A Brave New World of RNA-Binding Proteins

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

[1] European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

[2] Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom

[3] EMBL–Australia Collaborating Group, Department of Genome Sciences, The John Curtin School of Medical Research, The Australian National University, Acton (Canberra) ACT 2601, Australia

[4] Victor Chang Cardiac Research Institute, Darlinghurst (Sydney), NSW 2010, Australia

* Correspondence to MWH (hentze@embl.de) or TP (thomas.preiss@anu.edu.au)

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, ll. 203–206 'Miranda's speech'*

ABSTRACT

RNA-binding proteins (RBPs) are typically thought of as proteins that bind RNA through one or multiple globular RNA-binding domains (RBDs) to change the fate or function of the bound RNAs. Several hundreds of 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 classical RBDs. This review integrates these new RBPs and discusses their possible functions, including the notion that some may be regulated by RNA rather than exerting control over it.

INTRODUCTION

A 'classical' RNA-binding protein (RBP) participates in the formation of ribonucleoprotein (RNP) complexes that are principally involved in gene expression[1]. It 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], the K homology (KH)[4] or DEAD box helicase domains (**Fig. 1a**)[5]. These assertions represent decades of cumulative knowledge, including, 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 complex protein-RNA packaging without involvement of canonical RBDs. This suggests that unorthodox 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]. Implicit in this conventional view of RBPs is also that they should act in furtherance of the RNA's function[13]. A recent review used the metaphor of RBPs as the mRNA's clothes that will ensure that different mRNA domains (i.e. the 5' and 3' untranslated regions [UTRs] and the coding region [CDS]) are at times covered up or exposed, helping it to pass through different stages of its life[14]. The processes driving change of RNP composition have been likened to those involved in chromatin activity[15,16]. Accordingly, post-translational modifications (PTMs) of RBPs and epitranscriptomic modifications of RNA, together with the action of ATP-dependent RNA helicases, lead to dynamic RNP remodelling.

The above concepts have broad but clearly not universal applicability, as indicated by emerging evidence from several directions. First, multiple microscopically visible membrane-less RNP granules have been characterised 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 (SGs) 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 separations (LLPS), thought to be driven by intrinsically disordered regions (IDRs; protein regions lacking stable tertiary structure in their native state) of their constituent RBPs[19,20]. Their dynamicity and amorphous structure still remains puzzling, and well-defined functions remain to be assigned to the formation of these RNP bodies. Second, our realisation of the existence of a myriad of long noncoding RNAs (lncRNAs) triggered intense efforts to uncover their functions[21,22]. Many lncRNAs are presently

expected to act in the recruitment of transcription factors or chromatin-modifying complexes to genomic loci, or otherwise organise, scaffold or inhibit protein assemblies[23,24]. These latter scenarios tend to break with convention, indicating that an RNA may act on its protein binding partner rather than itself being the target of regulation (**Fig. 1b**). Finally, several unbiased approaches to identify RBPs proteome-wide 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, lacking either discernible RBDs or 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 covered by the reviews cited above and other recent works[26-28]. This review focuses primarily on the challenges posed by unconventional RBPs, their methods of discovery and emerging information about their modes of RNA-binding, RNA targets, as well as 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 nineties of several metabolic enzymes that engage in 'moonlighting' RNA-binding activity, it became apparent that the number and diverse nature of RBPs had been underestimated[29-32]. The list of unconventional RBPs grew incrementally over the decades, urging the development of methods to identify RBPs comprehensively in living cells.

Experimental approaches to catalogue RBPs

In vitro approaches were devised using either immobilised RNA probes or arrayed proteins (**Box 1**), identifying multiple novel RBPs[33-36]. More recently, RNA interactome capture was

developed as an *in vivo* approach focusing on native protein-RNA interactions (**Box 1**). It entails UV crosslinking of RBPs to RNA in live cells followed by collective capture of RNPs containing poly(A)$^+$ RNA on oligo(dT) beads, and identification of proteins by quantitative mass spectrometry (Q-MS)[37]. Two groups independently deployed RNA interactome capture to yield 860 and 791 proteins as the RBP repertoire of human HeLa and HEK293 cells, respectively[38,39]. Both RNA interactomes overlap considerably, with 543 shared RBPs, and enrichment for the gene ontology (GO) term "RNA binding" (*e.g.* 468 of a total of 860 proteins in the HeLa RNA interactome, based on annotation current at the time)[40], and detected the majority of the well-established RBPs. Classical RBDs such as the RRM, KH, DEAD box helicase and some zinc finger domains were also strongly enriched. Altogether this attested to the technical robustness of the approach. Beyond that, about half of the proteins in each RNA interactome lacked known RBDs, and hundreds had no previous relation 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, protein metabolism (chaperons and prolyl *cis-trans* isomerases), and others[38,39]. The discovery of RNA-binding activity of proteins involved in biological processes without apparent relation to RNA biology ('enigmRBPs') suggested unexplored regulatory layers between gene expression and other biological processes. About two dozen of these enigmRBPs were validated by orthogonal approaches[38,39,41], seven of them were analysed by immunoprecipitation followed by sequencing of bound RNA, demonstrating that these proteins specifically interact with distinct sets of poly(A)$^+$ RNAs[38,39] and exert defined biological functions.

RNA interactome capture was since applied to samples from diverse sources (**Fig. 2**), typically identifying several hundreds of active RBPs. Sources include several additional

human[38,39,42-44] and mouse cell lines[45-47], yeast[43,48,49], the unicellular parasites *Leishmania donovani*[50], *Plasmodium falciparum*[51] and *Trypanosoma brucei*[52], as well as plants[53-55], flies[56,57], worms[49] and fish[58]. The RBP sets of different origins each featured enrichment of RNA-related annotation (for example, **see Fig. S1a)**, and orthogonal methods were typically used to validate some of the unexpectedly discovered RBPs, including dual-fluorescence RNA-binding assay[39,41,49], crosslinking and immunoprecipitation (CLIP) followed by 5' radioactive labelling of RNA by T4 polynucleotide kinase[38,43,45,46,56,57], RT-PCR [49] or sequencing[38,39,43]. Using updated annotation, we compiled here all published RNA interactomes into RBP supersets for *Homo sapiens* (1914), *Mus musculus* (1393), *Saccharomyces cerevisiae* (1273), *Drosophila melanogaster* (777)*, Arabidopsis thaliana* (719) and *Caenorhabditis elegans* (593) (**Fig. 2a**, **Table S1**). These datasets represent a resource to mine for shared and selectively identified RBPs **(Fig. 2b-f**). While simple technical reasons will explain some of the differences in coverage and overlap, the outlines of shared 'core' RNA interactomes are emerging[43]. To illustrate this, we performed InParanoid[59,60] analysis to yield a high number of orthologue groups especially between mammals, which expectedly decreases in more evolutionary distant organisms (**Fig. 2g-h**). On the whole, widely shared RBPs tend to be more highly enriched in established RNA-related annotation than RBPs with cell or organismic context-dependent expression or activity.

Some specific observations from these datasets deserve attention. For example, the mouse embryonic stem cell (mESC) RNA interactome was found to be enriched in proteins with differential expression during differentiation, suggesting that RBPs are regulated during the transition from stem to differentiated cell[45]. In particular, the transcriptional network of the proto-oncogene Myc was found to be enriched among mESC RBPs, suggesting that RBPs contribute to Myc-dependent cell fate decisions. Moreover, mESC RBPs were significantly

upregulated during the third day of reprogramming into induced pluripotent stem cells (iPSC). RNA interactome capture was further combined with nuclear fractionation to determine the repertoire of nuclear RBPs of K562 myeloid leukaemia cells; the p53 interaction network was enriched among the newly discovered nuclear RBPs[61].

RNA modifying enzymes such as RNA tailing enzymes and 'writers' of epitranscriptomic modifications consistently form part of RNA interactome compendia. The cardiomyocyte RNA interactome contains 29 RBPs annotated for 5-methylcytosine (m^5C), N6-methyladenosine (m^6A) and pseudouridine modifications, as well as adenosine-to-inosine editing[46]. Such epitranscriptomic marks may affect RNA-protein interactions, or otherwise modulate RNA function[62,63] (**Box 2**).

Metabolic enzymes were recurrently identified as RBPs. Our meta-analysis revealed 71 such 'moonlighting' metabolic enzymes in humans, 104 in mouse and 132 in yeast RNA interactomes (**Table S1**), adding a plethora of new examples to a list of about twenty RNA-binding metabolic enzymes discovered by classic low-throughput studies[64-66]. Collectively, these dual RBP-enzymes cover a breadth of metabolism, with interesting differences in predominant pathways depending on the source material. For example, the RNA interactome of Huh-7 cells harbours numerous enzymes from glycolysis and other pathways of intermediary metabolism[43,49]. This likely relates to the important metabolic role of hepatocytes at the organismal level[43]. The HL-1 cardiomyocyte RNA interactome exhibits a high incidence particularly of mitochondrial metabolic enzymes[46], reflecting their high organelle content and energy needs. Many of these RNA-binding metabolic enzymes were validated by orthogonal approaches[39,43,44,46], including mitochondrial fractionation followed by T4 polynucleotide kinase (PNK) assay[46]. Interestingly, some of these enzymes interact with their own mRNA[49], perhaps hinting at the existence of negative feedback loops by cognate enzyme-mRNA interaction under conditions of substrate or co-factor deprivation,

as previously proposed[64-68]. Taken together, this suggests that RNA binding by metabolic enzymes is a common and widely conserved occurrence.

In silico approaches to catalogue RBPs

A set of 1,542 human RBPs (7.5% of the proteome) was defined based on computational analyses that required a protein to harbour known RBDs or other domain features characteristically found in proteins with RNA-related functions. This was complemented by manual curation to add missing but well-documented RBPs or proteins forming part of known RNPs, and to remove proteins with established RNA-unrelated functions[69]. Members of this curated set of RBPs tend to display ubiquitous expression profiles across tissues, suggesting housekeeping roles in cell biology, and the set overlaps well with the experimental human RNA interactomes (**Fig. S1b**). Notwithstanding its strengths, the approach might generate false positive results for i) proteins harbouring a classified RBD that serves a different functional role in that particular case[70], or ii) proteins interacting with RNA solely via protein-protein interactions with a direct RBP (as in the case of Y14 in the exon junction complex)[71]. A similar domain/function-search algorithm was recently applied to *Plasmodium falciparum*[51]. This study reported 924 RBPs, a surprisingly high fraction (18.1%) of the relatively small number of protein-coding genes of the malaria parasite.

The propensity of RBPs to interact with other RBPs, either directly or via bridging RNAs, was also exploited to identify novel RBPs[39,72,73]. The classification algorithm, referred to as "support vector machine obtained from neighbourhood associated RBPs" (SONAR), evaluates each protein of a proteome (or a dataset) against protein-protein interaction (PPI) data, and calculates its RBP classification score (RCS). The algorithm was trained on sets of human RBPs merged from available sources[39,69,74], where the "neighbourhood" of each protein was determined using the BioPlex PPI dataset[75], which includes many thousands of

experimentally determined protein-protein interactions. Sets of SONAR-predicted RBPs were established for human (1,784 proteins), *Drosophila melanogaster* (489) and *Saccharomyces cerevisiae* (745); these agree well with the experimentally determined RBP sets (**Fig. S1c,d**). SONAR can be readily applied to any organism as long as substantial PPI data are available. SONAR thus presents the same limitations as any proteomic-based approach, including the depth of the protein-protein interaction data available in a given organism. In addition, it may lead to false positives because proteins that interact with RBPs are not always RBPs themselves[71].

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. While a subset of 'house-keeping' RBPs might be constitutively and ubiquitously active[69], many RBPs have more restricted expression patterns and/or their RNA-binding activity may be regulated, *e.g.* by post-translational modifications[76-78], cofactor binding[79] or protein-protein interactions[80]. Moreover, some RBPs can 'sit idle' for lack of their RNA targets[81-83]. For example, cellular sensors against viral infection such as interferon-induced, double-stranded RNA-activated protein kinase (EIF2AK2 or PKR), retinoic acid-inducible gene-I (RIGI or DDX58) or toll-like receptors (TLRs) may only be activated by the presence of unusual RNA products derived from viral replication such as long double stranded RNA tracts[81] or triphosphate 5' ends[82,83] (Fig. 1b).

RNA interactome capture has been adapted to investigate responses of RNA-binding proteomes to physiological changes and environmental cues. This was recently applied to murine macrophages responding to lipopolysaccharide stimulation[47], primary mouse embryonic fibroblasts treated with etoposide[84], as well as to fruit fly and zebrafish embryos

at different stages of development[56,58]. During the maternal-to-zygotic transition (MZT), gene expression switches from maternally deposited mRNAs whose timed expression is orchestrated by RBPs, to a new transcriptome emerging after zygotic genome activation (ZGA)[85]. The dynamic changes in RBP activity expected during MZT were investigated in *Drosophila* embryos[56]. Using tandem mass tags (TMT)[86], biological triplicate samples derived from pre- (0-1 hours post fertilisation [hpf]) and post-ZGA embryos (4.5-5.5 hours) were analysed. Comparison between UV-irradiated *versus* non-irradiated pooled pre- and post-ZGA stage samples was used to determine the 'static' RNA interactome capture. In separate experiments, UV-irradiated samples from pre- and post-ZGA stages were directly compared for the identification of 'dynamic' RBPs. The former analysis yielded 523 proteins as high confidence fly embryo RBPs, whereas the latter identified 1131 proteins, 116 of which were differentially captured at 10% false discovery rate (FDR). 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 three classes of RBPs: class 1 (1015 proteins) showed no significant change 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) with 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'. The latter include eight known splicing factors, and seven of these bind RNA more avidly in pre-MZT embryos. These findings broadly concur with an analysis using mRNA expression and localisation as proxies to reveal that RBP expression peaks during the prezygotic phase and MZT, while transcription factors are highly expressed during the first zygotic waves and mid embryogenesis[57]. A similarly configured analysis of zebrafish embryos before (1.75 hpf)

and during zygotic genome activation (3 hpf) uncovered a repertoire of 227 RBPs active during vertebrate MZT. As expected, this set included many regulators of mRNA polyadenylation, translation and stability but proteins involved in RNA modification and pre-mRNA processing factors were also notably represented. 24 and 53 proteins were differentially active at pre-ZGA and ZGA, respectively, and appeared to be mostly 'dynamic binders'[58].

Comparative RNA interactome capture can thus be used to investigate dynamic changes in RNA-binding proteomes. It should thus be feasible to study a wide range of questions, from development and differentiation to signalling, metabolism, infection and other disease processes and drug effects.

UNORTHODOX RNA-BINDING DOMAINS

Many of the RBPs newly discovered by the different methods described above lack known RBDs, posing questions of how they interact with RNA. Tried-and-tested approaches exist to map RBDs within individual proteins, for example by mutagenesis combined with RNA-binding assays, such as electrophoretic mobility shift assay (EMSA)[87] or the CLIP-coupled PNK assay[45], however, high(er) throughput methods were required to identify the RBDs of hundreds of novel RBPs in a time-efficient manner. Thus, three approaches were developed, each using mass-spectrometry in a different configuration to identify protein regions that become crosslinked to RNA after UV exposure of live cells.

One approach focused on the purification and direct detection of the RNA-crosslinked peptides, whose mass is altered by the nucleic acid remnant (**Fig. 3a**)[42,88]. Data analysis was performed with RNP^{xl} [42], a custom-designed software to reduce the complexity of the search for peptide-nucleotide matching spectra. Applied to yeast RBPs, RNP^{xl} identified 376 crosslinked peptides corresponding to 133 unique crosslinking sites in 57 different

proteins, mostly canonical RBPs such as ribosomal proteins and proteins with RRM or KH domains[42]. A number of RNA-binding sites were identified in unorthodox RBPs, including the enzymes peptidyl-prolyl cis-trans isomerase (CPR1), enolase 1 (ENO1) and phosphoglycerate kinase (PGK1).

RBDmap adds a further digestion step with a protease that cleaves every 17 amino acids on average and a second round of oligo(dT) capture to the RNA interactome capture workflow[44]. After the second oligo(dT) capture, the covalently linked polypeptides are proteolysed by trypsin, generating an RNA-crosslinked peptide and a neighbouring peptide with native mass (Fig 3b). RBDmap employs the neighbouring peptides with native mass to assign RNA-binding sites (**Fig. 3b**)[44]. Applied to HeLa cells, RBDmap discovered 1,174 RNA-binding sites within 529 proteins[44], while more limited analysis of HL-1 cardiomyocytes still revealed 568 RNA-binding sites from 368 proteins[46]. RBDmap data display strong concordance with regular RNA interactome studies, reinforcing the identification of unorthodox RBPs even after two oligo(dT) purification steps and after extensive proteolytic treatment[44,46]. As expected, RBDmap "re-discovered" the classical RBDs such as the RRM-, KH-, or cold shock domains[2]. Notably, many of the reported RNA-binding sites mapped to globular proteins and domains lacking previous association with RNA binding. For HeLa cells these include the thioredoxin (TXN) fold, heat shock protein (HSP) 70 and HSP90 domains, 14-3-3, domain associated with zinc fingers (DZF), PDZ and NDR (complete list in[44]), many of which were validated by orthogonal approaches. Interestingly, the mapped RNA-binding sites showed enrichment for homologous regions across different proteins from the same family, and many mapped to enzymatic cores or protein-protein interaction surfaces, suggesting a possible interplay between these activities and RNA binding[44]. For HL-1 cells, RNA-binding regions were identified in 24 metabolic enzymes, 12 of which

mapped to di-nucleotide binding domains[46]. These data corroborate and expand the previously suggested RNA-binding activity of di-nucleotide-binding domains[32,65,66,89,90]. Unexpectedly, many of the identified RNA-binding regions mapped to IDRs, implicating them as predominant sites of protein interaction with RNA *in vivo* (see below). Another observation was that RNA-binding regions were enriched for regions mutated in Mendelian diseases, while natural variants distributed equally across binding and non-binding regions[44]. This suggests that numerous Mendelian diseases arise from altered RNA biology. Finally, the RNA-binding regions strongly overlap with known post-translational modification sites, including phosphorylation, acetylation and methylation[44]. This enrichment was not observed for protein regions lacking RNA-binding activity. Thus, posttranslational modifications may regulate RNA binding and RNP dynamics akin to chromatin remodelling.

Peptide cross-linking and affinity purification (pCLAP) is a recently described cousin of RBDmap, which 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 (Fig. 3b) [44].

A third approach, termed proteomic identification of RNA-binding regions (RBR-ID) identifies peptides with reduced intensity in UV-irradiated samples compared to non-irradiated controls (**Fig. 3c**)[92]. RBR-ID of nuclear proteins from mESCs detected 1,475 RNA-binding sites mapping to 803 mESC proteins with a 5% FDR. 47% of the RBR-ID proteins were present in the RNA interactome studies[38,39,43,45,61], while 865 previously known RBPs were missing. Interestingly, the 427 RBPs newly identified by RBR-ID were enriched for gene-regulatory and chromatin-associated functions. At the domain level, peptides with UV-dependent intensities (i.e. RNA-binding sites) mapped to RRM, KH, DEAD box and other classical RBDs, validating the approach. The RNA-binding sites of the

427 newly discovered RBPs surprisingly often mapped to chromatin-related domains such as the chromodomain and bromodomain, invoking a potential crosstalk between DNA and RNA. Some of the latter results confirmed prior candidate-based approaches characterizing the activity of chromatin remodelling proteins such as the catalytic subunit of the polycomb repressor complex 2 (PRC2) EZH2, which has been shown to interact with long non-coding RNA and nascent transcripts[93-95]; RBR-ID data suggest that this phenomenon might be broader than previously anticipated.

Together, RNP[x1], RBDmap and RBS-ID have greatly expanded the known repertoire of RBDs. However, we know little about the RNA targets of these novel RBDs or the function of these interactions. To address this, functional studies including determination of the specificity and affinity of these novel RBDs for their target sequences[96-100] will be required.

NOVEL RNA-BINDING MODES

The many novel RBPs and their non-typical RBDs raise pertinent questions regarding their biological function(s)[101]. Not every bimolecular collision that happens in cells should be assumed to be physiologically relevant. Which ranges for the affinities and RNA-binding specificities are to be expected? Concepts regarding specificity and non-specificity in RNA-protein interactions were recently reviewed[12]. A first relevant aspect is that indiscriminate RNA binding by RBPs is common and can be important to their function. For example, numerous proteins involved in mRNA translation (for example, the eukaryotic initiation factor [eIF]4F complex) and degradation need to be non-selective to fulfil their functions (**Fig. 4a vs b**). Similarly, the exon junction complex is deposited through protein-protein interactions[102] at a constant position upstream of splice junctions, and interacts with a plurality of RNA sequences[103] (Fig. 4c). Second, one must distinguish between 'biological specificity', the binding characteristics of RBPs *in vivo*, and 'intrinsic specificity', such as

one might determine by selection from a random pool of sequences *in vitro*. An interesting upshot of this distinction is that intrinsically specific RBPs may function biologically as less specific RBPs, when they bind their physiological RNA targets without much discrimination because these do not fall into the high-affinity/high-specificity range of their RNA-binding potential[12].

RNA Binding by Intrinsically Disordered Regions

As described above, IDRs are not only involved in aggregation of RNPs into granules by protein-protein interactions[19,20] but also directly engage in RNA binding[44,46,104,105]. RBPs are enriched in IDRs that are characterized by a low content of bulky hydrophobic amino acids, with the exception of tyrosine (Y), and a high proportion of small, polar or/and charged amino acids, particularly glycine (G), serine (S), arginine (R), lysine (K), glutamine (Q), glutamic acid (E) and aspartic acid (D)[44]. Interestingly, mutations in RBPs causing human Mendelian diseases occur with higher frequency within these RNA-binding IDRs than in globular domains, suggesting strong sequence constraints[40]. The occurrence of IDRs within RBPs appears to be conserved from yeast to human[43], often in the form of repeats such as RGG, YGG, SR, DE or KK[39,104]. A recent report proposes that the number of repeats within the IDRs of RBPs has expanded from yeast to human[43]. Because the number and identity of globular domains present within these proteins have remained the same, IDRs may represent a plastic component of RBPs that co-evolved with the increasing complexity of eukaryotic transcriptomes.

Around half of the RNA-binding sites reported by RBDmap (1,174) in HeLa cells mapped to IDRs, reflecting their prevalence as a mode of RNA binding[44]. 170 RBPs appeared to interact with RNA exclusively through IDRs, suggesting that these regions can suffice to mediate RNA binding. Amongst the arginine-rich motifs (ARM), RGG and SR repeats were

previously reported to bind RNA[104]. The discovery of numerous additional examples allowed assignment into sub-classes that differ by the lengths of their glycine linkers[44,106]. Nuclear magnetic resonance (NMR) analyses of human fragile X mental retardation protein (FMRP) showed that the positioning of the arginines is essential for selective binding of the RGG box to the guanidine-rich sequence in the sc1 mRNA[107], where the glycine linker orients the arginines geometrically to interact with the Watson-Crick nucleotide base pairs, which stack on two G-quadruplexes resulting from the protein-RNA co-folding (**Fig. 4d**). In addition, the glycine linker contributes to RNA binding by shape complementarity interactions at the interface between G-quadruplexes and Watson-Crick base pairs. Hence, the affinity and selectivity of RGG boxes for their target RNAs may be determined by the frequency of arginines and glycines.

A second RNA-binding family of IDRs involves aromatic residues, especially tyrosines (Y), that combine with glycines (G) and serines (S) forming [G/S]Y[G/S]. These motifs display a tendency to aggregate *in vitro*, inducing hydrogel formation and amyloid-like fibres, and engage in dynamic liquid-liquid phase separations *in vivo*[108,109]. Aromatic residues tend to form part of hydrophobic protein cores, but, when present at the protein surface, can interact with amino acids or nucleotides by stacking or hydrogen bonding[2]. When embedded in a glycine-rich context, the aromatic residue is particularly exposed, likely fostering its propensity to aggregate when interacting with similar protein motifs[108,109], or to bind RNA[44].

Finally, a heterogeneous set of linear motifs involving lysine (K) and, to a lesser extent, arginine (R) was also enriched within RBPs[44]. Interestingly, the stoichiometry and distances between the positively charged residues, as well as the combination of these with other amino acids, were conserved even across non-homologous proteins. Notably, the basic IDRs present in RBPs are similar to motifs in DNA-binding proteins. Here, the basic arms

can alter the DNA-binding of transcription factors by a large capture radius[110]. In this "monkey bar" model, transcription factors utilise their basic arms to reach distant DNA sites by "hopping" and "sliding" instead of 3D diffusion[110]. It is currently unknown whether basic arms may play similar roles in RBPs.

Thus, IDRs could represent malleable, potentially multifunctional RNA-binding motifs. Their RNA binding can range from highly specific to non-selective, and may promote co-folding upon interaction with their target RNAs[104,105,107,111] (**Fig. 4d**). Interestingly, the high sequence constraints of IDRs[40] enable RNA-binding regulation by reversible posttranslational modifications such as acetylation or phosphorylation[44]. In principle, these properties qualify IDRs as versatile modules for interaction with RNA, either alone or in cooperation with globular RBDs.

Shape Complementarity Interactions and Protein-Binding RNAs

Protein-RNA interactions are typically described as a process in which a protein harbours "sensors" (RBDs) to recognize and bind particular sequence and/or structural elements within its "target RNA" (Fig. 1a). However, synthetic RNA aptamers can bind proteins following the same molecular principles as those that enable proteins to bind RNA[112], suggesting that RNA can equally be the driving force mediating protein-RNA interactions (Fig. 1b). The 169 annotated ribosomal proteins follow 119 distinct domain architectures[69] but does this postulate an equal number of distinct RBDs? A more probable explanation appears to be that ribosomal proteins and rRNAs have co-evolved to interact with each other, where extended shape complementarity and the right spatial configuration of molecular interactions play a major role in forming a perfectly packed machinery: the ribosome[8]. Another example of intricate protein-RNA interactions is the spliceosome, where small nuclear (sn)RNAs interact with proteins to form the functionally active complex[11].

These interactions are enabled by the ability of RNA to fold into three-dimensional (3D)-structures resulting in complex surfaces with potential to interact with complementary (both in shape and biochemical properties) protein partners (Fig. 1b and 4e), as illustrated by tRNAs[113].

The fact that RNA can be more than just the "passive" partner in protein-RNA interactions is also illustrated by the 5' UTRs of viral RNAs, which evolved to interact with host proteins. For example, the internal ribosome entry site (IRES) of poliovirus binds the C-terminal moiety of the eukaryotic initiation factor (eIF)4G to recruit the 40S ribosomal subunit[114]. Similarly, the hepatitis C virus (HCV) IRES interacts directly with eIF3 and the 40S ribosomal subunit for translation initiation (**Fig. 4e**). A closer look at the HCV IRES-ribosome co-structure shows that the interaction is not mediated by well-defined protein regions endowed with RNA-binding activity; instead, the protein-RNA interface is large and displays strong shape-complementarity[115,116] (**Fig. 4e**).

Recently, two studies identified widespread RNA binding by chromatin-associated factors and DNA-binding proteins[61,92]. One of the proposed possibilities is that these factors interact with lncRNAs and nascent transcripts[93-95]. Plausibly, the lncRNAs themselves may 'drive' such interactions, and it might make sense to think of these RNAs as displaying protein-binding activity, rather than the other way around (**Fig. 1b**). Functionally, the RNA moiety can play different roles in this interaction, such as scaffolding protein-RNA complexes as illustrated by the viral IRES of HCV[116] (**Fig. 4e**) and the lncRNA nuclear enriched abundant transcript 1 (NEAT1) in paraspeckle formation[117] (**Fig. 4f**), or altering the activity of the bound protein as exemplified by RIGI and PKR and their interaction with intermediaries of viral replication[81-83] (**Fig. 4g**).

RNA Binding by Metabolic Enzymes

RNA interactome studies have persistently identified enzymes of intermediary metabolism as RBPs. Some of these enzyme RBP-RNA interactions appear to serve direct feedback gene regulation. For example, thymidylate synthase (TYMS), 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 the cytosolic aconitase/iron-regulatory protein 1 (IRP1) paradigm. To be active as an enzyme, the protein requires an iron sulfur cluster in its active site, which simultaneously precludes RNA binding[119]. In iron deficiency, this cluster is lacking and, in an open conformation, IRP1 controls the expression of proteins to increase iron uptake and decrease its storage, utilization and export[119,120] (**Fig. 4h**).

Other enzymes display more oblique links to metabolism when acting as RBPs. The glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH) oxidises its substrate to generate NADH, but it also has a diverse range of other cellular functions[121]. An important role for GAPDH as an RBP was discovered in T lymphocyte biology[90]. In resting T cells, which rely on oxidative phosphorylation for energy generation, GAPDH binds to AU-rich elements (AREs) in the 3'UTRs of cytokine mRNAs including γ-interferon mRNA and inhibits their translation. Following the metabolic switch to aerobic glycolysis upon T cell activation, GAPDH disengages from RNA binding, thus de-repressing cytokine production.

The high number of identified RNA-binding metabolic enzymes suggests that not all of these may have 'moonlighting' functions in post-transcriptional gene regulation as described above. Alternatively, (yet to be discovered) RNAs could affect their metabolic function. As discussed in more detail elsewhere[66], this could affect the enzyme's localisation or activity, *e.g.* by affecting an enzymatic side reaction, by allosteric control, or by providing a scaffold that organizes multi-enzyme complexes and even pathways.

Interestingly, the globular Rossmann-fold (R-f) domain has emerged as a common non-conventional RBD. The cardiomyocyte RNA-binding proteome includes an impressive 173 R-f proteins, and 29 of the 73 cardiomyocytic RNA-binding metabolic enzymes harbour at least one R-f domain[46]. The dinucleotide-binding R-f domain of oxidoreductases has long been discussed as a RNA-binding interface[32]. NAD+ and NADH have been reported to interfere with RNA binding by GAPDH, while cytokine mRNA sequences can inhibit the enzymatic activity of GAPDH *in vitro*[89]. Analysis of RBDmap data for 24 metabolic enzymes (including multiple R-f domain proteins) revealed diverse spatial relationships between the identified RNA contacts and previously characterized catalytically relevant regions. While there is a striking overlap for some examples, in other cases the mapped RNA contacts and the catalytic regions did not appear to overlap. Although this could partly reflect falsely negative assignments by RBDmap, the data also implicate possible roles of RNA in allosteric control or enzyme scaffolding[46].

Many enzymes are regulated allosterically by metabolites[66]. Conceivably, such control could also affect their RNA-binding activity. Furthermore, metabolism and metabolites could control enzyme-RNA interactions via 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]. More broadly, RBDmap identified RBDs as hotspots for PTMs, including tyrosine phosphorylation, methylation, acetylation and malonylation[44]. Thus, there is considerable scope for cross-talk between cellular metabolism and RNA binding[66,124].

CHARACTERISATION OF UNORTHODOX RBPs

A key step in characterising 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, either using carefully controlled *in vitro* methods[125] or preserving the context of the living cell[126]. Examples of the latter build on classic approaches such as RNP immunoprecipitation (RIP), potentially augmented by stabilisation of complexes through UV irradiation and/or chemically-induced covalent crosslinks (CLIP). These include photoactivatable-ribonucleoside-enhanced and individual-nucleotide-resolution CLIP (PAR-CLIP[97] and iCLIP[98], respectively). Enhanced CLIP (eCLIP; **Box 1**) is a variant of iCLIP with improved sensitivity and specificity[99,127]. CLIP-type studies have already helped to functionally characterise the biological roles of several unorthodox RBPs. These include metabolic enzymes (*e.g.* HSD17B10)[43], regulators of alternative splicing[128,129], the E3 ubiquitin/ISG15 ligase TRIM25[130,131], the nuclear cap binding protein 3 (NCBP3, previously known as C17orf85)[132], the FAST kinase domain-containing protein 2 (FASTKD2)[133], tropomyosin[134] and others.

The 3-hydroxyacyl-CoA dehydrogenase type-2 (HSD17B10) is a mitochondrial enzyme involved in the oxidation of isoleucine, branched-chain fatty acids, and xenobiotics as well as in the metabolism of sex hormones and neuroactive steroids[135]. 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, and the disease thus may be caused by a non-catalytic function of this protein[136]. A recent report identified HSD17B10 as a component of the mitochondrial RNase P complex, together with mitochondrial ribonuclease P protein 1 (TRMT10C) and 3 (MRPP3)[137]. HSD17B10 was identified as an RBP in several RNA interactomes, indicating that it binds directly to RNA[43]. Investigation of HSD17B10 by iCLIP[98] revealed an enrichment in mitochondrial RNAs[43]. Specifically, HSD17B10 preferentially interacts with the 5' ends of 15 of the 22 mitochondrial tRNAs, binding particularly at the D-stem, D-loop, anticodon stem and loop regions of these tRNAs. The results suggest that a metabolic

enzyme of the di-nucleotide-binding family, HSD17B10, plays a role in guiding RNase P to the ends of the mitochondrial tRNAs. The mutation R130C causes the classical cardiomyopathy and neuropathy phenotype associated with HSD17B10 dysfunction[136]. This mutant exhibits reduced binding to TRMT10C *in vitro*, although it retains the ability to form tetramers[138]. iCLIP data revealed that the R130C mutant also displays reduced RNA-binding activity[43], suggesting that the dysfunctional association with RNA contributes to the disease phenotype.

The HeLa RNA interactome catalogued four (out of six) members of the FAST kinase protein family as RBPs[39]. One of these, the mitochondrial FASTKD2, was recurrently identified in most human and mouse RNA interactomes[38,39,44-47,61]. Analysis of its binding partners by iCLIP revealed that FASTKD2 selectively associates with mitochondrial transcripts[133]. CRISPR-mediated depletion of FASTKD2 causes a strong reduction of its binding target 16S mitochondrial rRNA, in agreement with a recent report that FASTKD2 is important for the assembly of the mitochondrial ribosome[139]. Lack of FASTKD2 function also leads to a reduction in the levels of RNAs that it binds to, including those encoding the complex IV components COX1, COX2 and COX3, *cyb* and *nd6* mRNAs as well as 7S RNA and the prolyl-tRNA[133]. A nonsense mutation in FASTKD2 causes a hereditary neurological disorder[140]. iCLIP and functional assays thus suggest that this disorder is explained by defects in RNA binding causing altered mitochondrial protein synthesis and metabolism[133,139].

C17orf85 was catalogued as an RBP by RNA interactome studies[39], and more recently identified as a cap-binding protein that localizes to nuclear speckles, and renamed as NCBP3[132]. NCBP2 assembles with NCBP1 forming the canonical cap-binding complex (CBC) that binds to nuclear RNAs and plays important roles in RNA processing and export. NCBP1 knock down induced the expected retention of poly(A)$^+$ RNA in the nucleus, and

caused defects in cell proliferation[132]. However, almost no effect was observed when NCBP2 was depleted, suggesting that NCBP2 could be replaced by another nuclear factor enabled with similar cap-binding activity. NCBP3 harbours a predicted RRM and this domain suffices for *in vitro* binding to m^7GTP[132]. Similar to the poly(A)-specific ribonuclease (PARN)[141], NCBP3 interaction with m^7GTP is mediated by a tryptophan and two aspartic acids present at the RRM loops, and the mutation of these residues to alanine impairs the cap-binding activity of NCBP3[132]. Importantly, immunoprecipitation followed by mass spectrometry revealed that, like NCBP2, NCBP3 also interacts with NCBP1 forming part of an alternative nuclear cap-binding complex (CBC)[132].

Several members of the E3 ubiquitin/ISG15 tripartite motif ligase (TRIM) family have recurrently been catalogued as RBPs by RNA interactome studies. This includes TRIM25[39,45], which has additionally been validated as an RBP by orthogonal approaches and its RBD was identified by RBDmap[44,45,130]. TRIM25 harbours an N-terminal zinc finger of the RING family that is important for its ubiquitin ligase activity. Although no canonical RBD can readily be recognized, TRIM25 crosslinks very efficiently to RNA[45]. Recently, TRIM25 has been identified to interact with protein lin-28 homolog A (LIN28A) and terminal uridylyltransferase 4 (TUT4), factors involved in pre-miRNA polyuridylation, enhancing their activity[130]. Because pre-miRNA polyuridylation triggers miRNA decay[142], TRIM25 emerges as a regulator of miRNA biology. TRIM25 also appears to play a role as an RBP in virus-infected cells[131]. TRIM25 was identified as a ubiquitin ligase that triggers β-interferon (β-IFN) through stimulation of the antiviral factor RIGI[143]. Notably, dengue virus 2 (DENV2) genomic RNA is processed by the 5' to 3' exoribonuclease 1 (XRN1) until it gets stalled by a pseudoknot present at its 3' region, leading to the production of a shorter subgenomic RNA enabled with pathogenic activity[144]. A recent report showed that the role of this subgenomic RNA is to sequester TRIM25 preventing its enhancing activity on RIGI and

thus reducing β-IFN production (**Fig. 5a**)[131]. Hence, the interaction of DENV2 subgenomic RNAs with TRIM25 supports its capacity to counteract the antiviral response, implicating the unorthodox RBP TRIM25 in innate immunity against viruses.

Cyclins regulate the cell cycle by activating cyclin-dependent kinases (CDK). They have also sporadically been identified as unconventional RBPs in RNA interactome capture studies[38,45,46,56]. The Cyclins B and T are part of the RNA interactome of *Drosophila* embryos, and Cyclin B RNA-binding activity was validated by CLIP-PNK assay[56]. Cyclins (CCN) A2, L1 and T1 were identified in the nuclear RNA-binding proteome of mouse embryonic stem cells[92]. Mice with reduced CCNA2 expression are prone to tumour formation and chromosomal instability due to a predisposition to form lagging chromosomes and chromatin bridges[145]. This defect resulted from insufficient expression of the meiotic recombination 11 (Mre11) nuclease, apparently due to impaired translation. CCNA2 directly binds to two evolutionarily conserved regions in the 3'UTR of Mre11 mRNA, which appears to be necessary and sufficient for CCNA2 regulation (**Fig. 5b**)[145]. A C-terminal CCNA2 fragment lacking CDK-binding is both necessary and sufficient for RNA-binding, and its expression in mouse embryonic fibroblasts restored appropriate Mre11 synthesis. Interestingly, the CCNA2 RNA-interacting region binds eIF4A2. Taken together, the data identify an unexpected, CDK-independent function of CCNA2 as an RNA-binding protein to promote Mre11 mRNA translation, potentially through an interaction with eIF4A2.

OUTLOOK

What do we have to expect with so many new RBPs to be considered? Some might side with Miranda from Shakespeare's Tempest and marvel at these novel and goodly RBPs that populate the RNA interactome. Others may fear dystopia, as presented by Aldous Huxley's in his novel 'Brave New World', where newly discovered RBPs represent

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 mediates above background interaction with RNA without biological consequence for the protein or the RNAs that it interacts with. Quite remarkably, however, the list of unorthodox RBPs that fulfil professional roles continues to grow, as discussed above. Although better known for other biological functions and lacking in classical 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 RNA-protein interactions structurally. Likewise, it will be important to decipher the mechanisms by which these RNA-protein interactions are regulated, both by epitranscriptomic changes as well as by posttranslational modifications. An intriguing facet of this is the question of whether and how intrinsically disordered regions that bind RNA contribute to the formation of higher order assemblies by liquid-liquid phase separation and to understand the role of RNA in these transitions.

Might unorthodox RBPs be controlled by RNA? We have become accustomed to the view that protein functions are typically modulated by other proteins, but there is ample room for the possibility that the known biological function of a protein be altered by 'riboregulation', a change in the protein's function elicited by interaction with RNA (**Fig. 1b**). The RBPs PKR and RIGI can serve as examples of this class of proteins that could vastly expand. We already know much about RBPs, but future experiments are bound to surprise our intuition.

ACKNOWLEDGEMENTS

We dedicate this review to the memory of Bernd 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 (MWH), Medical Research Council Career Development Award MR/L019434/1 (AC), the EMBL Interdisciplinary Postdocs 2 EIPOD2/291772 program (TS) and the National Health and Medical Research Council APP1120483 (MWH and TP).

Box 1. Technical approaches to study RNA-protein interactions

System-wide identification of RBPs *in vitro*: In one approach, a series of immobilised RNA probes are used as bait, exposed to cellular extracts and binding proteins are identified by quantitative mass spectrometry (Q-MS) (**panel a**)[33]. Applied to combinations of different mRNA UTRs and human HeLa extract, this yielded a dozen proteins that differentially bound to the baits, several of these were novel RBPs. In a more recent iteration, incubation of a set of pre-miRNA baits with multiple different cell lysates yielded ~180 RBPs with distinct specificities[36]. In a second approach, arrayed proteins are provided as bait and incubated with labelled cellular RNA. RNA binding is determined by measuring the fluorescence intensity at each individual protein spot, in analogy to DNA microarrays (**panel b**). Two such proteome-wide screens identified 180[34] and 68[35] yeast RBPs, respectively. In a third approach, purified polyadenylated [poly(A)$^+$] cellular RNA was immobilized on oligo(dT) magnetic beads and, after incubation with cell extract, bound proteins were analysed by Q-MS[35]. 88 mostly highly abundant proteins were identified and 22 were known RBPs.

Identification of *in vivo* RBP repertoires by RNA interactome capture. In this approach, UV crosslinking of cultured cells or organisms covalently links proteins to RNA positioned at "zero distance". This is followed by denaturing cell lysis, collective capture of all RNPs

involving poly(A)$^+$ RNA on oligo(dT) beads, and identification of proteins by quantitative mass spectrometry (Q-MS)[37] **(panel c)**. One study used both conventional crosslinking (cCL) and photoactivatable ribonucleoside-enhanced crosslinking (PAR-CL)[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 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[97]. Another study also used PAR-CL, combining 4SU and 6-thioguanosine (6SG) labelling[38]. This latter study further exploited the U to C transitions occurring as a consequence of the crosslinking between 4SU and the RBP to analyse globally the footprints of the RNA interactome on the RNA. The protocol has been adapted to different model systems[43,48,49,53-57], can be used to monitor differential association of RBPs with RNA under different physiological conditions or in response to biological cues[47,56] as well as to identify RBPs in different subcellular compartments[61].

The eCLIP method. This approach is used to determine the footprints of a given RBP on its targets RNA with single nucleotide resolution[99,127]. UV-irradiation of live cells is followed by cell lysis and limited RNA digestion to fragment RNA. Protein-RNA complexes are immunoprecipitated with an antibody against the RBP under study. Then, the immunoprecipitated material is resolved by denaturing gel electrophoresis, transferred to a membrane. Segments of the membrane corresponding to RNPs are excised. The RNA is recovered and after cDNA library, the RNA regions bound by the RBP are identified by high throughput sequencing **(panel d)**. Since reverse-transcription often stalls at the site of the protein-RNA crosslink, eCLIP affords single nucleotide resolution. TAG-eCLIP broadens the scope of the approach, as it includes a CRISPR/Cas9-mediated insertion of a C-terminal

affinity tag into the endogenous RBP gene, bypassing the need for individual antibodies[147]. Together with several hundred antibodies against known RBPs that have now been tested for immunoprecipitation[74], this provides for a growing list of eCLIP datasets, which are accessible at https://www.encodeproject.org. Among the 122 proteins with available eCLIP data, 34 lack classical RBDs.

Box 2. Epitranscriptomic control of RNA binding. Progress has been swiftest with understanding m^{6}A function in diverse cellular contexts. The YT521-B homology (YTH) domain proteins were identified as m^{6}A 'readers' that specifically bind m^{6}A-containing mRNA regions to impact splicing, export, translation, or turnover[62,63]. The ensuing downstream effects are determined by the specific YTH protein that is recruited and on the mRNA context. The YTH-containing proteins YTHDC1, YTHDC2, YTHDF1, YTHDF2 and YTHDF3 were consistently found in both human and mouse RNA interactome datasets. The RNA-binding proteome of *A. thaliana* contains most plant YTH domain proteins, including cleavage and polyadenylation specificity factor 30 (AtCPSF30), possibly explaining a plant-specific link between m^{6}A and mRNA cleavage[54]. When YTHDF1 is recruited to m^{6}A sites in the 3'UTR of human mRNAs, it enhances their translation, likely through interactions with subunits of the translation initiation factor eIF3[148]**.** Conversely, when YTHDF2 binds to 5'UTR sites in the nucleus of murine embryonic fibroblasts, it blocks their demethylation in heat shock, thus facilitating selective translation in the cytoplasm by direct binding of eIF3 to these m^{6}A sites[149]. More generally, cytoplasmic YTHDF2 binds m^{6}A-containing mRNAs in human cells and promotes their relocation to processing bodies and degradation[150]. The latter function is critically involved in murine embryonic stem cell differentiation[151-153], facilitates maternal mRNA clearance, and the maternal-to-zygotic transition in zebrafish embryos[154]. Epitranscriptomic marks can also modulate RBP binding

indirectly by affecting RNA structure. For example, the nuclear protein HNRNPC responds to m^6A-operated 'structural switches' to gain access to thousands of its target sites in human nuclear RNAs[155]. HNRNPC preferentially binds to single-stranded U tracts, and m^6A can destabilise local RNA structure, making the U-tracts more accessible. Thus, when considering the determinants of RNP formation, epitranscriptomic changes need to be considered in addition to posttranslational RBP modifications.

FIGURES & LEGENDS

Figure 1. Functional crosstalk between proteins and RNA. Schematic representation of the biological consequences of protein-RNA interactions. (a) RNA-binding proteins (RBP) interact with RNA through defined RNA-binding domains to regulate various aspects of the RNA's function. (b) Inversely, the RNA binds to the RBP to affect the protein's fate.

Figure 2. Comparison of published RNA interactomes. The reported lists applying the most stringent criteria for RBPs identified by different studies and from various source materials were extracted, curated and their annotations were updated, and converted to base identifiers. The underlying data are provided in **Table S1**. (a) Supersets of RBPs identified by the combination of RBP detection studies in a given organism. These were compiled for *Mus musculus* (Mm), *Homo sapiens* (Hs), *Saccharomyces cerevisiae* (Sc), *Drosophila melanogaster* (Dm), *Arabidopsis thaliana* (At) and *Caenorhabditis elegans* (Ce). RBP sets for *Danio rerio* (Dr), *Trypanosoma brucei* (Tb), *Leishmania donovani* (Ld) and *Plasmodium falciparum* (Pf) were also added. (b-h) Venn Diagrams and UpSet plots[156] showing overlaps, set and intersections sizes between different RBP sets. Differences in technical approaches and data processing (for example, thresholds for detection and reproducibility) will affect these comparisons. (b) RBP repertoires as detected in human source material. RNA interactome capture (RIC; **Box 1**) with either conventional UV crosslinking (cCL) or photoactivatable ribonucleoside-enhanced crosslinking (PAR-CL) was applied to the following cell lines: cervical cancer HeLa[39], embryonic kidney HEK293[38,39], hepatocytic HuH7[43] and myeloid leukaemia K562 (serial nuclear RIC)[61]. HeLa cells were further subjected to RBDmap[44] and RNP^{xl}[42]. The human datasets show very high overlap, likely because of the preponderance of 'generic' cell lines as source material and related

experimental approaches. (c) Murine RBP repertoires. RIC was applied to primary embryonic fibroblasts (MEF ±etoposide)[84], embryonic stem cells (mESC, total cell and nuclear fraction), macrophages (RAW264.7 ±lipopolysaccharide)[47]. HL-1 cardiomyocytes were subjected to both RIC and RBDmap[44]. These datasets overlap less well, probably due to the use of more 'idiosyncratic' cell lines and inclusion of drug treatments. (d) Budding yeast RNA-binding proteomes. Two studies used either *in vitro* protein arrays or, in one case, oligo(dT) capture screens to identify RBPs[34],[35]. Three *in vivo* RNA interactomes were generated, either employing conventional UV crosslinking (cCL)[48,49] or photoactivatable ribonucleoside-enhanced crosslinking (PAR-CL)[43]. RNP^{xl} was used with two crosslinking approaches[42]. The diversity of technical approaches likely explains the big differences in coverage and overlap. (e) Fruit fly RBP sets. Two studies applied RNA interactome capture to *Drosophila melanogaster* embryos, using solely cCL[56] or both cCL and PAR-CL[57]. The focus here were embryos undergoing maternal-to-zygotic transition. This, together with differences in mass spectrometry approaches likely underlies the moderate overlap. (f) RNA interactomes captured from different plant sources, including cell suspension cultures and leaves[53], etiolated seedlings[54] and leaf mesophyll protoplasts[55]. Given the heterogeneous sources, the three datasets agree reasonably well with each other. The lower RBP identification rates suggest UV crosslinking limitations, likely due to the presence of a cell wall and/or UV-absorbing pigments. (g) Pairwise comparisons of InParanoid clusters between human (Hs), mouse (Mm) and yeast (Sc). The intersections between these sets constitute emerging mammalian or eukaryotic 'core' interactomes. (h) UpSet plot showing the overlap between the human superset of RBPs and human orthologues from *Mus musculus* (Mm), *Saccharomyces cerevisiae* (Se), *Drosophila melanogaster* (Dm), *Caenorhabditis elegans* (Ce) and *Arabidopsis thaliana* (At).

Figure 3. System-wide, high resolution identification of RNA-binding domains. Each of the three methods uses UV irradiation of live cells to establish covalent bonds at direct contact sites between RNA and protein (indicated by a star). Following cell lysis, the approaches diverge with regard to proteolysis, purification strategy and detection by quantitative mass spectrometry (Q-MS). (a) Purification and direct detection of RNA-crosslinked tryptic peptides[42,88]. Although simple in principle, this strategy is challenging due to the inefficiency of UV crosslinking and the heterogeneous mass contribution by the nucleic acid remnant, resulting in sub-stoichiometric amounts of peptides with hard-to-predict additional mass. To overcome this, covalently linked protein-RNA complexes are purified on oligo(dT) beads, by performing the initial steps of RNA interactome capture[37,43]. After digestion with trypsin and RNases, peptides crosslinked to remnants of RNA are further enriched using a TiO_2 matrix[157], prior to analysis by Q-MS and a complex search, performed by a custom-designed software termed RNP^{xl}, for peptide spectra with a defined mass shift caused by the nucleotide remnant. (b) Extrapolation of RNA-protein crosslink sites by RBDmap[44]. Again, RBPs are purified with $poly(A)^+$ RNA as per RNA interactome capture[37,43], but then digested with a protease that cleaves every 17 amino acids on average (LysC or ArgC), typically leaving peptides that still contain an internal trypsin cleavage site. RNA-linked peptides, termed RBDpep, are then recaptured on oligo(dT) beads, while those distant to the crosslink site are released into the supernatant. Both fractions are digested with trypsin and analysed by Q-MS. For RNA-bound material this leads to: i) a fragment with remnant RNA that will not be identified; and ii) a neighbouring fragment(s) with native mass (N-peptide). N-peptides that are enriched in the RNA-bound fraction are extended *in silico* to the next LysC/ArgC cleavage site to reconstitute the original RBDpep. (c) Proteomic identification of RNA-binding regions (RBR-ID)[92]. This approach directly exploits the mass shift of RNA-crosslinked peptides in conventional MS

analyses, by assigning RNA-binding activity to tryptic peptides (~9 amino acid length on average) with reproducibly reduced ion counts (peptide intensity) in UV-irradiated samples compared to non-irradiated controls.

The three approaches have some common limitations. UV crosslinking relies on a favourable geometry, nucleotide and amino acid composition at protein-RNA interfaces. Protein-RNA interactions with the phosphate backbone, for example, are non-optimal in this regard, and will be missed. Mass-spectrometric analyses are also influenced by the abundance of the peptides and their amino acid sequence (influencing optimal peptide size generated derived from tryptic cleavage and optimal 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 is 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, its high intra- and inter-experimental variability call for a sufficiently high number of technical and biological replicates to obtain high confidence results.

Figure 4. Modes of RNA binding. (a) Sche*matic rep*resentation of a RBP harbouring a classic RBD (RRM) interacting specifically with a specific RNA sequence in the context of a stem-loop[2]. (b) Depiction of the eukaryotic initiation factor (eIF)4F, which is composed of the cap-binding protein eIF4E, eIF4G and the helicase eIF4A. This complex associates with capped RNA in a sequence-independent manner to enable global initiation of translation[158]. (c) The exon junction complex is deposited by its interaction with CWC22 ~20 nucleotides upstream the exon-exon junction immediately after intron removal[102]. (d) The disordered RGG-motif present in FMR1 protein co-folds with its target RNA forming a tight electrostatic and shape-complementarity 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[116]. (f) The lncRNA NEAT1 hijacks the RBPs NONO, PSC1 and SFPQ in paraspeckles[117]. (g) PKR binds to double-stranded (ds)RNA, derived from viral replication. RNA binding promotes protein dimerization and autophosphorylation, activating PKR. In its active form, PKR phosphorylates eIF2α to block protein synthesis in infected cells[159]. (h) Aconitase 1 (ACO1) associates with an iron-sulphur cluster to catalyse the interconversion between citrate and isocitrate. Under low iron conditions, the iron-sulphur cluster is no longer synthesised and the apoprotein, iron regulatory protein 1 (IRP1), binds mRNAs encoding cellular factors involved in iron homeostasis, regulating their fate[119].

Figure 5. Examples of the biological roles of unorthodox RBPs. (a) Schematic representation of (a) TRIM25 hijacking by the subgenomic RNA of DENV to reduce β-interferon (β-IFN) synthesis[131]. DENV genomic RNA degradation by XRN1 leads to the generation of a subgenomic RNA due to XRN1 stalling at a pseudoknot present at the 3' region of this RNA. The subgenomic DENV RNA recruits TRIM25 but not RIGI, which requires the presence of tri-phosphate 5' ends to interact, thus hijacking TRIM25 in RNPs lacking its molecular partner in the β-IFN pathway, RIGI. (b) Cdk-independent function of CCNA2 as an RNA-binding protein. CCNA2 directly binds to two evolutionarily conserved regions in the 3'UTR of Mre11 mRNA and can to promote Mre11 mRNA translation, potentially through an interaction with eIF4A2[145].

GLOSSARY

Ribonucleoprotein (RNP): A complex formed between one or several RNAs and proteins.

RNA-binding domain (RBD): A subdomain within a protein that mediates direct interaction with RNA.

RNA-recognition motif (RRM): One of the most common RBDs, consisting of ~90 amino acids that fold into two α-helices packed against a four-stranded β-sheet. Typically, the residues in the β–sheet interact with RNA.

HnRNP K homology (KH) domain: Another common RBD, consisting of ~70 amino acids that fold 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 protein: RNA helicases exhibiting two highly similar domains that resemble the bacterial recombinase A. ATP binds in a cleft between the two helicase domains, while RNA binds across both domains on the opposite side. The DEAD box is one of nine conserved sequence motifs in this class of proteins and contains the amino acid sequence Asp-Glu-Ala-Asp (DEAD).

Epitranscriptome: A collective term for all chemically diverse RNA modifications that exist within a transcriptome. In analogy to the epigenome, the term further implies that many such modifications (e.g. methylation, pseudouridylation) are deliberately placed and serve regulatory roles.

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

Processing (P-)bodies: Microscopically visible foci present in the cytoplasm of eukaryotic cells. P-bodies contain mRNAs and many components of mRNA silencing and turnover.

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

Liquid-liquid phase separations (LLPS): A (bio-)physical process whereby non-membrane-bound compartments (e.g. SGs, P-bodies) are formed as phase-separated, liquid-like droplets within cells.

Intrinsically disordered regions (IDRs): Features within a native protein that lack stable secondary or tertiary structure and thus appear unfolded.

Long noncoding RNAs (lncRNAs): RNAs longer than 200 nucleotides without ascribed protein-coding potential. The arbitrary length criterium is used to distinguish lncRNAs from small noncoding RNA types, such as e.g. microRNAs or tRNAs.

UV crosslinking: A method using ultraviolet light irradiation to covalently connect proteins and RNA when positioned in very close proximity ('zero distance') to each other. UV crosslinking can be applied *in vitro* or in living cells.

Zinc finger domain: A protein domain involving cysteines (C) and histidines (H) that coordinates zinc cations. Zinc finger domains are classified based on the order and frequency of C and H, and can mediate interactions with DNA, RNA or proteins depending on the subclass.

InParanoid analyses: Method for detecting orthologues and in-paralog clusters across different, often distant species.

BioPlex PPI dataset: Comprehensive collection of protein-protein interaction networks generated by experimental approaches.

Electrophoretic mobility shift assay (EMSA): A method to study protein interactions with nucleic acids *in vitro*. Cell extracts or purified proteins are incubated with radiolabelled (or fluorescently labelled) nucleic acids, and the resulting complexes are resolved through a native gel. If the nucleic acid is bound by a protein, the protein will retard its mobility through the gel compared to the unbound probe.

G-quadruplex: RNA structure involving two or more stacks formed by a planar array of four guanine bases (G). It requires the coordination of a monovalent cation of an appropriate radius, such as potassium.

RNA aptamer: Relatively short and often highly folded short RNA molecule that was experimentally selected for specific, high affinity interactions with proteins or other molecules.

RNase P complex: Ribonuclease complex for processing of precursor t-RNA

Rossmann-fold (R-f): A protein domain characterised by the presence of up to seven mostly parallel β-strands combined with connecting α-helices. This domain is typically found in proteins that bind nucleotides, such as metabolic enzymes with di-nucleotide (NAD, NADP, FAD) binding activity.

Uridylation: A biochemical reaction mediated by uridylyltransferases (e.g. TUTase) that involves the addition of multiple uridines to the 3' end of RNA molecules. This modification is typically a signal for RNA degradation.

REFERENCES

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., Loerke, J., Budkevich, T. V., Yamamoto, K., Schmidt, A., Penczek, P. A., . . . Spahn, C. M. 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*, (2017).
27 Rissland, O. S. The organization and regulation of mRNA-protein complexes. *Wiley Interdiscip Rev RNA* **8**, (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 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).

30 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).
31 Chu, E., Voeller, D., Koeller, D. M., Drake, J. C., Takimoto, C. H., Maley, G. F., . . . Allegra, C. J. Identification of an RNA binding site for human thymidylate synthase. *Proc Natl Acad Sci U S A* **90**, 517-521, (1993).
32 Hentze, M. W. Enzymes as RNA-binding proteins: a role for (di)nucleotide-binding domains? *Trends Biochem Sci* **19**, 101-103, (1994).
33 Butter, F., Scheibe, M., Morl, M. & Mann, M. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci U S A* **106**, 10626-10631, (2009).
34 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).
35 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).
36 Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiss, J. L., Eichner, N., . . . Meister, G. A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis. *Mol Cell* **66**, 270-284 e213, (2017).
37 Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L. M., . . . Hentze, M. W. System-wide identification of RNA-binding proteins by interactome capture. *Nat Protoc* **8**, 491-500, (2013).
38 Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., . . . Landthaler, M. The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol Cell* **46**, 674-690, (2012).
39 Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., . . . Hentze, M. W. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393-1406, (2012).
40 Castello, A., Fischer, B., Hentze, M. W. & Preiss, T. RNA-binding proteins in Mendelian disease. *Trends Genet* **29**, 318-327, (2013).
41 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).
42 Kramer, K., Sachsenberg, T., Beckmann, B. M., Qamar, S., Boon, K. L., Hentze, M. W., . . . Urlaub, H. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat Methods* **11**, 1064-1070, (2014).
43 Beckmann, B. M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A. M., . . . Hentze, M. W. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat Commun* **6**, 10127, (2015).
44 Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A. M., Foehr, S., . . . Hentze, M. W. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell* **63**, 696-710, (2016).
45 Kwon, S. C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K. T., . . . Kim, V. N. The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol* **20**, 1122–1130, (2013).
46 Liao, Y., Castello, A., Fischer, B., Leicht, S., Foehr, S., Frese, C. K., . . . Preiss, T. The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. *Cell Rep* **16**, 1456-1469, (2016).

47 Liepelt, A., Naarmann-de Vries, I. S., Simons, N., Eichelbaum, K., Foehr, S., Archer, S. K., . . . Ostareck-Lederer, A. Identification of RNA-binding proteins in macrophages by interactome capture. *Mol Cell Proteomics* **15**, 2699-2674, (2016).
48 Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat Struct Mol Biol* **20**, 127-133, (2013).
49 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).
50 Nandan, D., Thomas, S. A., Nguyen, A., Moon, K. M., Foster, L. J. & Reiner, N. E. Comprehensive Identification of mRNA-Binding Proteins of Leishmania donovani by Interactome Capture. *PLoS One* **12**, e0170068, (2017).
51 Bunnik, E. M., Batugedara, G., Saraf, A., Prudhomme, J., Florens, L. & Le Roch, K. G. The mRNA-bound proteome of the human malaria parasite Plasmodium falciparum. *Genome Biol* **17**, 147, (2016).
52 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).
53 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).
54 Reichel, M., Liao, Y., Rettel, M., Ragan, C., Evers, M., Alleaume, A. M., . . . Millar, A. A. In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. *Plant Cell* **28**, 2435-2452, (2016).
55 Zhang, Z., Boonen, K., Ferrari, P., Schoofs, L., Janssens, E., van Noort, V., . . . Geuten, K. UV crosslinked mRNA-binding proteins captured from leaf mesophyll protoplasts. *Plant Methods* **12**, 42, (2016).
56 Sysoev, V. O., Fischer, B., Frese, C. K., Gupta, I., Krijgsveld, J., Hentze, M. W., . . . Ephrussi, A. Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *Nat Commun* **7**, 12128, (2016).
57 Wessels, H. H., Imami, K., Baltz, A. G., Kolinski, M., Beldovskaya, A., Selbach, M., . . . Landthaler, M. The mRNA-bound proteome of the early fly embryo. *Genome Res* **26**, 1000-1009, (2016).
58 Despic, V., Dejung, M., Gu, M., Krishnan, J., Zhang, J., Herzel, L., . . . Neugebauer, K. M. Dynamic RNA-protein interactions underlie the zebrafish maternal-to-zygotic transition. *Genome Res*, (2017).
59 Sonnhammer, E. L. & Ostlund, G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res* **43**, D234-239, (2015).
60 O'Brien, K. P., Remm, M. & Sonnhammer, E. L. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res* **33**, D476-480, (2005).
61 Conrad, T., Albrecht, A. S., de Melo Costa, V. R., Sauer, S., Meierhofer, D. & Orom, U. A. Serial interactome capture of the human cell nucleus. *Nat Commun* **7**, 11212, (2016).
62 Gilbert, W. V., Bell, T. A. & Schaening, C. Messenger RNA modifications: Form, distribution, and function. *Science* **352**, 1408-1412, (2016).
63 Hoernes, T. P. & Erlacher, M. D. Translating the epitranscriptome. *Wiley Interdiscip Rev RNA* **8**, (2017).
64 Ciesla, J. Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? *Acta Biochim Pol* **53**, 11-32, (2006).
65 Hentze, M. W. & Preiss, T. The REM phase of gene regulation. *Trends Biochem Sci* **35**, 423-426, (2010).

66 Castello, A., Hentze, M. W. & Preiss, T. Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. *Trends Endocrinol Metab* **26**, 746-757, (2015).
67 Chu, E. & Allegra, C. J. The role of thymidylate synthase as an RNA binding protein. *Bioessays* **18**, 191-198, (1996).
68 Liu, J., Schmitz, J. C., Lin, X., Tai, N., Yan, W., Farrell, M., . . . Chu, E. Thymidylate synthase as a translational regulator of cellular gene expression. *Biochim Biophys Acta* **1587**, 174-182, (2002).
69 Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829-845, (2014).
70 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 40 S ribosomal subunit. *J Biol Chem* **282**, 8165-8174, (2007).
71 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).
72 Klass, D. M., Scheibe, M., Butter, F., Hogan, G. J., Mann, M. & Brown, P. O. Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in Saccharomyces cerevisiae. *Genome Res* **23**, 1028-1038, (2013).
73 Brannan, K. W., Jin, W., Huelga, S. C., Banks, C. A., Gilmore, J. M., Florens, L., . . . Yeo, G. W. SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. *Mol Cell* **64**, 282-293, (2016).
74 Sundararaman, B., Zhan, L., Blue, S. M., Stanton, R., Elkins, K., Olson, S., . . . Yeo, G. W. Resources for the Comprehensive Discovery of Functional RNA Elements. *Mol Cell* **61**, 903-913, (2016).
75 Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., . . . Gygi, S. P. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425-440, (2015).
76 Mukhopadhyay, R., Ray, P. S., Arif, A., Brady, A. K., Kinter, M. & Fox, P. L. DAPK-ZIPK-L13a axis constitutes a negative-feedback module regulating inflammatory gene expression. *Mol Cell* **32**, 371-382, (2008).
77 Arif, A., Jia, J., Mukhopadhyay, R., Willard, B., Kinter, M. & Fox, P. L. Two-site phosphorylation of EPRS coordinates multimodal regulation of noncanonical translational control activity. *Mol Cell* **35**, 164-180, (2009).
78 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 U S A* **108**, 1415-1420, (2011).
79 Clingman, C. C., Deveau, L. M., Hay, S. A., Genga, R. M., Shandilya, S. M., Massi, F. & Ryder, S. P. Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. *eLife*, e02848, (2014).
80 Haghighat, A. & Sonenberg, N. eIF4G dramatically enhances the binding of eIF4E to the mRNA 5'-cap structure. *J Biol Chem* **272**, 21677-21680, (1997).
81 Dabo, S. & Meurs, E. F. dsRNA-dependent protein kinase PKR and its role in stress, signaling and HCV infection. *Viruses* **4**, 2598-2635, (2012).
82 Habjan, M. & Pichlmair, A. Cytoplasmic sensing of viral nucleic acids. *Curr Opin Virol* **11**, 31-37, (2015).
83 Rehwinkel, J. RNA sensing: the more RIG-I the merrier? *EMBO Rep* **14**, 751-752, (2013).

84 Boucas, J., Fritz, C., Schmitt, A., Riabinska, A., Thelen, L., Peifer, M., . . . Reinhardt, H. C. 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).
85 Lasko, P. Posttranscriptional regulation in Drosophila oocytes and early embryos. *Wiley Interdiscip Rev RNA* **2**, 408-416, (2011).
86 Altelaar, A. F., Frese, C. K., Preisinger, C., Hennrich, M. L., Schram, A. W., Timmers, H. T., . . . Mohammed, S. Benchmarking stable isotope labeling based quantitative proteomics. *J Proteomics* **88**, 14-26, (2013).
87 Fernandez-Chamorro, J., Pineiro, D., Gordon, J. M., Ramajo, J., Francisco-Velilla, R., Macias, M. J. & Martinez-Salas, E. 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 Proteomics* **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., Curtis, J. D., Maggi, L. B., Jr., Faubert, B., Villarino, A. V., O'Sullivan, D., . . . Pearce, E. L. 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., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B. A. & Bonasio, R. High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Mol Cell* **64**, 416-430, (2016).
93 Kaneko, S., Li, G., Son, J., Xu, C. F., Margueron, R., Neubert, T. A. & Reinberg, D. 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., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., . . . Hughes, T. R. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172-177, (2013).
97 Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., . . . Tuschl, T. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141, (2010).
98 König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., . . . Ule, J. 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., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., . . . Yeo, G. W. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508-514, (2016).

100 Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P. A. & Burge, C. B. 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., Huppertz, I., Attig, J., Konig, J., Wang, Z., Hauer, C., . . . Ule, J. 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., Kuryavyi, V., Darnell, J. C., Serganov, A., Majumdar, A., Ilin, S., . . . Patel, D. J. 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., Kato, M., Xie, S., Wu, L. C., Mirzaei, H., Pei, J., . . . McKnight, S. L. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768-779, (2012).
109 Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., . . . McKnight, S. L. 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., Janjic, N., Jarvis, T., Schneider, D., Walker, J. J., Wilcox, S. K. & Zichi, D. Aptamers and the RNA world, past and present. *Cold Spring Harb Perspect Biol* **4**, (2012).
113 Shi, H. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA* **6**, 1091-1105, (2000).
114 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 U S A* **106**, 9197-9202, (2009).
115 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).
116 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).
117 Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A. & Lawrence, J. B. 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., Koeller, D. M., Casey, J. L., Drake, J. C., Chabner, B. A., Elwood, P. C., . . . Allegra, C. J. Autoregulation of human thymidylate synthase messenger RNA translation by thymidylate synthase. *Proc Natl Acad Sci U S A* **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., Selezneva, A. I., Dupuy, J., Volbeda, A., Fontecilla-Camps, J. C., Theil, E. C. & Volz, K. 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., Redondo-Horcajo, M., Magan-Marchal, N., Lagares, D., Martinez-Ruiz, A., Kleinert, H. & Lamas, S. 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., Van, F., P., Pratt, G. A., Wang, X., Wei, X., Blue, S. M. & Yeo, G. W. 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., Nowak, J. S., Zuo, J., Rappsilber, J., Spoel, S. H. & Michlewski, G. Trim25 Is an RNA-Specific Activator of Lin28a/TuT4-Mediated Uridylation. *Cell Rep* **9**, 1265-1272, (2014).
131 Manokaran, G., Finol, E., Wang, C., Gunaratne, J., Bahl, J., Ong, E. Z., . . . Ooi, E. E. Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness. *Science* **350**, 217-221, (2015).
132 Gebhardt, A., Habjan, M., Benda, C., Meiler, A., Haas, D. A., Hein, M. Y., . . . Pichlmair, A. mRNA export through an additional cap-binding complex consisting of NCBP1 and NCBP3. *Nat Commun* **6**, 8192, (2015).
133 Popow, J., Alleaume, A. M., Curk, T., Schwarzl, T., Sauer, S. & Hentze, M. W. FASTKD2 is an RNA-binding protein required for mitochondrial RNA processing and translation. *RNA* **21**, 1873-1884, (2015).
134 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).

135 Yang, S. Y., He, X. Y., Isaacs, C., Dobkin, C., Miller, D. & Philipp, M. Roles of 17beta-hydroxysteroid dehydrogenase type 10 in neurodegenerative disorders. *J Steroid Biochem Mol Biol* **143**, 460-472, (2014).
136 Rauschenberger, K., Scholer, K., Sass, J. O., Sauer, S., Djuric, Z., Rumig, C., . . . Zschocke, J. 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).
137 Holzmann, J., Frank, P., Loffler, E., Bennett, K. L., Gerner, C. & Rossmanith, W. RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135**, 462-474, (2008).
138 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).
139 Antonicka, H. & Shoubridge, E. A. Mitochondrial RNA Granules Are Centers for Posttranscriptional RNA Processing and Ribosome Biogenesis. *Cell Rep*, (2015).
140 Ghezzi, D., Saada, A., D'Adamo, P., Fernandez-Vizarra, E., Gasparini, P., Tiranti, V., . . . Zeviani, M. FASTKD2 nonsense mutation in an infantile mitochondrial encephalomyopathy associated with cytochrome c oxidase deficiency. *Am J Hum Genet* **83**, 415-423, (2008).
141 Nagata, T., Suzuki, S., Endo, R., Shirouzu, M., Terada, T., Inoue, M., . . . Yokoyama, S. The RRM domain of poly(A)-specific ribonuclease has a noncanonical binding site for mRNA cap analog recognition. *Nucleic Acids Res* **36**, 4754-4767, (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., Shin, Y. C., Joo, C. H., Urano, T., Liang, C., Sun, L., . . . Jung, J. U. TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **446**, 916-920, (2007).
144 Chapman, E. G., Costantino, D. A., Rabe, J. L., Moon, S. L., Wilusz, J., Nix, J. C. & Kieft, J. S. The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science* **344**, 307-310, (2014).
145 Kanakkanthara, A., Jeganathan, K. B., Limzerwala, J. F., Baker, D. J., Hamada, M., Nam, H. J., . . . van Deursen, J. M. Cyclin A2 is an RNA binding protein that controls Mre11 mRNA translation. *Science* **353**, 1549-1552, (2016).
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., Gelboin-Burkhart, C., Wang, R., Pratt, G. A., Blue, S. M. & Yeo, G. W. CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA binding proteins. *Methods* **118**, 50-59, (2016).
148 Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., . . . He, C. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388-1399, (2015).
149 Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S. R. & Qian, S. B. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591-594, (2015).
150 Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., . . . He, C. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-120, (2014).
151 Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z. & Zhao, J. C. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**, 191-198, (2014).

152 Batista, P. J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., . . . Chang, H. Y. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707-719, (2014).
153 Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., . . . Hanna, J. H. Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* **347**, 1002-1006, (2015).
154 Zhao, B. S., Wang, X., Beadell, A. V., Lu, Z., Shi, H., Kuuspalu, A., . . . He, C. m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **542**, 475-478, (2017).
155 Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M. & Pan, T. 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 Altelaar, A. F., Munoz, J. & Heck, A. J. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* **14**, 35-48, (2013).
158 Prevot, D., Darlix, J. L. & Ohlmann, T. Conducting the initiation of protein synthesis: the role of eIF4G. *Biol Cell* **95**, 141-156, (2003).
159 Balachandran, S. & Barber, G. N. PKR in innate immunity, cancer, and viral oncolysis. *Methods Mol Biol* **383**, 277-301, (2007).

SUPPLEMENTARY INFORMATION

Supplementary Figure 1. Comparison of RNA interactome capture data with in silico methods to identify RNA-binding proteins (RBPs). Venn diagrams showing overlaps of the human RNA interactome datasets with a curated list of RBPs[1] (a) or with SONAR-identified RBPs[2] (b). (c) Venn diagram showing the overlap between the *Saccharomyces cerevisiae* RNA interactome superset[3-5] with the protein array-based, in vitro approach to identify RBPs[6,7] and SONAR-identified RBPs in yeast[2]. (d) Functional annotation of the mouse, human and yeast RBP supersets (all identified RBPs in mouse (Mm)[8-12], human (H)[4,13-17] and yeast (Sc) studies) [3-7,15]. Domain classifications [2,14] were combined to annotate proteins with known RNA-binding domains. Current versions of GO terms include the RBPs of early interactomes[13,14].

Supplementary Table 1. Curated table of RBP supersets identified in *Homo sapiens* [4,13-17], *Mus musculus*)[8,9,11,12], *Saccharomyces cerevisiae*[3-5], *Drosophila melanogaster*[18,19], *Arabidopsis thaliana*[20-22] and *Caenorhabditis elegans*[5] using RNA interactome capture. RNA-binding proteomes generated with RNA interactome capture are compiled in Table S1. Due to ambiguous or outdated identifier (ID) annotations, our reported protein numbers, which are based on updated annotations and curation, occasionally differ from those in the original publications. We used the most stringent cutoffs provided by the original publications. Also note that the current gene ontology (GO) term “RNA binding” already includes the initial RNA interactome studies[13,14]. For human, mouse and yeast, ‘Metabolism’ has been annotated as ‘true’, when the respective RBP was listed under ‘metabolism’ in ‘Reactome’ (version 60). The ‘Enzyme’ annotation is ‘true’ when the protein is listed in in the six enzyme commission groups (EC 1-6) from the IntEnz database (release May 2017). “Metabolic enzyme” is “true” when the RBP fulfils our definition of metabolic enzyme. ‘Metabolic enzyme’ is an ambiguous term to define, and different authors have applied different criteria. For the purpose of this review, we generated a working list by mapping RBPs to “Metabolism” in the Reactome database (version 60)[23]. The resulting proteins were manually curated by removing genes whose protein products are not classified in the six enzyme commission groups (EC 1-6) from the IntEnz database (release May 2017)[24], but retaining the subunits of ATP synthase and the respiratory chain complexes on the list of metabolic enzymes.

REFERENCES

1 Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829-845, (2014).
2 Brannan, K. W., Jin, W., Huelga, S. C., Banks, C. A., Gilmore, J. M., Florens, L., . . . Yeo, G. W. SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. *Mol Cell* **64**, 282-293, (2016).
3 Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat Struct Mol Biol* **20**, 127-133, (2013).
4 Beckmann, B. M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A. M., . . . Hentze, M. W. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat Commun* **6**, 10127, (2015).
5 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).
6 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).
7 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).
8 Kwon, S. C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K. T., . . . Kim, V. N. The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol* **20**, 1122–1130, (2013).
9 Boucas, J., Fritz, C., Schmitt, A., Riabinska, A., Thelen, L., Peifer, M., . . . Reinhardt, H. C. 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).
10 He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B. A. & Bonasio, R. High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Mol Cell* **64**, 416-430, (2016).
11 Liao, Y., Castello, A., Fischer, B., Leicht, S., Foehr, S., Frese, C. K., . . . Preiss, T. The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. *Cell Rep* **16**, 1456-1469, (2016).
12 Liepelt, A., Naarmann-de Vries, I. S., Simons, N., Eichelbaum, K., Foehr, S., Archer, S. K., . . . Ostareck-Lederer, A. Identification of RNA-binding proteins in macrophages by interactome capture. *Mol Cell Proteomics* **15**, 2699-2674, (2016).
13 Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., . . . Landthaler, M. The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol Cell* **46**, 674-690, (2012).
14 Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., . . . Hentze, M. W. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393-1406, (2012).
15 Kramer, K., Sachsenberg, T., Beckmann, B. M., Qamar, S., Boon, K. L., Hentze, M. W., . . . Urlaub, H. Photo-cross-linking and high-resolution

mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat Methods* **11**, 1064-1070, (2014).

16 Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A. M., Foehr, S., . . . Hentze, M. W. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell* **63**, 696-710, (2016).

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

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

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

20 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).

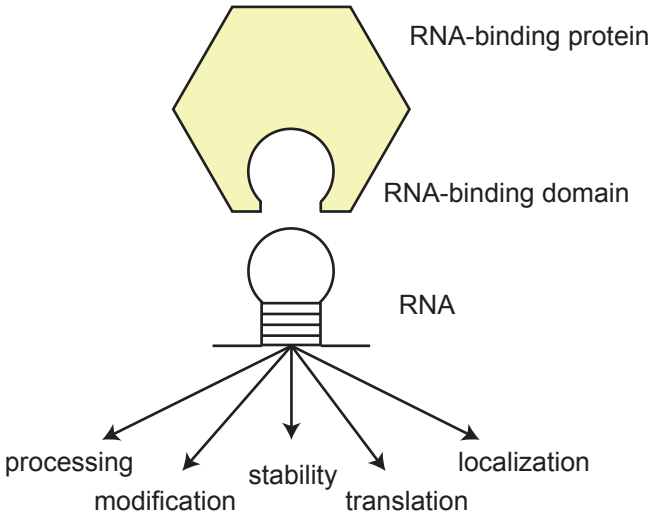
21 Reichel, M., Liao, Y., Rettel, M., Ragan, C., Evers, M., Alleaume, A. M., . . . Millar, A. A. In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. *Plant Cell* **28**, 2435-2452, (2016).

22 Zhang, Z., Boonen, K., Ferrari, P., Schoofs, L., Janssens, E., van Noort, V., . . . Geuten, K. UV crosslinked mRNA-binding proteins captured from leaf mesophyll protoplasts. *Plant Methods* **12**, 42, (2016).

23 Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., . . . D'Eustachio, P. The Reactome pathway knowledgebase. *Nucleic Acids Res* **42**, D472-477, (2014).

24 Fleischmann, A., Darsow, M., Degtyarenko, K., Fleischmann, W., Boyce, S., Axelsen, K. B., . . . Apweiler, R. IntEnz, the integrated relational enzyme database. *Nucleic Acids Res* **32**, D434-437, (2004).

a RBP acting on RNA



b RNA acting on RBP

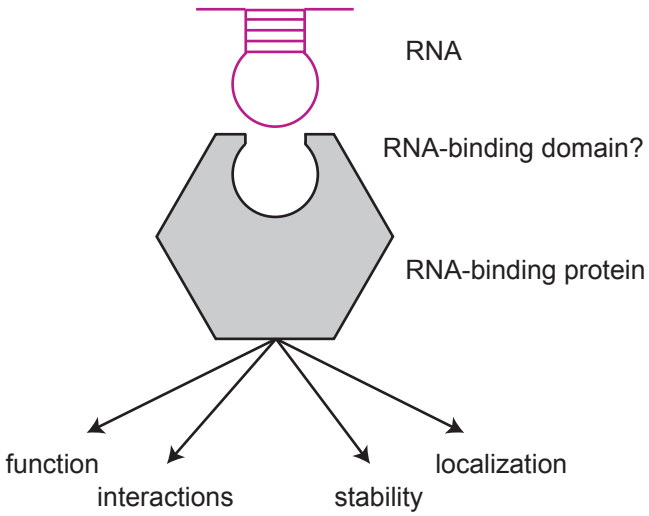


Figure 1

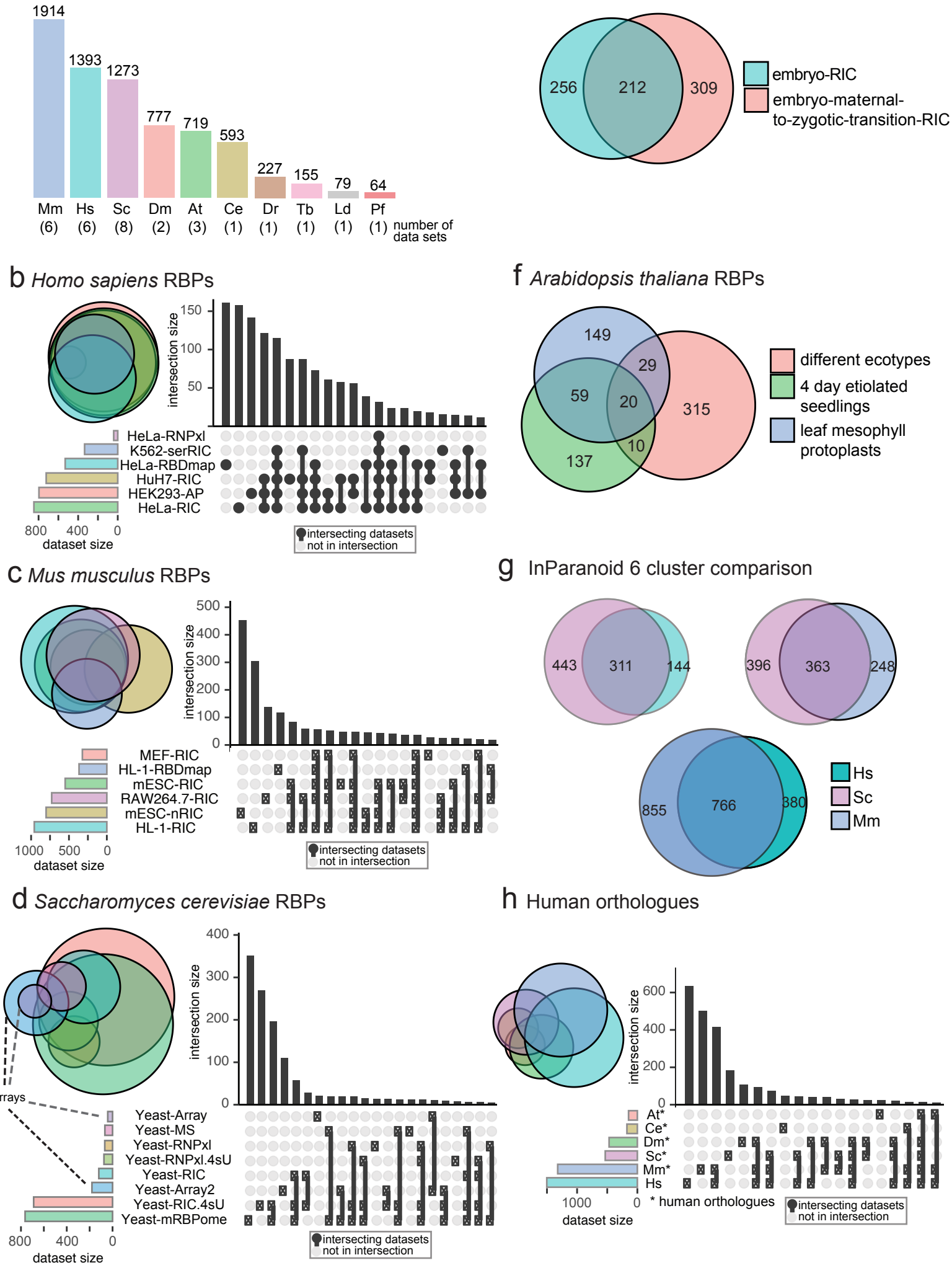

a Identified RNA−binding proteins
1914
1393
1273
777
719
593
227
155
79
64
Mm (6)
Hs (6)
Sc (8)
Dm (2)
At (3)
Ce (1)
Dr (1)
Tb (1)
Ld (1)
Pf (1)
number of data sets
b Homo sapiens RBPs
intersection size
150
100
50
HeLa-RNPxl
K562-serRIC
HeLa-RBDmap
HuH7-RIC
HEK293-AP
HeLa-RIC
800
400
0
dataset size
intersecting datasets
not in intersection
c Mus musculus RBPs
intersection size
500
400
300
200
100
0
MEF-RIC
HL-1-RBDmap
mESC-RIC
RAW264.7-RIC
mESC-nRIC
HL-1-RIC
1000
500
0
dataset size
intersecting datasets
not in intersection
d Saccharomyces cerevisiae RBPs
intersection size
400
300
200
100
0
Arrays
Yeast-Array
Yeast-MS
Yeast-RNPxl
Yeast-RNPxl.4sU
Yeast-RIC
Yeast-Array2
Yeast-RIC.4sU
Yeast-mRBPome
800
400
0
dataset size
intersecting datasets
not in intersection
e Drosophila melanogaster RBPs
256
212
309
embryo-RIC
embryo-maternal-to-zygotic-transition-RIC
f Arabidopsis thaliana RBPs
149
29
59
20
315
10
137
different ecotypes
4 day etiolated seedlings
leaf mesophyll protoplasts
g InParanoid 6 cluster comparison
443
311
144
396
363
248
855
766
380
Hs
Sc
Mm
h Human orthologues
intersection size
600
400
200
0
At*
Ce*
Dm*
Sc*
Mm*
Hs
1000
0
dataset size
* human orthologues
intersecting datasets
not in intersection

Figure 2

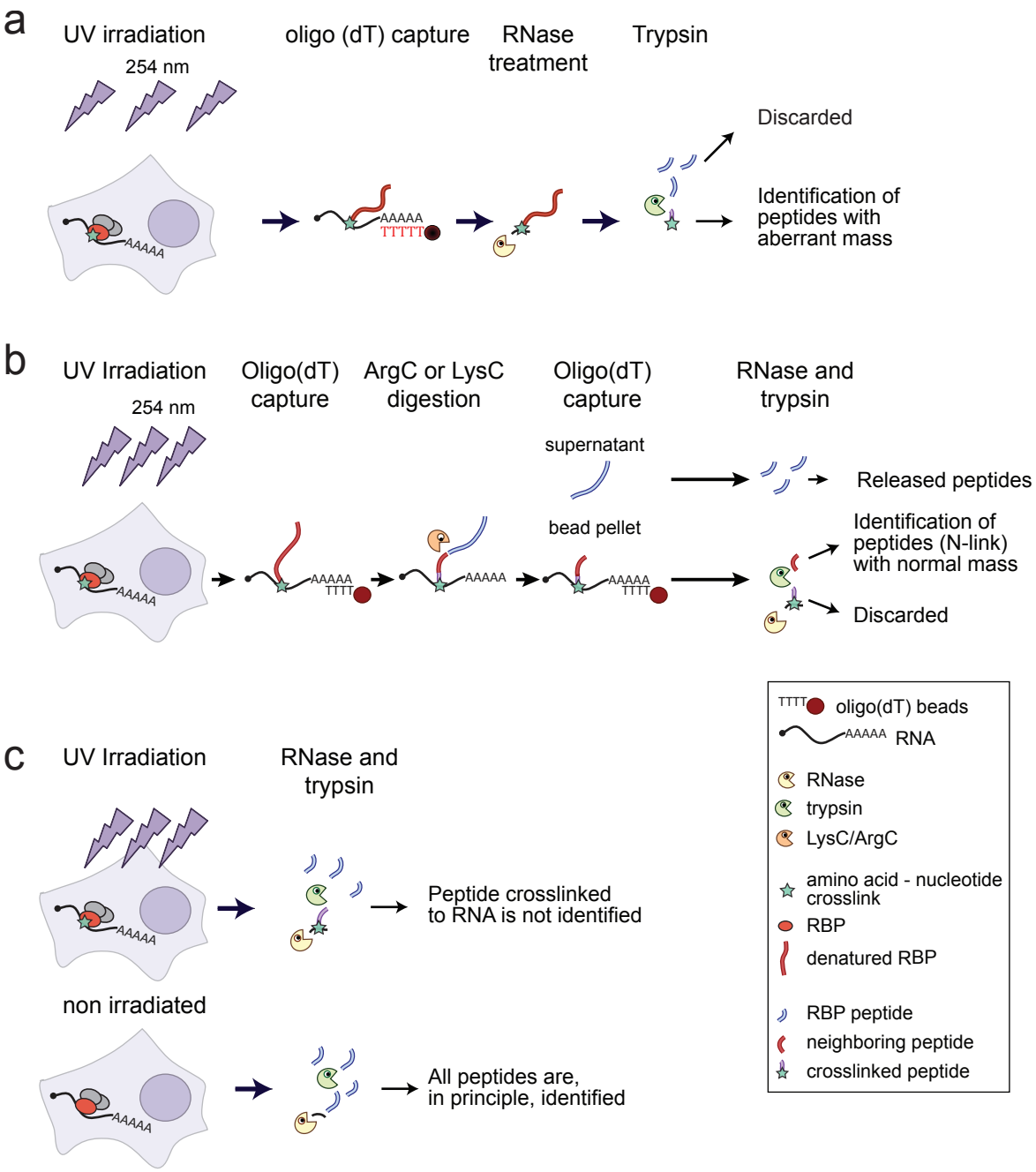

a
UV irradiation
254 nm
oligo (dT) capture
RNase treatment
Trypsin
Discarded
Identification of peptides with aberrant mass
AAAAA
TTTTT
b
UV Irradiation
254 nm
Oligo(dT) capture
ArgC or LysC digestion
Oligo(dT) capture
RNase and trypsin
supernatant
Released peptides
bead pellet
Identification of peptides (N-link) with normal mass
Discarded
c
UV Irradiation
RNase and trypsin
Peptide crosslinked to RNA is not identified
non irradiated
All peptides are, in principle, identified
TTTT oligo(dT) beads
AAAAA RNA
RNase
trypsin
LysC/ArgC
amino acid - nucleotide crosslink
RBP
denatured RBP
RBP peptide
neighboring peptide
crosslinked peptide

Figure 3

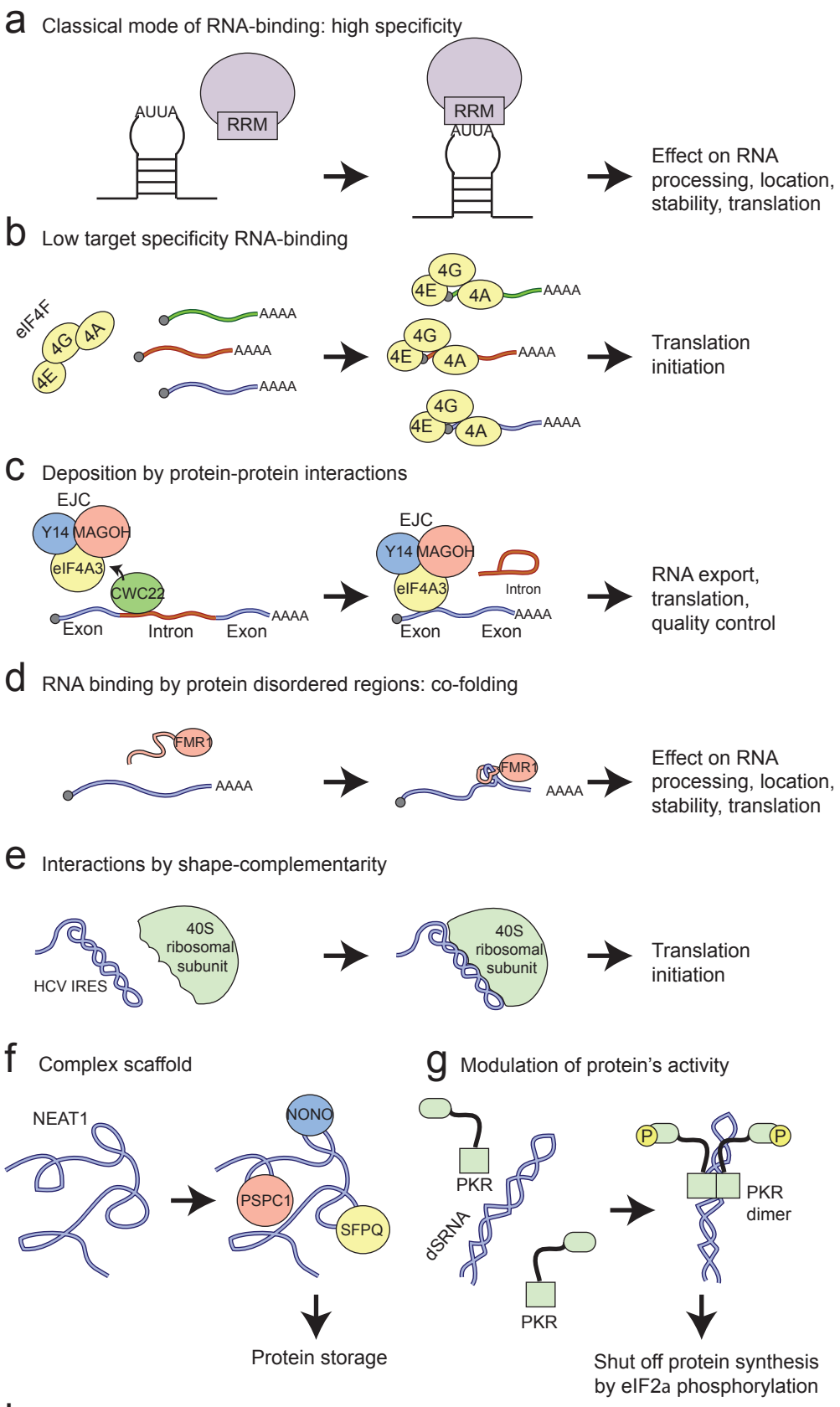

a
Classical mode of RNA-binding: high specificity
AUUA
RRM
RRM
AUUA
Effect on RNA processing, location, stability, translation
b
Low target specificity RNA-binding
eIF4F
4G
4A
4E
AAAA
AAAA
AAAA
4G
4E
4A
AAAA
4G
4E
4A
AAAA
Translation initiation
c
Deposition by protein-protein interactions
EJC
Y14
MAGOH
eIF4A3
CWC22
Exon
Intron
Exon
AAAA
EJC
Y14
MAGOH
eIF4A3
Intron
AAAA
Exon
Exon
RNA export, translation, quality control
d
RNA binding by protein disordered regions: co-folding
FMR1
AAAA
FMR1
AAAA
Effect on RNA processing, location, stability, translation
e
Interactions by shape-complementarity
HCV IRES
40S ribosomal subunit
40S ribosomal subunit
Translation initiation
f
Complex scaffold
NEAT1
NONO
PSPC1
SFPQ
Protein storage
g
Modulation of protein's activity
PKR
dsRNA
PKR
P
P
PKR dimer
Shut off protein synthesis by eIF2a phosphorylation
h
Moonlighting RNA binding activity: regulation by metabolites
High iron levels
ACO1
iron sulphur cluster
Low iron levels
IRP1
RNA stability translation

Figure 4

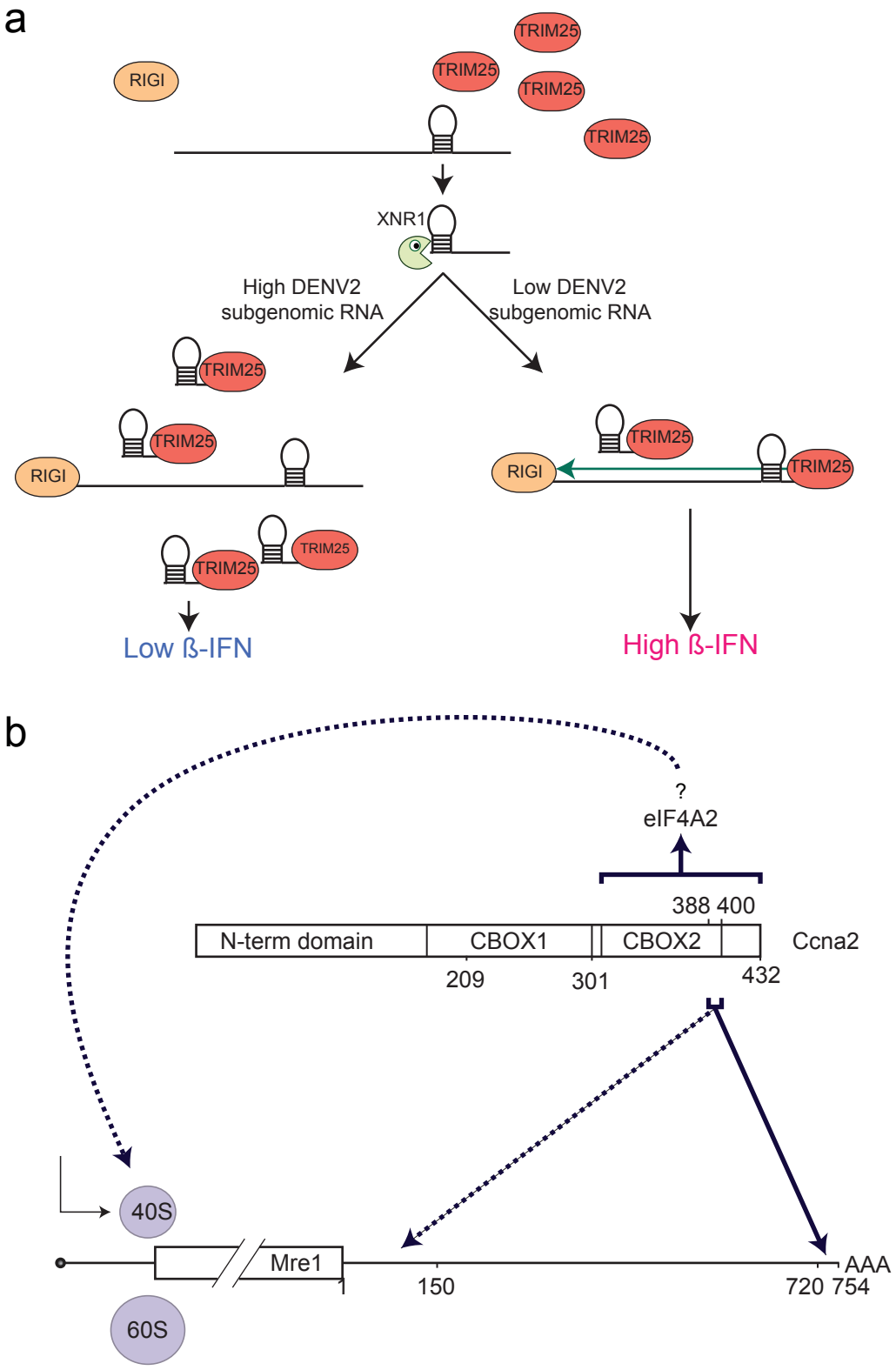

a
TRIM25
TRIM25
TRIM25
TRIM25
RIGI
XNR1
High DENV2
subgenomic RNA
Low DENV2
subgenomic RNA
TRIM25
TRIM25
RIGI
TRIM25
RIGI
TRIM25
TRIM25
TRIM25
Low ß-IFN
High ß-IFN
b
?
eIF4A2
388 400
N-term domain
CBOX1
CBOX2
Ccna2
209
301
432
40S
Mre1
AAA
1
150
720 754
60S

Figure 5

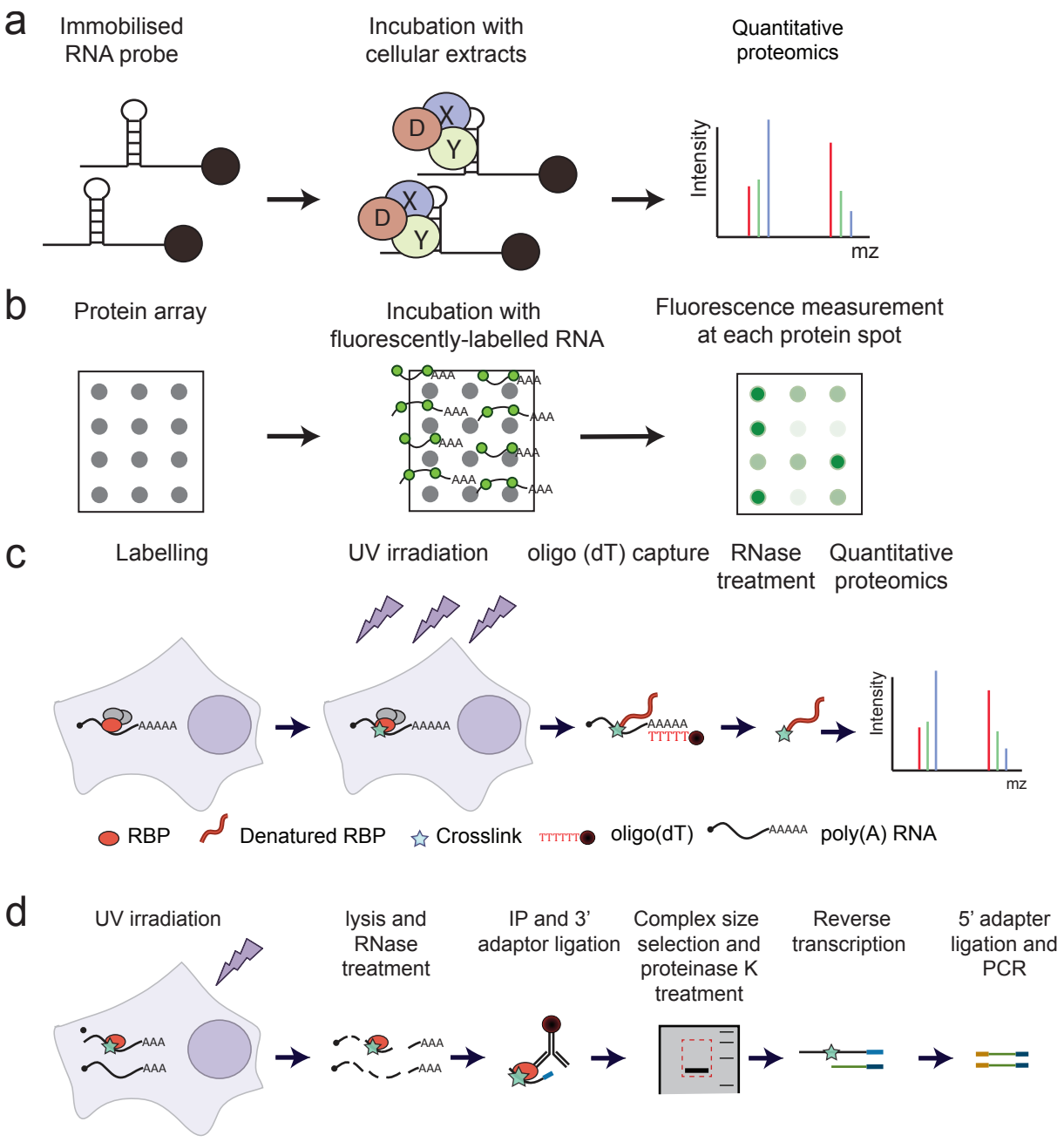
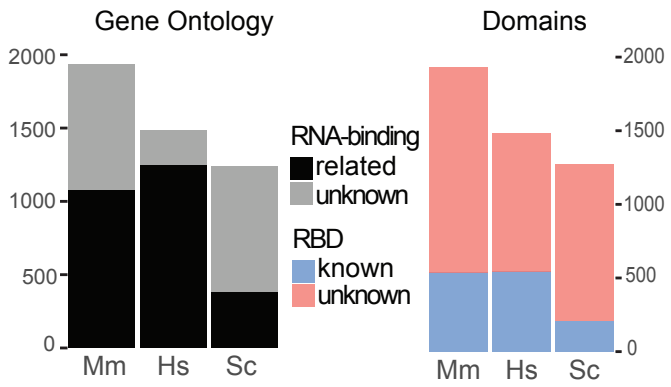
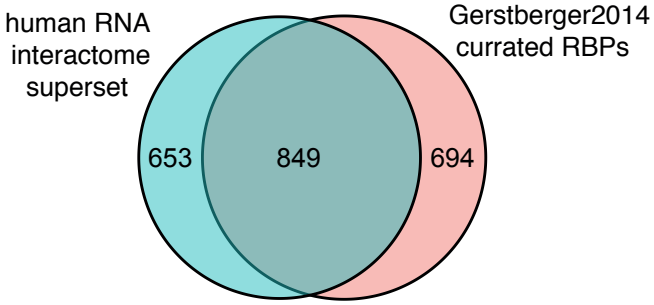


Figure for Box 1

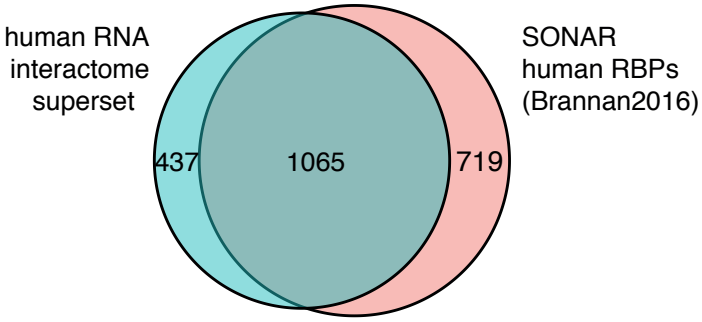
a Functional analysis of supersets



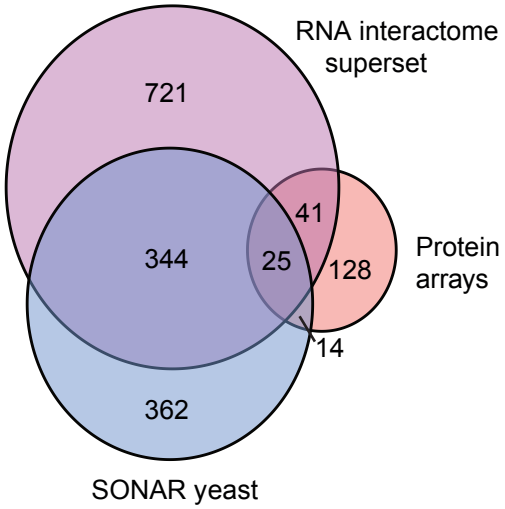
b Overlap of human superset with Gerstberger2014 (Total 2169)



c Overlap human superset with SONAR



d Yeast RNA interactome superset vs. SONAR and protein array



Supplementary Figure 1