A brave new world of RNA-binding proteins

Matthias W. Hentze[1], Alfredo Castello[2], Thomas Schwarzl[1] and Thomas Preiss[3,4]

Abstract | RNA-binding proteins (RBPs) are typically thought of as proteins that bind RNA through one or multiple globular RNA-binding domains (RBDs) and change the fate or function of the bound RNAs. Several hundred such RBPs have been discovered and investigated over the years. Recent proteome-wide studies have more than doubled the number of proteins implicated in RNA binding and uncovered hundreds of additional RBPs lacking conventional RBDs. In this Review, we discuss these new RBPs and the emerging understanding of their unexpected modes of RNA binding, which can be mediated by intrinsically disordered regions, protein–protein interaction interfaces and enzymatic cores, among others. We also discuss the RNA targets and molecular and cellular functions of the new RBPs, as well as the possibility that some RBPs may be regulated by RNA rather than regulate RNA.

RNA recognition motif (RRM). An RNA-binding domain of ~90 amino acids that folds into two α-helices packed against a four-stranded β-sheet, which interact with RNA.

[1]*European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany.*
[2]*Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK*
[3]*European Molecular Biology Laboratory–Australia Collaborating Group, Department of Genome Sciences, The John Curtin School of Medical Research, The Australian National University, Acton (Canberra) Australian Capital Territory 2601, Australia.*
[4]*Victor Chang Cardiac Research Institute, Darlinghurst (Sydney), New South Wales 2010, Australia.*

Correspondence to M.W.H. or T.P.
hentze@embl.de; thomas.preiss@anu.edu.au

doi:10.1038/nrm.2017.130
Published online 17 Jan 2018

O wonder!
How many goodly creatures are there here!
How beauteous mankind is! O brave new world,
That has such people in't.

William Shakespeare, *The Tempest*, Act V, Scene I, 'Miranda's speech'

A 'conventional' RNA-binding protein (RBP) participates in the formation of ribonucleoprotein (RNP) complexes that are principally involved in gene expression[1] (FIG. 1a). The RBP does so by binding to sequence and/or structural motifs in RNA via modular combinations of a limited set of structurally well-defined RNA-binding domains (RBDs)[2] such as the RNA recognition motif (RRM)[3], hnRNP K homology domain (KH)[4] or DEAD box helicase domain[5]. These assertions represent decades of cumulative knowledge, which includes cellular, biochemical and structural data. However, recent advances in determining the structures of large RNP machines such as the ribosome[6–8] and spliceosome[9–11] reveal the existence of complex protein–RNA interactions that do not require canonical RBDs. This finding suggests that such unconventional RNA binding is a broader phenomenon than previously anticipated.

A widely held assumption is that RBPs with high affinity and/or specificity for their targets are more likely to have (ascertainable) biological functions[12]. Also implicit in this conventional view of RBPs is that they should act to alter the fate or function of the RNA[13]. In a recent review, RBPs were described as the 'mRNA's clothes', which ensure that different mRNA regions — 5′ and 3′ untranslated regions (UTRs) and the coding region — are at times covered or exposed, thereby helping the mRNA to progress through the different stages of its life[14]. The processes that drive changes in RNP composition have been likened to those involved in chromatin dynamics[15,16]. Accordingly, post-translational modifications (PTMs) of RBPs, the epitranscriptome and the action of ATP-dependent RNA helicases lead to dynamic RNP remodelling.

The concept that RBPs regulate various aspects of RNA function has broad but not universal applicability, as indicated by emerging evidence from several sources. First, multiple microscopically visible, membrane-less RNP granules have been characterized in different cell types and cellular compartments[17,18]. These include Cajal bodies and paraspeckles in the nucleus, as well as processing (P^-) bodies and stress granules in the cytoplasm. The term 'granule' is somewhat of a misnomer as several of these RNP bodies have now been shown to form by liquid–liquid phase separation thought to be driven by the intrinsically disordered regions (IDRs) of their constituent RBPs[19,20]. Their dynamic composition and amorphous structure still remain puzzling, and well-defined functions remain to be assigned to the formation of these RNP bodies. Second, the discovery of the existence of a myriad of long non-coding RNAs (lncRNAs) triggered intense efforts to uncover their functions[21,22]. Many lncRNAs are currently assumed to participate in the recruitment of transcription factors or chromatin-modifying complexes to chromatin or otherwise organize, scaffold or inhibit protein complexes[23,24]. These functions break with convention by indicating that RNAs may regulate

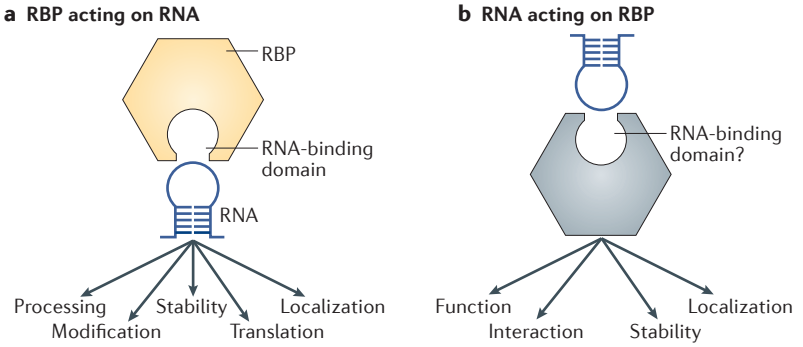


Figure 1 | **Functional crosstalk between proteins and RNA. a** | An RNA-binding protein (RBP) can interact with RNA through defined RNA-binding domains to regulate RNA metabolism and function. **b** | Inversely, the RNA can bind to the RBP to affect its fate and function.

hnRNP K homology domain (KH). An RNA-binding domain of ~70 amino acids that folds into three α-helices packed against a three-stranded β-sheet. RNA binds to a hydrophobic cleft formed between two core α-helices and a GXXG loop that interconnects them.

DEAD box helicase
RNA helicases with two highly similar domains that resemble the bacterial recombinase A and contain the conserved sequence Asp-Glu-Ala-Asp (DEAD). RNA binds across both helicase domains.

Epitranscriptome
The collective, chemically diverse RNA modifications found in a transcriptome. Many of the modifications serve regulatory roles.

Cajal bodies
Subnuclear membrane-less structures involved in multiple aspects of nuclear RNA metabolism.

Paraspeckles
Ribonucleoprotein particles of poorly defined function in the nucleoplasm of mammalian cells.

Processing (P$^-$) bodies
Microscopically visible foci in the cytoplasm of eukaryotic cells that contain mRNAs and mRNA silencing and turnover factors.

Stress granules
Cytoplasmic aggregates of stalled translation initiation complexes in eukaryotic cells that are induced by different forms of cellular stress.

RBP function rather than be regulated by RBPs (FIG. 1b). Finally, several unbiased approaches to identify RBPs throughout the proteome have been developed recently, yielding a growing collection of RNA-binding proteomes from multiple organisms and cellular contexts. These compendia persistently identify large numbers of novel RBPs[25] that defy convention by lacking discernible RBDs, established cellular functions that link them to RNA biology in a straightforward fashion or both.

The features and functions of classical RBPs have been expertly discussed in the reviews cited above and other recent publications[26–29]. In this Review, we focus primarily on the challenges posed by unconventional RBPs, their methods of discovery and the emerging understanding of their modes of RNA binding, RNA targets and their molecular and cellular functions. We will contrast and integrate these with what we know about classical RBPs.

The era of RNA interactomes

Since the discovery in the early 1990s of several metabolic enzymes 'moonlighting' in RNA-binding activity, it has become apparent that the number and diverse nature of RBPs have been underestimated[30–33]. The list of unconventional RBPs has grown incrementally over the decades, urging the development of methods to identify RBPs comprehensively in living cells.

Experimental approaches to cataloguing RNA-binding proteins. *In vitro* methods that used either immobilized RNA probes or arrayed proteins identified multiple novel RBPs[34–37]. More recently, RNA interactome capture (RIC) was developed as an *in vivo* method that focuses on native protein–RNA interactions (BOX 1). It entails ultraviolet crosslinking of RBPs to RNA in live cells, followed by collective capture of RNPs with polyadenylated (poly(A)) RNA on oligo(dT) beads and identification of proteins by quantitative mass spectrometry (Q-MS)[38]. RIC yielded 860 and 791 RBPs from human HeLa and HEK293 cells, respectively[39,40]. Both RNA interactomes overlap considerably, with 543 shared RBPs and enrichment for the gene ontology term 'RNA binding' (REF. 41). The majority of well-established RBPs were detected, and classical RBDs such as the RRM, KH, DEAD box helicase and some zinc-finger domains were strongly enriched, attesting to the technical robustness of the method.

About half of the proteins in each RNA interactome lacked known RBDs, and hundreds had no known relationship to RNA biology. Interestingly, both studies revealed common biological roles and molecular functions among the newly discovered proteins, including intermediary metabolism, cell-cycle progression, antiviral response, spindle organization and protein metabolism (chaperons and prolyl *cis-trans* isomerases), among others[39,40]. The discovery of RNA-binding activity in proteins involved in biological processes with no apparent relation to RNA biology ('enigmRBPs') suggested the existence of unexplored interactions between gene expression and other biological processes. About two dozen of these enigmRBPs were validated by orthogonal approaches[39,40,42]; seven of them were analysed by immunoprecipitation followed by sequencing of bound RNA, demonstrating that these proteins specifically interact with distinct sets of RNAs[39,40] and exert defined biological functions.

RIC was since applied to samples from diverse sources (FIG. 2), typically identifying hundreds of active RBPs. Sources include several additional human[39,40,43–45] and mouse cell lines[46–48], budding yeast[44,49,50], the unicellular parasites *Leishmania donovani*[51], *Plasmodium falciparum*[52] and *Trypanosoma brucei*[53], as well as plants[54–56], flies[57,58], worms[50] and fish[59]. The RBP sets of different origins each featured enrichment of RNA-related annotation (Supplementary information S1 (figure)), and orthogonal methods were typically used to validate some of the unexpectedly discovered RBPs, including immunoprecipitation of GFP–RBP fusion proteins and detection of co-isolated poly(A) RNA with fluorescent oligo(dT) probes[40,42,50], crosslinking and immunoprecipitation (CLIP) followed by 5′ radioactive labelling of RNA by T4 polynucleotide kinase (PNK)[39,44,46,47,57,58], reverse transcription PCR[50] or sequencing[39,40,44]. Using updated annotation, we compiled here all published RNA interactomes into RBP supersets for *Homo sapiens* (1,914 RBPs in total), *Mus musculus* (1,393), *Saccharomyces cerevisiae* (1,273), *Drosophila melanogaster* (777), *Arabidopsis thaliana* (719) and *Caenorhabditis elegans* (593) (FIG. 2a; Supplementary information S2 (table)). These data sets represent a resource to mine shared and selectively identified RBPs (FIG. 2b–f). Analysis of the RBP supersets identified common, eukaryotic 'core' RBPs[44]. To illustrate this further, we performed InParanoid analysis[60,61], which yielded a high number of orthologue groups, especially among mammals, expectedly decreasing in evolutionary more distant organisms (FIG. 2g,h). In general, annotations of evolutionarily conserved RBPs tend to be more RNA-related than those of RBPs with cell type-specific or organism-specific expression or activity.

The analyses of RBP data sets provided novel insights into the functionality of RBPs. For example, the mouse embryonic stem cell (mESC) RNA interactome was enriched in proteins with differential expression during differentiation, suggesting that RBPs are regulated

during the transition from a pluripotent to a differentiated cell[46]. In particular, components of the transcriptional network of the proto-oncogene MYC were enriched in mESC RBPs, suggesting that RBPs contribute to the implementation of MYC-dependent cell fate decisions. Furthermore, mESC RBPs were significantly upregulated during the first 3 days of reprogramming of mouse embryonic fibroblasts (MEFs) into induced pluripotent stem cells[46]. RIC in combination with nuclear fractionation identified the repertoire of nuclear RBPs of K562 myeloid leukaemia cells; of these, the newly discovered RBPs were enriched for components of the p53 interaction network[62].

RNA-modifying enzymes also consistently form part of the RNA interactome compendia. For example, the cardiomyocyte RNA interactome contains 29 RBPs

Liquid–liquid phase separation
A (bio)physical process whereby membrane-less compartments are formed within cells as phase-separated, liquid-like droplets.

Intrinsically disordered regions
(IDRs). Areas within native proteins that lack stable secondary or tertiary structure and thus appear unfolded.

Long non-coding RNAs
(lncRNAs). RNAs longer than 200 nucleotides without annotated protein-coding potential.

Ultraviolet crosslinking
A method that uses ultraviolet light irradiation *in vitro* or in living cells to covalently connect proteins and RNA that are positioned in very close proximity to each other.

Zinc-finger domains
Zinc-binding protein domains that can mediate interactions with DNA, RNA or proteins, depending on their subclass.

Intermediary metabolism
A collective term for metabolic processes that convert nutrients into cellular components.

InParanoid analysis
A method for detecting orthologues and in-paralogue gene clusters across different, often distant species.

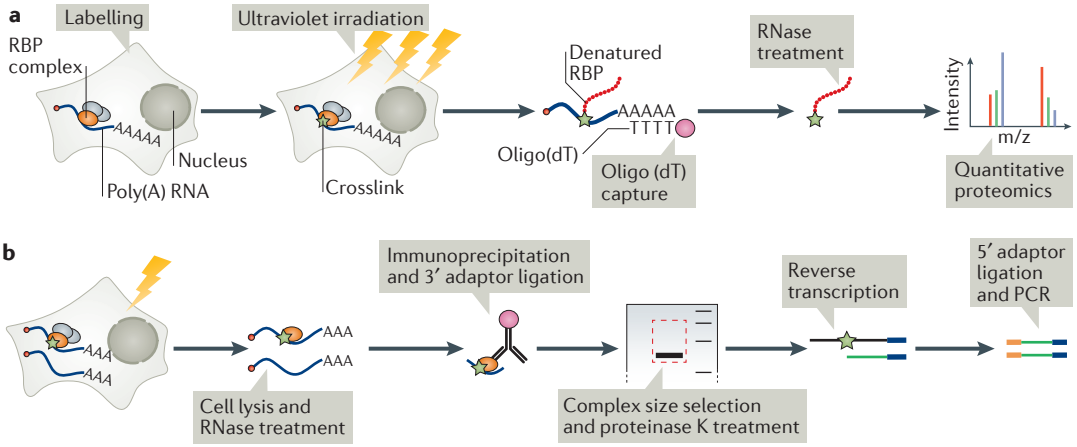
Box 1 | Technical approaches to studying protein–RNA interactions

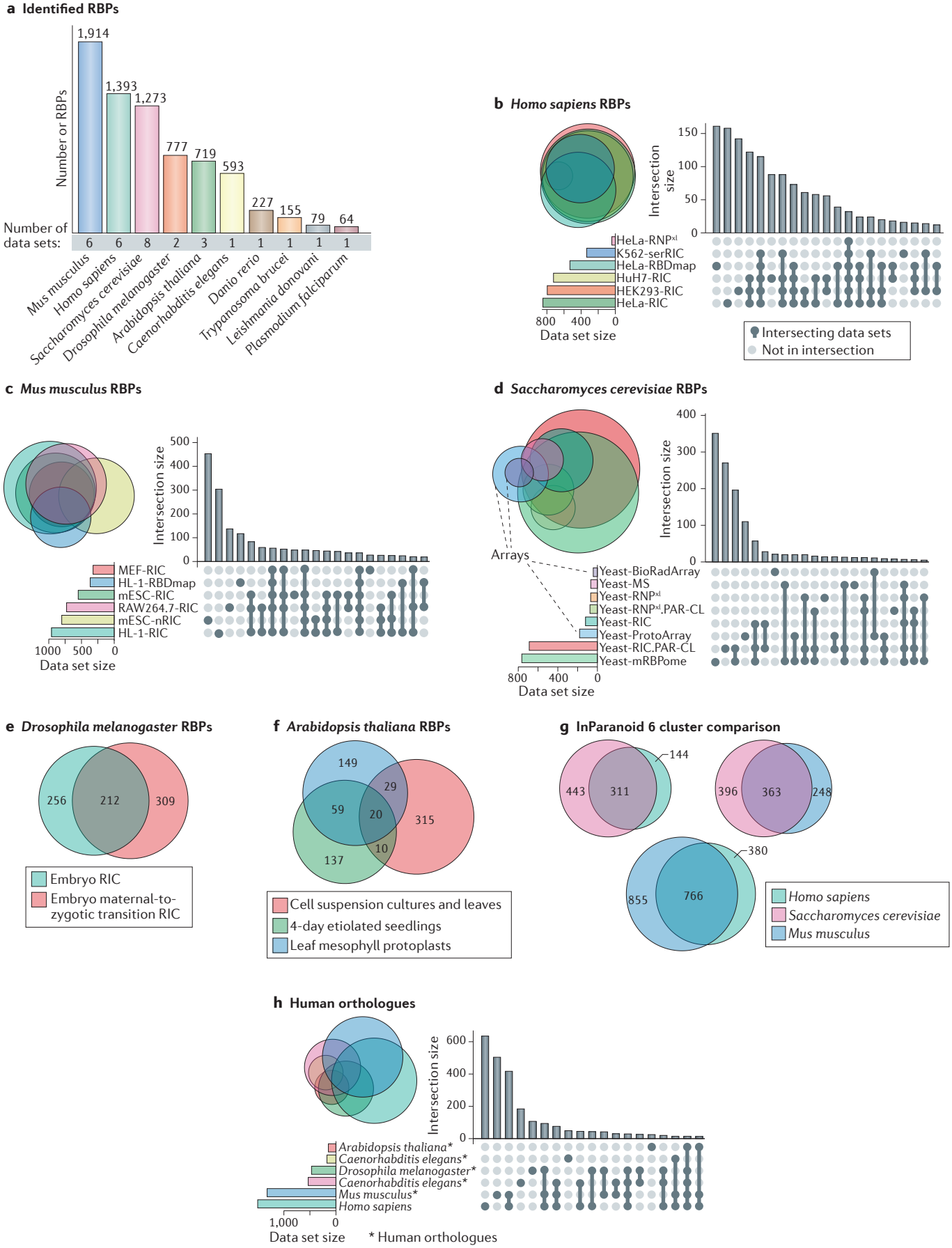
Here, we discuss several *in vitro* and *in vivo* approaches for system-wide analysis of protein–RNA interactions.

System-wide identification of RNA-binding proteins in vitro. One approach for identifying RNA-binding proteins (RBPs) *in vitro* uses immobilized RNA probes as bait that are incubated with cellular extracts and subjected to quantitative mass spectrometry (Q-MS) to identify the RBPs[34]. The use of mRNA untranslated regions as bait identified a dozen proteins in HeLa extract that differentially bound to the bait, several of which were novel RBPs[34]. Similarly, incubation of a set of precursor microRNA bait with multiple different cell lysates yielded ~180 RBPs with distinct specificities[37]. In a second approach, arrayed proteins are used as bait and incubated with fluorescently labelled cellular RNA. RNA binding is determined by measuring the fluorescence intensity at each individual protein spot, analogous to DNA microarrays. Two such proteome-wide screens identified 180 (REF. 34) and 68 (REF. 35) yeast RBPs, respectively. In a third approach, purified polyadenylated (poly(A)) cellular RNA is immobilized on oligo(dT) magnetic beads and after incubation with cell extract, bound proteins were analysed by Q-MS[36]. This approach identified 88 mostly highly abundant proteins, of which 22 were known RBPs.

Identification of RNA-binding protein repertoires in vivo by RNA interactome capture. In this approach, ultraviolet light irradiation of cultured cells or organisms covalently links proteins to RNA positioned in direct proximity. This irradiation is followed by denaturing cell lysis, collective capture of all ribonucleoproteins formed with poly(A) RNA on oligo(dT) beads and identification of proteins by Q-MS[38] (see the figure, part **a**). Both conventional crosslinking and photoactivatable ribonucleoside-enhanced crosslinking (PAR-CL) can be used[97]. The first relies on the natural excitability of nucleoside bases by 254 nm ultraviolet light, which generates short-lived free radicals that attack amino acids in close proximity, thereby forming covalent bonds[146]. By contrast, PAR-CL utilizes the nucleoside analogue 4-thiouridine (4SU), which is taken up by cultured cells and incorporated into nascent RNAs. Crosslinking is then achieved by irradiation with ultraviolet light at 365 nm (REF. 97). Another study used PAR-CL with 4SU in combination with 6-thioguanosine labelling[39], exploiting the U-to-C transitions occurring as a consequence of crosslinking between 4SU and the RBP to globally analyse the RNA interactome. The protocol has been adapted to different model systems[44,49,50,54–58] and can be used to monitor differential association of RBPs with RNA under different physiological conditions or in response to biological cues[48,57], as well as to identify RBPs in different subcellular compartments[62].

Enhanced crosslinking immunoprecipitation (eCLIP). This method is used to determine the binding footprints of a given RBP on its target RNAs with single-nucleotide resolution[99,127]. Ultraviolet light irradiation of live cells is followed by cell lysis and limited digestion to fragment RNA. Protein–RNA complexes are immunoprecipitated with an antibody against the RBP under study. The immunoprecipitated material is resolved by denaturing gel electrophoresis, the RNA is recovered and subjected to reverse transcription into cDNA and the regions bound by the RBP are identified by high-throughput sequencing (see the figure, part **b**). Because reverse transcription often stalls at the site of the protein–RNA crosslink, eCLIP offers single-nucleotide resolution. TAG–eCLIP includes a CRISPR–Cas9-mediated insertion of a carboxy-terminal affinity tag into the endogenous RBP gene, thereby bypassing the need for RBP-specific antibodies[147]. Combined with the use of several hundred immunoprecipitation-tested antibodies against known RBPs[75], the eCLIP method has produced RNA target sets for a growing number of RBPs, which are accessible at https://www.encodeproject.org. Among the 122 proteins with available eCLIP data, 34 lack classical RNA-binding domains.





UpSet plot
A plot used to visualize the total size and overlaps of various data sets.

BioPlex PPI data set
A comprehensive collection of protein–protein interaction networks generated by experimental approaches.

annotated for 5-methylcytosine, N^6-methyladenosine and pseudouridine modifications, as well as adenosine-to-inosine editing[47]. Such RNA modifications can affect RNA–protein interactions or modulate RNA function[63,64] (BOX 2).

Metabolic enzymes are recurrently identified as RBPs. Our meta-analysis revealed 71 such metabolic enzymes moonlighting as RBPs in humans, 104 in mouse and 132 in yeast RNA interactomes (Supplementary information S2 (table)), thereby substantially expanding a list of about 20 RNA-binding metabolic enzymes discovered previously by classic, low-throughput techniques[65–67]. Collectively, these dual RBP–metabolic enzymes represent a breadth of metabolic pathways, with interesting differences in the predominating pathways depending on the source material. For example, the RNA interactome of the human hepatocyte HuH7 cells includes numerous enzymes that function in glycolysis and other pathways of intermediary metabolism[44,50]; this likely relates to the important metabolic role of hepatocytes[44]. The HL-1 cardiomyocyte RNA interactome exhibits a high proportion of mitochondrial metabolic enzymes[47], reflecting their high mitochondrial content and energy needs. Many of these RNA-binding metabolic enzymes were validated by orthogonal approaches[40,44,45,47], including mitochondrial fractionation followed by PNK assay[47]. Interestingly, some of these enzymes interact with their own mRNA[50], perhaps hinting at the existence of negative feedback loops of cognate enzyme–mRNA interactions under conditions of substrate or cofactor deprivation, as previously proposed[65–69]. Thus, RNA binding by metabolic enzymes appears to be common and widely conserved.

◀ Figure 2 | **Comparison of published RNA interactomes.** We stringently curated and updated the annotations of RNA-binding proteins (RBPs) identified from various sources, listed in Supplementary information S2 (table). **a** | Supersets of RBPs identified by the combination of different RBP detection studies in different cell lines and organisms. (**b-h**) Venn diagrams and UpSet plots[156] showing the intersections between different RBP sets. **b** | Human RBPs. RNA interactome capture (RIC; BOX 1) with either conventional ultraviolet light crosslinking (cCL) or photoactivatable ribonucleoside-enhanced crosslinking (PAR-CL) was applied to cervical cancer (HeLa)[40], embryonic kidney (HEK293) [39,40], hepatocyte (HuH7)[44] and myeloid leukaemia (K562) [62] cells lines. HeLa cells were separately also subjected to RBDmap[45] and RNP^{xl} (REF. 43). The human data sets highly overlap, likely because of the prevalence of long-established cell lines as source material. **c** | Murine RBPs. RIC was applied to primary mouse embryonic fibroblasts (MEFs)[85], mouse embryonic stem cells (mESCs) and macrophages (RAW264.7)[48]. HL-1 cardiomyocytes were subjected to both RIC and RBDmap[45]. **d** | Budding yeast RBPs. Two studies used either *in vitro* protein arrays or oligo(dT) capture screens to identify RBPs[35,36], and three *in vivo* RNA interactomes were generated by using cCL[49,50] or PAR-CL[44]. RNP^{xl} was used with two crosslinking approaches[43]. The diversity of the applied technical approaches likely explains the disparity in coverage and overlap. **e** | Fruitfly RBPs. RIC was applied through the use of cCL[57], or both cCL and PAR-CL[58] on embryos undergoing the maternal-to-zygotic transition. Together with differences in mass spectrometry (MS) approaches, this likely underlies the moderate overlap. **f** | Plant RBPs. RIC was performed on different plant sources, including cell-suspension cultures and leaves[54], etiolated seedlings[55] and leaf mesophyll protoplasts[56]. Given the heterogeneous sources, the three data sets agree reasonably well with each other. The lower RBP identification rates suggest ultraviolet crosslinking limitations, likely because of the presence of a cell wall and/or ultraviolet-absorbing pigments. **g** | Pairwise comparisons of InParanoid analysis clusters between humans, mice and yeast. **h** | UpSet plot showing the overlap between the human superset of RBPs and human orthologues. nRIC, nuclear RIC.

Computational approaches to cataloguing RNA-binding proteins. Recently, a set of 1,542 human RBPs (7.5% of the proteome) was defined through the use of computational analyses that required a protein to harbour known RBDs or other domain features characteristic of proteins with RNA-related functions[70]. This approach was complemented by manual curation to add missing but well-documented RBPs or protein components of known RNPs and to remove proteins with established RNA-unrelated functions. Members of this set of RBPs tend to be ubiquitously expressed across tissues, suggesting that they have housekeeping roles. This set overlaps well with the experimentally determined human RNA interactomes (Supplementary information S1 (figure)). Nevertheless, this computational approach might generate false-positive results for two types of proteins: those with classified RBDs that perform non-RNA-binding functions[71] or those that interact with RNA indirectly through interactions with direct RBPs, such as Y14 (also known as RBM8A) of the exon junction complex[72]. A similar domain-search or function-search algorithm was recently applied to *P. falciparum* and yielded 924 RBPs[52], which is a surprisingly high fraction (18.1%) of the relatively small number of protein-coding genes of this malaria-causing parasite.

The propensity of RBPs to interact with other RBPs, either directly or through bridging RNAs, was exploited to identify novel RBPs[40,73,74]. The classification algorithm, termed 'support vector machine obtained from neighbourhood associated RBPs' (SONAR), evaluates each protein against protein–protein interaction (PPI) data and calculates its RBP classification score. The algorithm was trained on sets of human RBPs from available sources[40,70,75], where the interaction partners of each protein were determined through the use of the BioPlex PPI data set[76], which includes many thousands of experimentally determined PPIs. Sets of SONAR-predicted RBPs were established for human (1,784 proteins), *D. melanogaster* (489) and *S. cerevisiae* (745); these agree well with the experimentally determined RBP sets (Supplementary information S1 (figure)). SONAR can be readily applied to any organism as long as substantial PPI data are available; thus, SONAR has the same limitations as any proteomic-based approach, including the depth of the PPI data that are available. In addition, it may produce false positives because proteins that interact with RBPs are not always RBPs themselves[72].

The plasticity of RNA-binding proteomes

Biology is dynamic: the binding of RBPs to RNA constantly changes, and the composition of RNA interactomes is context-dependent and responds to stimuli. Whereas a subset of 'housekeeping' RBPs might be constitutively and ubiquitously active[70], many RBPs have more restricted expression patterns and/or their RNA-binding activity may be regulated, for example, by PTMs[77–79], cofactor binding[80] or PPIs[81]. Moreover, some RBPs can 'sit idle' for lack of their RNA targets[82–84]. For example, cellular sensors against viral infection such as interferon-induced, double-stranded RNA-activated

Box 2 | **Global control of RNA binding by RNA modifications**

N^6-Methyladenosine (m^6A) is the most abundant internal modification in eukaryotic mRNAs, with a diverse and growing number of assigned cellular functions. YT521-B homology (YTH) domain proteins are m^6A 'readers'; that is, they specifically recognize and bind m^6A-containing mRNA regions to affect mRNA splicing, nuclear export, translation or turnover[63,64]. The precise effects of m^6A binding are determined by the YTH protein recruited and by the mRNA context of the modification. The YTH domain-containing proteins YTHDC1, YTHDC2, YTHDF1, YTHDF2 and YTHDF3 are consistently featured in both human and mouse RNA interactome data sets. The RNA-binding proteome of *Arabidopsis thaliana* contains most of the plant's YTH domain proteins, including 30-kDa cleavage and polyadenylation specificity factor 30 (CPSF30), possibly explaining the existence of a plant-specific link between m^6A and mRNA cleavage[55]. Recruitment of YTHDF1 to m^6A sites in the 3′ untranslated region (UTR) of human mRNAs enhances their translation, likely through interactions with subunits of eukaryotic translation initiation factor 3 (eIF3)[148]. Conversely, in murine embryonic fibroblasts, YTHDF2 binding to the 5′ UTRs of heat-shock response mRNAs blocks their demethylation during heat shock, thus facilitating selective translation in the cytoplasm through direct binding of eIF3 to the m^6A sites[149]. More generally, cytoplasmic YTHDF2 binds m^6A-containing mRNAs in human cells and promotes their relocation to processing bodies and their degradation[150]; this is crucial for murine embryonic stem cell differentiation[151–153] and facilitates maternal mRNA clearance and the maternal-to-zygotic transition in zebrafish embryos[154]. RNA modifications can also modulate binding of indirect readers by affecting RNA structure. For example, heterogeneous nuclear ribonucleoproteins C1/C2 (HNRNPC) responds to m^6A-induced 'structural switches' to gain access to thousands of its target sites in human nuclear RNAs[155]. Specifically, HNRNPC preferentially binds to single-strand U-tracts; m^6A can destabilize the local RNA structure and expose the U-tracts to binding by HNRNPC. Thus, when considering the determinants of ribonucleoprotein formation, epitranscriptomic changes need to be considered in addition to post-translational modifications of RNA-binding proteins.

protein kinase (PKR), retinoic acid-inducible gene I protein (RIGI; also known as DDX58) or Toll-like receptors may be activated only by the presence of unusual RNA products derived from viral replication, such as long double-stranded RNA[82] or triphosphate 5′ ends[83,84].

RIC has been adapted to investigate the changes in RNA-binding proteomes in response to physiological and environmental cues. It was applied to murine macrophages responding to lipopolysaccharide stimulation[48], to primary MEFs treated with the DNA-damage inducing agent etoposide[85] and to fruitfly and zebrafish embryos at different stages of development[57,59]. During the maternal-to-zygotic transition (MZT), the origin of gene expression switches from maternal mRNAs, the timed expression of which is determined by RBPs, to embryonic transcripts emerging after zygotic genome activation (ZGA)[86]. The dynamic changes in RBP activity expected during MZT were investigated in *D. melanogaster* embryos by analysing pre-ZGA (0–1 hour post-fertilization (hpf)) and post-ZGA embryos (4.5–5.5 hpf)[57]. First, a comparison between ultraviolet light-irradiated versus non-irradiated pre-ZGA embryos pooled with post-ZGA embryos was used to determine the developmentally 'constitutive' RNA interactome. The repertoire of developmentally 'dynamic' RBPs was determined by comparing ultraviolet light-irradiated samples from pre-ZGA embryos with samples from ultraviolet light-irradiated post-ZGA embryos. The former analysis yielded 523 high-confidence constitutive RBPs, whereas the latter identified 1,131 RBPs, 116 of which were high-confidence dynamic RBPs[57].

To determine whether differential RNA binding was due to alterations in protein levels, the total proteomes were also determined in parallel. Comparison of the RNA interactome and whole proteome data revealed the existence of three classes of RBPs: class 1 (1,015 proteins) showed no significant change during MZT in either RNA binding or total abundance; class 2 (78 proteins) showed commensurate changes in both parameters, suggesting that differential RNA binding was due to altered protein levels; and class 3 (38 proteins) showed a clear change in RNA binding without a corresponding change in RBP abundance, implying a modulation in the ability of these proteins to interact with RNA. These RBPs were thus dubbed 'dynamic binders'; they include eight known splicing factors, of which seven bind RNA more avidly in pre-MZT embryos[57]. These findings broadly concur with an analysis of expression and localization of RBP-encoding mRNAs as proxies to RBP levels, which revealed that RBP expression peaks during the prezygotic and MZT phases whereas transcription factors are highly expressed during ZGA and mid-embryogenesis[58]. A similarly configured analysis of zebrafish embryos before (1.75 hpf) and during ZGA (3 hpf) uncovered 227 RBPs active during vertebrate MZT[59]. As expected, the set included many regulators of mRNA polyadenylation, translation and stability; proteins involved in RNA modification and pre-mRNA processing were also notably represented. Of the 227 RBPs, 24 and 53 proteins were differentially active before ZGA and during ZGA, respectively, and appeared to be mostly dynamic binders.

Thus, comparative RIC can be used to investigate dynamic changes in RNA-binding proteomes and enable the study of a wide range of biological processes, from development and differentiation to signalling, metabolism, infection and other disease-related processes and the effects of drugs.

Systematic identification of RNA-binding domains

Many of the newly discovered RBPs lack known RNA-binding domains, which raises the question of how they interact with RNA. Tried-and-tested low-throughput methods exist to map RBDs in proteins, for example, by mutagenesis combined with RNA-binding assays, such as electrophoretic mobility shift assay (EMSA)[87] or the CLIP-coupled PNK assay[46], but high(er)-throughput methods were required to efficiently identify the RBDs of hundreds of novel RBPs. Several such approaches were developed, each employing mass spectrometry in a different configuration to identify protein regions that become crosslinked to RNA following the exposure of live cells to ultraviolet light irradiation.

One approach focuses on the purification and direct detection of the RNA-crosslinked peptides, the mass of which is altered by the nucleic acid remnant[43,88] (FIG. 3a), followed by data analysis with RNP^{xl} (REF. 43). Applied to yeast RBPs, RNP^{xl} identified crosslinked peptides corresponding to 57 different proteins, mostly canonical RBPs such as ribosomal proteins and proteins with RRM or KH domains[43]. A number of RNA-binding sites were identified in RBPs that lacked known

Maternal-to-zygotic transition
(MZT). The phase in embryonic development during which control by maternally derived factors ceases and the zygotic genome is activated.

Electrophoretic mobility shift assay
(EMSA). A method to study protein–nucleic acid interactions *in vitro* by resolving a labelled nucleic acid probe and its binding proteins on the basis of the reduced mobility of the probe–protein complexes through a nondenaturing gel.

RNP^{xl}
Custom-designed software to facilitate the identification of mass spectra derived from a peptide crosslinked to a nucleotide.

RBDs (unconventional RBPs), including the enzymes peptidyl-prolyl *cis-trans* isomerase, enolase 1 (also known as alpha-enolase) and phosphoglycerate kinase.

A second method, RBDmap, adds to the RIC workflow another digestion step with a protease that cleaves every 17 amino acids on average and a second round of oligo(dT) to capture unconventional RBPs[45]. After the second oligo(dT) capture, the covalently linked polypeptides are cleaved by trypsin to generate an RNA-crosslinked peptide and a neighbouring peptide with native mass (FIG. 3b). RBDmap detects these neighbouring peptides, which have native mass, and computationally extrapolates the RNA-binding sites[45]. Applied to HeLa cells, RBDmap discovered 1,174 RNA-binding sites in 529 proteins[45]; a more limited analysis of HL-1 cardiomyocytes revealed 568 RNA-binding sites in 368 proteins[47]. RBDmap data are in strong concordance with those produced by regular RNA interactome methods, thereby confirming the RNA-binding activity of hundreds of unconventional RBPs[45,47]. As expected, RBDmap identified conventional RBDs such as the RRM, KH and cold-shock domains[2].

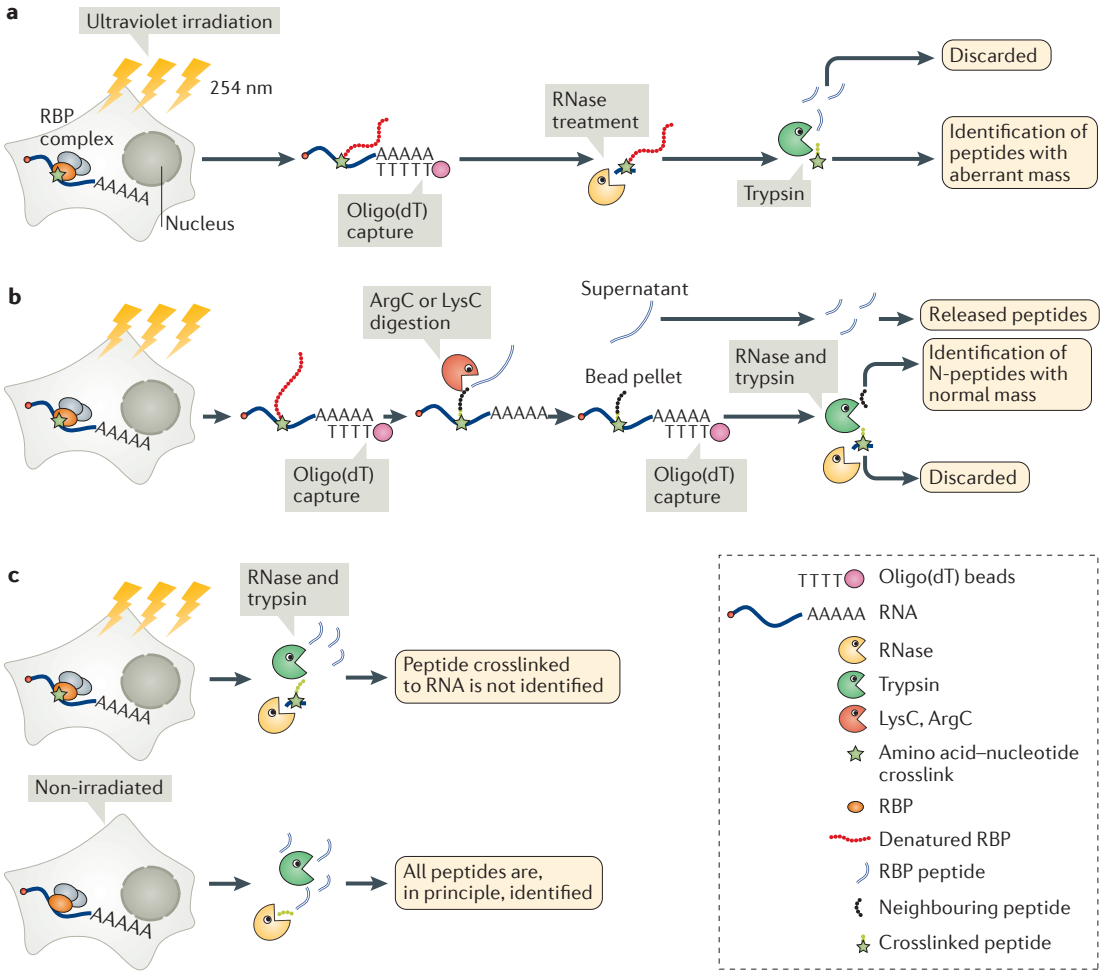


Figure 3 | **Global high-resolution identification of RNA-binding domains.** Three methods are illustrated that use ultraviolet irradiation of live cells to establish covalent bonds at direct contact sites between RNA and proteins. Following cell lysis (not shown), the methods diverge with regard to proteolysis, purification and detection by quantitative mass spectrometry (Q-MS). **a** | Purification and direct detection of RNA-crosslinked tryptic peptides[43,88] is challenging owing to the inefficiency of ultraviolet crosslinking and the heterogeneous mass contribution by the nucleic acid remnants. To overcome this, covalently linked protein–RNA complexes are purified on oligo(dT) beads. After digestion with trypsin and RNases, peptides crosslinked to RNA remnants are further enriched before their analysis by Q-MS and a search by the custom-designed software RNP^{xl}. **b** | In RBDmap[45], RNA-binding proteins (RBPs) are purified with poly(A) RNA and digested with a protease (LysC or ArgC) that cleaves every 17 amino acids on average, typically leaving peptides that still contain a trypsin cleavage site. The RNA-linked peptides, termed RBDpeps, are then recaptured on oligo(dT) beads, whereas those distant to the crosslink site are released into the supernatant. Both fractions are digested with trypsin and analysed by Q-MS. RNA-bound material comprises peptides with remnant RNA that will not be identified and their neighbouring fragment(s) with native mass (N-peptide(s)). N-Peptides that are enriched in the RNA-bound fraction are extended *in silico* to the next LysC or ArgC cleavage site to reconstitute the original RBDpep. **c** | Proteomic identification of RNA-binding regions (RBR-ID)[92] uses mass spectrometry and exploits the mass shift of RNA-crosslinked peptides by assigning RNA-binding activity to tryptic peptides (about nine amino acids on average) with reproducible under-representation in ultraviolet-irradiated samples compared with non-irradiated controls.

Unexpectedly, many of the identified RNA-binding regions mapped to IDRs, implicating them as prominent sites of protein–RNA interactions *in vivo* (see below). Other RNA-binding regions mapped to globular domains that lacked previous association with RNA binding. In HeLa cells, these include the thioredoxin fold, the heat shock 70 kDa protein 1 (HSP70-1; also known as HSPA1A) and HSP90 domains, the 14-3-3 domain, the domain associated with zinc-fingers (DZF), the PSD95–Dlg1–zo-1 (PDZ) domain and the nuclear dbf2-related (NDR) domain (complete list in REF. 45), many of which were validated by orthogonal approaches. Interestingly, the mapped RNA-binding sites showed enrichment for homologous regions from different proteins of the same family, and many mapped to enzymatic cores or PPI surfaces, suggesting an interplay between these activities and RNA binding[45]. In HL-1 cells, RNA-binding regions were identified in 24 metabolic enzymes, 12 of which mapped to dinucleotide-binding domains[47]. These data corroborate and expand the previously suggested RNA-binding activity of dinucleotide-binding domains[33,66,67,89,90].

Another observation was that mutations that cause monogenic diseases were enriched in RNA-binding regions, whereas natural sequence variants were distributed equally across binding and nonbinding regions[45]. This observation suggests that numerous monogenic diseases arise from altered RNA binding. Finally, RNA-binding regions strongly overlap with known sites of PTM, including phosphorylation, acetylation and methylation[45]. This enrichment was not observed for protein regions lacking RNA-binding activity. Thus, PTMs may regulate RNA binding and RNP dynamics, akin to chromatin remodelling.

Peptide crosslinking and affinity purification (pCLAP) is a recently described cousin of RBDmap that implements the first protease treatment directly after lysis, requiring only one oligo(dT) capture round[91]. The trade-off is that pCLAP does not quantify the peptides in the released fraction. This reference is used in RBDmap to determine with high confidence the protein regions engaged in RNA binding[45] (FIG. 3b).

Finally, proteomic identification of RNA-binding regions (RBR-ID) identifies under-represented peptides in ultraviolet light-irradiated samples compared with non-irradiated controls[92] (FIG. 3c). RBR-ID of nuclear proteins from mESCs detected 1,475 RNA-binding sites with a 5% false discovery rate, and many mapped to RRM, KH, DEAD box and other conventional RBDs, thereby validating the method[92]. Of 803 mESC proteins identified as RBPs by RBR-ID, 376 (47%) were previously found in the RNA interactome studies[39,40,44,46,62]. Remarkably, the 427 RBPs that were newly identified by RBR-ID showed enrichment for gene-regulatory and chromatin-associated functions, and their RNA-binding sites frequently mapped to chromatin-related domains such as the chromodomain and bromodomain. These discoveries indicated a potential crosstalk between chromatin and RNA, expanding on previous candidate-based approaches, which characterized the activity of chromatin remodelling proteins such as histone-lysine *N*-methyltransferase EZH2. EZH2 is the catalytic subunit of polycomb repressor complex 2 and has been shown to interact with lncRNA and nascent transcripts[93–95]. The RBR-ID data suggest that such interactions are more common than previously anticipated.

These approaches have some common limitations. Ultraviolet crosslinking relies on a favourable geometry and nucleotide and amino acid composition at protein–RNA interfaces. For example, protein–RNA interactions with the phosphate backbone are nonoptimal in this regard and will likely be missed. Mass spectrometry analyses are also influenced by the abundance of the peptides and their amino acid sequence (affecting the size of the tryptic peptides and their mass-to-charge ratio). The RNP^{xl} workflow offers the unique advantage of single amino acid resolution but at the expense of a limited sensitivity and the need for specialized proteomic analyses. RBDmap and pCLAP are less complex to implement and more sensitive but at the price of a lower resolution (~17 amino acids). The conceptually very straightforward RBR-ID gives an intermediate resolution (~9 amino acids); however, ultraviolet-dependent loss of peptides may be due, in some instances, to peptide interactions with mononucleotides, dinucleotides or other molecules absorbing at 312 nm, as well as to high intraexperimental and interexperimental variability. Nevertheless, RNP^{x1}, RBDmap, pCLAP and RBS-ID have greatly expanded the known repertoire of RBPs and their nontypical RBDs. However, we still know little about the RNA targets of the novel RBPs or the function of these interactions. To address this, functional studies that include determination of the specificity and affinity of these novel RBDs for their target sequences[96–100] will be required.

Novel types of RNA binding

The many novel RBPs and their nontypical RBDs raise important questions about their biological functions[101] (FIG. 4). Not every bimolecular collision in a cell should be assumed to be physiologically relevant. Which ranges of affinities and RNA-binding specificities are to be expected of these RBPs? Concepts regarding specificity and nonspecificity in RNA–protein interactions were recently reviewed[12]. The first aspect of RNA–protein interactions to consider is that indiscriminate RNA binding by RBPs is common and can be important to their function. For example, numerous proteins involved in mRNA translation and degradation need to be non-selective to fulfil their functions (FIG. 4b). Similarly, the exon junction complex is deposited on nascent transcripts through PPIs[102] at a fixed position upstream of splice junctions and interacts with a plurality of RNAs[103] (FIG. 4c). Second, one must distinguish between 'biological specificity' — the binding characteristics of RBPs *in vivo* — and 'intrinsic specificity', which can be determined *in vitro* by selection from a random pool of RNA sequences. An interesting upshot of this distinction is that intrinsically specific RBPs may function *in vivo* as less-specific RBPs when binding to their physiological RNA targets because these RNAs do not fall into the high-affinity and/or high-specificity range of their RNA-binding potential[12].

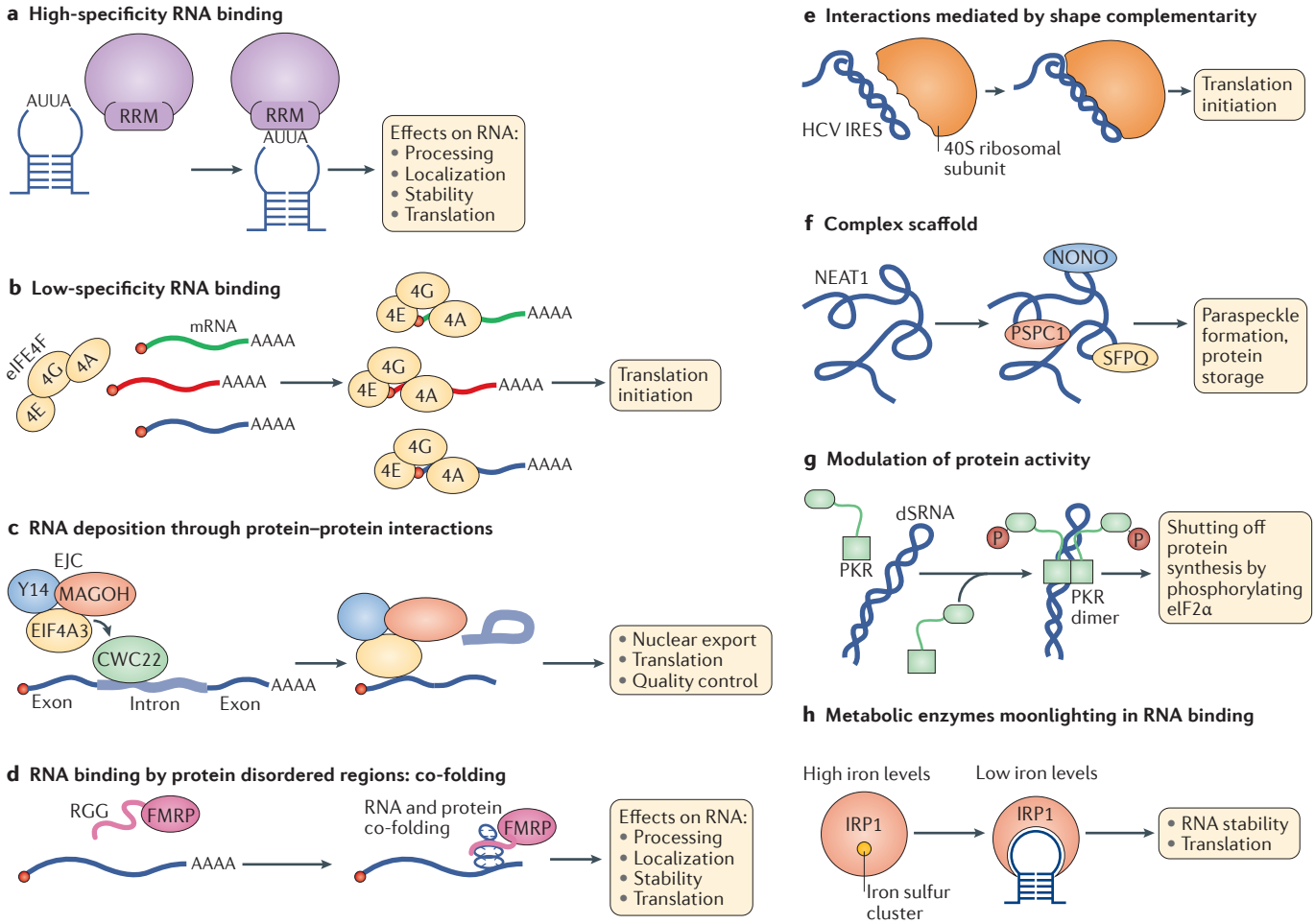


Figure 4 | **Modes of RNA binding. a** | An RNA-binding protein (RBP) harbouring a classic RNA-binding domain such as the RNA recognition motif (RRM) can interact with high specificity with an RNA sequence in the context of a stem–loop[2]. **b** | The eukaryotic translation initiation factor 4F (eIF4F) complex is composed of the cap-binding proteins eIF4E (4E) and eIF4G (4G) and the helicase eIF4A (4A). This complex associates with capped RNA in a sequence-independent manner to enable translation initiation[157]. **c** | The exon junction complex (EJC) is deposited nonselectively on nascent transcripts by its interaction with the splicing factor CWC22 (complexed with CEF1 22) about 20 nucleotides upstream of the exon–exon junction, immediately following intron removal[102]. **d** | The intrinsically disordered Arg–Gly–Gly (RGG) repeat motif of fragile X mental retardation protein (FMRP) co-folds with its target RNA, forming a tight electrostatic and shape-complementation-driven interaction[107]. **e** | The internal ribosome entry site (IRES) of hepatitis C virus (HCV) interacts directly with the ribosome through a complex interaction mode that involves shape complementarity between the IRES and the 40S ribosome subunit[115]. **f** | The long non-coding RNA nuclear enriched abundant transcript 1 (NEAT1) sequesters the RBPs non-POU domain-containing octamer-binding protein (NONO), paraspeckle component 1 (PSPC1) and splicing factor, proline- and glutamine-rich (SFPQ) to form paraspeckles[117]. **g** | Interferon-induced, double-stranded RNA-activated protein kinase (PKR) binds to double-stranded RNA (dsRNA) derived from viral replication. Binding RNA promotes PKR dimerization, autophosphorylation and activation. Active PKR phosphorylates eIF2α to block protein synthesis in infected cells[158]. **h** | Iron-regulatory protein 1 (IRP1) associates with an iron–sulfur cluster to catalyse the interconversion between citrate and isocitrate. In conditions of low iron levels, the iron–sulfur cluster is no longer synthesized and IRP1 binds mRNAs that encode cellular factors involved in iron homeostasis, thereby regulating their fate[119]. eIF4A3, eukaryotic initiation factor 4A-III; MAGOH, protein mago nashi homolog; Y14, RBP Y14.

RNA binding by intrinsically disordered protein regions. IDRs are not only involved in aggregation of RNPs into granules through PPIs[19,20] but they also directly engage in RNA binding[45,47,104,105]. RBPs are enriched in IDRs that are characterized by a low content of bulky hydrophobic amino acids, with the exception of Tyr, and a high proportion of small, polar and/or charged amino acids, particularly Gly, Ser, Arg, Lys, Gln, Glu and Asp[45]. Interestingly, mutations in RBPs that cause human monogenic diseases occur with higher frequency in RNA-binding IDRs than in globular domains, suggesting that IDRs are subjected to strong sequence constraints[41]. The occurrence of IDRs within RBPs appears to be conserved from yeast to humans[44], often in the form of repeats such as RGG, YGG, SR, DE or KK[40,104]. A recent report proposed that, as the number of repeats in IDRs of RBPs has expanded from yeast to humans whereas the number and identity of globular domains in these proteins have remained the same, IDRs may represent a plastic component of RBPs that co-evolved with the increasing complexity of eukaryotic transcriptomes[44].

Around half of the 1,174 RNA-binding sites reported by RBDmap in HeLa cells mapped to IDRs, reflecting

their prevalence as a mode of RNA binding[45]. One-hundred seventy RBPs appeared to interact with RNA exclusively through IDRs, suggesting that these regions suffice for mediating RNA binding. Among the Arg-rich motifs, RGG and SR repeats were previously reported to bind RNA[104]. The discovery of distinct RGG motifs allowed their assignment into subclasses that differ by the lengths of their glycine linkers[45,106]. Nuclear magnetic resonance analyses of human fragile X mental retardation protein 1 (FMRP; also known as FMR1) showed that the positioning of the Arg residues is essential for the selective binding of the RGG motif to the guanine-rich sequence of the sc1 mRNA[107]. Gly repeats surrounding the Arg residues have an important role in orienting these positively charged residues for interaction with the Watson–Crick base pairs, which stack on two adjacent RNA G-quadruplexes that form as a result of the protein–RNA co-folding (FIG. 4d). The high flexibility of the Gly repeats contributes to RNA binding by allowing RGG conformational adaptation to the shape of its target RNA. Hence, the affinity and selectivity of RGG motifs for their target RNAs may be determined by the frequency and order of Arg and Gly residues.

A second type of RNA-binding IDRs comprises aromatic residues, especially Tyr, which combine with Gly and Ser in forming [G/S]Y[G/S] motifs. These motifs have a tendency to aggregate *in vitro*, which induces the formation of hydrogels and amyloid-like fibres, and to engage in dynamic liquid–liquid phase separation *in vivo*[108,109]. Aromatic residues tend to form part of the hydrophobic protein cores but, when present at the protein surface, they can interact with amino acids or nucleotides through stacking or hydrogen bonding[2]. When embedded within Gly-rich sequences, the aromatic residue is particularly exposed, which likely fosters its propensity to aggregate when interacting with similar protein motifs[108,109] or to bind RNA[45].

Finally, a heterogeneous set of linear motifs comprising Lys and, to a lesser extent, Arg was also enriched in RBPs[45]. Interestingly, the stoichiometry and distances between these positively charged residues, as well as their combination with other amino acids, were conserved even across nonhomologous proteins. Notably, such basic IDRs in RBPs are similar to motifs in DNA-binding proteins, where the basic arms can alter the DNA-binding properties of transcription factors by providing them with a large capture radius[110]. In this 'monkey bar' model, transcription factors utilize their basic arms to reach distant DNA sites by 'hopping' and 'sliding' instead of 3D diffusion. It is currently unknown whether basic arms may have similar roles in RBPs.

Thus, IDRs could represent malleable, potentially multifunctional RNA-binding motifs. Their RNA-binding capacity can range from highly specific to nonselective and may promote protein–RNA co-folding upon their interaction with target RNAs[104,105,107,111] (FIG. 4d). Interestingly, the high sequence constraints of IDRs[41] can enable the regulation of RNA-binding by reversible PTMs such as acetylation or phosphorylation[45]. In principle, these properties qualify IDRs as versatile modules for interaction with RNA, either alone or in cooperation with globular RBDs.

Shape complementarity-based interactions and RNAs acting on RNA-binding proteins. Protein–RNA interactions are typically described as mediated by a protein with 'sensors' (RBDs) that recognize and bind particular sequences and/or structures in a target RNA (FIG. 1a). However, synthetic RNA aptamers can bind proteins according to the same principles that enable proteins to bind RNA[112], suggesting that RNA is a driving force in mediating protein–RNA interactions (FIG. 1b). An example of a major cellular machinery that is based on such intricate protein–RNA interactions is the spliceosome. Here, small nuclear RNAs fold into 3D structures, which form complex surfaces that interact with complementary (in both shape and biochemical properties) protein partners to drive the assembly of the functionally active spliceosome[9–11].

Similarly, ribosomal RNAs (rRNAs) have a major role in ribosome assembly by engaging the ribosomal proteins within the complex architecture of the ribosome[6–8]. Most of the 169 annotated ribosomal proteins directly bind to rRNAs, yet the majority lacks conventional RBDs and instead features 119 distinct domain architectures[70]. Ribosomal proteins and rRNAs appear to have co-evolved to interact with each other; therefore, instead of classic RBD-driven contacts, shape complementarity and the right spatial configuration of molecular interactions form the perfectly assembled machinery of the ribosome[8].

The 5′ UTRs of certain viral RNAs have evolved complex shapes to interact with the cellular translation machinery to direct protein synthesis. For example, the internal ribosome entry site (IRES) of poliovirus binds to the carboxy-terminal moiety of eukaryotic translation initiation factor 4G1 (eIF4G1) to recruit the 40S ribosomal subunit and initiate translation[113]. Similarly, the hepatitis C virus (HCV) IRES interacts directly with eIF3 and the 40S ribosomal subunit to enable translation initiation (FIG. 4e). Again, the HCV IRES–ribosome co-structure reveals that the interaction is not mediated by well-defined protein regions endowed with RNA-binding activity; instead, the protein–RNA interface is large and has strong shape complementarity between the IRES and 40S subunit[114–116] (FIG. 4e).

Recently, two studies identified widespread RNA binding by chromatin-associated factors and DNA-binding proteins[62,92], which could interact with lncRNAs and nascent transcripts[93–95]. Plausibly, the lncRNAs themselves drive these interactions, and it might make sense to think of them as RNAs with protein-binding activity rather than the other way around (FIG. 1b). The RNA moieties could have different functions in these interactions: scaffolding protein–RNA complexes, as in the case of the viral IRESs[115] (FIG. 4e); sequestering or coordinating proteins, as in the case of the lncRNA nuclear enriched abundant transcript 1 (NEAT1), which is essential for the formation of paraspeckles[117] (FIG. 4f); or altering the activity of the bound protein, as exemplified in the interactions of RIGI and PKR with intermediates of viral replication[82–84] (FIG. 4g).

G-Quadruplexes
Nucleic acid structures made of two or more stacks of planar arrays of four guanine bases.

RNA aptamers
Relatively short and often highly folded RNA molecules, which are selected for specific, high-affinity interactions with proteins or other molecules.

RNA binding by metabolic enzymes. RNA interactome studies have persistently identified enzymes of intermediary metabolism as RBPs. Some of these enzyme–RNA interactions appear to function as direct gene regulation feedbacks. For example, thymidylate synthase, which is an enzyme that catalyses the formation of dTMP from dUMP binds to its own mRNA and inhibits its translation when dUMP levels are low[118]. A more indirect form of feedback regulation is exerted by cytoplasmic iron-regulatory protein 1 (IRP1; also known as ACO1). To be active as an enzyme, the protein requires an iron–sulfur cluster in its active site, which precludes RNA binding[119]. In conditions of iron deficiency, the cluster is lacking and, in its open conformation, IRP1 regulates the expression of proteins to increase iron uptake and decrease iron storage, utilization and export[119,120] (FIG. 4h).

Other enzymes have more indirect links to metabolism when acting as RBPs. The glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH) oxidizes its substrate to reduce nicotinamide adenine dinucleotide, but it also has a diverse range of other cellular functions[121]. GAPDH has an important role as an RBP in T lymphocytes[90]. In resting T cells, which rely on oxidative phosphorylation for energy generation, GAPDH binds to AU-rich elements in the 3′ UTRs of cytokine mRNAs, including interferon-γ (IFNγ) mRNA, and inhibits their translation. Upon the metabolic switch to aerobic glycolysis following T cell activation, GAPDH disengages from RNA, thus de-repressing cytokine production.

The high number of identified RNA-binding metabolic enzymes suggests that not all of these have moonlighting functions in post-transcriptional gene regulation. Alternatively, their ability to interact with RNA could serve yet to be discovered RNAs in affecting their metabolic function. As discussed in more detail elsewhere[67], RNA binding could modulate the localization or activity of an enzyme, for example, by affecting an enzymatic side reaction, by allosteric control or by providing a scaffold that organizes multienzyme complexes and even pathways.

Interestingly, the globular Rossmann-fold (R-f) domain has emerged as a common unconventional RBD. The cardiomyocyte RNA-binding proteome includes 173 R-f-containing proteins, and 29 of the 73 cardiomyocyte RNA-binding metabolic enzymes harbour at least 1 R-f domain[47]. For example, the dinucleotide-binding R-f domain of oxidoreductase enzymes has long been considered to represent an RNA-binding interface[33]. Similarly, nicotinamide adenine dinucleotides interfere with RNA binding by GAPDH, and cytokine mRNA sequences can inhibit the enzymatic activity of GAPDH *in vitro*[89]. Analysis of RBDmap data of 24 metabolic enzymes (including multiple R-f domain proteins) revealed diverse spatial connections between the identified RNA contacts and previously characterized, catalytically relevant regions. However, in many cases, the mapped RNA contacts and the catalytic regions did not appear to overlap. Although this could partly reflect false-negative assignments by RBDmap, the data suggest that RNAs have roles in allosteric control of enzymes or in enzyme scaffolding[47].

Many enzymes are allosterically regulated by metabolites[67]. Conceivably, this could also affect their RNA-binding activity. Furthermore, metabolism could control enzyme–RNA interactions through metabolite-driven PTMs. For example, *S*-glutathionylation blocks the RNA-binding activity of GAPDH[122]. Many metabolic enzymes are acetylated, which requires sufficient concentrations of acetyl-CoA[123]. RBDmap identified RBDs as hot spots for PTMs, including tyrosine phosphorylation, methylation, acetylation and malonylation[45]. Together, the data point to considerable crosstalk between cellular metabolism and RNA binding[67,124].

The roles of unconventional RNA-binding proteins

A key step in characterizing the molecular and cellular function of novel RBPs is to identify their RNA targets. Several high-throughput sequencing-based methods to achieve this have emerged in recent years, including carefully controlled *in vitro* methods[125] and methods that preserve the context of the living cell[126]. Some of the latter are based on classic approaches, such as RNP immunoprecipitation (RIP), with optional stabilization of RNP complexes through ultraviolet light irradiation and/or chemically induced covalent crosslinks (CLIP). These include photoactivatable-ribonucleoside-enhanced CLIP (PAR-CLIP)[97] and individual-nucleotide-resolution CLIP (iCLIP)[98]. Enhanced CLIP (eCLIP; BOX 1) is a variant of iCLIP with improved sensitivity and specificity[99,127]. CLIP-type studies have already helped to characterize the biological roles of several unconventional RBPs. These include metabolic enzymes such as 3-hydroxyacyl-CoA dehydrogenase type 2 (HSD17B10)[44], regulators of alternative splicing[128,129], the E3 ubiquitin and ISG15 ligase TRIM25 (REFS 130–132), nuclear cap-binding protein subunit 3 (NCBP3; previously known as C17orf85)[133], FAST kinase domain-containing protein 2, mitochondrial (FASTKD2)[134], tropomyosin[135] and others.

HSD17B10 is a mitochondrial enzyme involved in the oxidation of isoleucine, branched-chain fatty acids and xenobiotics and in the metabolism of sex hormones and neuroactive steroids[136]. Mutations in HSD17B10 cause a hereditary mitochondrial cardiomyopathy and neuropathy syndrome (OMIM 300438). Interestingly, the severity of the disorder does not correlate with the loss of enzymatic activity; thus, the disease may be caused by a noncatalytic function of this protein[137]. HSD17B10 is a component of the mitochondrial RNase P complex, together with mitochondrial ribonuclease P protein 1 (TRMT10C) and mitochondrial ribonuclease P protein 3 (REF. 138). HSD17B10 was identified as an RBP in several RNA interactomes, indicating that it directly binds RNA[44]. Furthermore, iCLIP revealed that HSD17B10 preferentially binds the 5′ ends of 15 of the 22 mitochondrial tRNAs, particularly at the D-stem, D-loop, anticodon stem and loop regions of these tRNAs[44]. Thus, as a metabolic enzyme of the dinucleotide-binding family, HSD17B10 has a role in guiding RNase P to the ends of mitochondrial tRNAs. HSD17B10 bearing the mutation R130C, which causes the classical cardiomyopathy and neuropathy phenotype

Rossmann-fold
(R-f). A protein domain with up to seven mostly parallel β-strands combined with connecting α-helices. Typically found in proteins that bind nucleotides.

RNase P complex
An RNase complex that processes precursor tRNA.

Speckles
Nucleoplasmic granules located at interchromatin regions that are enriched in splicing factors

Polyuridylation
The addition of multiple uridines to the 3′ end of RNA molecules by uridylyltransferases as a signal for RNA degradation.

associated with HSD17B10 dysfunction[137], exhibits reduced TRMT10C binding *in vitro*[139], and iCLIP data revealed that the R130C mutant also displays reduced RNA-binding activity[44], suggesting that dysfunctional HSD17B10 association with RNA contributes to the phenotype of the disease.

The HeLa cell RNA interactome catalogued four (out of six) members of the FAST kinase protein family as RBPs[40]. One of these, the mitochondrial FASTKD2, was recurrently identified in most human and mouse RNA interactomes[39,40,45–48,62]. Analysis of its binding partners by iCLIP revealed that FASTKD2 selectively associates with mitochondrial transcripts[134]. Depletion of FASTKD2 caused a strong reduction in the levels of its binding target, 16S mitochondrial rRNA, consistent with the importance of FASTKD2 for mitochondrial ribosome biogenesis[140]. Similarly, lack of FASTKD2 activity leads to a reduction in the levels of other mitochondrial RNAs it binds, including the mRNAs encoding cytochrome *c* oxidase subunit 1 (COX1; also known as MTCO1), COX2, COX3, mitochondrially encoded cytochrome *b* and NADH-ubiquinone oxidoreductase chain 6, as well as the non-coding RNAs 7S RNA and prolyl-tRNA[134]. A nonsense mutation in FASTKD2 causes a hereditary neurological disorder[141]. Functional assays in combination with iCLIP suggested that this disorder is caused by defects in RNA binding, which result in altered mitochondrial protein synthesis and metabolism[134,140].

NCBP3 was catalogued as an RBP in RNA interactome studies[40] and was more recently identified as a cap-binding protein that localizes to nuclear speckles[133]. NCBP1 and NCBP2 form the canonical cap-binding complex that binds to nuclear RNAs and is important for mRNA processing and export. Although NCBP1 depletion resulted in the expected retention of poly(A) RNA in the nucleus, NCBP2 depletion had almost no effect, suggesting that NCBP2 is replaced by another nuclear factor with similar cap-binding activity[133]. NCBP3 has an RRM that is sufficient *in vitro* for binding the 7-methylguanosine cap. The interaction of NCBP3 with 7-methylguanosine is mediated by a tryptophan and two aspartic acids at the RRM loops, and mutating these residues to alanine impairs the cap-binding activity of NCBP3 (REF. 133). Importantly, immunoprecipitation followed by mass spectrometry revealed that, similar to NCBP2, NCBP3 also interacts with NCBP1, thereby forming part of an alternative nuclear cap-binding complex[133].

Several members of the TRIM protein family have recurrently been catalogued as RBPs in RNA interactome studies, including TRIM25 (REFS 40,46), which crosslinks very efficiently with RNA[46]. TRIM25 lacks a canonical RBD and instead binds RNA through its PRY–SPRY domain[45,46,130]. Recently, TRIM25 was shown to interact with and activate protein lin-28 homologue A and terminal uridylyltransferase 4, which are involved in precursor microRNA (pre-miRNA) polyuridylation[130]. Because pre-miRNA polyuridylation triggers miRNA decay[142], TRIM25 emerges as an miRNA regulator. Furthermore, a CLIP study revealed a broad spectrum of TRIM25 cellular RNA targets, prominently mRNAs and lncRNAs; TRIM25 displayed some preference for GC-rich sequences and sites located in the 3′ UTR of mRNAs[132]. A role for TRIM25 in regulating its RNA targets was not apparent in this study. Instead, interaction of TRIM25 with RNA promoted its E3 ubiquitin ligase activity, representing an additional example of a protein activity regulated by RNA[45,132].

TRIM25–RNA interactions are also important in dengue virus-infected cells[131]. TRIM25 triggers the production of IFNβ by ubiquitylation of the antiviral factor RIGI[143]. The genomic RNA of dengue virus 2 is processed

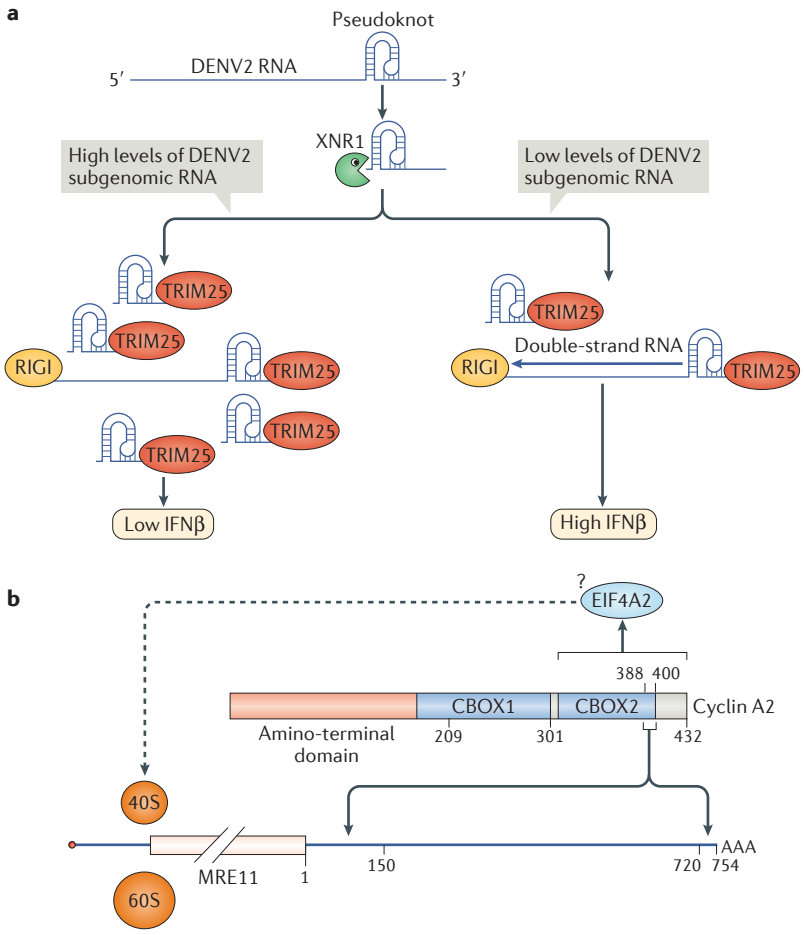


Figure 5 | **Biological roles of unconventional RNA-binding proteins. a** | Tripartite motif-containing protein 25 (TRIM25) is sequestered by the subgenomic RNA of dengue virus 2 (DENV2) to reduce interferon-β (IFNβ) synthesis[131]. DENV2 genomic RNA degradation by 5′–3′ exoribonuclease 1 (XRN1) proceeds until XRN1 stalls at a pseudoknot at the 3′ region, leading to the generation of subgenomic RNA. The subgenomic RNA molecules recruit TRIM25 but not its partner in the IFNβ pathway, retinoic acid-inducible gene I protein (RIGI), which requires the presence of a triphosphate at the 5′ end for binding (not shown); this sequesters TRIM25 in ribonucleoproteins that lack RIGI, leading to the inhibition of the IFNβ response. **b** | Cyclin-dependent, kinase-independent function of cyclin A2 as an RNA-binding protein. Cyclin A2 directly binds to two evolutionarily conserved regions in the 3′ untranslated region of the meiotic recombination 11 homologue 1 (MRE11) mRNA (nucleotides 1–150 and 720–754; marked with arrows) to promote MRE11 mRNA translation. Deletion analysis identified amino acids 302–432 of cyclin A2 as both necessary and sufficient for RNA binding and promotion of the appropriate MRE11 synthesis. The same region also interacts with eukaryotic translation initiation factor 4A-II (eIF4A2), suggesting it promotes translation initiation (dashed line)[145]. Mutational analysis of the cyclin box 2 (CBOX2) domain identified an α-helix (amino acids 388–400) with solvent, positive and polar characteristics as crucial for RNA binding.

by 5′–3′ exoribonuclease 1 (XRN1) until it is stalled by a pseudoknot at the 3′ region, leading to the production of a shorter subgenomic RNA with pathogenic activity[144]. The subgenomic RNA sequesters TRIM25 and prevents the activation of RIGI, thereby reducing IFNβ production[131] (FIG. 5a). Hence, the interaction of dengue virus 2 subgenomic RNAs with TRIM25 supports its capacity to counteract the antiviral response, implicating the unconventional RBP TRIM25 in the innate immunity response against viruses.

Cyclins regulate the cell cycle by activating cyclin-dependent kinases (CDKs). They have also sporadically been identified as unconventional RBPs in RIC studies[39,46,47,57]. Cyclin B and cyclin T are part of the RNA interactome of *D. melanogaster* embryos, and the cyclin B RNA-binding activity was validated by CLIP–PNK assay[57]. Cyclins A2, L1 and T1 were identified in the nuclear RNA-binding proteome of mouse embryonic stem cells[92]. Mice with reduced cyclin A2 expression are prone to tumour formation and chromosomal instability owing to a predisposition to form lagging chromosomes and chromatin bridges during cell division[145]. This defect resulted from insufficient expression of the double-strand break repair protein MRE11 apparently because of impaired translation. Cyclin A2 directly binds to two evolutionarily conserved regions in the 3′ UTR of MRE11 mRNA, which appears to be necessary and sufficient for MRE11 expression[145] (FIG. 5b). A carboxy-terminal fragment of cyclin A2 that lacks CDK binding is both necessary and sufficient for RNA binding, and its expression in MEFs restored appropriate MRE11 synthesis. Interestingly, the RNA-interacting region of cyclin A2 also binds the translation initiation factor eIF4A2. Taken together, the data identify an unexpected, CDK-independent function of cyclin A2 as an RBP that promotes MRE11 mRNA translation, potentially through an interaction with eIF4A2.

Future perspective

What can we expect from the discovery of so many new RBPs? Some might side with Miranda from Shakespeare's *The Tempest* and marvel at these novel and goodly RBPs that populate the RNA interactome. Others might think of Huxley's brave new world and fear dystopia, considering the newly discovered RBPs as nonconformist misfits lacking biological function. Which roles do these new RBPs play? Some may indeed play none, having been discovered on the basis of a biophysical property that confers above-background interaction with RNA without biological relevance. Quite remarkably, however, the list of unconventional RBPs with newly discovered biological roles continues to grow. Although better known for other biological functions and lacking conventional RBDs, they 'moonlight' as RBPs and affect RNA fate, akin to the functions of orthodox RBPs (FIG. 1a). It will be illuminating to study these protein–RNA interactions structurally. Likewise, it will be important to decipher the mechanisms by which these protein–RNA interactions are regulated, through both RNA modifications and protein PTMs. An intriguing question is whether and how protein IDRs that bind RNA contribute to the formation of higher-order cellular assemblies through liquid–liquid phase separation and if so, what the role of RNA is in these transitions.

Finally, could unconventional RBPs be controlled by RNA? We have become accustomed to the view that protein functions are modulated by other proteins, but there is ample room for considering the possibility that the known biological function of a protein can be altered through 'riboregulation' — in other words, that a change in protein function can be elicited by its interaction with RNA (FIG. 1b). The RBPs PKR, RIGI and TRIM25 can serve as examples of this class of proteins. We already know much about RBPs, but future experiments are bound to surpass our expectations.

1. Dreyfuss, G., Kim, V. N. & Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* **3**, 195–205 (2002).
2. Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **8**, 479–490 (2007).
3. Clery, A., Blatter, M. & Allain, F. H. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290–298 (2008).
4. Valverde, R., Edwards, L. & Regan, L. Structure and function of KH domains. *FEBS J.* **275**, 2712–2726 (2008).
5. Linder, P. & Jankowsky, E. From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* **12**, 505–516 (2011).
6. Ramakrishnan, V. The ribosome emerges from a black box. *Cell* **159**, 979–984 (2014).
7. Steitz, T. A. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **9**, 242–253 (2008).
8. Behrmann, E. *et al.* Structural snapshots of actively translating human ribosomes. *Cell* **161**, 845–857 (2015).
9. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121 (2014).
10. Papasaikas, P. & Valcarcel, J. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem. Sci.* **41**, 33–45 (2016).
11. Plaschka, C., Lin, P. C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature* **546**, 617–621 (2017).
12. Jankowsky, E. & Harris, M. E. Specificity and nonspecificity in RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* **16**, 533–544 (2015).
13. Gebauer, F., Preiss, T. & Hentze, M. W. From *cis*-regulatory elements to complex RNPs and back. *Cold Spring Harb. Perspect. Biol.* **4**, a012245 (2012).
14. Singh, G., Pratt, G., Yeo, G. W. & Moore, M. J. The clothes make the mRNA: past and present trends in mRNP fashion. *Annu. Rev. Biochem.* **84**, 325–354 (2015).
15. Lee, S. R. & Lykke-Andersen, J. Emerging roles for ribonucleoprotein modification and remodeling in controlling RNA fate. *Trends Cell Biol.* **23**, 504–510 (2013).
16. Chen, C. Y. & Shyu, A. B. Emerging mechanisms of mRNP remodeling regulation. Wiley Interdiscip *Rev. RNA* **5**, 713–722 (2014).
17. Anderson, P. & Kedersha, N. RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nat. Rev. Mol. Cell Biol.* **10**, 430–436 (2009).
18. Buchan, J. R. mRNP granules. Assembly, function, and connections with disease. *RNA Biol.* **11**, 1019–1030 (2014).
19. Protter, D. S. & Parker, R. Principles and properties of stress granules. *Trends Cell Biol.* **26**, 668–679 (2016).
20. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
21. Gloss, B. S. & Dinger, M. E. The specificity of long noncoding RNA expression. *Biochim. Biophys. Acta* **1859**, 16–22 (2016).
22. Geisler, S. & Coller, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **14**, 699–712 (2013).
23. Hudson, W. H. & Ortlund, E. A. The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* **15**, 749–760 (2014).
24. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
25. Beckmann, B. M., Castello, A. & Medenbach, J. The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflugers Arch.* **468**, 1029–1040 (2016).
26. Gehring, N. H., Wahle, E. & Fischer, U. Deciphering the mRNP code: RNA-bound determinants of post-transcriptional gene regulation. *Trends Biochem. Sci.* **42**, 369–382 (2017).
27. Rissland, O. S. The organization and regulation of mRNA-protein complexes. *Wiley Interdiscip. Rev. RNA* **8**, 1369 (2017).
28. Muller-McNicoll, M. & Neugebauer, K. M. How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat. Rev. Genet.* **14**, 275–287 (2013).
29. Helder, S., Blythe, A. J., Bond, C. S. & Mackay, J. P. Determinants of affinity and specificity in RNA-binding proteins. *Curr. Opin. Struct. Biol.* **38**, 83–91 (2016).
30. Hentze, M. W. & Argos, P. Homology between IRE-BP, a regulatory RNA-binding protein, aconitase, and isopropylmalate isomerase. *Nucleic Acids Res.* **19**, 1739–1740 (1991).

31. Rouault, T. A., Stout, C. D., Kaptain, S., Harford, J. B. & Klausner, R. D. Structural relationship between an iron-regulated RNA-binding protein (IRE-BP) and aconitase: functional implications. *Cell* **64**, 881–883 (1991).
32. Chu, E. *et al.* Identification of an RNA binding site for human thymidylate synthase. *Proc. Natl Acad. Sci. USA* **90**, 517–521 (1993).
33. Hentze, M. W. Enzymes as RNA-binding proteins: a role for (di)nucleotide-binding domains? *Trends Biochem. Sci.* **19**, 101–103 (1994).
34. Butter, F., Scheibe, M., Morl, M. & Mann, M. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc. Natl Acad. Sci. USA* **106**, 10626–10631 (2009).
35. Scherrer, T., Mittal, N., Janga, S. C. & Gerber, A. P. A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS ONE* **5**, e15499 (2010).
36. Tsvetanova, N. G., Klass, D. M., Salzman, J. & Brown, P. O. Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*. *PLoS ONE* **5**, e12671 (2010).
37. Treiber, T. *et al.* A compendium of RNA-binding proteins that regulate microRNA biogenesis. *Mol. Cell* **66**, 270–284.e13 (2017).
38. Castello, A. *et al.* System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.* **8**, 491–500 (2013).
39. Baltz, A. G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
This is one of the two pioneer RIC studies. In this work, RIC was applied to HEK293 cells identifying 791 RBPs, many of which were previously not known to bind RNA.
40. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
This is also one of the two pioneer RIC studies. In this work, RIC was applied to HeLa cells identifying 860 RBPs, many of which were previously not known to bind RNA.
41. Castello, A., Fischer, B., Hentze, M. W. & Preiss, T. RNA-binding proteins in Mendelian disease. *Trends Genet.* **29**, 318–327 (2013).
42. Strein, C., Alleaume, A. M., Rothbauer, U., Hentze, M. W. & Castello, A. A versatile assay for RNA-binding proteins in living cells. *RNA* **20**, 721–731 (2014).
43. Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11**, 1064–1070 (2014).
This article presents the development of a method that assigns RNA-binding sites within RBPs by direct identification of peptides crosslinked to a nucleotide.
44. Beckmann, B. M. *et al.* The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun.* **6**, 10127 (2015).
RIC was applied in this study to human hepatoma cells (HuH7) and *S. cerevisiae*. The RNA targets of one of the discovered RBPs, HSD17B10, were identified by iCLIP. This protein binds mitochondrial tRNAs, and a mutation-causing disease abrogates the interaction with RNA.
45. Castello, A. *et al.* Comprehensive identification of RNA-binding domains in human cells. *Mol. Cell* **63**, 696–710 (2016).
This article describes RBDmap, a method for proteome-wide identification of RBDs. Applied to HeLa cells, RBDmap revealed 1,174 RNA-binding sites in 529 RBPs.
46. Kwon, S. C. *et al.* The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1122–1130 (2013).
47. Liao, Y. *et al.* The cardiomyocyte RNA-binding proteome: links to intermediary metabolism and heart disease. *Cell Rep.* **16**, 1456–1469 (2016).
48. Liepelt, A. *et al.* Identification of RNA-binding proteins in macrophages by interactome capture. *Mol. Cell Proteom.* **15**, 2699–2674 (2016).
49. Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat. Struct. Mol. Biol.* **20**, 127–133 (2013).
50. Matia-Gonzalez, A. M., Laing, E. E. & Gerber, A. P. Conserved mRNA-binding proteomes in eukaryotic organisms. *Nat. Struct. Mol. Biol.* **22**, 1027–1033 (2015).
51. Nandan, D. *et al.* Comprehensive identification of mRNA-binding proteins of *Leishmania donovani* by interactome capture. *PLoS ONE* **12**, e0170068 (2017).
52. Bunnik, E. M. *et al.* The mRNA-bound proteome of the human malaria parasite *Plasmodium falciparum*. *Genome Biol.* **17**, 147 (2016).
53. Lueong, S., Merce, C., Fischer, B., Hoheisel, J. D. & Erben, E. D. Gene expression regulatory networks in *Trypanosoma brucei*: insights into the role of the mRNA-binding proteome. *Mol. Microbiol.* **100**, 457–471 (2016).
54. Marondedze, C., Thomas, L., Serrano, N. L., Lilley, K. S. & Gehring, C. The RNA-binding protein repertoire of *Arabidopsis thaliana*. *Sci. Rep.* **6**, 29766 (2016).
55. Reichel, M. *et al.* In Planta Determination of the mRNA-Binding Proteome of *Arabidopsis* Etiolated Seedlings. *Plant Cell* **28**, 2435–2452 (2016).
56. Zhang, Z. *et al.* UV crosslinked mRNA-binding proteins captured from leaf mesophyll protoplasts. *Plant Methods* **12**, 42 (2016).
57. Sysoev, V. O. *et al.* Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in *Drosophila*. *Nat. Commun.* **7**, 12128 (2016).
This study is the first quantitative RIC analysis to detect differential RBP activity. Here, 116 'dynamic' RBPs were identified during *D. melanogaster* early embryogenesis.
58. Wessels, H. H. *et al.* The mRNA-bound proteome of the early fly embryo. *Genome Res.* **26**, 1000–1009 (2016).
59. Despic, V. *et al.* Dynamic RNA-protein interactions underlie the zebrafish maternal-to-zygotic transition. *Genome Res.* **27**, 1184–1194 (2017).
60. Sonnhammer, E. L. & Ostlund, G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* **43**, D234–D239 (2015).
61. O'Brien, K. P., Remm, M. & Sonnhammer, E. L. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.* **33**, D476–D480 (2005).
62. Conrad, T. *et al.* Serial interactome capture of the human cell nucleus. *Nat. Commun.* **7**, 11212 (2016).
63. Gilbert, W. V., Bell, T. A. & Schaening, C. Messenger RNA modifications: form, distribution, and function. *Science* **352**, 1408–1412 (2016).
64. Hoernes, T. P. & Erlacher, M. D. Translating the epitranscriptome. *Wiley Interdiscip. Rev. RNA* **8**, e1375 (2017).
65. Ciesla, J. Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? *Acta Biochim. Pol.* **53**, 11–32 (2006).
66. Hentze, M. W. & Preiss, T. The REM phase of gene regulation. *Trends Biochem. Sci.* **35**, 423–426 (2010).
67. Castello, A., Hentze, M. W. & Preiss, T. Metabolic enzymes enjoying new partnerships as RNA-binding proteins. *Trends Endocrinol. Metab.* **26**, 746–757 (2015).
68. Chu, E. & Allegra, C. J. The role of thymidylate synthase as an RNA binding protein. *Bioessays* **18**, 191–198 (1996).
69. Liu, J. *et al.* Thymidylate synthase as a translational regulator of cellular gene expression. *Biochim. Biophys. Acta* **1587**, 174–182 (2002).
70. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
This study used a data mining approach to compile a catalogue of RBPs employing domain composition and known roles of proteins to classify them as RBPs.
71. ElAntak, L., Tzakos, A. G., Locker, N. & Lukavsky, P. J. Structure of eIF3b RNA recognition motif and its interaction with eIF3j: structural insights into the recruitment of eIF3b to the 40S ribosomal subunit. *J. Biol. Chem.* **282**, 8165–8174 (2007).
72. Bono, F., Ebert, J., Lorentzen, E. & Conti, E. The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* **126**, 713–725 (2006).
73. Klass, D. M. *et al.* Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in *Saccharomyces cerevisiae*. *Genome Res.* **23**, 1028–1038 (2013).
74. Brannan, K. W. *et al.* SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. *Mol. Cell* **64**, 282–293 (2016).
SONAR uses protein–protein interaction databases to predict potential RBPs by exploiting the tendency of RBPs to interact with other RBPs.
75. Sundararaman, B. *et al.* Resources for the Comprehensive Discovery of Functional RNA Elements. *Mol. Cell* **61**, 903–913 (2016).
76. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425–440 (2015).
77. Mukhopadhyay, R. *et al.* DAPK-ZIPK-L13a axis constitutes a negative-feedback module regulating inflammatory gene expression. *Mol. Cell* **32**, 371–382 (2008).
78. Arif, A. *et al.* Two-site phosphorylation of EPRS coordinates multimodal regulation of noncanonical translational control activity. *Mol. Cell* **35**, 164–180 (2009).
79. Arif, A., Jia, J., Moodt, R. A., DiCorleto, P. E. & Fox, P. L. Phosphorylation of glutamyl-prolyl tRNA synthetase by cyclin-dependent kinase 5 dictates transcript-selective translational control. *Proc. Natl Acad. Sci. USA* **108**, 1415–1420 (2011).
80. Clingman, C. C. *et al.* Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. *eLife* **3**, e02848 (2014).
81. Haghighat, A. & Sonenberg, N. eIF4G dramatically enhances the binding of eIF4E to the mRNA 5′-cap structure. *J. Biol. Chem.* **272**, 21677–21680 (1997).
82. Dabo, S. & Meurs, E. F. dsRNA-dependent protein kinase PKR and its role in stress, signaling and HCV infection. *Viruses* **4**, 2598–2635 (2012).
83. Habjan, M. & Pichlmair, A. Cytoplasmic sensing of viral nucleic acids. *Curr. Opin. Virol.* **11**, 31–37 (2015).
84. Rehwinkel, J. RNA sensing: the more RIG-I the merrier? *EMBO Rep.* **14**, 751–752 (2013).
85. Boucas, J. *et al.* Label-free protein-RNA interactome analysis identifies Khsrp signaling downstream of the p38/Mk2 kinase complex as a critical modulator of cell cycle progression. *PLoS ONE* **10**, e0125745 (2015).
86. Lasko, P. Posttranscriptional regulation in *Drosophila* oocytes and early embryos. *Wiley Interdiscip. Rev. RNA* **2**, 408–416 (2011).
87. Fernandez-Chamorro, J. *et al.* Identification of novel non-canonical RNA-binding sites in Gemin5 involved in internal initiation of translation. *Nucleic Acids Res.* **42**, 5742–5754 (2014).
88. Schmidt, C., Kramer, K. & Urlaub, H. Investigation of protein-RNA interactions by mass spectrometry — techniques and applications. *J. Proteom.* **75**, 3478–3494 (2012).
89. Nagy, E. & Rigby, W. F. Glyceraldehyde-3-phosphate dehydrogenase selectively binds AU-rich RNA in the NAD^+-binding region (Rossmann fold). *J. Biol. Chem.* **270**, 2755–2763 (1995).
90. Chang, C. H. *et al.* Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* **153**, 1239–1251 (2013).
91. Mullari, M., Lyon, D., Jensen, L. J. & Nielsen, M. L. Specifying RNA-binding regions in proteins by peptide cross-linking and affinity purification. *J. Proteome Res.* **16**, 2762–2772 (2017).
92. He, C. *et al.* High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Mol. Cell* **64**, 416–430 (2016).
This study used a proteomic-based method to assign RNA-binding sites to RBPs by identifying peptides for which the intensity is reduced after ultraviolet irradiation. The study detected 1,475 RNA-binding sites mapping to 803 proteins.
93. Kaneko, S. *et al.* Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev.* **24**, 2615–2620 (2010).
94. Kaneko, S., Son, J., Bonasio, R., Shen, S. S. & Reinberg, D. Nascent RNA interaction keeps PRC2 activity poised and in check. *Genes Dev.* **28**, 1983–1988 (2014).
95. Kaneko, S., Son, J., Shen, S. S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1258–1264 (2013).
96. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).
97. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
98. König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).

99. Van Nostrand, E. L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
This study used an enhanced version of the iCLIP method as well as ultraviolet crosslinking, RNase treatment and RNA sequencing to determine the binding sites of RBPs across the cellular transcriptome. This has been applied to many RBPs under the same experimental conditions, offering a rich source of information on protein–RNA interactions.
100. Lambert, N. *et al.* RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol. Cell* **54**, 887–900 (2014).
101. Riley, K. J. & Steitz, J. A. The "Observer Effect" in genome-wide surveys of protein-RNA interactions. *Mol. Cell* **49**, 601–604 (2013).
102. Steckelberg, A. L., Boehm, V., Gromadzka, A. M. & Gehring, N. H. CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep.* **2**, 454–461 (2012).
103. Haberman, N. *et al.* Insights into the design and interpretation of iCLIP experiments. *Genome Biol.* **18**, 7 (2017).
104. Jarvelin, A. I., Noerenberg, M., Davis, I. & Castello, A. The new (dis)order in RNA regulation. *Cell Commun. Signal* **14**, 9 (2016).
105. Basu, S. & Bahadur, R. P. A structural perspective of RNA recognition by intrinsically disordered proteins. *Cell. Mol. Life Sci.* **73**, 4075–4084 (2016).
106. Thandapani, P., O'Connor, T. R., Bailey, T. L. & Richard, S. Defining the RGG/RG motif. *Mol. Cell* **50**, 613–623 (2013).
107. Phan, A. T. *et al.* Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.* **18**, 796–804 (2011).
108. Han, T. W. *et al.* Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779 (2012).
109. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
110. Vuzman, D., Azia, A. & Levy, Y. Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J. Mol. Biol.* **396**, 674–684 (2010).
111. Casu, F., Duggan, B. M. & Hennig, M. The arginine-rich RNA-binding motif of HIV-1 Rev is intrinsically disordered and folds upon RRE binding. *Biophys. J.* **105**, 1004–1017 (2013).
112. Gold, L. *et al.* Aptamers and the RNA world, past and present. *Cold Spring Harb. Perspect. Biol.* **4**, a003582 (2012).
113. de Breyne, S., Yu, Y., Unbehaun, A., Pestova, T. V. & Hellen, C. U. Direct functional interaction of initiation factor eIF4G with type 1 internal ribosomal entry sites. *Proc. Natl Acad. Sci. USA* **106**, 9197–9202 (2009).
114. Sizova, D. V., Kolupaeva, V. G., Pestova, T. V., Shatsky, I. N. & Hellen, C. U. Specific interaction of eukaryotic translation initiation factor 3 with the 5' nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. *J. Virol.* **72**, 4775–4782 (1998).
115. Quade, N., Boehringer, D., Leibundgut, M., van den Heuvel, J. & Ban, N. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-A resolution. *Nat. Commun.* **6**, 7646 (2015).
116. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* http://dx.doi.org/10.1038/nrm.2017.103 (2017).
117. Clemson, C. M. *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717–726 (2009).
118. Chu, E. *et al.* Autoregulation of human thymidylate synthase messenger RNA translation by thymidylate synthase. *Proc. Natl Acad. Sci. USA* **88**, 8977–8981 (1991).
119. Muckenthaler, M. U., Rivella, S., Hentze, M. W. & Galy, B. A. Red carpet for iron metabolism. *Cell* **168**, 344–361 (2017).
120. Walden, W. E. *et al.* Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science* **314**, 1903–1908 (2006).
121. Tristan, C., Shahani, N., Sedlak, T. W. & Sawa, A. The diverse functions of GAPDH: views from different subcellular compartments. *Cell Signal.* **23**, 317–323 (2011).
122. Rodriguez-Pascual, F. *et al.* Glyceraldehyde-3-phosphate dehydrogenase regulates endothelin-1 expression by a novel, redox-sensitive mechanism involving mRNA stability. *Mol. Cell. Biol.* **28**, 7139–7155 (2008).
123. Xing, S. & Poirier, Y. The protein acetylome and the regulation of metabolism. *Trends Plant Sci.* **17**, 423–430 (2012).
124. Arif, W., Datar, G. & Kalsotra, A. Intersections of post-transcriptional gene regulatory mechanisms with intermediary metabolism. *Biochim. Biophys. Acta* **1860**, 349–362 (2017).
125. Campbell, Z. T. & Wickens, M. Probing RNA-protein networks: biochemistry meets genomics. *Trends Biochem. Sci.* **40**, 157–164 (2015).
126. König, J., Zarnack, K., Luscombe, N. M. & Ule, J. Protein-RNA interactions: new genomic technologies and perspectives. *Nat. Rev. Genet.* **13**, 77–83 (2011).
127. Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA binding proteins. *bioRxiv*, 179648 (2017).
128. Papasaikas, P., Tejedor, J. R., Vigevani, L. & Valcarcel, J. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol. Cell* **57**, 7–22 (2015).
129. Tejedor, J. R., Papasaikas, P. & Valcarcel, J. Genome-wide identification of Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. *Mol. Cell* **57**, 23–38 (2015).
130. Choudhury, N. R. *et al.* Trim25 is an RNA-specific activator of Lin28a/TuT4-mediated uridylation. *Cell Rep.* **9**, 1265–1272 (2014).
131. Manokaran, G. *et al.* Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness. *Science* **350**, 217–221 (2015).
This work shows that the XRN1-generated, subgenomic RNA of dengue virus interferes with IFNβ induction by sequestering TRIM25, preventing the ubiquitination of RIGI by this protein.
132. Choudhury, N. R. *et al.* RNA-binding activity of TRIM25 is mediated by its PRY/SPRY domain and is required for ubiquitination. *BMC Biol.* **15**, 105 (2017).
In this paper, TRIM25 target RNAs are uncovered by CLIP analysis. TRIM25 displays a preference for guanine-rich tracks present at the 3′ UTR of hundreds of mRNAs. TRIM25 E3 ubiquitin ligase activity is stimulated in the interaction with RNA.
133. Gebhardt, A. *et al.* mRNA export through an additional cap-binding complex consisting of NCBP1 and NCBP3. *Nat. Commun.* **6**, 8192 (2015).
In this study, NCBP3 mediates RNA export through its interaction with NCBP1 following a similar mechanism as NCBP2. NCBP3 is an RBP newly discovered by the RIC studies.
134. Popow, J. *et al.* FASTKD2 is an RNA-binding protein required for mitochondrial RNA processing and translation. *RNA* **21**, 1873–1884 (2015).
This work provides an iCLIP analysis of the noncanonical RBP FASTKD2, showing that FASTKD2 binds mitochondrial RNAs and regulates their metabolism.
135. Gaspar, I., Sysoev, V., Komissarov, A. & Ephrussi, A. An RNA-binding atypical tropomyosin recruits kinesin-1 dynamically to oskar mRNPs. *EMBO J.* **36**, 319–333 (2017).
136. Yang, S. Y. *et al.* Roles of 17beta-hydroxysteroid dehydrogenase type 10 in neurodegenerative disorders. *J. Steroid Biochem. Mol. Biol.* **143**, 460–472 (2014).
137. Rauschenberger, K. *et al.* A non-enzymatic function of 17beta-hydroxysteroid dehydrogenase type 10 is required for mitochondrial integrity and cell survival. *EMBO Mol. Med.* **2**, 51–62 (2010).
138. Holzmann, J. *et al.* RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135**, 462–474 (2008).
139. Vilardo, E. & Rossmanith, W. Molecular insights into HSD10 disease: impact of SDR5C1 mutations on the human mitochondrial RNase P complex. *Nucleic Acids Res.* **43**, 6649 (2015).
140. Antonicka, H. & Shoubridge, E. A. Mitochondrial RNA granules are centers for posttranscriptional rna processing and ribosome biogenesis. *Cell Rep.* **10**, 920–932 (2015).
141. Ghezzi, D. *et al.* FASTKD2 nonsense mutation in an infantile mitochondrial encephalomyopathy associated with cytochrome *c* oxidase deficiency. *Am. J. Hum. Genet.* **83**, 415–423 (2008).
142. Faehnle, C. R., Walleshauser, J. & Joshua-Tor, L. Mechanism of Dis3l2 substrate recognition in the Lin28-let-7 pathway. *Nature* **514**, 252–256 (2014).
143. Gack, M. U. *et al.* TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **446**, 916–920 (2007).
144. Chapman, E. G. *et al.* The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science* **344**, 307–310 (2014).
145. Kanakkanthara, A. *et al.* Cyclin A2 is an RNA binding protein that controls Mre11 mRNA translation. *Science* **353**, 1549–1552 (2016).
This study finds that cyclin A2 directly binds the 3′ UTR of the MRE11 mRNA (marked with arrows) to promote MRE11 mRNA translation.
146. Pashev, I. G., Dimitrov, S. I. & Angelov, D. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem. Sci.* **16**, 323–326 (1991).
147. Van Nostrand, E. L. *et al.* CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA binding proteins. *Methods* **118**, 50–59 (2016).
148. Wang, X. *et al.* N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**, 1388–1399 (2015).
149. Zhou, J. *et al.* Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594 (2015).
150. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).
151. Wang, Y. *et al.* N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* **16**, 191–198 (2014).
152. Batista, P. J. *et al.* m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707–719 (2014).
153. Geula, S. *et al.* Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).
154. Zhao, B. S. *et al.* m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **542**, 475–478 (2017).
155. Liu, N. *et al.* N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).
156. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).
157. Prevot, D., Darlix, J. L. & Ohlmann, T. Conducting the initiation of protein synthesis: the role of eIF4G. *Biol. Cell* **95**, 141–156 (2003).
158. Balachandran, S. & Barber, G. N. PKR in innate immunity, cancer, and viral oncolysis. *Methods Mol. Biol.* **383**, 277–301 (2007).

Acknowledgements
The authors dedicate this Review to the memory of B. Fischer, who sadly passed away while this Review was in preparation. The authors are grateful to the members of their laboratories for helpful discussions throughout. The authors also acknowledge research funding from the European Research Council (ERC-2011-ADG_20110310; M.W.H.), a UK Medical Research Council Career Development Award (MR/L019434/1; A.C), the European Molecular Biology Laboratory Interdisciplinary Postdoctoral 2 Programme (EIPOD2/291772; T.S.) and the Australian National Health and Medical Research Council (APP1120483; M.W.H. and T.P.).

Author contributions
All authors researched data for the article, made substantial contributions to the discussion of content, wrote the article and reviewed and/or edited the manuscript before submission.

Competing interests statement
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SUPPLEMENTARY INFORMATION
S1 (figure) | S2 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF