltz et al., 2012; Castello et al., 2012). There are multiple excellent reviews on RIC and we refer readers to these for details (Hentze et al., 2018; Licatalosi, Ye, & Jankowsky, 2019). Further adaptations of the technology have led to the characterization of distinct RNP species by using specific oligonucleotides for the isolation of RNA–protein complexes (Table 2). Methods for the capture of the non-polyadenylated RNA proteome are now completing the catalogue of proteins interacting with RNAs (Table 2). It is reassuring that most known RBPs were captured in these studies while revealing a plethora of proteins interacting with RNA that were previously not assigned to be RBPs and/or carry no

TABLE 2 Commonly used methods to validate individual protein–RNA interactions

Method	Interaction type	Material	Description	Qualifier	Example reference
GFP-trap	In vivo	GFP-tagged cell line	• In vivo UV-crosslinking • Immunoprecipitation (GFP-Trap) • Hybridization with fluorescent oligo-d(T) probe • Quantitation probe fluorescence/GFP	High-throughput, limited to poly(A)+	Strein, Alleaume, Rothbauer, Hentze, and Castello (2014)
Polynucleotide kinase (PNK) assay	In vivo	Epitope-tagged cell line	• In vivo UV-crosslinking • Immunoprecipitation • RNase digest and radioactive 5′ end labeling • SDS-PAGE and autoradiography	Fast and easy	Bressin et al. (2019)
Electrophoretic mobility shift assay (EMSA)	In vitro	Purified protein	• Protein purification • Incubation with RNA probe • Non-denaturing electrophoresis • Detection of a bandshift (Coomassie)	The classic, but can be technically demanding	Fillebeen, Wilkinson, and Pantopoulos (2014)
Fluorescence anisotropy	In vitro	Purified protein	• Protein purification • FA measurement in presence and absence of RNA probe	High-throughput, measures binding constants with high precision, requires dedicated equipment	Mao et al. (2006)
Isothermal titration calorimetry	In vitro	Purified protein	• Protein purification • Gradual titration of RNA ligand into sample	Measures binding constants with high precision, requires dedicated equipment	Recht, Ryder, and Williamson (2008)
Nitrocellulose filtration	In vitro	Purified protein	• Protein purification • Incubation with radioactively labeled RNA probe • Retention of protein and bound RNA on nitrocellulose filter • Scintillation counting	Fast and easy	Rio (2012)
NMR	In vitro	Purified protein	• Protein purification • NMR spectra in presence and absence of RNA	Can assess correct protein folding, for example, of an RNA-binding mutant, requires dedicated equipment	Schlundt et al. (2014)

known RBDs. Detailed biochemical and structural characterization of most novel RNA binders is still lacking, including the verification of the domains mediating RNA binding by independent methods. The novel RNA interactors that were characterized in more detail paint a picture of diverse modes of RNA binding among the noncanonical RBPs (Figure 3). In some proteins, the RNA binding capacity has been assigned to a structured domain with additional functions as exemplified by metabolic enzymes where RNA binding often involves a Rossman fold, a globular domain commonly responsible for interactions with nucleotide co-factors such as ATP/GTP or NAD(P)+/FAD (Liao et al., 2016). On the

FIGURE 4 The choice of RIC data normalization determines our perspective on RBPs. A representative RIC dataset from yeast was normalized to either a non-crosslinked (noCL) control (left panel), or to cellular protein abundances (right panel). Proteins annotated with a classical RBD (RRM, KH, dsRNA, Piwi, DEAD, Pumilio, CSD, zinc finger-CCCH) or any of the nonclassical RBDs identified in the initial human interactome or with the RBDmap approach (Castello et al., 2012, 2016) are highlighted in black and turquoise, respectively. Normalization of RIC data to a non-crosslinked control will yield the RNA's view of the RNA–protein interaction landscape (ochre). The normalization to protein abundances reveals the relative RNA binding activity of RNA interactors identified in the experiment (black). When combined, these viewpoints reflect the richness and the complexity of the RNA–protein interaction landscape
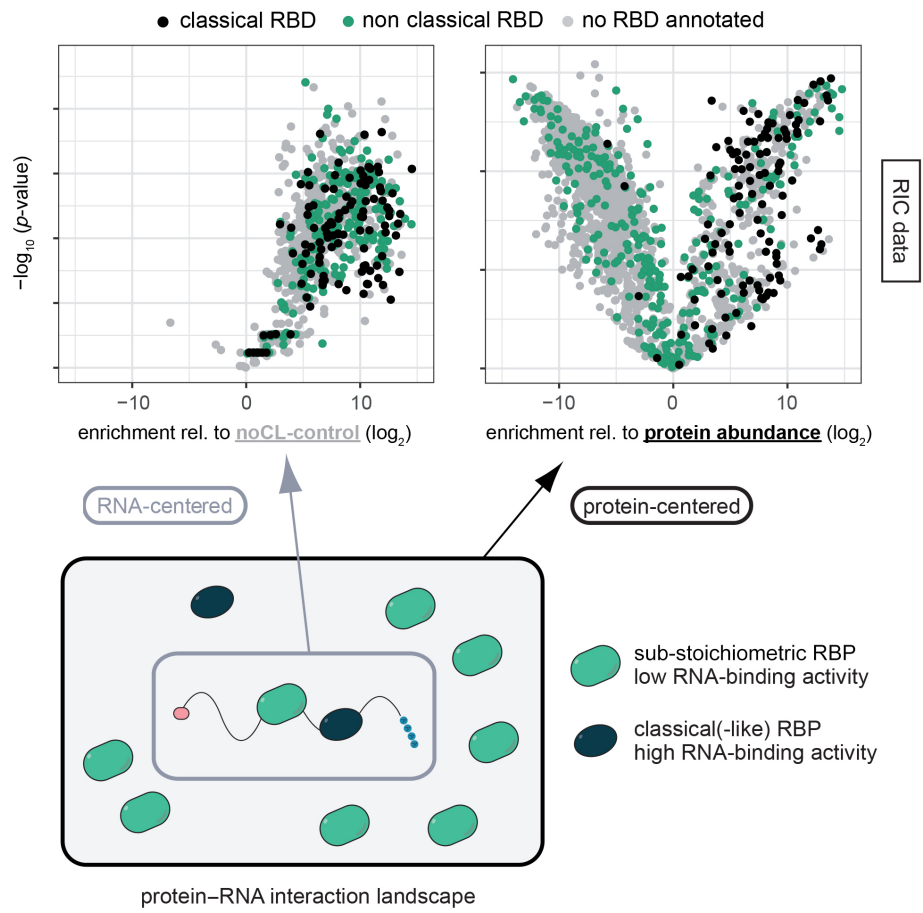
other hand, isocitrate dehydrogenase 1 (IDH1) provides an example where the RNA binding and catalytic activities seem to reside in different sites and do not compete with each other (Liu et al., 2019). Further characterization is required to determine if the RNA binding region of IDH1 and other noncanonical RBPs actually comprise new RBDs. Examples of well-defined novel RBDs do already exist, as demonstrated by DUF2373/WKF domain of C7orf50 (Trendel et al., 2019). In contrast to canonical RBPs, RNA interaction regions in noncanonical RBPs do not always have a fixed structure, hence called intrinsically disordered regions (IDRs; Jarvelin, Noerenberg, Davis, & Castello, 2016). In fact, IDRs are significantly overrepresented within RNA interactors lacking canonical RBDs. The disordered regions can be grouped into arginine-serine (RS)-rich, arginine-glycine (RG)-rich and other basic sequences and they can mediate both specific and nonspecific interactions with RNA (Jarvelin et al., 2016). The lack of well-defined domains poses novel challenges in characterizing the RNA–protein interactions such as prediction of RNA targets. For example, the computational prediction of RBP sites for well-defined RBDs is already difficult enough and is likely to be a great challenge in the case of noncanonical RBPs.

With growing numbers of novel RNA-binders—and in particular with the recurrent detection in RNA interactomes of proteins with well-described primary functions that are not RNA-related—comes the debate of how prevalent the observed interactions are in the cellular context. A very simple means to estimate in vivo RNA-binding activities of RBPs is the normalization of RIC data to cellular protein abundances, which has not been carried out in most of the initial RIC studies where the degree of protein enrichment was instead determined relative to a non-crosslinked control (Baltz et al., 2012; Beckmann et al., 2015; Castello et al., 2012; Conrad et al., 2016; Kwon et al., 2013; Matia-Gonzalez et al., 2015; Mitchell et al., 2013). RNA binders can be discriminated based on their enrichment in the RIC experiment relative to the input proteome (Kilchert et al. 2019) (Figure 4). Notably, the behavior of RBPs is by no means uniform, and spans a continuum of RNA binding activity. Proteins that harbor classical RBDs (such as an RRM or a KH-domain) tend to have a high in vivo RNA binding activity. In stark contrast, many novel RBPs—including some that carry non-classical RBDs identified with targeted approaches (Castello et al., 2016; Liao et al., 2016)—exhibit low in vivo RNA-binding activity and their binding to RNA is expected to be sub-stoichiometric (Figure 4). To some degree, the behavior of individual RBD-containing proteins can be extrapolated to the non-classical RBDs they harbor, and serve to

categorize them. For some domains, all or most proteins that contain the domain may turn out to have high in vivo RNA binding activity—these are "classical-like" in behavior and likely represent "professional" RBDs. Conversely, if a particular domain is systematically found in proteins whose occupancy on RNA is low, the proposed RBDs represents sub-stoichiometric RNA binding characteristics (Figure 4, lower panel). Sub-stoichiometric RBDs either have low affinity to RNA or may have regulated RNA binding activity. Such proteins are exemplified by NAD$^+$-dependent dehydrogenases that have been consistently detected in interactomes from various species (Baltz et al., 2012; Beckmann et al., 2015; Castello et al., 2012; Kramer et al., 2014; Kilhert et al. 2019). These protein were shown to bind to RNA via their dinucleotide-binding pocket (Castello et al., 2016; Liao et al., 2016), suggesting that they associate with RNA when they are idle, possibly to achieve negative feedback regulation of metabolic pathways. Likewise, cyclophilins, a class of proline cis/trans isomerases, show increased association with RNA upon virus infection (Garcia-Moreno et al., 2019). RIC studies have also provided evidence that some protein domains are "adaptive" in RNA binding. This is the case for any domain that usually mediates protein–protein interactions but can be "tweaked" to bind to RNA, for example, by the incorporation of additional basic amino acids. Such behavior has been proposed for the WD40 domain, which—if present within proteins shown to interact with RNA—is more likely to contain lysines and arginines than otherwise (Castello et al., 2012). It is important to point out that while from the protein's perspective many nonclassical RBPs may only bind to RNA at sub-stoichiometric levels, this does not mean that these interactions are negligible in quantity. In fact, as nonclassical RBPs can be very abundant, the likelihood that an individual RNA molecule interacts with an RBP with a low RNA binding activity may be just as high as to be bound by classical RBP that is expressed at low levels. It was precisely this "RNA-centered" side of the picture that was revealed by the original RIC studies, and that has fascinated the RNA community (Figure 4, left panel). If we now combine this view with the complementing "protein-centered" perspective, which takes information on RBP binding activities into account, we can begin to arrive at a holistic understanding of the protein–RNA interaction landscape.

2.3 | RNA binding domains in combination

A large proportion of canonical RBPs contain more than one RBD that have potential to modulate each other's RNA binding activity. The consensus target sequences for RBPs or individual RBDs can be determined systematically by in vitro SELEX (Systematic evolution of ligands by exponential enrichment) and other related methods evaluating RNA binding specificity or they can derived from the RNA crosslinking footprints in CLIP-seq data (Kishore et al., 2011; Reid et al., 2009; Riley et al., 2014). The target sequences of RBPs are usually short, degenerate RNA sequences and overall there has been a good correlation between SELEX and CLIP-seq data (Änkö & Neugebauer, 2012). However, the consensus RNA binding sequences of RBDs give a one-dimensional view on RNA–protein interactions and complementation by mechanistic and structural studies is critical to assess the contribution of each RBD within the protein or even larger macromolecular complex to the net RNA binding affinity, specificity and target site selection. The well-studied RRMs provide an example of how the combination of RBDs affects the RNA recognition by the RBPs. Multiple copies of RRMs or RRMs in combination with other RBDs are frequently found in canonical RBPs. In some cases such as the splicing factor Sex-lethal, the interaction between RRMs is induced by RNA binding, each of the RRMs contributing to the overall affinity and specificity upon RNA recognition (Handa et al., 1999). The polypyrimidine track binding protein (PTBP1) has four RRMs of which RRM 1 and 2 are independent of each other when not bound to RNA but RRM 3 and 4 interact already in the free state (Vitali et al., 2006). Systematic structural and biochemical characterization has revealed that the cooperation between RRM 3 and 4 bring together distant sequences of target RNA molecules inducing looping of the RNA (Lamichhane et al., 2010). The multidomain RBP IGF2BP3 combines two RRMs with four KH-domains to recognize distinct, appropriately spaced short sequence elements in a large sequence window, thus ensuring target selectivity (Schneider et al., 2019). The first evidence of the cooperation between an RRM and a zinc finger domain came from the SR protein SRSF7 (Cavaloc et al., 1999). SRSF7 is highly similar in amino acid sequence to SRSF3, another member of the SR protein family (Shepard & Hertel, 2009). Based on SELEX, structural and CLIP-seq studies SRSF7 binds to a purine-rich sequence whereas SRSF3 lacking the zinc knuckle binds a pyrimidine-rich sequence (Änkö et al., 2012; Cavaloc et al., 1999; Hargous et al., 2006; Muller-McNicoll et al., 2016). When the zinc knuckle of SRSF7 was mutated the RNA selectivity of SRSF7 became similar to the binding preference of SRSF3 (Cavaloc et al., 1999), demonstrating that the zinc knuckle modifies the RNA binding specificity of SRSF7. Reconstitution and structural analysis of mammalian CPSF have revealed that the polyadenylation signal is recognized in a combinatorial manner by the zinc finger domains 2 and 3 of CPSF30, and also the WD40 domain and N-terminus of

WDR33 (Clerici, Faini, Muckenfuss, Aebersold, & Jinek, 2018; Schonemann et al., 2014; Sun et al., 2018). Based on proteome-wide studies, additional components of the complex crosslink strongly with RNA and may be involved in recognition of auxiliary motifs (Castello et al., 2012; Kilhert et al. 2019). Another example is provided by hnRNP C. The combination of RNA binding footprints based on CLIP-seq and previously identified solution structures suggests that the RRMs and bZLMs (bZIP-like RBD) within hnRNP C oligomers cooperate in defining the RNA processing outcome of target RNAs (Konig et al., 2010; McAfee, Shahied-Milam, Soltaninassab, & LeStourgeon, 1996; Whitson, LeStourgeon, & Krezel, 2005). Although a single RBD is enough to interact with RNA, these examples demonstrate how frequently RBDs cooperate in RNA binding, giving clues on how different domains influence RNA target selection and RNA processing outcome. However, what remains unaddressed is why only a subset of the potential RNA targets are bound in a given cell at the given time. We know from accumulating data including RBP structures, CLIP-seq and biochemical studies that most RBPs do not bind to RNA promiscuously but show high target site specificity. Yet, their consensus target sites are usually short and degenerate and can be found frequently across the transcriptome. When accumulating new RBP structures, CLIP-seq and RIC data sets the key task remains to dissect what the role of each of the RBDs is not only in the biochemical aspects of the RNA–protein interaction but also in the RNA target selection that gives rise to the cellular significance of regulated RNA processing.

Interestingly, some RBPs with well-defined RBDs also contain IDRs that may interact with RNA. Do the IDRs or other regions of the protein influence RNA binding through direct contacts with RNA? SR proteins serve a good case study when looking at the RNA binding of canonical RNA binders with IDRs. SR proteins contain one or two RRM(s) that make specific interactions with RNA as well as a disordered RS domain (Änkö, 2014; Haynes & Iakoucheva, 2006; Shepard & Hertel, 2009). Domain swapping experiments on a few selected target RNAs demonstrated that the RRM is sufficient to determine RNA binding specificity at least on the studied RNAs (Sapra et al., 2009). However, the RS domain of SR proteins has been proposed to make direct contacts with RNA which is in accordance with the observed RNA binding capacity of RS-rich IDRs (Hertel & Graveley, 2005; Jarvelin et al., 2016; Shen & Green, 2006, 2007). With the technological advances in gene editing now allowing the alteration of protein properties at relative ease and CLIP-seq methods becoming routine approaches, it would be interesting to address the role of IDR in canonical RBPs by mapping the binding sites of modified SR proteins. Does the RS domain play a modulatory role on all or a subset of RNA targets? Another interesting and likely possibility is that the RS domain (i.e., the IDR) affects other properties of the proteins such as propensity to form condensates through LLPS (Jarvelin et al., 2016; Uversky, 2017; Wheeler & Hyman, 2018). SR proteins are no exception in this regard, and the contribution of noncanonical RBDs to the canonical RBD activity is an intriguing area of investigation.

3 | FROM PARTS LISTS TO COMPREHENSIVE UNDERSTANDING OF RNP FUNCTIONS

Genome-wide sequencing technologies have enabled the mapping of the RNA targets and detailed RNA binding sites of many RBPs. This has revealed complex networks of interactions, in most cases a single RBP binding to hundreds or thousands of RNAs. Inverse approaches (RIC and related methods) have identified RNP components and the multi-condition RIC studies have started addressing the dynamics of the RNP complexes (Tables 1 and 2). These studies—although conducted with mixed populations of cells and RNP species—suggest that individual RNA molecules interact with at least tens of proteins at any given time. However, the number and identity of all bound proteins has not been determined for any individual RNA to date. With the rapid development of single-cell and single-molecule techniques, we will likely see such data emerge in the near future. A much more daunting task will be to understand how the individually detected RNA–protein interactions affect the RNP as a whole. We have only started to investigate the RNA–protein pairs that were identified using the global methods and we still know surprisingly little on how the individual proteins of an RNP impact the activity of each other once at the RNA. We know much more about the recruitment of proteins to multiprotein RNA processing machineries, the step-wise assembly of the spliceosome serving as a prime example (Gornemann, Kotovic, Hujer, & Neugebauer, 2005; Listerman, Sapra, & Neugebauer, 2006). Conceptually, it seems obvious that the RNP components have synergistic or antagonistic effects, and there are examples of functional pairs. For instance, for the Sxl-Unr translation regulatory complex involved in dosage compensation in *Drosophila*, cooperative complex formation of the two RBPs was shown to increase the affinity to the target RNA by a 1000-fold (Hennig et al., 2014). Considering that the RNP composition is likely very dynamic in the cellular environment, studying the mechanistic interactions within the network of RNP components is a great challenge. One approach has been

"in vitro iCLIP" where crosslinking of a recombinant protein, U2AF2/U2AF65, to purified RNA was compared to the protein's in vivo crosslinking behavior to assess the extent to which its RNA binding is modulated by trans-acting factors (Sutandy et al., 2018). The identification of a whole host of novel RNA binders has complicated the task. In most cases, the jury is still out to resolve whether the novel RBPs without a typical RBDs interact with RNA to promote different steps of RNA metabolism or whether some of these interactions represent novel regulatory mechanisms through RNA–protein interactions. Examples of both types of interactions exist as is discussed below, suggesting that cells have evolved diverse ways to use RNA–protein interactions and formations of RNPs in controlling cellular processes.

3.1 | RNA–protein interactions—Structure is the function

The discovery of LLPS droplets or condensates (aka membraneless organelles) as an important organizing principle in cells has rekindled interest in researching RNPs (Brangwynne et al., 2009; Shin et al., 2018). Many of the LLPS domains in cells contain RNA in complex with proteins, RNPs forming non-membrane bound macromolecular condensates (also called granules) that concentrate specific sets of RNAs and regulatory proteins (Lin, Protter, Rosen, & Parker, 2015). Recent studies have demonstrated that RNA is an architectural element that can affect the composition and size of the condensates composed of RNA and protein (Garcia-Jove Navarro et al., 2019). The idea of RNA as a structural component is not new. The discovery of long noncoding RNAs (lncRNA) led to the identification of cellular sub-compartments where RNA acts as glue to bring various protein components together. One of the best characterized examples of such structures is the paraspeckle that is a nuclear body built around the lncRNA *NEAT1* (Bond & Fox, 2009; Fox et al., 2002). The *NEAT1* RNA provides a scaffold for the binding of the core paraspeckle proteins PSF/SFPQ, NONO, and PSPC1. The presence of NEAT1 and the RNA–protein interactions it mediates are critical for the maintenance and integrity of the paraspeckles (Fox et al., 2018). Paraspeckles are one of the few known subcellular structures where the RNA is absolutely essential for its formation but many other cellular "bodies" contain RNAs that help bringing together the constituents of the subcellular structure. The lncRNA *MALAT1* (aka *NEAT2*) is found in the nuclear speckles enriched in various RNA processing factors (Hutchinson et al., 2007; Ji et al., 2003). Unlike *NEAT1*, *MALAT1* is not required for the structural integrity of the speckles (Nakagawa et al., 2012). However, *MALAT1* can be found in most CLIP-seq datasets likely reflecting its high abundance in the nucleus but also bona fide interaction with RBPs. Cajal bodies are another sub-nuclear compartment build around the protein Coilin but containing a range of ncRNAs, in particular small nuclear RNAs. Coilin is an intrinsically disordered protein with no defined motifs forming abundant interactions with the RNAs localizing to Cajal bodies. Although individual RNAs per se may not be essential for the Cajal body structure, the interactions Coilin makes with the various RNAs likely are (Machyna et al., 2014; Machyna, Heyn, & Neugebauer, 2013). Formation of the RNP granules is functionally significant in gene regulation, for instance, by co-localizing RNA with their processing enzymes. Cajal bodies serve as an example of the in vivo significance of subcellular condensates. During zebrafish development, the formation of Cajal bodies promote the assembly of snRNPs by overcoming a rate-limiting step in their biogenesis (Strzelecka et al., 2010). On the contrary, the aberrant formation of RNP granules is linked to human disease as is highlighted by the enhanced understanding of the pathological mechanisms involved in neurodegenerative diseases amyotrophic lateral sclerosis and frontotemporal dementia (Mandrioli, Mediani, Alberti, & Carra, 2019).

3.2 | Protein binding RNAs

The more comprehensive picture of the abundant RNA–protein interactions in cells has demonstrated that RNA is not always in the receiving end as a target of modulation by RBPs. Instead, the roles can be reversed. The polycomb repressive complex 2 (PRC2)—a histone methyltransferase required for epigenetic silencing during development—is one of the best studied example of proteins whose activity is modulated by RNA (Davidovich & Cech, 2015). PRC2 consists of four core subunits (EZH2, SUZ12, EED, and YY1) as well as multiple accessory subunits and binds to G-rich RNA (Yu, Lee, Oksuz, Stafford, & Reinberg, 2019). Recently, the RNA-interaction domain of PRC2 was mapped to a patch within its allosteric regulatory site adjacent to the methyltransferase center (Long et al., 2017; Zhang et al., 2019). Interestingly, RNA binding by PRC2 inhibits its enzymatic activity (Cifuentes-Rojas et al., 2014; Kaneko, Son, Bonasio, Shen, & Reinberg, 2014), indicating that it is the RNA that regulates the function of the protein complex. Similar to PRC2, the enzymatic activity of Toll-like receptors (TLRs) are controlled by RNA binding. Pathogen associated-molecular patterns

including single- or double-stranded RNA derived from pathogens such as bacteria, fungi, parasites, and viruses and can be recognized by TLRs, leading to the activation of the TLR signaling cascade and expression of inflammation-related genes (Kawai & Akira, 2010). Intriguingly, also cellular miRNAs can acts as ligands of TLRs and initiate an immune response (Bayraktar et al., 2019). It is tempting to speculate that some of the metabolic enzymes identified in the RIC datasets as RNA binders may represent further examples of interactions where the RNA binds to the protein to modulate its function rather than vice versa.

3.3 | Functional validation of noncanonical RNA binding proteins

The logical next step following the extensive mapping of RNA binding sites and constituents of RNP complexes is to start functionally validating these interactions. By itself, RIC data reveals whether or not a protein associates with RNA, but it will not tell which parts of the protein are involved in RNA binding. In part, RBDs can be inferred from the proteome-wide RIC data if they are significantly overrepresented in the RNA interactome. However, this requires a sufficient number of annotations of a given domain to reach statistical significance. This approach also performs poorly on "adaptive" RBDs, completely fails in identifying unusual domains and "one-of-a-kind" RNA binders. To determine the function of a newly identified RNA binding event, classical mutagenesis studies may prove powerful as it is critical to separate the different RNA processing steps as well as identify the protein domains responsible for regulation. However, for the rational design of an RNA binding mutant, the knowledge of the domain—or ideally the amino acid residues—that mediate binding to RNA is crucial. RBDmap and related approaches can help in identification of RNA binding regions in proteins without canonical RBDs (Castello et al., 2016; Liao et al., 2016; Zhang et al., 2019). Here, RNA–protein complexes are subjected to limited proteolysis between two rounds of RNA pull-down and the generated peptides are mapped back to the polypeptide chain to determine the RBD. Alternative approaches use mass spectrometry to identify the precise location of the RNA–protein crosslinks (He et al., 2016; Kramer et al., 2014; Richter, Hsiao, Plessmann, & Urlaub, 2009; Shchepachev et al., 2019; Winz, Peil, Turowski, Rappsilber, & Tollervey, 2019). The identification of photo-crosslinks by mass spectrometry does not require a separate experiment, but can be added to the classical RIC work-flow in a modular fashion as enrichment of crosslinked peptides is carried out at the stage of the tryptic digest.

After the validation and identification of RNA binding regions, the greatest challenge is to identify the cellular functions of the newly identified RBP. As the number of RNA binders awaiting characterization is enormous, the comparative RIC studies may come handy in identifying processes where individual RBPs are recruited to distinct RNAs. The power of investigating the dynamics of RNPs in response to external cues has already been exemplified in the context of development, cellular adaptation to stress and viral infection (Despic et al., 2017; Garcia-Moreno et al., 2019; Marondedze et al., 2016; Shchepachev et al., 2019; Sysoev et al., 2016). In addition to the identification of condition-specific RBPs, comparative RIC has the potential to pinpoint regulated RNA binding events. When normalized to protein abundances, a comparative RIC experiment may reveal changes in the RNA occupancy of individual RBPs, which can be, for instance, a consequence of signal-dependent posttranslational modifications on the RBD. The KH-domain containing Sam68 is a known case of an RBP whose affinity to RNA can be modulated both negatively and positively by phosphorylation and acetylation, respectively (Babic, Jakymiw, & Fujita, 2004; Derry et al., 2000), and we expect other cases like this will be identified through targeted RIC experiments.

4 | CONCLUSIONS AND FUTURE PERSPECTIVE

The scope of known RNA–protein interactions in cellular regulation has exploded in the recent years. The life and function of all RNAs are delicately controlled by RBPs through various RNA biogenesis steps. Each regulatory interaction has the potential to modify the gene expression output of cells. Disruptions in RNA regulation have dire consequences for the cell as is seen in developmental defects and diseases including cancers in humans where the underlying cause involve RBPs or their target sites within RNAs (Brinegar & Cooper, 2016; Cooper, Wan, & Dreyfuss, 2009; Corbett, 2018; Montes, Sanford, Comiskey, & Chandler, 2019; Pereira, Billaud, & Almeida, 2017; Ratnadiwakara et al., 2018; Sterne-Weiler & Sanford, 2014). Recent studies have shown that the significance of RNA–protein interactions does not stop at RNA biogenesis. The cells take advantage of these interactions in forming subcellular structures and regulating the activity of various proteins with functions unrelated to RNA processing. This new direction in RNP biology has emerged largely through various RNP capture methods enabling a global view of all proteins in complex with cellular

RNAs. We are now at the stage where we need to start interpreting and functionally validating the RNA–protein interaction networks. This is a great but exciting challenge as understanding the comprehensive composition of individual RNPs will provide further answers to long-standing questions of combinatorial control in RNA regulation. This will shed light into how single RBPs can regulate such a multitude of processes. Furthermore, technology developments such as the CRISPR (clusters of regularly interspaced short palindromic repeats)/Cas9 technology (Jinek et al., 2012) is a prime example how detailed understanding of the nucleotide–protein interactions not only give insights into how cells work but also may have unexpected applications with far reaching implications.

CONFLICT OF INTEREST
The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS
Cornelia Kilchert: Conceptualization; investigation; visualization; and writing-original draft. **Katja Sträßer**: Conceptualization; methodology; and writing-original draft. **Vladislav Kunetsky**: Visualization. **Minna-Liisa Änkö**: Conceptualization; investigation; methodology; and writing-original draft.

ORCID
Minna-Liisa Änkö https://orcid.org/0000-0003-0446-3566

REFERENCES
Änkö, M.-L. (2014). Regulation of gene expression programmes by serine–arginine rich splicing factors. *Seminars in Cell & Developmental Biology*, *32*, 11–21.

Änkö, M.-L., Muller-McNicoll, M., Brandl, H., Curk, T., Gorup, C., Henry, I., ... Neugebauer, K. (2012). The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biology*, *13*, R17.

Änkö, M.-L., & Neugebauer, K. M. (2012). RNA protein interactions in vivo: Global gets specific. *Trends in Biochemical Sciences*, *37*, 255–262.

Asencio, C., Chatterjee, A., & Hentze, M. W. (2018). Silica-based solid-phase extraction of cross-linked nucleic acid-bound proteins. *Life Science Alliance*, *1*, e201800088.

Aubol, B. E., Serrano, P., Fattet, L., Wuthrich, K., & Adams, J. A. (2018). Molecular interactions connecting the function of the serine–arginine-rich protein SRSF1 to protein phosphatase 1. *Journal of Biological Chemistry*, *293*, 16751–16760.

Auweter, S. D., Fasan, R., Reymond, L., Underwood, J. G., Black, D. L., Pitsch, S., & Allain, F. H. T. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO Journal*, *25*, 163–173.

Babic, I., Jakymiw, A., & Fujita, D. J. (2004). The RNA binding protein Sam68 is acetylated in tumor cell lines, and its acetylation correlates with enhanced RNA binding activity. *Oncogene*, *23*, 3781–3789.

Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., ... Landthaler, M. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular Cell*, *46*, 674–690.

Bao, X., Guo, X., Yin, M., Tariq, M., Lai, Y., Kanwal, S., ... Esteban, M. A. (2018). Capturing the interactome of newly transcribed RNA. *Nature Methods*, *15*, 213–220.

Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., ... Frey, B. J. (2010). Deciphering the splicing code. *Nature*, *465*, 53–59.

Barnes, M., van Rensburg, G., Li, W. M., Mehmood, K., Mackedenski, S., Chan, C. M., ... Lee, C. H. (2015). Molecular insights into the coding region determinant-binding protein-RNA interaction through site-directed mutagenesis in the heterogeneous nuclear ribonucleoprotein-K-homology domains. *Journal of Biological Chemistry*, *290*, 625–639.

Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R., & Rosch, P. (1995). Structural studies of HIV-1 Tat protein. *Journal of Molecular Biology*, *247*, 529–535.

Bayraktar, R., Bertilaccio, M. T. S., & Calin, G. A. (2019). The interaction between two worlds: MicroRNAs and toll-like receptors. *Frontiers in Immunology*, *10*, 1053.

Beckmann, B. M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A. M., ... Hentze, M. W. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nature Communications*, *6*, 10127.

Bhat, V. D., McCann, K. L., Wang, Y., Fonseca, D. R., Shukla, T., Alexander, J. C., ... Campbell, Z. T. (2019). Engineering a conserved RNA regulatory protein repurposes its biological function in vivo. *eLife*, *8*, e4378.

Bond, C. S., & Fox, A. H. (2009). Paraspeckles: Nuclear bodies built on long noncoding RNA. *The Journal of Cell Biology*, *186*, 637–644.

Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., ... Hyman, A. A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, *324*, 1729–1732.

Bressin, A., Schulte-Sasse, R., Figini, D., Urdaneta, E. C., Beckmann, B. M., & Marsico, A. (2019). TriPepSVM: De novo prediction of RNA-binding proteins based on short amino acid motifs. *Nucleic Acids Research*, *47*, 4406–4417.

Bretschneider, H., Gandhi, S., Deshwar, A. G., Zuberi, K., & Frey, B. J. (2018). COSSMO: Predicting competitive alternative splice site selection using deep learning. *Bioinformatics*, *34*, i429–i437.

Brinegar, A. E., & Cooper, T. A. (2016). Roles for RNA-binding proteins in development and disease. *Brain Research*, *1647*, 1–8.

Bunnik, E. M., Batugedara, G., Saraf, A., Prudhomme, J., Florens, L., & Le Roch, K. G. (2016). The mRNA-bound proteome of the human malaria parasite *Plasmodium falciparum*. *Genome Biology*, *17*, 147.

Burd, C. G., & Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science*, *265*, 615–621.

Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., ... Patel, D. J. (1998). Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. *Nature Structural Biology*, *5*, 203–212.

Campbell, Z. T., Valley, C. T., & Wickens, M. (2014). A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nature Structural & Molecular Biology*, *21*, 732–738.

Casanal, A., Kumar, A., Hill, C. H., Easter, A. D., Emsley, P., Degliesposti, G., ... Passmore, L. A. (2017). Architecture of eukaryotic mRNA 3'-end processing machinery. *Science*, *358*, 1056–1059.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., ... Hentze, M. W. (2012). Insights into RNA biology from an Atlas of mammalian mRNA-binding proteins. *Cell*, *149*, 1393–1406.

Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A. M., Foehr, S., ... Hentze, M. W. (2016). Comprehensive identification of RNA-binding domains in human cells. *Molecular Cell*, *63*, 696–710.

Cavaloc, Y., Bourgeois, C. F., Kister, L., & Stevenin, J. (1999). The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA*, *5*, 468–483.

Cifuentes-Rojas, C., Hernandez, A. J., Sarma, K., & Lee, J. T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. *Molecular Cell*, *55*, 171–185.

Clerici, M., Faini, M., Muckenfuss, L. M., Aebersold, R., & Jinek, M. (2018). Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. *Nature Structural & Molecular Biology*, *25*, 135–138.

Clery, A., Blatter, M., & Allain, F. H. (2008). RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology*, *18*, 290–298.

Clingman, C. C., Deveau, L. M., Hay, S. A., Genga, R. M., Shandilya, S. M., Massi, F., & Ryder, S. P. (2014). Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. *eLife*, *3*, e02848.

Conrad, T., Albrecht, A. S., de Melo Costa, V. R., Sauer, S., Meierhofer, D., & Orom, U. A. (2016). Serial interactome capture of the human cell nucleus. *Nature Communications*, *7*, 11212.

Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., & Hughes, T. R. (2011). RBPDB: A database of RNA-binding specificities. *Nucleic Acids Research*, *39*, D301–D308.

Cooper, T. A., Wan, L., & Dreyfuss, G. (2009). RNA and Disease. *Cell*, *136*, 777–793.

Corbett, A. H. (2018). Post-transcriptional regulation of gene expression and human disease. *Current Opinion in Cell Biology*, *52*, 96–104.

Courchaine, E. M., Lu, A., & Neugebauer, K. M. (2016). Droplet organelles? *EMBO Journal*, *35*, 1603–1612.

Davidovich, C., & Cech, T. R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *RNA*, *21*, 2007–2022.

Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., ... Cherry, J. M. (2018). The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Research*, *46*, D794–D801.

Deo, R. C., Bonanno, J. B., Sonenberg, N., & Burley, S. K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell*, *98*, 835–845.

Derry, J. J., Richard, S., Valderrama Carvajal, H., Ye, X., Vasioukhin, V., Cochrane, A. W., ... Tyner, A. L. (2000). Sik (BRK) phosphorylates Sam68 in the nucleus and negatively regulates its RNA binding ability. *Molecular and Cellular Biology*, *20*, 6114–6126.

Despic, V., Dejung, M., Gu, M., Krishnan, J., Zhang, J., Herzel, L., ... Neugebauer, K. M. (2017). Dynamic RNA–protein interactions underlie the zebrafish maternal-to-zygotic transition. *Genome Research*, *27*, 1184–1194.

Dong, S., Wang, Y., Cassidy-Amstutz, C., Lu, G., Bigler, R., Jezyk, M. R., ... Wang, Z. (2011). Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. *Journal of Biological Chemistry*, *286*, 26732–26742.

Drino, A., & Schaefer, M. R. (2018). RNAs, phase separation, and membrane-less organelles: Are post-transcriptional modifications modulating organelle dynamics? *Bioessays*, *40*, e1800085.

Filipovska, A., Razif, M. F., Nygard, K. K., & Rackham, O. (2011). A universal code for RNA recognition by PUF proteins. *Nature Chemical Biology*, *7*, 425–427.

Fillebeen, C., Wilkinson, N., & Pantopoulos, K. (2014). Electrophoretic mobility shift assay (EMSA) for the study of RNA–protein interactions: The IRE/IRP example. *Journal of Visualized Experiments*, *2014*, 52230.

Fox, A. H., Lam, Y. W., Leung, A. K. L., Lyon, C. E., Andersen, J., Mann, M., & Lamond, A. I. (2002). Paraspeckles: A novel nuclear domain. *Current Biology*, *12*, 13–25.

Fox, A. H., Nakagawa, S., Hirose, T., & Bond, C. S. (2018). Paraspeckles: Where long noncoding RNA meets phase separation. *Trends in Biochemical Sciences*, *43*, 124–135.

Garcia-Jove Navarro, M., Kashida, S., Chouaib, R., Souquere, S., Pierron, G., Weil, D., & Gueroui, Z. (2019). RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates. *Nature Communications*, *10*, 3230.

Garcia-Moreno, M., Noerenberg, M., Ni, S., Jarvelin, A. I., Gonzalez-Almela, E., Lenz, C. E., ... Castello, A. (2019). System-wide profiling of RNA-binding proteins uncovers key regulators of virus infection. *Molecular Cell*, *74*, 196–211.e111.

Gehring, N. H., Wahle, E., & Fischer, U. (2017). Deciphering the mRNP Code: RNA-bound determinants of post-transcriptional gene regulation. *Trends in Biochemical Sciences*, *42*, 369–382.

Gerstberger, S., Hafner, M., & Tuschl, T. (2014). A census of human RNA-binding proteins. *Nature Reviews. Genetics*, *15*, 829–845.

Giudice, G., Sanchez-Cabo, F., Torroja, C., & Lara-Pezzi, E. (2016). ATtRACT—A database of RNA-binding proteins and associated motifs. *Database*, *2016*.

Giulietti, M., Piva, F., D'Antonio, M., D'Onorio De Meo, P., Paoletti, D., Castrignano, T., ... Pesole, G. (2013). SpliceAid-F: A database of human splicing factors and their RNA-binding sites. *Nucleic Acids Research*, *41*, D125–D131.

Gornemann, J., Kotovic, K. M., Hujer, K., & Neugebauer, K. M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Molecular Cell*, *19*, 53–63.

Grishin, N. V. (2001). KH domain: One motif, two folds. *Nucleic Acids Research*, *29*, 638–643.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., ... Tuschl, T. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, *141*, 129–141.

Hall, T. M. (2005). Multiple modes of RNA recognition by zinc finger proteins. *Current Opinion in Structural Biology*, *15*, 367–373.

Hamann, F., Enders, M., & Ficner, R. (2019). Structural basis for RNA translocation by DEAH-box ATPases. *Nucleic Acids Research*, *47*, 4349–4362.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., ... Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature*, *398*, 579–585.

Hargous, Y., Hautbergue, G. M., Tintaru, A. M., Skrisovska, L., Golovanov, A. P., Stevenin, J., ... Allain, F. H. (2006). Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO Journal*, *25*, 5126–5137.

Haynes, C., & Iakoucheva, L. M. (2006). Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Research*, *34*, 305–312.

He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B. A., & Bonasio, R. (2016). High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Molecular Cell*, *64*, 416–430.

Hennig, J., Militti, C., Popowicz, G. M., Wang, I., Sonntag, M., Geerlof, A., ... Sattler, M. (2014). Structural basis for the assembly of the Sxl-Unr translation regulatory complex. *Nature*, *515*, 287–290.

Hentze, M. W., Castello, A., Schwarzl, T., & Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nature Reviews. Molecular Cell Biology*, *19*, 327–341.

Hertel, K. J., & Graveley, B. R. (2005). RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends in Biochemical Sciences*, *30*, 115–118.

Herzel, L., Straube, K., & Neugebauer, K. M. (2018). Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Research*, *28*, 1008–1019.

Hieronymus, H., & Silver, P. A. (2003). Genome-wide analysis of RNA–protein interactions illustrates specificity of the mRNA export machinery. *Nature Genetics*, *33*, 155–161.

Hug, N., Longman, D., & Cáceres, J. F. (2016). Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research*, *44*, 1483–1495.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L. E., Sibley, C. R., Sugimoto, Y., ... Ule, J. (2014). iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods*, *65*, 274–287.

Hutchinson, J. N., Ensminger, A. W., Clemson, C. M., Lynch, C. R., Lawrence, J. B., & Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*, *8*, 39.

Jarmoskaite, I., & Russell, R. (2014). RNA helicase proteins as chaperones and remodelers. *Annual Review of Biochemistry*, *83*, 697–725.

Jarvelin, A. I., Noerenberg, M., Davis, I., & Castello, A. (2016). The new (dis)order in RNA regulation. *Cell Communication and Signaling*, *14*, 9.

Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P. M., ... Müller-Tidow, C. (2003). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, *22*, 8031–8041.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*, 816–821.

Kaneko, S., Son, J., Bonasio, R., Shen, S. S., & Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check. *Genes & Development*, *28*, 1983–1988.

Kawai, T., & Akira, S. (2010). The role of pattern-recognition receptors in innate immunity: Update on toll-like receptors. *Nature Immunology*, *11*, 373–384.

Kilchert, C., Kecman, T., Priest, E., Hester, S., Kus, K., Castello, A., Mohammed, S., Vasiljeva, L. (2019) System-wide analyses of the fission yeast poly(A)+ RNA interactome reveal insights into organisation and function of RNA-protein complexes, *bioRxiv*, 748194.

Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., & Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature Methods*, *8*, 559–564.

Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annual Review of Biochemistry*, *79*, 213–231.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., ... Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Structural & Molecular Biology*, *17*, 909–915.

Kramer, K., Sachsenberg, T., Beckmann, B. M., Qamar, S., Boon, K. L., Hentze, M. W., ... Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nature Methods*, *11*, 1064–1070.

Kuwasako, K., Nameki, N., Tsuda, K., Takahashi, M., Sato, A., Tochio, N., ... Muto, Y. (2017). Solution structure of the first RNA recognition motif domain of human spliceosomal protein SF3b49 and its mode of interaction with a SF3b145 fragment. *Protein Science*, *26*, 280–291.

Kwon, S. C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K. T., ... Kim, V. N. (2013). The RNA-binding protein repertoire of embryonic stem cells. *Nature Structural & Molecular Biology*, *20*, 1122–1130.

Laity, J. H., Lee, B. M., & Wright, P. E. (2001). Zinc finger proteins: New insights into structural and functional diversity. *Current Opinion in Structural Biology*, *11*, 39–46.

Lamichhane, R., Daubner, G. M., Thomas-Crusells, J., Auweter, S. D., Manatschal, C., Austin, K. S., ... Rueda, D. (2010). RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 4105–4110.

Letunic, I., & Bork, P. (2017). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*, *46*, D493–D496.

Liao, Y., Castello, A., Fischer, B., Leicht, S., Foehr, S., Frese, C. K., ... Preiss, T. (2016). The cardiomyocyte RNA-binding proteome: Links to intermediary metabolism and heart disease. *Cell Reports*, *16*, 1456–1469.

Liao, J.-Y., Yang, B., Zhang, Y.-C., Wang, X.-J., Ye, Y., Peng, J.-W., ... Yin, D. (2019). EuRBPDB: a comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA binding proteins (RBPs). *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz823

Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., ... Darnell, R. B. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, *456*, 464–469.

Licatalosi, D. D., Ye, X., & Jankowsky, E. (2019). Approaches for measuring the dynamics of RNA–protein interactions. *WIREs RNA*, *11*, e1565.

Liepelt, A., Naarmann-de Vries, I. S., Simons, N., Eichelbaum, K., Fohr, S., Archer, S. K., ... Ostareck-Lederer, A. (2016). Identification of RNA-binding proteins in macrophages by interactome capture. *Molecular & Cellular Proteomics*, *15*, 2699–2714.

Lin, Y., Protter, D. S., Rosen, M. K., & Parker, R. (2015). Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Molecular Cell*, *60*, 208–219.

Linder, P., & Jankowsky, E. (2011). From unwinding to clamping—The DEAD box RNA helicase family. *Nature Reviews. Molecular Cell Biology*, *12*, 505–516.

Listerman, I., Sapra, A. K., & Neugebauer, K. M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nature Structural & Molecular Biology*, *13*, 815–822.

Liu, L., Li, T., Song, G., He, Q., Yin, Y., Lu, J. Y., ... Shen, X. (2019). Insight into novel RNA-binding activities via large-scale analysis of lncRNA-bound proteome and IDH1-bound transcriptome. *Nucleic Acids Research*, *47*, 2244–2262.

Long, Y., Bolanos, B., Gong, L., Liu, W., Goodrich, K. J., Yang, X., ... Liu, X. (2017). Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. *eLife*, *6*, e31558.

Lu, D., Searles, M. A., & Klug, A. (2003). Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature*, *426*, 96–100.

Lueong, S., Merce, C., Fischer, B., Hoheisel, J. D., & Erben, E. D. (2016). Gene expression regulatory networks in *Trypanosoma brucei*: Insights into the role of the mRNA-binding proteome. *Molecular Microbiology*, *100*, 457–471.

Lukong, K. E., & Richard, S. (2003). Sam68, the KH domain-containing superSTAR. *Biochimica et Biophysica Acta*, *1653*, 73–86.

Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. *Nature Reviews. Molecular Cell Biology*, *8*, 479–490.

Machyna, M., Heyn, P., & Neugebauer, K. M. (2013). Cajal bodies: Where form meets function. *WIREs RNA*, *4*, 17–34.

Machyna, M., Kehr, S., Straube, K., Kappei, D., Buchholz, F., Butter, F., ... Neugebauer, K. M. (2014). The coilin interactome identifies hundreds of small noncoding RNAs that traffic through Cajal bodies. *Molecular Cell*, *56*, 389–399.

Mandrioli, J., Mediani, L., Alberti, S., & Carra, S. (2019). ALS and FTD: Where RNA metabolism meets protein quality control. *Seminars in Cell and Developmental Biology*. [Epub ahead of print]

Maniatis, T., & Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature*, *416*, 499–506.

Mao, C., Flavin, K. G., Wang, S., Dodson, R., Ross, J., & Shapiro, D. J. (2006). Analysis of RNA–protein interactions by a microplate-based fluorescence anisotropy assay. *Analytical Biochemistry*, *350*, 222–232.

Marondedze, C., Thomas, L., Serrano, N. L., Lilley, K. S., & Gehring, C. (2016). The RNA-binding protein repertoire of *Arabidopsis thaliana*. *Scientific Reports*, *6*, 29766.

Matia-Gonzalez, A. M., Laing, E. E., & Gerber, A. P. (2015). Conserved mRNA-binding proteomes in eukaryotic organisms. *Nature Structural & Molecular Biology*, *22*, 1027–1033.

Matus-Ortega, M. E., Regonesi, M. E., Pina-Escobedo, A., Tortora, P., Deho, G., & Garcia-Mena, J. (2007). The KH and S1 domains of *Escherichia coli* polynucleotide phosphorylase are necessary for autoregulation and growth at low temperature. *Biochimica et Biophysica Acta*, *1769*, 194–203.

McAfee, J. G., Shahied-Milam, L., Soltaninassab, S. R., & LeStourgeon, W. M. (1996). A major determinant of hnRNP C protein binding to RNA is a novel bZIP-like RNA binding domain. *RNA*, *2*, 1139–1152.

Meinel, D. M., & Strasser, K. (2015). Co-transcriptional mRNP formation is coordinated within a molecular mRNP packaging station in *S. cerevisiae*. *Bioessays*, *37*, 666–677.

Michlewski, G., & Caceres, J. F. (2019). Post-transcriptional control of miRNA biogenesis. *RNA*, *25*, 1–16.

Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., ... Lee, J. T. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science*, *349*, aab2276.

Mitchell, S. F., Jain, S., She, M., & Parker, R. (2013). Global analysis of yeast mRNPs. *Nature Structural & Molecular Biology*, *20*, 127–133.

Montes, M., Sanford, B. L., Comiskey, D. F., & Chandler, D. S. (2019). RNA splicing and disease: Animal models to therapies. *Trends in Genetics*, *35*, 68–87.

Mukherjee, J., Hermesh, O., Eliscovich, C., Nalpas, N., Franz-Wachtel, M., Macek, B., & Jansen, R. P. (2019). β-Actin mRNA interactome mapping by proximity biotinylation. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 12863–12872.

Muller-McNicoll, M., Botti, V., de Jesus Domingues, A. M., Brandl, H., Schwich, O. D., Steiner, M. C., ... Neugebauer, K. M. (2016). SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes & Development*, *30*, 553–566.

Muller-McNicoll, M., & Neugebauer, K. M. (2013). How cells get the message: Dynamic assembly and function of mRNA-protein complexes. *Nature Reviews Genetics*, *14*, 275–287.

Myrick, L. K., Hashimoto, H., Cheng, X., & Warren, S. T. (2015). Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. *Human Molecular Genetics*, *24*, 1733–1740.

Nagata, T., Suzuki, S., Endo, R., Shirouzu, M., Terada, T., Inoue, M., ... Yokoyama, S. (2008). The RRM domain of poly(A)-specific ribonuclease has a noncanonical binding site for mRNA cap analog recognition. *Nucleic Acids Research*, *36*, 4754–4767.

Nakagawa, S., Ip, J. Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T., & Prasanth, K. V. (2012). Malat1 is not an essential component of nuclear speckles in mice. *RNA*, *18*, 1487–1499.

Nussbacher, J. K., & Yeo, G. W. (2018). Systematic discovery of RNA binding proteins that regulate microRNA levels. *Molecular Cell*, *69*, 1005–1016.e1007.

Oddone, A., Lorentzen, E., Basquin, J., Gasch, A., Rybin, V., Conti, E., & Sattler, M. (2007). Structural and biochemical characterization of the yeast exosome component Rrp40. *EMBO Reports*, *8*, 63–69.

Pereira, B., Billaud, M., & Almeida, R. (2017). RNA-binding proteins in cancer: Old players and new actors. *Trends in Cancer*, *3*, 506–528.

Petoukhov, M. V., Monie, T. P., Allain, F. H. T., Matthews, S., Curry, S., & Svergun, D. I. (2006). Conformation of polypyrimidine tract binding protein in solution. *Structure*, *14*, 1021–1027.

Pintacuda, G., Young, A. N., & Cerase, A. (2017). Function by structure: Spotlights on Xist long non-coding RNA. *Frontiers in Molecular Biosciences*, *4*, 90.

Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992). Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science*, *257*, 76–80.

Ratnadiwakara, M., Mohenska, M., & Anko, M. L. (2018). Splicing factors as regulators of miRNA biogenesis—Links to human disease. *Seminars in Cell and Developmental Biology*, *79*, 113–122.

Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., ... Lührmann, R. (2016). Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science*, *353*, 1399–1405.

Recht, M. I., Ryder, S. P., & Williamson, J. R. (2008). Monitoring assembly of ribonucleoprotein complexes by isothermal titration calorimetry. *Methods in Molecular Biology*, *488*, 117–127.

Reichel, M., Liao, Y., Rettel, M., Ragan, C., Evers, M., Alleaume, A. M., ... Millar, A. A. (2016). In planta determination of the mRNA-binding proteome of Arabidopsis etiolated seedlings. *The Plant Cell*, *28*, 2435–2452.

Reid, D. C., Chang, B. L., Gunderson, S. I., Alpert, L., Thompson, W. A., & Fairbrother, W. G. (2009). Next-generation SELEX identifies sequence and structural determinants of splicing factor binding in human pre-mRNA sequence. *RNA*, *15*, 2385–2397.

Richter, F. M., Hsiao, H. H., Plessmann, U., & Urlaub, H. (2009). Enrichment of protein–RNA crosslinks from crude UV-irradiated mixtures for MS analysis by on-line chromatography using titanium dioxide columns. *Biopolymers*, *91*, 297–309.

Riley, T. R., Slattery, M., Abe, N., Rastogi, C., Liu, D., Mann, R. S., & Bussemaker, H. J. (2014). SELEX-seq: A method for characterizing the complete repertoire of binding site preferences for transcription factor complexes. *Methods in Molecular Biology*, *1196*, 255–278.

Rio, D. C. (2012). Filter-binding assay for analysis of RNA–protein interactions. *Cold Spring Harbor Protocols*, *2012*, 1078–1081.

Rogell, B., Fischer, B., Rettel, M., Krijgsveld, J., Castello, A., & Hentze, M. W. (2017). Specific RNP capture with antisense LNA/DNA mixmers. *RNA*, *23*, 1290–1302.

Saldana-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jacome-Lopez, K., Nora, E. P., ... Reinberg, D. (2019). RNA interactions are essential for CTCF-mediated genome organization. *Molecular Cell*, *76*, 412–422.e5.

Sapra, A. K., Änkö, M.-L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., ... Neugebauer, K. M. (2009). SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. *Molecular Cell*, *34*, 179–190.

Sawicka, K., Bushell, M., Spriggs, K. A., & Willis, A. E. (2008). Polypyrimidine-tract-binding protein: A multifunctional RNA-binding protein. *Biochemical Society Transactions*, *36*, 641–647.

Scheiba, R. M., de Opakua, A. I., Diaz-Quintana, A., Cruz-Gallardo, I., Martinez-Cruz, L. A., Martinez-Chantar, M. L., ... Diaz-Moreno, I. (2014). The C-terminal RNA binding motif of HuR is a multi-functional domain leading to HuR oligomerization and binding to U-rich RNA targets. *RNA Biology*, *11*, 1250–1261.

Schlundt, A., Heinz, G. A., Janowski, R., Geerlof, A., Stehle, R., Heissmeyer, V., ... Sattler, M. (2014). Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nature Structural & Molecular Biology*, *21*, 671–678.

Schneider, T., Hung, L. H., Aziz, M., Wilmen, A., Thaum, S., Wagner, J., ... Bindereif, A. (2019). Combinatorial recognition of clustered RNA elements by the multidomain RNA-binding protein IMP3. *Nature Communications*, *10*, 2266.

Schonemann, L., Kuhn, U., Martin, G., Schafer, P., Gruber, A. R., Keller, W., ... Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: Recognition of the polyadenylation signal by WDR33. *Genes & Development*, *28*, 2381–2393.

Schuller, J. M., Falk, S., Fromm, L., Hurt, E., & Conti, E. (2018). Structure of the nuclear exosome captured on a maturing preribosome. *Science*, *360*, 219–222.

Shchepachev, V., Bresson, S., Spanos, C., Petfalski, E., Fischer, L., Rappsilber, J., & Tollervey, D. (2019). Defining the RNA interactome by total RNA-associated protein purification. *Molecular Systems Biology*, *15*, e8689.

Shen, H., & Green, M. R. (2006). RS domains contact splicing signals and promote splicing by a commonmechanism in yeast through humans. *Genes & Development*, *20*, 1755–1765.

Shen, H., & Green, M. R. (2007). RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. *Nature Structural & Molecular Biology*, *14*, 597–603.

Shepard, P., & Hertel, K. (2009). The SR protein family. *Genome Biology*, *10*, 242.

Shin, Y., Chang, Y. C., Lee, D. S. W., Berry, J., Sanders, D. W., Ronceray, P., ... Brangwynne, C. P. (2018). Liquid nuclear condensates mechanically sense and restructure the genome. *Cell*, *175*, 1481–1491.e1413.

Siomi, H., Choi, M., Siomi, M. C., Nussbaum, R. L., & Dreyfuss, G. (1994). Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell*, *77*, 33–39.

Sterne-Weiler, T., & Sanford, J. (2014). Exon identity crisis: Disease-causing mutations that disrupt the splicing code. *Genome Biology*, *15*, 201.

Strein, C., Alleaume, A. M., Rothbauer, U., Hentze, M. W., & Castello, A. (2014). A versatile assay for RNA-binding proteins in living cells. *RNA*, *20*, 721–731.

Strzelecka, M., Trowitzsch, S., Weber, G., Luhrmann, R., Oates, A. C., & Neugebauer, K. M. (2010). Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis. *Nature Structural & Molecular Biology*, *17*, 403–409.

Sun, Y., Zhang, Y., Hamilton, K., Manley, J. L., Shi, Y., Walz, T., & Tong, L. (2018). Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E1419–E1428.

Sutandy, F. X. R., Ebersberger, S., Huang, L., Busch, A., Bach, M., Kang, H. S., ... König, J. (2018). In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. *Genome Research*, *28*, 699–713.

Sysoev, V. O., Fischer, B., Frese, C. K., Gupta, I., Krijgsveld, J., Hentze, M. W., ... Ephrussi, A. (2016). Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *Nature Communications*, *7*, 12128.

Tadesse, H., Deschenes-Furry, J., Boisvenue, S., & Cote, J. (2008). KH-type splicing regulatory protein interacts with survival motor neuron protein and is misregulated in spinal muscular atrophy. *Human Molecular Genetics*, *17*, 506–524.

Teplova, M., Malinina, L., Darnell, J. C., Song, J., Lu, M., Abagyan, R., ... Patel, D. J. (2011). Protein–RNA and protein–protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. *Structure*, *19*, 930–944.

Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiss, J. L., Eichner, N., ... Meister, G. (2017). A compendium of RNA-binding proteins that regulate microRNA biogenesis. *Molecular Cell*, *66*, 270–284.e213.

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M. W., & Krijgsveld, J. (2019). The human RNA-binding proteome and its dynamics during translational arrest. *Cell*, *176*, 391–403.e319.

Ule, J., Jensen, K., Mele, A., & Darnell, R. B. (2005). CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods*, *37*, 376–386.

Urdaneta, E. C., Vieira-Vieira, C. H., Hick, T., Wessels, H. H., Figini, D., Moschall, R., ... Beckmann, B. M. (2019). Purification of cross-linked RNA-protein complexes by phenol–toluol extraction. *Nature Communications*, *10*, 990.

Uversky, V. N. (2017). Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current Opinion in Structural Biology*, *44*, 18–30.

Valverde, R., Edwards, L., & Regan, L. (2008). Structure and function of KH domains. *FEBS Journal*, *275*, 2712–2726.

Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., ... Yeo, G. W. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature Methods*, *13*, 508–514.

Vitali, F., Henning, A., Oberstrass, F. C., Hargous, Y., Auweter, S. D., Erat, M., & Allain, F. H.-T. (2006). Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO Journal*, *25*, 150–162.

Wan, F., Anderson, D. E., Barnitz, R. A., Snow, A., Bidere, N., Zheng, L., ... Lenardo, M. J. (2007). Ribosomal protein S3: A KH domain subunit in NF-kappaB complexes that mediates selective gene regulation. *Cell*, *131*, 927–939.

Wan, R., Yan, C., Bai, R., Huang, G., & Shi, Y. (2016). Structure of a yeast catalytic step I spliceosome at 3.4 A resolution. *Science*, *353*, 895–904.

Wang, M., Oge, L., Perez-Garcia, M. D., Hamama, L., & Sakr, S. (2018). The PUF protein family: Overview on PUF RNA targets, biological functions, and post-transcriptional regulation. *International Journal of Molecular Sciences*, *19*, E410.

Wang, W., Maucuer, A., Gupta, A., Manceau, V., Thickman, K. R., Bauer, W. J., ... Kielkopf, C. L. (2013). Structure of phosphorylated SF1 bound to U2AF(6)(5) in an essential splicing factor complex. *Structure*, *21*, 197–208.

Wang, X., McLachlan, J., Zamore, P. D., & Hall, T. M. (2002). Modular recognition of RNA by a human pumilio-homology domain. *Cell*, *110*, 501–512.

Wang, Z., Qiu, H., He, J., Liu, L., Xue, W., Fox, A., ... Xu, J. (2019). The emerging roles of hnRNPK. *Journal of Cellular Physiology*, 1995-2008.

Wessels, H. H., Imami, K., Baltz, A. G., Kolinski, M., Beldovskaya, A., Selbach, M., ... Landthaler, M. (2016). The mRNA-bound proteome of the early fly embryo. *Genome Research*, *26*, 1000–1009.

Wheeler, R. J., & Hyman, A. A. (2018). Controlling compartmentalization by non-membrane-bound organelles. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*, 20170193.

Whitson, S. R., LeStourgeon, W. M., & Krezel, A. M. (2005). Solution structure of the symmetric coiled coil tetramer formed by the oligomerization domain of hnRNP C: Implications for biological function. *Journal of Molecular Biology*, *350*, 319–337.

Wickramasinghe, V. O., & Laskey, R. A. (2015). Control of mammalian gene expression by selective mRNA export. *Nature Reviews. Molecular Cell Biology*, *16*, 431–442.

Winz, M. L., Peil, L., Turowski, T. W., Rappsilber, J., & Tollervey, D. (2019). Molecular interactions between Hel2 and RNA supporting ribosome-associated quality control. *Nature Communications*, *10*, 563.

Wu, M., Nilsson, P., Henriksson, N., Niedzwiecka, A., Lim, M. K., Cheng, Z., ... Song, H. (2009). Structural basis of m(7)GpppG binding to poly(A)-specific ribonuclease. *Structure*, *17*, 276–286.

Yu, J. R., Lee, C. H., Oksuz, O., Stafford, J. M., & Reinberg, D. (2019). PRC2 is high maintenance. *Genes & Development*, *33*, 903–935.

Zhang, Q., McKenzie, N. J., Warneford-Thomson, R., Gail, E. H., Flanigan, S. F., Owen, B. M., ... Davidovich, C. (2019). RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nature Structural & Molecular Biology*, *26*, 237–247.

Zhang, Y., Madl, T., Bagdiul, I., Kern, T., Kang, H. S., Zou, P., ... Sattler, M. (2013). Structure, phosphorylation and U2AF65 binding of the N-terminal domain of splicing factor 1 during 3′-splice site recognition. *Nucleic Acids Research*, *41*, 1343–1354.

Zhao, Y. Y., Mao, M. W., Zhang, W. J., Wang, J., Li, H. T., Yang, Y., ... Wu, J. W. (2018). Expanding RNA binding specificity and affinity of engineered PUF domains. *Nucleic Acids Research*, *46*, 4771–4782.