



Hentze, M. W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology*, 19(5), pp. 327-341.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/219002/>

Deposited on: 12 August 2020

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

1 **A Brave New World of RNA-Binding Proteins**

2

3 Matthias W. Hentze^{1*}, Alfredo Castello², Thomas Schwarzl¹ and Thomas Preiss^{3,4*}

4

5

6 ¹ European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

7 ² Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU,

8 United Kingdom

9 ³ EMBL–Australia Collaborating Group, Department of Genome Sciences, The John Curtin

10 School of Medical Research, The Australian National University, Acton (Canberra) ACT

11 2601, Australia

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

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

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

5 — *William Shakespeare, The Tempest, Act V, Scene I, ll. 203–206 'Miranda's speech'*

6

7

8 **ABSTRACT**

9 RNA-binding proteins (RBPs) are typically thought of as proteins that bind RNA through one
10 or multiple globular RNA-binding domains (RBDs) to change the fate or function of the
11 bound RNAs. Several hundreds of such RBPs have been discovered and investigated over
12 the years. Recent proteome-wide studies have more than doubled the number of proteins
13 implicated in RNA binding and uncovered hundreds of additional RBPs lacking classical
14 RBDs. This review integrates these new RBPs and discusses their possible functions,
15 including the notion that some may be regulated by RNA rather than exerting control over it.

16

17

18 **INTRODUCTION**

19 A 'classical' RNA-binding protein (RBP) participates in the formation of ribonucleoprotein
20 (RNP) complexes that are principally involved in gene expression¹. It does so by binding to
21 sequence and/or structural motifs in RNA, via modular combinations of a limited set of
22 structurally well-defined RNA-binding domains (RBDs)² such as the RNA recognition motif
23 (RRM)³, the K homology (KH)⁴ or DEAD box helicase domains (**Fig. 1a**)⁵. These assertions
24 represent decades of cumulative knowledge, including, cellular, biochemical and structural
25 data. However, recent advances in determining the structures of large RNP machines such

1 as the ribosome⁶⁻⁸ and spliceosome⁹⁻¹¹ reveal complex protein-RNA packaging without
2 involvement of canonical RBDs. This suggests that unorthodox RNA binding is a broader
3 phenomenon than previously anticipated.

4 A widely held assumption is that RBPs with high affinity and/or specificity for their targets
5 are more likely to have (ascertainable) biological functions¹². Implicit in this conventional
6 view of RBPs is also that they should act in furtherance of the RNA's function¹³. A recent
7 review used the metaphor of RBPs as the mRNA's clothes that will ensure that different
8 mRNA domains (i.e. the 5' and 3' untranslated regions [UTRs] and the coding region [CDS])
9 are at times covered up or exposed, helping it to pass through different stages of its life¹⁴.
10 The processes driving change of RNP composition have been likened to those involved in
11 chromatin activity^{15,16}. Accordingly, post-translational modifications (PTMs) of RBPs and
12 epitranscriptomic modifications of RNA, together with the action of ATP-dependent RNA
13 helicases, lead to dynamic RNP remodelling.

14 The above concepts have broad but clearly not universal applicability, as indicated by
15 emerging evidence from several directions. First, multiple microscopically visible
16 membrane-less RNP granules have been characterised in different cell types and cellular
17 compartments^{17,18}. These include Cajal bodies and paraspeckles in the nucleus as well as
18 processing (P-) bodies and stress granules (SGs) in the cytoplasm. The term granule is
19 somewhat of a misnomer as several of these RNP bodies have now been shown to form by
20 liquid-liquid phase separations (LLPS), thought to be driven by intrinsically disordered
21 regions (IDRs; protein regions lacking stable tertiary structure in their native state) of their
22 constituent RBPs^{19,20}. Their dynamicity and amorphous structure still remains puzzling, and
23 well-defined functions remain to be assigned to the formation of these RNP bodies.
24 Second, our realisation of the existence of a myriad of long noncoding RNAs (lncRNAs)
25 triggered intense efforts to uncover their functions^{21,22}. Many lncRNAs are presently

1 expected to act in the recruitment of transcription factors or chromatin-modifying complexes
2 to genomic loci, or otherwise organise, scaffold or inhibit protein assemblies^{23,24}. These
3 latter scenarios tend to break with convention, indicating that an RNA may act on its protein
4 binding partner rather than itself being the target of regulation (**Fig. 1b**). Finally, several
5 unbiased approaches to identify RBPs proteome-wide have been developed recently,
6 yielding a growing collection of RNA-binding proteomes from multiple organisms and
7 cellular contexts. These compendia persistently identify large numbers of novel RBPs²⁵ that
8 defy convention, lacking either discernible RBDs or established cellular functions that link
9 them to RNA biology in a straightforward fashion, or both.

10 The features and functions of classical RBPs have been expertly covered by the reviews
11 cited above and other recent works²⁶⁻²⁸. This review focuses primarily on the challenges
12 posed by unconventional RBPs, their methods of discovery and emerging information about
13 their modes of RNA-binding, RNA targets, as well as their molecular and cellular functions.
14 We will contrast and integrate these with what we know about classical RBPs.

15

16 **THE ERA OF RNA INTERACTOMES**

17 Since the discovery in the early nineties of several metabolic enzymes that engage in
18 ‘moonlighting’ RNA-binding activity, it became apparent that the number and diverse nature
19 of RBPs had been underestimated²⁹⁻³². The list of unconventional RBPs grew incrementally
20 over the decades, urging the development of methods to identify RBPs comprehensively in
21 living cells.

22

23 *Experimental approaches to catalogue RBPs*

24 *In vitro* approaches were devised using either immobilised RNA probes or arrayed proteins
25 (**Box 1**), identifying multiple novel RBPs³³⁻³⁶. More recently, RNA interactome capture was

1 developed as an *in vivo* approach focusing on native protein-RNA interactions (**Box 1**). It
2 entails UV crosslinking of RBPs to RNA in live cells followed by collective capture of RNPs
3 containing poly(A)⁺ RNA on oligo(dT) beads, and identification of proteins by quantitative
4 mass spectrometry (Q-MS)³⁷. Two groups independently deployed RNA interactome
5 capture to yield 860 and 791 proteins as the RBP repertoire of human HeLa and HEK293
6 cells, respectively^{38,39}. Both RNA interactomes overlap considerably, with 543 shared
7 RBPs, and enrichment for the gene ontology (GO) term “RNA binding” (*e.g.* 468 of a total of
8 860 proteins in the HeLa RNA interactome, based on annotation current at the time)⁴⁰, and
9 detected the majority of the well-established RBPs. Classical RBDs such as the RRM, KH,
10 DEAD box helicase and some zinc finger domains were also strongly enriched. Altogether
11 this attested to the technical robustness of the approach. Beyond that, about half of the
12 proteins in each RNA interactome lacked known RBDs, and hundreds had no previous
13 relation to RNA biology. Interestingly, both studies revealed common biological roles and
14 molecular functions among the newly discovered proteins, including intermediary
15 metabolism, cell-cycle progression, antiviral response, spindle organization, protein
16 metabolism (chaperons and prolyl *cis-trans* isomerases), and others^{38,39}. The discovery of
17 RNA-binding activity of proteins involved in biological processes without apparent relation to
18 RNA biology (‘enigmRBPs’) suggested unexplored regulatory layers between gene
19 expression and other biological processes. About two dozen of these enigmRBPs were
20 validated by orthogonal approaches^{38,39,41}, seven of them were analysed by
21 immunoprecipitation followed by sequencing of bound RNA, demonstrating that these
22 proteins specifically interact with distinct sets of poly(A)⁺ RNAs^{38,39} and exert defined
23 biological functions.

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

1 human^{38,39,42-44} and mouse cell lines⁴⁵⁻⁴⁷, yeast^{43,48,49}, the unicellular parasites *Leishmania*
2 *donovani*⁵⁰, *Plasmodium falciparum*⁵¹ and *Trypanosoma brucei*⁵², as well as plants⁵³⁻⁵⁵,
3 flies^{56,57}, worms⁴⁹ and fish⁵⁸. The RBP sets of different origins each featured enrichment of
4 RNA-related annotation (for example, **see Fig. S1a**), and orthogonal methods were typically
5 used to validate some of the unexpectedly discovered RBPs, including dual-fluorescence
6 RNA-binding assay^{39,41,49}, crosslinking and immunoprecipitation (CLIP) followed by 5'
7 radioactive labelling of RNA by T4 polynucleotide kinase^{38,43,45,46,56,57}, RT-PCR⁴⁹ or
8 sequencing^{38,39,43}. Using updated annotation, we compiled here all published RNA
9 interactomes into RBP supersets for *Homo sapiens* (1914), *Mus musculus* (1393),
10 *Saccharomyces cerevisiae* (1273), *Drosophila melanogaster* (777), *Arabidopsis thaliana*
11 (719) and *Caenorhabditis elegans* (593) (**Fig. 2a, Table S1**). These datasets represent a
12 resource to mine for shared and selectively identified RBPs (**Fig. 2b-f**). While simple
13 technical reasons will explain some of the differences in coverage and overlap, the outlines
14 of shared 'core' RNA interactomes are emerging⁴³. To illustrate this, we performed
15 InParanoid^{59,60} analysis to yield a high number of orthologue groups especially between
16 mammals, which expectedly decreases in more evolutionary distant organisms (**Fig. 2g-h**).
17 On the whole, widely shared RBPs tend to be more highly enriched in established RNA-
18 related annotation than RBPs with cell or organismic context-dependent expression or
19 activity.

20 Some specific observations from these datasets deserve attention. For example, the mouse
21 embryonic stem cell (mESC) RNA interactome was found to be enriched in proteins with
22 differential expression during differentiation, suggesting that RBPs are regulated during the
23 transition from stem to differentiated cell⁴⁵. In particular, the transcriptional network of the
24 proto-oncogene Myc was found to be enriched among mESC RBPs, suggesting that RBPs
25 contribute to Myc-dependent cell fate decisions. Moreover, mESC RBPs were significantly

1 upregulated during the third day of reprogramming into induced pluripotent stem cells
2 (iPSC). RNA interactome capture was further combined with nuclear fractionation to
3 determine the repertoire of nuclear RBPs of K562 myeloid leukaemia cells; the p53
4 interaction network was enriched among the newly discovered nuclear RBPs⁶¹.

5 RNA modifying enzymes such as RNA tailing enzymes and ‘writers’ of epitranscriptomic
6 modifications consistently form part of RNA interactome compendia. The cardiomyocyte
7 RNA interactome contains 29 RBPs annotated for 5-methylcytosine (m⁵C), N6-
8 methyladenosine (m⁶A) and pseudouridine modifications, as well as adenosine-to-inosine
9 editing⁴⁶. Such epitranscriptomic marks may affect RNA-protein interactions, or otherwise
10 modulate RNA function^{62,63} (**Box 2**).

11 Metabolic enzymes were recurrently identified as RBPs. Our meta-analysis revealed 71
12 such ‘moonlighting’ metabolic enzymes in humans, 104 in mouse and 132 in yeast RNA
13 interactomes (**Table S1**), adding a plethora of new examples to a list of about twenty RNA-
14 binding metabolic enzymes discovered by classic low-throughput studies⁶⁴⁻⁶⁶. Collectively,
15 these dual RBP-enzymes cover a breadth of metabolism, with interesting differences in
16 predominant pathways depending on the source material. For example, the RNA
17 interactome of Huh-7 cells harbours numerous enzymes from glycolysis and other
18 pathways of intermediary metabolism^{43,49}. This likely relates to the important metabolic role
19 of hepatocytes at the organismal level⁴³. The HL-1 cardiomyocyte RNA interactome exhibits
20 a high incidence particularly of mitochondrial metabolic enzymes⁴⁶, reflecting their high
21 organelle content and energy needs. Many of these RNA-binding metabolic enzymes were
22 validated by orthogonal approaches^{39,43,44,46}, including mitochondrial fractionation followed
23 by T4 polynucleotide kinase (PNK) assay⁴⁶. Interestingly, some of these enzymes interact
24 with their own mRNA⁴⁹, perhaps hinting at the existence of negative feedback loops by
25 cognate enzyme-mRNA interaction under conditions of substrate or co-factor deprivation,

1 as previously proposed⁶⁴⁻⁶⁸. Taken together, this suggests that RNA binding by metabolic
2 enzymes is a common and widely conserved occurrence.

3

4 *In silico approaches to catalogue RBPs*

5 A set of 1,542 human RBPs (7.5% of the proteome) was defined based on computational
6 analyses that required a protein to harbour known RBDs or other domain features
7 characteristically found in proteins with RNA-related functions. This was complemented by
8 manual curation to add missing but well-documented RBPs or proteins forming part of
9 known RNPs, and to remove proteins with established RNA-unrelated functions⁶⁹. Members
10 of this curated set of RBPs tend to display ubiquitous expression profiles across tissues,
11 suggesting housekeeping roles in cell biology, and the set overlaps well with the
12 experimental human RNA interactomes (**Fig. S1b**). Notwithstanding its strengths, the
13 approach might generate false positive results for i) proteins harbouring a classified RBD
14 that serves a different functional role in that particular case⁷⁰, or ii) proteins interacting with
15 RNA solely via protein-protein interactions with a direct RBP (as in the case of Y14 in the
16 exon junction complex)⁷¹. A similar domain/function-search algorithm was recently applied
17 to *Plasmodium falciparum*⁵¹. This study reported 924 RBPs, a surprisingly high fraction
18 (18.1%) of the relatively small number of protein-coding genes of the malaria parasite.

19 The propensity of RBPs to interact with other RBPs, either directly or via bridging RNAs,
20 was also exploited to identify novel RBPs^{39,72,73}. The classification algorithm, referred to as
21 “support vector machine obtained from neighbourhood associated RBPs” (SONAR),
22 evaluates each protein of a proteome (or a dataset) against protein-protein interaction (PPI)
23 data, and calculates its RBP classification score (RCS). The algorithm was trained on sets
24 of human RBPs merged from available sources^{39,69,74}, where the “neighbourhood” of each
25 protein was determined using the BioPlex PPI dataset⁷⁵, which includes many thousands of

1 experimentally determined protein-protein interactions. Sets of SONAR-predicted RBPs
2 were established for human (1,784 proteins), *Drosophila melanogaster* (489) and
3 *Saccharomyces cerevisiae* (745); these agree well with the experimentally determined RBP
4 sets (**Fig. S1c,d**). SONAR can be readily applied to any organism as long as substantial
5 PPI data are available. SONAR thus presents the same limitations as any proteomic-based
6 approach, including the depth of the protein-protein interaction data available in a given
7 organism. In addition, it may lead to false positives because proteins that interact with
8 RBPs are not always RBPs themselves⁷¹.

9

10 *The plasticity of RNA-binding proteomes*

11 Biology is dynamic. The binding of RBPs to RNA constantly changes, and the composition
12 of RNA interactomes is context-dependent and responds to stimuli. While a subset of
13 'house-keeping' RBPs might be constitutively and ubiquitously active⁶⁹, many RBPs have
14 more restricted expression patterns and/or their RNA-binding activity may be regulated, e.g.
15 by post-translational modifications⁷⁶⁻⁷⁸, cofactor binding⁷⁹ or protein-protein interactions⁸⁰.
16 Moreover, some RBPs can 'sit idle' for lack of their RNA targets⁸¹⁻⁸³. For example, cellular
17 sensors against viral infection such as interferon-induced, double-stranded RNA-activated
18 protein kinase (EIF2AK2 or PKR), retinoic acid-inducible gene-I (RIGI or DDX58) or toll-like
19 receptors (TLRs) may only be activated by the presence of unusual RNA products derived
20 from viral replication such as long double stranded RNA tracts⁸¹ or triphosphate 5' ends^{82,83}
21 (Fig. 1b).

22 RNA interactome capture has been adapted to investigate responses of RNA-binding
23 proteomes to physiological changes and environmental cues. This was recently applied to
24 murine macrophages responding to lipopolysaccharide stimulation⁴⁷, primary mouse
25 embryonic fibroblasts treated with etoposide⁸⁴, as well as to fruit fly and zebrafish embryos

1 at different stages of development^{56,58}. During the maternal-to-zygotic transition (MZT),
2 gene expression switches from maternally deposited mRNAs whose timed expression is
3 orchestrated by RBPs, to a new transcriptome emerging after zygotic genome activation
4 (ZGA)⁸⁵. The dynamic changes in RBP activity expected during MZT were investigated in
5 *Drosophila* embryos⁵⁶. Using tandem mass tags (TMT)⁸⁶, biological triplicate samples
6 derived from pre- (0-1 hours post fertilisation [hpf]) and post-ZGA embryos (4.5-5.5 hours)
7 were analysed. Comparison between UV-irradiated *versus* non-irradiated pooled pre- and
8 post-ZGA stage samples was used to determine the 'static' RNA interactome capture. In
9 separate experiments, UV-irradiated samples from pre- and post-ZGA stages were directly
10 compared for the identification of 'dynamic' RBPs. The former analysis yielded 523 proteins
11 as high confidence fly embryo RBPs, whereas the latter identified 1131 proteins, 116 of
12 which were differentially captured at 10% false discovery rate (FDR). To determine whether
13 differential RNA binding was due to alterations in protein levels, the total proteomes were
14 also determined in parallel. Comparison of the RNA interactome and whole proteome data
15 revealed three classes of RBPs: class 1 (1015 proteins) showed no significant change in
16 either RNA binding or total abundance; class 2 (78 proteins) showed commensurate
17 changes in both parameters, suggesting that differential RNA binding was due to altered
18 protein levels; and, class 3 (38 proteins) with a clear change in RNA binding without a
19 corresponding change in RBP abundance, implying a modulation in the ability of these
20 proteins to interact with RNA. These RBPs were thus dubbed 'dynamic binders'. The latter
21 include eight known splicing factors, and seven of these bind RNA more avidly in pre-MZT
22 embryos. These findings broadly concur with an analysis using mRNA expression and
23 localisation as proxies to reveal that RBP expression peaks during the prezygotic phase
24 and MZT, while transcription factors are highly expressed during the first zygotic waves and
25 mid embryogenesis⁵⁷. A similarly configured analysis of zebrafish embryos before (1.75 hpf)

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

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

11

12 **UNORTHODOX RNA-BINDING DOMAINS**

13 Many of the RBPs newly discovered by the different methods described above lack known
14 RBDs, posing questions of how they interact with RNA. Tried-and-tested approaches exist
15 to map RBDs within individual proteins, for example by mutagenesis combined with RNA-
16 binding assays, such as electrophoretic mobility shift assay (EMSA)⁸⁷ or the CLIP-coupled
17 PNK assay⁴⁵, however, high(er) throughput methods were required to identify the RBDs of
18 hundreds of novel RBPs in a time-efficient manner. Thus, three approaches were
19 developed, each using mass-spectrometry in a different configuration to identify protein
20 regions that become crosslinked to RNA after UV exposure of live cells.

21 One approach focused on the purification and direct detection of the RNA-crosslinked
22 peptides, whose mass is altered by the nucleic acid remnant (**Fig. 3a**)^{42,88}. Data analysis
23 was performed with RNP^{xl}⁴², a custom-designed software to reduce the complexity of the
24 search for peptide-nucleotide matching spectra. Applied to yeast RBPs, RNP^{xl} identified
25 376 crosslinked peptides corresponding to 133 unique crosslinking sites in 57 different

1 proteins, mostly canonical RBPs such as ribosomal proteins and proteins with RRM or KH
2 domains⁴². A number of RNA-binding sites were identified in unorthodox RBPs, including
3 the enzymes peptidyl-prolyl cis-trans isomerase (CPR1), enolase 1 (ENO1) and
4 phosphoglycerate kinase (PGK1).

5 RBDmap adds a further digestion step with a protease that cleaves every 17 amino acids
6 on average and a second round of oligo(dT) capture to the RNA interactome capture
7 workflow⁴⁴. After the second oligo(dT) capture, the covalently linked polypeptides are
8 proteolysed by trypsin, generating an RNA-crosslinked peptide and a neighbouring peptide
9 with native mass (Fig 3b). RBDmap employs the neighbouring peptides with native mass to
10 assign RNA-binding sites (**Fig. 3b**)⁴⁴. Applied to HeLa cells, RBDmap discovered 1,174
11 RNA-binding sites within 529 proteins⁴⁴, while more limited analysis of HL-1 cardiomyocytes
12 still revealed 568 RNA-binding sites from 368 proteins⁴⁶. RBDmap data display strong
13 concordance with regular RNA interactome studies, reinforcing the identification of
14 unorthodox RBPs even after two oligo(dT) purification steps and after extensive proteolytic
15 treatment^{44,46}. As expected, RBDmap “re-discovered” the classical RBDs such as the RRM-
16 , KH-, or cold shock domains². Notably, many of the reported RNA-binding sites mapped to
17 globular proteins and domains lacking previous association with RNA binding. For HeLa
18 cells these include the thioredoxin (TXN) fold, heat shock protein (HSP) 70 and HSP90
19 domains, 14-3-3, domain associated with zinc fingers (DZF), PDZ and NDR (complete list
20 in⁴⁴), many of which were validated by orthogonal approaches. Interestingly, the mapped
21 RNA-binding sites showed enrichment for homologous regions across different proteins
22 from the same family, and many mapped to enzymatic cores or protein-protein interaction
23 surfaces, suggesting a possible interplay between these activities and RNA binding⁴⁴. For
24 HL-1 cells, RNA-binding regions were identified in 24 metabolic enzymes, 12 of which

1 mapped to di-nucleotide binding domains⁴⁶. These data corroborate and expand the
2 previously suggested RNA-binding activity of di-nucleotide-binding domains^{32,65,66,89,90}.
3 Unexpectedly, many of the identified RNA-binding regions mapped to IDRs, implicating
4 them as predominant sites of protein interaction with RNA *in vivo* (see below). Another
5 observation was that RNA-binding regions were enriched for regions mutated in Mendelian
6 diseases, while natural variants distributed equally across binding and non-binding
7 regions⁴⁴. This suggests that numerous Mendelian diseases arise from altered RNA
8 biology. Finally, the RNA-binding regions strongly overlap with known post-translational
9 modification sites, including phosphorylation, acetylation and methylation⁴⁴. This enrichment
10 was not observed for protein regions lacking RNA-binding activity. Thus, posttranslational
11 modifications may regulate RNA binding and RNP dynamics akin to chromatin remodelling.
12 Peptide cross-linking and affinity purification (pCLAP) is a recently described cousin of
13 RBDmap, which implements the first protease treatment directly after lysis, requiring only
14 one oligo(dT) capture round⁹¹. The trade-off is that pCLAP does not quantify the peptides in
15 the released fraction. This reference is used in RBDmap to determine with high confidence
16 the protein regions engaged in RNA binding (Fig. 3b)⁴⁴.
17 A third approach, termed proteomic identification of RNA-binding regions (RBR-ID)
18 identifies peptides with reduced intensity in UV-irradiated samples compared to non-
19 irradiated controls (**Fig. 3c**)⁹². RBR-ID of nuclear proteins from mESCs detected 1,475
20 RNA-binding sites mapping to 803 mESC proteins with a 5% FDR. 47% of the RBR-ID
21 proteins were present in the RNA interactome studies^{38,39,43,45,61}, while 865 previously
22 known RBPs were missing. Interestingly, the 427 RBPs newly identified by RBR-ID were
23 enriched for gene-regulatory and chromatin-associated functions. At the domain level,
24 peptides with UV-dependent intensities (i.e. RNA-binding sites) mapped to RRM, KH,
25 DEAD box and other classical RBDs, validating the approach. The RNA-binding sites of the

1 427 newly discovered RBPs surprisingly often mapped to chromatin-related domains such
2 as the chromodomain and bromodomain, invoking a potential crosstalk between DNA and
3 RNA. Some of the latter results confirmed prior candidate-based approaches characterizing
4 the activity of chromatin remodelling proteins such as the catalytic subunit of the polycomb
5 repressor complex 2 (PRC2) EZH2, which has been shown to interact with long non-coding
6 RNA and nascent transcripts⁹³⁻⁹⁵; RBR-ID data suggest that this phenomenon might be
7 broader than previously anticipated.

8 Together, RNP^{x1}, RBDmap and RBS-ID have greatly expanded the known repertoire of
9 RBDs. However, we know little about the RNA targets of these novel RBDs or the function
10 of these interactions. To address this, functional studies including determination of the
11 specificity and affinity of these novel RBDs for their target sequences⁹⁶⁻¹⁰⁰ will be required.

12

13 **NOVEL RNA-BINDING MODES**

14 The many novel RBPs and their non-typical RBDs raise pertinent questions regarding their
15 biological function(s)¹⁰¹. Not every bimolecular collision that happens in cells should be
16 assumed to be physiologically relevant. Which ranges for the affinities and RNA-binding
17 specificities are to be expected? Concepts regarding specificity and non-specificity in RNA-
18 protein interactions were recently reviewed¹². A first relevant aspect is that indiscriminate
19 RNA binding by RBPs is common and can be important to their function. For example,
20 numerous proteins involved in mRNA translation (for example, the eukaryotic initiation
21 factor [eIF]4F complex) and degradation need to be non-selective to fulfil their functions
22 (**Fig. 4a vs b**). Similarly, the exon junction complex is deposited through protein-protein
23 interactions¹⁰² at a constant position upstream of splice junctions, and interacts with a
24 plurality of RNA sequences¹⁰³ (Fig. 4c). Second, one must distinguish between 'biological
25 specificity', the binding characteristics of RBPs *in vivo*, and 'intrinsic specificity', such as

1 one might determine by selection from a random pool of sequences *in vitro*. An interesting
2 upshot of this distinction is that intrinsically specific RBPs may function biologically as less
3 specific RBPs, when they bind their physiological RNA targets without much discrimination
4 because these do not fall into the high-affinity/high-specificity range of their RNA-binding
5 potential¹².

6

7 *RNA Binding by Intrinsically Disordered Regions*

8 As described above, IDRs are not only involved in aggregation of RNPs into granules by
9 protein-protein interactions^{19,20} but also directly engage in RNA binding^{44,46,104,105}. RBPs are
10 enriched in IDRs that are characterized by a low content of bulky hydrophobic amino acids,
11 with the exception of tyrosine (Y), and a high proportion of small, polar or/and charged
12 amino acids, particularly glycine (G), serine (S), arginine (R), lysine (K), glutamine (Q),
13 glutamic acid (E) and aspartic acid (D)⁴⁴. Interestingly, mutations in RBPs causing human
14 Mendelian diseases occur with higher frequency within these RNA-binding IDRs than in
15 globular domains, suggesting strong sequence constraints⁴⁰. The occurrence of IDRs within
16 RBPs appears to be conserved from yeast to human⁴³, often in the form of repeats such as
17 RGG, YGG, SR, DE or KK^{39,104}. A recent report proposes that the number of repeats within
18 the IDRs of RBPs has expanded from yeast to human⁴³. Because the number and identity
19 of globular domains present within these proteins have remained the same, IDRs may
20 represent a plastic component of RBPs that co-evolved with the increasing complexity of
21 eukaryotic transcriptomes.

22 Around half of the RNA-binding sites reported by RBDmap (1,174) in HeLa cells mapped to
23 IDRs, reflecting their prevalence as a mode of RNA binding⁴⁴. 170 RBPs appeared to
24 interact with RNA exclusively through IDRs, suggesting that these regions can suffice to
25 mediate RNA binding. Amongst the arginine-rich motifs (ARM), RGG and SR repeats were

1 previously reported to bind RNA¹⁰⁴. The discovery of numerous additional examples
2 allowed assignment into sub-classes that differ by the lengths of their glycine linkers^{44,106}.
3 Nuclear magnetic resonance (NMR) analyses of human fragile X mental retardation protein
4 (FMRP) showed that the positioning of the arginines is essential for selective binding of the
5 RGG box to the guanidine-rich sequence in the sc1 mRNA¹⁰⁷, where the glycine linker
6 orients the arginines geometrically to interact with the Watson-Crick nucleotide base pairs,
7 which stack on two G-quadruplexes resulting from the protein-RNA co-folding (**Fig. 4d**). In
8 addition, the glycine linker contributes to RNA binding by shape complementarity
9 interactions at the interface between G-quadruplexes and Watson-Crick base pairs. Hence,
10 the affinity and selectivity of RGG boxes for their target RNAs may be determined by the
11 frequency of arginines and glycines.

12 A second RNA-binding family of IDRs involves aromatic residues, especially tyrosines (Y),
13 that combine with glycines (G) and serines (S) forming [G/S]Y[G/S]. These motifs display a
14 tendency to aggregate *in vitro*, inducing hydrogel formation and amyloid-like fibres, and
15 engage in dynamic liquid-liquid phase separations *in vivo*^{108,109}. Aromatic residues tend to
16 form part of hydrophobic protein cores, but, when present at the protein surface, can
17 interact with amino acids or nucleotides by stacking or hydrogen bonding². When
18 embedded in a glycine-rich context, the aromatic residue is particularly exposed, likely
19 fostering its propensity to aggregate when interacting with similar protein motifs^{108,109}, or to
20 bind RNA⁴⁴.

21 Finally, a heterogeneous set of linear motifs involving lysine (K) and, to a lesser extent,
22 arginine (R) was also enriched within RBPs⁴⁴. Interestingly, the stoichiometry and distances
23 between the positively charged residues, as well as the combination of these with other
24 amino acids, were conserved even across non-homologous proteins. Notably, the basic
25 IDRs present in RBPs are similar to motifs in DNA-binding proteins. Here, the basic arms

1 can alter the DNA-binding of transcription factors by a large capture radius¹¹⁰. In this
2 “monkey bar” model, transcription factors utilise their basic arms to reach distant DNA sites
3 by “hopping” and “sliding” instead of 3D diffusion¹¹⁰. It is currently unknown whether basic
4 arms may play similar roles in RBPs.

5 Thus, IDRs could represent malleable, potentially multifunctional RNA-binding motifs. Their
6 RNA binding can range from highly specific to non-selective, and may promote co-folding
7 upon interaction with their target RNAs^{104,105,107,111} (**Fig. 4d**). Interestingly, the high
8 sequence constraints of IDRs⁴⁰ enable RNA-binding regulation by reversible
9 posttranslational modifications such as acetylation or phosphorylation⁴⁴. In principle, these
10 properties qualify IDRs as versatile modules for interaction with RNA, either alone or in
11 cooperation with globular RBDs.

12 13 *Shape Complementarity Interactions and Protein-Binding RNAs*

14 Protein-RNA interactions are typically described as a process in which a protein harbours
15 “sensors” (RBDs) to recognize and bind particular sequence and/or structural elements
16 within its “target RNA” (Fig. 1a). However, synthetic RNA aptamers can bind proteins
17 following the same molecular principles as those that enable proteins to bind RNA¹¹²,
18 suggesting that RNA can equally be the driving force mediating protein-RNA interactions
19 (Fig. 1b). The 169 annotated ribosomal proteins follow 119 distinct domain architectures⁶⁹
20 but does this postulate an equal number of distinct RBDs? A more probable explanation
21 appears to be that ribosomal proteins and rRNAs have co-evolved to interact with each
22 other, where extended shape complementarity and the right spatial configuration of
23 molecular interactions play a major role in forming a perfectly packed machinery: the
24 ribosome⁸. Another example of intricate protein-RNA interactions is the spliceosome, where
25 small nuclear (sn)RNAs interact with proteins to form the functionally active complex¹¹.

1 These interactions are enabled by the ability of RNA to fold into three-dimensional (3D)-
2 structures resulting in complex surfaces with potential to interact with complementary (both
3 in shape and biochemical properties) protein partners (Fig. 1b and 4e), as illustrated by
4 tRNAs¹¹³.

5 The fact that RNA can be more than just the “passive” partner in protein-RNA interactions is
6 also illustrated by the 5' UTRs of viral RNAs, which evolved to interact with host proteins.
7 For example, the internal ribosome entry site (IRES) of poliovirus binds the C-terminal
8 moiety of the eukaryotic initiation factor (eIF)4G to recruit the 40S ribosomal subunit¹¹⁴.
9 Similarly, the hepatitis C virus (HCV) IRES interacts directly with eIF3 and the 40S
10 ribosomal subunit for translation initiation (**Fig. 4e**). A closer look at the HCV IRES-
11 ribosome co-structure shows that the interaction is not mediated by well-defined protein
12 regions endowed with RNA-binding activity; instead, the protein-RNA interface is large and
13 displays strong shape-complementarity^{115,116} (**Fig. 4e**).

14 Recently, two studies identified widespread RNA binding by chromatin-associated factors
15 and DNA-binding proteins^{61,92}. One of the proposed possibilities is that these factors
16 interact with lncRNAs and nascent transcripts⁹³⁻⁹⁵. Plausibly, the lncRNAs themselves may
17 ‘drive’ such interactions, and it might make sense to think of these RNAs as displaying
18 protein-binding activity, rather than the other way around (**Fig. 1b**). Functionally, the RNA
19 moiety can play different roles in this interaction, such as scaffolding protein-RNA
20 complexes as illustrated by the viral IRES of HCV¹¹⁶ (**Fig. 4e**) and the lncRNA nuclear
21 enriched abundant transcript 1 (NEAT1) in paraspeckle formation¹¹⁷ (**Fig. 4f**), or altering the
22 activity of the bound protein as exemplified by RIGI and PKR and their interaction with
23 intermediaries of viral replication⁸¹⁻⁸³ (**Fig. 4g**).

24

25 *RNA Binding by Metabolic Enzymes*

1 RNA interactome studies have persistently identified enzymes of intermediary metabolism
2 as RBPs. Some of these enzyme RBP-RNA interactions appear to serve direct feedback
3 gene regulation. For example, thymidylate synthase (TYMS), an enzyme that catalyses the
4 formation of dTMP from dUMP, binds to its own mRNA and inhibits its translation when
5 dUMP levels are low¹¹⁸. A more indirect form of feedback regulation is exerted by the
6 cytosolic aconitase/iron-regulatory protein 1 (IRP1) paradigm. To be active as an enzyme,
7 the protein requires an iron sulfur cluster in its active site, which simultaneously precludes
8 RNA binding¹¹⁹. In iron deficiency, this cluster is lacking and, in an open conformation, IRP1
9 controls the expression of proteins to increase iron uptake and decrease its storage,
10 utilization and export^{119,120} (**Fig. 4h**).

11 Other enzymes display more oblique links to metabolism when acting as RBPs. The
12 glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH) oxidises its
13 substrate to generate NADH, but it also has a diverse range of other cellular functions¹²¹.
14 An important role for GAPDH as an RBP was discovered in T lymphocyte biology⁹⁰. In
15 resting T cells, which rely on oxidative phosphorylation for energy generation, GAPDH
16 binds to AU-rich elements (AREs) in the 3'UTRs of cytokine mRNAs including γ -interferon
17 mRNA and inhibits their translation. Following the metabolic switch to aerobic glycolysis
18 upon T cell activation, GAPDH disengages from RNA binding, thus de-repressing cytokine
19 production.

20 The high number of identified RNA-binding metabolic enzymes suggests that not all of
21 these may have 'moonlighting' functions in post-transcriptional gene regulation as
22 described above. Alternatively, (yet to be discovered) RNAs could affect their metabolic
23 function. As discussed in more detail elsewhere⁶⁶, this could affect the enzyme's
24 localisation or activity, e.g. by affecting an enzymatic side reaction, by allosteric control, or
25 by providing a scaffold that organizes multi-enzyme complexes and even pathways.

1 Interestingly, the globular Rossmann-fold (R-f) domain has emerged as a common non-
2 conventional RBD. The cardiomyocyte RNA-binding proteome includes an impressive 173
3 R-f proteins, and 29 of the 73 cardiomyocytic RNA-binding metabolic enzymes harbour at
4 least one R-f domain⁴⁶. The dinucleotide-binding R-f domain of oxidoreductases has long
5 been discussed as a RNA-binding interface³². NAD⁺ and NADH have been reported to
6 interfere with RNA binding by GAPDH, while cytokine mRNA sequences can inhibit the
7 enzymatic activity of GAPDH *in vitro*⁸⁹. Analysis of RBDmap data for 24 metabolic enzymes
8 (including multiple R-f domain proteins) revealed diverse spatial relationships between the
9 identified RNA contacts and previously characterized catalytically relevant regions. While
10 there is a striking overlap for some examples, in other cases the mapped RNA contacts and
11 the catalytic regions did not appear to overlap. Although this could partly reflect falsely
12 negative assignments by RBDmap, the data also implicate possible roles of RNA in
13 allosteric control or enzyme scaffolding⁴⁶.

14 Many enzymes are regulated allosterically by metabolites⁶⁶. Conceivably, such control
15 could also affect their RNA-binding activity. Furthermore, metabolism and metabolites could
16 control enzyme-RNA interactions via metabolite-driven PTMs. For example, S-
17 glutathionylation blocks the RNA-binding activity of GAPDH¹²². Many metabolic enzymes
18 are acetylated, which requires sufficient concentrations of acetyl-CoA¹²³. More broadly,
19 RBDmap identified RBDs as hotspots for PTMs, including tyrosine phosphorylation,
20 methylation, acetylation and malonylation⁴⁴. Thus, there is considerable scope for cross-talk
21 between cellular metabolism and RNA binding^{66,124}.

22

23 **CHARACTERISATION OF UNORTHODOX RBPs**

24 A key step in characterising the molecular and cellular function of novel RBPs is to identify
25 their RNA targets. Several high-throughput sequencing-based methods to achieve this have

1 emerged in recent years, either using carefully controlled *in vitro* methods¹²⁵ or preserving
2 the context of the living cell¹²⁶. Examples of the latter build on classic approaches such as
3 RNP immunoprecipitation (RIP), potentially augmented by stabilisation of complexes
4 through UV irradiation and/or chemically-induced covalent crosslinks (CLIP). These include
5 photoactivatable-ribonucleoside-enhanced and individual-nucleotide-resolution CLIP (PAR-
6 CLIP⁹⁷ and iCLIP⁹⁸, respectively). Enhanced CLIP (eCLIP; **Box 1**) is a variant of iCLIP with
7 improved sensitivity and specificity^{99,127}. CLIP-type studies have already helped to
8 functionally characterise the biological roles of several unorthodox RBPs. These include
9 metabolic enzymes (*e.g.* HSD17B10)⁴³, regulators of alternative splicing^{128,129}, the E3
10 ubiquitin/ISG15 ligase TRIM25^{130,131}, the nuclear cap binding protein 3 (NCBP3, previously
11 known as C17orf85)¹³², the FAST kinase domain-containing protein 2 (FASTKD2)¹³³,
12 tropomyosin¹³⁴ and others.

13 The 3-hydroxyacyl-CoA dehydrogenase type-2 (HSD17B10) is a mitochondrial enzyme
14 involved in the oxidation of isoleucine, branched-chain fatty acids, and xenobiotics as well
15 as in the metabolism of sex hormones and neuroactive steroids¹³⁵. Mutations in HSD17B10
16 cause a hereditary mitochondrial cardiomyopathy and neuropathy syndrome (OMIM
17 300438). Interestingly, the severity of the disorder does not correlate with the loss of
18 enzymatic activity, and the disease thus may be caused by a non-catalytic function of this
19 protein¹³⁶. A recent report identified HSD17B10 as a component of the mitochondrial
20 RNase P complex, together with mitochondrial ribonuclease P protein 1 (TRMT10C) and 3
21 (MRPP3)¹³⁷. HSD17B10 was identified as an RBP in several RNA interactomes, indicating
22 that it binds directly to RNA⁴³. Investigation of HSD17B10 by iCLIP⁹⁸ revealed an
23 enrichment in mitochondrial RNAs⁴³. Specifically, HSD17B10 preferentially interacts with
24 the 5' ends of 15 of the 22 mitochondrial tRNAs, binding particularly at the D-stem, D-loop,
25 anticodon stem and loop regions of these tRNAs. The results suggest that a metabolic

1 enzyme of the di-nucleotide-binding family, HSD17B10, plays a role in guiding RNase P to
2 the ends of the mitochondrial tRNAs. The mutation R130C causes the classical
3 cardiomyopathy and neuropathy phenotype associated with HSD17B10 dysfunction¹³⁶. This
4 mutant exhibits reduced binding to TRMT10C *in vitro*, although it retains the ability to form
5 tetramers¹³⁸. iCLIP data revealed that the R130C mutant also displays reduced RNA-
6 binding activity⁴³, suggesting that the dysfunctional association with RNA contributes to the
7 disease phenotype.

8 The HeLa RNA interactome catalogued four (out of six) members of the FAST kinase
9 protein family as RBPs³⁹. One of these, the mitochondrial FASTKD2, was recurrently
10 identified in most human and mouse RNA interactomes^{38,39,44-47,61}. Analysis of its binding
11 partners by iCLIP revealed that FASTKD2 selectively associates with mitochondrial
12 transcripts¹³³. CRISPR-mediated depletion of FASTKD2 causes a strong reduction of its
13 binding target 16S mitochondrial rRNA, in agreement with a recent report that FASTKD2 is
14 important for the assembly of the mitochondrial ribosome¹³⁹. Lack of FASTKD2 function
15 also leads to a reduction in the levels of RNAs that it binds to, including those encoding the
16 complex IV components COX1, COX2 and COX3, *cyb* and *nd6* mRNAs as well as 7S RNA
17 and the polyl-tRNA¹³³. A nonsense mutation in FASTKD2 causes a hereditary neurological
18 disorder¹⁴⁰. iCLIP and functional assays thus suggest that this disorder is explained by
19 defects in RNA binding causing altered mitochondrial protein synthesis and
20 metabolism^{133,139}.

21 C17orf85 was catalogued as an RBP by RNA interactome studies³⁹, and more recently
22 identified as a cap-binding protein that localizes to nuclear speckles, and renamed as
23 NCBP3¹³². NCBP2 assembles with NCBP1 forming the canonical cap-binding complex
24 (CBC) that binds to nuclear RNAs and plays important roles in RNA processing and export.
25 NCBP1 knock down induced the expected retention of poly(A)⁺ RNA in the nucleus, and

1 caused defects in cell proliferation¹³². However, almost no effect was observed when
2 NCBP2 was depleted, suggesting that NCBP2 could be replaced by another nuclear factor
3 enabled with similar cap-binding activity. NCBP3 harbours a predicted RRM and this
4 domain suffices for *in vitro* binding to m⁷GTP¹³². Similar to the poly(A)-specific ribonuclease
5 (PARN)¹⁴¹, NCBP3 interaction with m⁷GTP is mediated by a tryptophan and two aspartic
6 acids present at the RRM loops, and the mutation of these residues to alanine impairs the
7 cap-binding activity of NCBP3¹³². Importantly, immunoprecipitation followed by mass
8 spectrometry revealed that, like NCBP2, NCBP3 also interacts with NCBP1 forming part of
9 an alternative nuclear cap-binding complex (CBC)¹³².

10 Several members of the E3 ubiquitin/ISG15 tripartite motif ligase (TRIM) family have
11 recurrently been catalogued as RBPs by RNA interactome studies. This includes
12 TRIM25^{39,45}, which has additionally been validated as an RBP by orthogonal approaches
13 and its RBD was identified by RBDmap^{44,45,130}. TRIM25 harbours an N-terminal zinc finger
14 of the RING family that is important for its ubiquitin ligase activity. Although no canonical
15 RBD can readily be recognized, TRIM25 crosslinks very efficiently to RNA⁴⁵. Recently,
16 TRIM25 has been identified to interact with protein lin-28 homolog A (LIN28A) and terminal
17 uridylyltransferase 4 (TUT4), factors involved in pre-miRNA polyuridylation, enhancing their
18 activity¹³⁰. Because pre-miRNA polyuridylation triggers miRNA decay¹⁴², TRIM25 emerges
19 as a regulator of miRNA biology. TRIM25 also appears to play a role as an RBP in virus-
20 infected cells¹³¹. TRIM25 was identified as a ubiquitin ligase that triggers β -interferon (β -
21 IFN) through stimulation of the antiviral factor RIGI¹⁴³. Notably, dengue virus 2 (DENV2)
22 genomic RNA is processed by the 5' to 3' exoribonuclease 1 (XRN1) until it gets stalled by
23 a pseudoknot present at its 3' region, leading to the production of a shorter subgenomic
24 RNA enabled with pathogenic activity¹⁴⁴. A recent report showed that the role of this
25 subgenomic RNA is to sequester TRIM25 preventing its enhancing activity on RIGI and

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

4 Cyclins regulate the cell cycle by activating cyclin-dependent kinases (CDK). They have
5 also sporadically been identified as unconventional RBPs in RNA interactome capture
6 studies^{38,45,46,56}. The Cyclins B and T are part of the RNA interactome of *Drosophila*
7 embryos, and Cyclin B RNA-binding activity was validated by CLIP-PNK assay⁵⁶. Cyclins
8 (CCN) A2, L1 and T1 were identified in the nuclear RNA-binding proteome of mouse
9 embryonic stem cells⁹². Mice with reduced CCNA2 expression are prone to tumour
10 formation and chromosomal instability due to a predisposition to form lagging chromosomes
11 and chromatin bridges¹⁴⁵. This defect resulted from insufficient expression of the meiotic
12 recombination 11 (Mre11) nuclease, apparently due to impaired translation. CCNA2 directly
13 binds to two evolutionarily conserved regions in the 3'UTR of Mre11 mRNA, which appears
14 to be necessary and sufficient for CCNA2 regulation (**Fig. 5b**)¹⁴⁵. A C-terminal CCNA2
15 fragment lacking CDK-binding is both necessary and sufficient for RNA-binding, and its
16 expression in mouse embryonic fibroblasts restored appropriate Mre11 synthesis.
17 Interestingly, the CCNA2 RNA-interacting region binds eIF4A2. Taken together, the data
18 identify an unexpected, CDK-independent function of CCNA2 as an RNA-binding protein to
19 promote Mre11 mRNA translation, potentially through an interaction with eIF4A2.

20

21 **OUTLOOK**

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

1 nonconformist misfits lacking biological function. Which roles do these new RBPs play?
2 Some may indeed play none, having been discovered on the basis of a biophysical property
3 that mediates above background interaction with RNA without biological consequence for
4 the protein or the RNAs that it interacts with. Quite remarkably, however, the list of
5 unorthodox RBPs that fulfil professional roles continues to grow, as discussed above.
6 Although better known for other biological functions and lacking in classical RBDs, they
7 ‘moonlight’ as RBPs and affect RNA fate, akin to the functions of orthodox RBPs (**Fig. 1a**).
8 It will be illuminating to study these RNA-protein interactions structurally. Likewise, it will be
9 important to decipher the mechanisms by which these RNA-protein interactions are
10 regulated, both by epitranscriptomic changes as well as by posttranslational modifications.
11 An intriguing facet of this is the question of whether and how intrinsically disordered regions
12 that bind RNA contribute to the formation of higher order assemblies by liquid-liquid phase
13 separation and to understand the role of RNA in these transitions.
14 Might unorthodox RBPs be controlled by RNA? We have become accustomed to the view
15 that protein functions are typically modulated by other proteins, but there is ample room for
16 the possibility that the known biological function of a protein be altered by ‘riboregulation’, a
17 change in the protein’s function elicited by interaction with RNA (**Fig. 1b**). The RBPs PKR
18 and RIGI can serve as examples of this class of proteins that could vastly expand. We
19 already know much about RBPs, but future experiments are bound to surprise our intuition.

20

21

22 **ACKNOWLEDGEMENTS**

23 We dedicate this review to the memory of Bernd Fischer, who sadly passed away while this
24 review was in preparation.

1 The authors are grateful to the members of their laboratories for helpful discussions
2 throughout. The authors also acknowledge research funding from the European Research
3 Council ERC-2011-ADG_20110310 (MWH), Medical Research Council Career
4 Development Award MR/L019434/1 (AC), the EMBL Interdisciplinary Postdocs 2
5 EIPOD2/291772 program (TS) and the National Health and Medical Research Council
6 APP1120483 (MWH and TP).

7

8 **Box 1. Technical approaches to study RNA-protein interactions**

9 **System-wide identification of RBPs *in vitro*:** In one approach, a series of immobilised
10 RNA probes are used as bait, exposed to cellular extracts and binding proteins are
11 identified by quantitative mass spectrometry (Q-MS) (**panel a**)³³. Applied to combinations of
12 different mRNA UTRs and human HeLa extract, this yielded a dozen proteins that
13 differentially bound to the baits, several of these were novel RBPs. In a more recent
14 iteration, incubation of a set of pre-miRNA baits with multiple different cell lysates yielded
15 ~180 RBPs with distinct specificities³⁶. In a second approach, arrayed proteins are provided
16 as bait and incubated with labelled cellular RNA. RNA binding is determined by measuring
17 the fluorescence intensity at each individual protein spot, in analogy to DNA microarrays
18 (**panel b**). Two such proteome-wide screens identified 180³⁴ and 68³⁵ yeast RBPs,
19 respectively. In a third approach, purified polyadenylated [poly(A)⁺] cellular RNA was
20 immobilized on oligo(dT) magnetic beads and, after incubation with cell extract, bound
21 proteins were analysed by Q-MS³⁵. 88 mostly highly abundant proteins were identified and
22 22 were known RBPs.

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

1 involving poly(A)⁺ RNA on oligo(dT) beads, and identification of proteins by quantitative
2 mass spectrometry (Q-MS)³⁷ (**panel c**). One study used both conventional crosslinking
3 (cCL) and photoactivatable ribonucleoside-enhanced crosslinking (PAR-CL)⁹⁷. The first
4 relies on the natural excitability of nucleoside bases by 254 nm ultraviolet light, which
5 generates short-lived, free radicals that attack amino acids in close proximity forming
6 covalent bonds¹⁴⁶. By contrast, PAR-CL utilizes the nucleoside analogue 4-thiouridine
7 (4SU), which is taken up by cultured cells and incorporated into nascent RNAs.
8 Crosslinking is then achieved by irradiation with ultraviolet light at 365 nm⁹⁷. Another study
9 also used PAR-CL, combining 4SU and 6-thioguanosine (6SG) labelling³⁸. This latter study
10 further exploited the U to C transitions occurring as a consequence of the crosslinking
11 between 4SU and the RBP to analyse globally the footprints of the RNA interactome on the
12 RNA. The protocol has been adapted to different model systems^{43,48,49,53-57}, can be used to
13 monitor differential association of RBPs with RNA under different physiological conditions or
14 in response to biological cues^{47,56} as well as to identify RBPs in different subcellular
15 compartments⁶¹.

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

1 affinity tag into the endogenous RBP gene, bypassing the need for individual antibodies¹⁴⁷.
2 Together with several hundred antibodies against known RBPs that have now been tested
3 for immunoprecipitation⁷⁴, this provides for a growing list of eCLIP datasets, which are
4 accessible at <https://www.encodeproject.org>. Among the 122 proteins with available eCLIP
5 data, 34 lack classical RBDs.

6
7 **Box 2. Epitranscriptomic control of RNA binding.** Progress has been swiftest with
8 understanding m⁶A function in diverse cellular contexts. The YT521-B homology (YTH)
9 domain proteins were identified as m⁶A 'readers' that specifically bind m⁶A-containing
10 mRNA regions to impact splicing, export, translation, or turnover^{62,63}. The ensuing
11 downstream effects are determined by the specific YTH protein that is recruited and on the
12 mRNA context. The YTH-containing proteins YTHDC1, YTHDC2, YTHDF1, YTHDF2 and
13 YTHDF3 were consistently found in both human and mouse RNA interactome datasets.
14 The RNA-binding proteome of *A. thaliana* contains most plant YTH domain proteins,
15 including cleavage and polyadenylation specificity factor 30 (AtCPSF30), possibly
16 explaining a plant-specific link between m⁶A and mRNA cleavage⁵⁴. When YTHDF1 is
17 recruited to m⁶A sites in the 3'UTR of human mRNAs, it enhances their translation, likely
18 through interactions with subunits of the translation initiation factor eIF3¹⁴⁸. Conversely,
19 when YTHDF2 binds to 5'UTR sites in the nucleus of murine embryonic fibroblasts, it blocks
20 their demethylation in heat shock, thus facilitating selective translation in the cytoplasm by
21 direct binding of eIF3 to these m⁶A sites¹⁴⁹. More generally, cytoplasmic YTHDF2 binds
22 m⁶A-containing mRNAs in human cells and promotes their relocation to processing bodies
23 and degradation¹⁵⁰. The latter function is critically involved in murine embryonic stem cell
24 differentiation¹⁵¹⁻¹⁵³, facilitates maternal mRNA clearance, and the maternal-to-zygotic
25 transition in zebrafish embryos¹⁵⁴. Epitranscriptomic marks can also modulate RBP binding

1 indirectly by affecting RNA structure. For example, the nuclear protein HNRNPC responds
2 to m⁶A-operated 'structural switches' to gain access to thousands of its target sites in
3 human nuclear RNAs¹⁵⁵. HNRNPC preferentially binds to single-stranded U tracts, and m⁶A
4 can destabilise local RNA structure, making the U-tracts more accessible. Thus, when
5 considering the determinants of RNP formation, epitranscriptomic changes need to be
6 considered in addition to posttranslational RBP modifications.

7

1 FIGURES & LEGENDS

2

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

7

8 **Figure 2. Comparison of published RNA interactomes.** The reported lists applying the
9 most stringent criteria for RBPs identified by different studies and from various source
10 materials were extracted, curated and their annotations were updated, and converted to
11 base identifiers. The underlying data are provided in **Table S1**. (a) Supersets of RBPs
12 identified by the combination of RBP detection studies in a given organism. These were
13 compiled for *Mus musculus* (Mm), *Homo sapiens* (Hs), *Saccharomyces cerevisiae* (Sc),
14 *Drosophila melanogaster* (Dm), *Arabidopsis thaliana* (At) and *Caenorhabditis elegans* (Ce).
15 RBP sets for *Danio rerio* (Dr), *Trypanosoma brucei* (Tb), *Leishmania donovani* (Ld) and
16 *Plasmodium falciparum* (Pf) were also added. (b-h) Venn Diagrams and UpSet plots¹⁵⁶
17 showing overlaps, set and intersections sizes between different RBP sets. Differences in
18 technical approaches and data processing (for example, thresholds for detection and
19 reproducibility) will affect these comparisons. (b) RBP repertoires as detected in human
20 source material. RNA interactome capture (RIC; **Box 1**) with either conventional UV
21 crosslinking (cCL) or photoactivatable ribonucleoside-enhanced crosslinking (PAR-CL) was
22 applied to the following cell lines: cervical cancer HeLa³⁹, embryonic kidney HEK293^{38,39},
23 hepatocytic HuH7⁴³ and myeloid leukaemia K562 (serial nuclear RIC)⁶¹. HeLa cells were
24 further subjected to RBDmap⁴⁴ and RNP^{x142}. The human datasets show very high overlap,
25 likely because of the preponderance of 'generic' cell lines as source material and related

1 experimental approaches. (c) Murine RBP repertoires. RIC was applied to primary
2 embryonic fibroblasts (MEF ±etoposide)⁸⁴, embryonic stem cells (mESC, total cell and
3 nuclear fraction), macrophages (RAW264.7 ±lipopolysaccharide)⁴⁷. HL-1 cardiomyocytes
4 were subjected to both RIC and RBDmap⁴⁴. These datasets overlap less well, probably due
5 to the use of more 'idiosyncratic' cell lines and inclusion of drug treatments. (d) Budding
6 yeast RNA-binding proteomes. Two studies used either *in vitro* protein arrays or, in one
7 case, oligo(dT) capture screens to identify RBPs^{34, 35}. Three *in vivo* RNA interactomes were
8 generated, either employing conventional UV crosslinking (cCL)^{48,49} or photoactivatable
9 ribonucleoside-enhanced crosslinking (PAR-CL)⁴³. RNP^{xl} was used with two crosslinking
10 approaches⁴². The diversity of technical approaches likely explains the big differences in
11 coverage and overlap. (e) Fruit fly RBP sets. Two studies applied RNA interactome capture
12 to *Drosophila melanogaster* embryos, using solely cCL⁵⁶ or both cCL and PAR-CL⁵⁷. The
13 focus here were embryos undergoing maternal-to-zygotic transition. This, together with
14 differences in mass spectrometry approaches likely underlies the moderate overlap. (f)
15 RNA interactomes captured from different plant sources, including cell suspension cultures
16 and leaves⁵³, etiolated seedlings⁵⁴ and leaf mesophyll protoplasts⁵⁵. Given the
17 heterogeneous sources, the three datasets agree reasonably well with each other. The
18 lower RBP identification rates suggest UV crosslinking limitations, likely due to the
19 presence of a cell wall and/or UV-absorbing pigments. (g) Pairwise comparisons of
20 InParanoid clusters between human (Hs), mouse (Mm) and yeast (Sc). The intersections
21 between these sets constitute emerging mammalian or eukaryotic 'core' interactomes. (h)
22 UpSet plot showing the overlap between the human superset of RBPs and human
23 orthologues from *Mus musculus* (Mm), *Saccharomyces cerevisiae* (Se), *Drosophila*
24 *melanogaster* (Dm), *Caenorhabditis elegans* (Ce) and *Arabidopsis thaliana* (At).

25

1 **Figure 3. System-wide, high resolution identification of RNA-binding domains.** Each
2 of the three methods uses UV irradiation of live cells to establish covalent bonds at direct
3 contact sites between RNA and protein (indicated by a star). Following cell lysis, the
4 approaches diverge with regard to proteolysis, purification strategy and detection by
5 quantitative mass spectrometry (Q-MS). (a) Purification and direct detection of RNA-
6 crosslinked tryptic peptides^{42,88}. Although simple in principle, this strategy is challenging
7 due to the inefficiency of UV crosslinking and the heterogeneous mass contribution by the
8 nucleic acid remnant, resulting in sub-stoichiometric amounts of peptides with hard-to-
9 predict additional mass. To overcome this, covalently linked protein-RNA complexes are
10 purified on oligo(dT) beads, by performing the initial steps of RNA interactome capture^{37,43}.
11 After digestion with trypsin and RNases, peptides crosslinked to remnants of RNA are
12 further enriched using a TiO₂ matrix¹⁵⁷, prior to analysis by Q-MS and a complex search,
13 performed by a custom-designed software termed RNP^{xl}, for peptide spectra with a defined
14 mass shift caused by the nucleotide remnant. (b) Extrapolation of RNA-protein crosslink
15 sites by RBDmap⁴⁴. Again, RBPs are purified with poly(A)⁺ RNA as per RNA interactome
16 capture^{37,43}, but then digested with a protease that cleaves every 17 amino acids on
17 average (LysC or ArgC), typically leaving peptides that still contain an internal trypsin
18 cleavage site. RNA-linked peptides, termed RBDpep, are then recaptured on oligo(dT)
19 beads, while those distant to the crosslink site are released into the supernatant. Both
20 fractions are digested with trypsin and analysed by Q-MS. For RNA-bound material this
21 leads to: i) a fragment with remnant RNA that will not be identified; and ii) a neighbouring
22 fragment(s) with native mass (N-peptide). N-peptides that are enriched in the RNA-bound
23 fraction are extended *in silico* to the next LysC/ArgC cleavage site to reconstitute the
24 original RBDpep. (c) Proteomic identification of RNA-binding regions (RBR-ID)⁹². This
25 approach directly exploits the mass shift of RNA-crosslinked peptides in conventional MS

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

4 The three approaches have some common limitations. UV crosslinking relies on a
5 favourable geometry, nucleotide and amino acid composition at protein-RNA interfaces.
6 Protein-RNA interactions with the phosphate backbone, for example, are non-optimal in this
7 regard, and will be missed. Mass-spectrometric analyses are also influenced by the
8 abundance of the peptides and their amino acid sequence (influencing optimal peptide size
9 generated derived from tryptic cleavage and optimal mass-to-charge ratio). The RNP^{xl}
10 workflow offers the unique advantage of single amino acid resolution, but at the expense of
11 a limited sensitivity and the need for specialized proteomic analyses. RBDmap is less
12 complex to implement and more sensitive, but at the price of a lower resolution (~17 amino
13 acids). The conceptually very straightforward RBR-ID gives an intermediate resolution (~9
14 amino acids), however, its high intra- and inter-experimental variability call for a sufficiently
15 high number of technical and biological replicates to obtain high confidence results.

16

17 **Figure 4. Modes of RNA binding.** (a) *Schematic representation* of a RBP harbouring a
18 classic RBD (RRM) interacting specifically with a specific RNA sequence in the context of a
19 stem-loop². (b) Depiction of the eukaryotic initiation factor (eIF)4F, which is composed of
20 the cap-binding protein eIF4E, eIF4G and the helicase eIF4A. This complex associates with
21 capped RNA in a sequence-independent manner to enable global initiation of translation¹⁵⁸.
22 (c) The exon junction complex is deposited by its interaction with CWC22 ~20 nucleotides
23 upstream the exon-exon junction immediately after intron removal¹⁰². (d) The disordered
24 RGG-motif present in FMR1 protein co-folds with its target RNA forming a tight electrostatic
25 and shape-complementarity driven interaction¹⁰⁷. (e) The internal ribosome entry site

1 (IRES) of hepatitis C virus (HCV), interacts directly with the ribosome through a complex
2 interaction mode that involves shape-complementarity¹¹⁶. (f) The lncRNA NEAT1 hijacks
3 the RBPs NONO, PSC1 and SFPQ in paraspeckles¹¹⁷. (g) PKR binds to double-stranded
4 (ds)RNA, derived from viral replication. RNA binding promotes protein dimerization and
5 autophosphorylation, activating PKR. In its active form, PKR phosphorylates eIF2 α to block
6 protein synthesis in infected cells¹⁵⁹. (h) Aconitase 1 (ACO1) associates with an iron-
7 sulphur cluster to catalyse the interconversion between citrate and isocitrate. Under low iron
8 conditions, the iron-sulphur cluster is no longer synthesised and the apoprotein, iron
9 regulatory protein 1 (IRP1), binds mRNAs encoding cellular factors involved in iron
10 homeostasis, regulating their fate¹¹⁹.

11

12 **Figure 5. Examples of the biological roles of unorthodox RBPs.** (a) Schematic
13 representation of (a) TRIM25 hijacking by the subgenomic RNA of DENV to reduce β -
14 interferon (β -IFN) synthesis¹³¹. DENV genomic RNA degradation by XRN1 leads to the
15 generation of a subgenomic RNA due to XRN1 stalling at a pseudoknot present at the 3'
16 region of this RNA. The subgenomic DENV RNA recruits TRIM25 but not RIGI, which
17 requires the presence of tri-phosphate 5' ends to interact, thus hijacking TRIM25 in RNPs
18 lacking its molecular partner in the β -IFN pathway, RIGI. (b) Cdk-independent function of
19 CCNA2 as an RNA-binding protein. CCNA2 directly binds to two evolutionarily conserved
20 regions in the 3'UTR of Mre11 mRNA and can to promote Mre11 mRNA translation,
21 potentially through an interaction with eIF4A2¹⁴⁵.

22

23 **GLOSSARY**

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

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

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

6 **HnRNP K homology (KH) domain:** Another common RBD, consisting of ~70 amino acids
7 that fold into three α -helices packed against a three-stranded β -sheet. RNA binds to a
8 hydrophobic cleft formed between two core α -helices and a GXXG loop that interconnects
9 them.

10 **DEAD box protein:** RNA helicases exhibiting two highly similar domains that resemble the
11 bacterial recombinase A. ATP binds in a cleft between the two helicase domains, while
12 RNA binds across both domains on the opposite side. The DEAD box is one of nine
13 conserved sequence motifs in this class of proteins and contains the amino acid sequence
14 Asp-Glu-Ala-Asp (DEAD).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

15

16

17 REFERENCES

18

- 19 1 Dreyfuss, G., Kim, V. N. & Kataoka, N. Messenger-RNA-binding proteins and the
20 messages they carry. *Nat Rev Mol Cell Biol* **3**, 195-205, (2002).
- 21 2 Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for
22 efficient function. *Nat Rev Mol Cell Biol* **8**, 479-490, (2007).
- 23 3 Clery, A., Blatter, M. & Allain, F. H. RNA recognition motifs: boring? Not quite. *Curr*
24 *Opin Struct Biol* **18**, 290-298, (2008).
- 25 4 Valverde, R., Edwards, L. & Regan, L. Structure and function of KH domains. *FEBS*
26 *J* **275**, 2712-2726, (2008).
- 27 5 Linder, P. & Jankowsky, E. From unwinding to clamping - the DEAD box RNA
28 helicase family. *Nat Rev Mol Cell Biol* **12**, 505-516, (2011).
- 29 6 Ramakrishnan, V. The ribosome emerges from a black box. *Cell* **159**, 979-984,
30 (2014).
- 31 7 Steitz, T. A. A structural understanding of the dynamic ribosome machine. *Nat Rev*
32 *Mol Cell Biol* **9**, 242-253, (2008).

1 8 Behrmann, E., Loerke, J., Budkevich, T. V., Yamamoto, K., Schmidt, A., Penczek, P.
2 A., . . . Spahn, C. M. Structural snapshots of actively translating human ribosomes.
3 *Cell* **161**, 845-857, (2015).

4 9 Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol*
5 **15**, 108-121, (2014).

6 10 Papasaikas, P. & Valcarcel, J. The Spliceosome: The Ultimate RNA Chaperone and
7 Sculptor. *Trends Biochem Sci* **41**, 33-45, (2016).

8 11 Plaschka, C., Lin, P. C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature*
9 **546**, 617-621, (2017).

10 12 Jankowsky, E. & Harris, M. E. Specificity and nonspecificity in RNA-protein
11 interactions. *Nat Rev Mol Cell Biol* **16**, 533-544, (2015).

12 13 Gebauer, F., Preiss, T. & Hentze, M. W. From cis-regulatory elements to complex
13 RNPs and back. *Cold Spring Harb Perspect Biol* **4**, a012245, (2012).

14 14 Singh, G., Pratt, G., Yeo, G. W. & Moore, M. J. The Clothes Make the mRNA: Past
15 and Present Trends in mRNP Fashion. *Annu Rev Biochem* **84**, 325-354, (2015).

16 15 Lee, S. R. & Lykke-Andersen, J. Emerging roles for ribonucleoprotein modification
17 and remodeling in controlling RNA fate. *Trends Cell Biol* **23**, 504-510, (2013).

18 16 Chen, C. Y. & Shyu, A. B. Emerging mechanisms of mRNP remodeling regulation.
19 *Wiley Interdiscip Rev RNA* **5**, 713-722, (2014).

20 17 Anderson, P. & Kedersha, N. RNA granules: post-transcriptional and epigenetic
21 modulators of gene expression. *Nat Rev Mol Cell Biol* **10**, 430-436, (2009).

22 18 Buchan, J. R. mRNP granules. Assembly, function, and connections with disease.
23 *RNA Biol* **11**, 1019-1030, (2014).

24 19 Protter, D. S. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell*
25 *Biol* **26**, 668-679, (2016).

26 20 Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and
27 regulation. *Nat Rev Mol Cell Biol* **16**, 18-29, (2015).

28 21 Gloss, B. S. & Dinger, M. E. The specificity of long noncoding RNA expression.
29 *Biochim Biophys Acta* **1859**, 16-22, (2016).

30 22 Geisler, S. & Coller, J. RNA in unexpected places: long non-coding RNA functions in
31 diverse cellular contexts. *Nat Rev Mol Cell Biol* **14**, 699-712, (2013).

32 23 Hudson, W. H. & Ortlund, E. A. The structure, function and evolution of proteins that
33 bind DNA and RNA. *Nat Rev Mol Cell Biol* **15**, 749-760, (2014).

34 24 Cech, T. R. & Steitz, J. A. The noncoding RNA revolution-trashing old rules to forge
35 new ones. *Cell* **157**, 77-94, (2014).

36 25 Beckmann, B. M., Castello, A. & Medenbach, J. The expanding universe of
37 ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions.
38 *Pflugers Arch* **468**, 1029-1040, (2016).

39 26 Gehring, N. H., Wahle, E. & Fischer, U. Deciphering the mRNP Code: RNA-Bound
40 Determinants of Post-Transcriptional Gene Regulation. *Trends Biochem Sci*, (2017).

41 27 Rissland, O. S. The organization and regulation of mRNA-protein complexes. *Wiley*
42 *Interdiscip Rev RNA* **8**, (2017).

43 28 Muller-McNicoll, M. & Neugebauer, K. M. How cells get the message: dynamic
44 assembly and function of mRNA-protein complexes. *Nat Rev Genet* **14**, 275-287,
45 (2013).

46 29 Hentze, M. W. & Argos, P. Homology between IRE-BP, a regulatory RNA-binding
47 protein, aconitase, and isopropylmalate isomerase. *Nucleic Acids Res* **19**, 1739-
48 1740, (1991).

- 1 30 Rouault, T. A., Stout, C. D., Kaptain, S., Harford, J. B. & Klausner, R. D. Structural
2 relationship between an iron-regulated RNA-binding protein (IRE-BP) and aconitase:
3 functional implications. *Cell* **64**, 881-883, (1991).
- 4 31 Chu, E., Voeller, D., Koeller, D. M., Drake, J. C., Takimoto, C. H., Maley, G. F., . . .
5 Allegra, C. J. Identification of an RNA binding site for human thymidylate synthase.
6 *Proc Natl Acad Sci U S A* **90**, 517-521, (1993).
- 7 32 Hentze, M. W. Enzymes as RNA-binding proteins: a role for (di)nucleotide-binding
8 domains? *Trends Biochem Sci* **19**, 101-103, (1994).
- 9 33 Butter, F., Scheibe, M., Morl, M. & Mann, M. Unbiased RNA-protein interaction
10 screen by quantitative proteomics. *Proc Natl Acad Sci U S A* **106**, 10626-10631,
11 (2009).
- 12 34 Scherrer, T., Mittal, N., Janga, S. C. & Gerber, A. P. A screen for RNA-binding
13 proteins in yeast indicates dual functions for many enzymes. *PLoS One* **5**, e15499,
14 (2010).
- 15 35 Tsvetanova, N. G., Klass, D. M., Salzman, J. & Brown, P. O. Proteome-wide search
16 reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*. *PLoS One*
17 **5**, e12671., (2010).
- 18 36 Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiss, J. L., Eichner, N., . . .
19 Meister, G. A Compendium of RNA-Binding Proteins that Regulate MicroRNA
20 Biogenesis. *Mol Cell* **66**, 270-284 e213, (2017).
- 21 37 Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L. M., . . .
22 Hentze, M. W. System-wide identification of RNA-binding proteins by interactome
23 capture. *Nat Protoc* **8**, 491-500, (2013).
- 24 38 Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y.,
25 Schueler, M., . . . Landthaler, M. The mRNA-Bound Proteome and Its Global
26 Occupancy Profile on Protein-Coding Transcripts. *Mol Cell* **46**, 674-690, (2012).
- 27 39 Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., . . .
28 Hentze, M. W. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding
29 Proteins. *Cell* **149**, 1393-1406, (2012).
- 30 40 Castello, A., Fischer, B., Hentze, M. W. & Preiss, T. RNA-binding proteins in
31 Mendelian disease. *Trends Genet* **29**, 318-327, (2013).
- 32 41 Strein, C., Alleaume, A. M., Rothbauer, U., Hentze, M. W. & Castello, A. A versatile
33 assay for RNA-binding proteins in living cells. *RNA* **20**, 721-731, (2014).
- 34 42 Kramer, K., Sachsenberg, T., Beckmann, B. M., Qamar, S., Boon, K. L., Hentze, M.
35 W., . . . Urlaub, H. Photo-cross-linking and high-resolution mass spectrometry for
36 assignment of RNA-binding sites in RNA-binding proteins. *Nat Methods* **11**, 1064-
37 1070, (2014).
- 38 43 Beckmann, B. M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.
39 M., . . . Hentze, M. W. The RNA-binding proteomes from yeast to man harbour
40 conserved enigmRBPs. *Nat Commun* **6**, 10127, (2015).
- 41 44 Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A. M., Foehr, S., . . .
42 Hentze, M. W. Comprehensive Identification of RNA-Binding Domains in Human
43 Cells. *Mol Cell* **63**, 696-710, (2016).
- 44 45 Kwon, S. C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K. T., . . . Kim, V. N.
45 The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol* **20**,
46 1122-1130, (2013).
- 47 46 Liao, Y., Castello, A., Fischer, B., Leicht, S., Foehr, S., Frese, C. K., . . . Preiss, T.
48 The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and
49 Heart Disease. *Cell Rep* **16**, 1456-1469, (2016).

1 47 Liepelt, A., Naarmann-de Vries, I. S., Simons, N., Eichelbaum, K., Foehr, S., Archer,
2 S. K., . . . Ostareck-Lederer, A. Identification of RNA-binding proteins in
3 macrophages by interactome capture. *Mol Cell Proteomics* **15**, 2699-2674, (2016).

4 48 Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat*
5 *Struct Mol Biol* **20**, 127-133, (2013).

6 49 Matia-Gonzalez, A. M., Laing, E. E. & Gerber, A. P. Conserved mRNA-binding
7 proteomes in eukaryotic organisms. *Nat Struct Mol Biol* **22**, 1027-1033, (2015).

8 50 Nandan, D., Thomas, S. A., Nguyen, A., Moon, K. M., Foster, L. J. & Reiner, N. E.
9 Comprehensive Identification of mRNA-Binding Proteins of *Leishmania donovani* by
10 Interactome Capture. *PLoS One* **12**, e0170068, (2017).

11 51 Bunnik, E. M., Batugedara, G., Saraf, A., Prudhomme, J., Florens, L. & Le Roch, K.
12 G. The mRNA-bound proteome of the human malaria parasite *Plasmodium*
13 *falciparum*. *Genome Biol* **17**, 147, (2016).

14 52 Lueong, S., Merce, C., Fischer, B., Hoheisel, J. D. & Erben, E. D. Gene expression
15 regulatory networks in *Trypanosoma brucei*: insights into the role of the mRNA-
16 binding proteome. *Mol Microbiol* **100**, 457-471, (2016).

17 53 Marondedze, C., Thomas, L., Serrano, N. L., Lilley, K. S. & Gehring, C. The RNA-
18 binding protein repertoire of *Arabidopsis thaliana*. *Sci Rep* **6**, 29766, (2016).

19 54 Reichel, M., Liao, Y., Rettel, M., Ragan, C., Evers, M., Alleaume, A. M., . . . Millar, A.
20 A. In Planta Determination of the mRNA-Binding Proteome of *Arabidopsis* Etiolated
21 Seedlings. *Plant Cell* **28**, 2435-2452, (2016).

22 55 Zhang, Z., Boonen, K., Ferrari, P., Schoofs, L., Janssens, E., van Noort, V., . . .
23 Geuten, K. UV crosslinked mRNA-binding proteins captured from leaf mesophyll
24 protoplasts. *Plant Methods* **12**, 42, (2016).

25 56 Sysoev, V. O., Fischer, B., Frese, C. K., Gupta, I., Krijgsveld, J., Hentze, M. W., . . .
26 Ephrussi, A. Global changes of the RNA-bound proteome during the maternal-to-
27 zygotic transition in *Drosophila*. *Nat Commun* **7**, 12128, (2016).

28 57 Wessels, H. H., Imami, K., Baltz, A. G., Kolinski, M., Beldovskaya, A., Selbach, M., .
29 . . Landthaler, M. The mRNA-bound proteome of the early fly embryo. *Genome Res*
30 **26**, 1000-1009, (2016).

31 58 Despic, V., Dejung, M., Gu, M., Krishnan, J., Zhang, J., Herzog, L., . . . Neugebauer,
32 K. M. Dynamic RNA-protein interactions underlie the zebrafish maternal-to-zygotic
33 transition. *Genome Res*, (2017).

34 59 Sonnhammer, E. L. & Ostlund, G. InParanoid 8: orthology analysis between 273
35 proteomes, mostly eukaryotic. *Nucleic Acids Res* **43**, D234-239, (2015).

36 60 O'Brien, K. P., Remm, M. & Sonnhammer, E. L. Inparanoid: a comprehensive
37 database of eukaryotic orthologs. *Nucleic Acids Res* **33**, D476-480, (2005).

38 61 Conrad, T., Albrecht, A. S., de Melo Costa, V. R., Sauer, S., Meierhofer, D. & Orom,
39 U. A. Serial interactome capture of the human cell nucleus. *Nat Commun* **7**, 11212,
40 (2016).

41 62 Gilbert, W. V., Bell, T. A. & Schaening, C. Messenger RNA modifications: Form,
42 distribution, and function. *Science* **352**, 1408-1412, (2016).

43 63 Hoernes, T. P. & Erlacher, M. D. Translating the epitranscriptome. *Wiley Interdiscip*
44 *Rev RNA* **8**, (2017).

45 64 Ciesla, J. Metabolic enzymes that bind RNA: yet another level of cellular regulatory
46 network? *Acta Biochim Pol* **53**, 11-32, (2006).

47 65 Hentze, M. W. & Preiss, T. The REM phase of gene regulation. *Trends Biochem Sci*
48 **35**, 423-426, (2010).

1 66 Castello, A., Hentze, M. W. & Preiss, T. Metabolic Enzymes Enjoying New
2 Partnerships as RNA-Binding Proteins. *Trends Endocrinol Metab* **26**, 746-757,
3 (2015).

4 67 Chu, E. & Allegra, C. J. The role of thymidylate synthase as an RNA binding protein.
5 *Bioessays* **18**, 191-198, (1996).

6 68 Liu, J., Schmitz, J. C., Lin, X., Tai, N., Yan, W., Farrell, M., . . . Chu, E. Thymidylate
7 synthase as a translational regulator of cellular gene expression. *Biochim Biophys*
8 *Acta* **1587**, 174-182, (2002).

9 69 Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins.
10 *Nat Rev Genet* **15**, 829-845, (2014).

11 70 ElAntak, L., Tzakos, A. G., Locker, N. & Lukavsky, P. J. Structure of eIF3b RNA
12 recognition motif and its interaction with eIF3j: structural insights into the recruitment
13 of eIF3b to the 40 S ribosomal subunit. *J Biol Chem* **282**, 8165-8174, (2007).

14 71 Bono, F., Ebert, J., Lorentzen, E. & Conti, E. The crystal structure of the exon
15 junction complex reveals how it maintains a stable grip on mRNA. *Cell* **126**, 713-725,
16 (2006).

17 72 Klass, D. M., Scheibe, M., Butter, F., Hogan, G. J., Mann, M. & Brown, P. O.
18 Quantitative proteomic analysis reveals concurrent RNA-protein interactions and
19 identifies new RNA-binding proteins in *Saccharomyces cerevisiae*. *Genome Res* **23**,
20 1028-1038, (2013).

21 73 Brannan, K. W., Jin, W., Huelga, S. C., Banks, C. A., Gilmore, J. M., Florens, L., . . .
22 Yeo, G. W. SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale
23 Protein-Protein Interactomes. *Mol Cell* **64**, 282-293, (2016).

24 74 Sundararaman, B., Zhan, L., Blue, S. M., Stanton, R., Elkins, K., Olson, S., . . . Yeo,
25 G. W. Resources for the Comprehensive Discovery of Functional RNA Elements.
26 *Mol Cell* **61**, 903-913, (2016).

27 75 Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., . . . Gygi,
28 S. P. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell*
29 **162**, 425-440, (2015).

30 76 Mukhopadhyay, R., Ray, P. S., Arif, A., Brady, A. K., Kinter, M. & Fox, P. L. DAPK-
31 ZIPK-L13a axis constitutes a negative-feedback module regulating inflammatory
32 gene expression. *Mol Cell* **32**, 371-382, (2008).

33 77 Arif, A., Jia, J., Mukhopadhyay, R., Willard, B., Kinter, M. & Fox, P. L. Two-site
34 phosphorylation of EPRS coordinates multimodal regulation of noncanonical
35 translational control activity. *Mol Cell* **35**, 164-180, (2009).

36 78 Arif, A., Jia, J., Moodt, R. A., DiCorleto, P. E. & Fox, P. L. Phosphorylation of
37 glutamyl-prolyl tRNA synthetase by cyclin-dependent kinase 5 dictates transcript-
38 selective translational control. *Proc Natl Acad Sci U S A* **108**, 1415-1420, (2011).

39 79 Clingman, C. C., Deveau, L. M., Hay, S. A., Genga, R. M., Shandilya, S. M., Massi,
40 F. & Ryder, S. P. Allosteric inhibition of a stem cell RNA-binding protein by an
41 intermediary metabolite. *eLife*, e02848, (2014).

42 80 Haghghat, A. & Sonenberg, N. eIF4G dramatically enhances the binding of eIF4E to
43 the mRNA 5'-cap structure. *J Biol Chem* **272**, 21677-21680, (1997).

44 81 Dabo, S. & Meurs, E. F. dsRNA-dependent protein kinase PKR and its role in stress,
45 signaling and HCV infection. *Viruses* **4**, 2598-2635, (2012).

46 82 Habjan, M. & Pichlmair, A. Cytoplasmic sensing of viral nucleic acids. *Curr Opin Virol*
47 **11**, 31-37, (2015).

48 83 Rehwinkel, J. RNA sensing: the more RIG-I the merrier? *EMBO Rep* **14**, 751-752,
49 (2013).

- 1 84 Boucas, J., Fritz, C., Schmitt, A., Riabinska, A., Thelen, L., Peifer, M., . . . Reinhardt,
2 H. C. Label-Free Protein-RNA Interactome Analysis Identifies Khsp Signaling
3 Downstream of the p38/Mk2 Kinase Complex as a Critical Modulator of Cell Cycle
4 Progression. *PLoS One* **10**, e0125745, (2015).
- 5 85 Lasko, P. Posttranscriptional regulation in *Drosophila* oocytes and early embryos.
6 *Wiley Interdiscip Rev RNA* **2**, 408-416, (2011).
- 7 86 Altelaar, A. F., Frese, C. K., Preisinger, C., Hennrich, M. L., Schram, A. W.,
8 Timmers, H. T., . . . Mohammed, S. Benchmarking stable isotope labeling based
9 quantitative proteomics. *J Proteomics* **88**, 14-26, (2013).
- 10 87 Fernandez-Chamorro, J., Pineiro, D., Gordon, J. M., Ramajo, J., Francisco-Velilla,
11 R., Macias, M. J. & Martinez-Salas, E. Identification of novel non-canonical RNA-
12 binding sites in Gemin5 involved in internal initiation of translation. *Nucleic Acids Res*
13 **42**, 5742-5754, (2014).
- 14 88 Schmidt, C., Kramer, K. & Urlaub, H. Investigation of protein-RNA interactions by
15 mass spectrometry--Techniques and applications. *J Proteomics* **75**, 3478-3494,
16 (2012).
- 17 89 Nagy, E. & Rigby, W. F. Glyceraldehyde-3-phosphate dehydrogenase selectively
18 binds AU-rich RNA in the NAD(+)-binding region (Rossmann fold). *J Biol Chem* **270**,
19 2755-2763, (1995).
- 20 90 Chang, C. H., Curtis, J. D., Maggi, L. B., Jr., Faubert, B., Villarino, A. V., O'Sullivan,
21 D., . . . Pearce, E. L. Posttranscriptional control of T cell effector function by aerobic
22 glycolysis. *Cell* **153**, 1239-1251, (2013).
- 23 91 Mullari, M., Lyon, D., Jensen, L. J. & Nielsen, M. L. Specifying RNA-Binding Regions
24 in Proteins by Peptide Cross-Linking and Affinity Purification. *J Proteome Res* **16**,
25 2762-2772, (2017).
- 26 92 He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B.
27 A. & Bonasio, R. High-Resolution Mapping of RNA-Binding Regions in the Nuclear
28 Proteome of Embryonic Stem Cells. *Mol Cell* **64**, 416-430, (2016).
- 29 93 Kaneko, S., Li, G., Son, J., Xu, C. F., Margueron, R., Neubert, T. A. & Reinberg, D.
30 Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-
31 regulates its binding to ncRNA. *Genes Dev* **24**, 2615-2620, (2010).
- 32 94 Kaneko, S., Son, J., Bonasio, R., Shen, S. S. & Reinberg, D. Nascent RNA
33 interaction keeps PRC2 activity poised and in check. *Genes Dev* **28**, 1983-1988,
34 (2014).
- 35 95 Kaneko, S., Son, J., Shen, S. S., Reinberg, D. & Bonasio, R. PRC2 binds active
36 promoters and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol*
37 **20**, 1258-1264, (2013).
- 38 96 Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., . . .
39 Hughes, T. R. A compendium of RNA-binding motifs for decoding gene regulation.
40 *Nature* **499**, 172-177, (2013).
- 41 97 Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., . . .
42 Tuschl, T. Transcriptome-wide identification of RNA-binding protein and microRNA
43 target sites by PAR-CLIP. *Cell* **141**, 129-141, (2010).
- 44 98 König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., . . . Ule, J. iCLIP
45 reveals the function of hnRNP particles in splicing at individual nucleotide resolution.
46 *Nat Struct Mol Biol* **17**, 909-915, (2010).
- 47 99 Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y.,
48 Sundararaman, B., . . . Yeo, G. W. Robust transcriptome-wide discovery of RNA-
49 binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508-514,
50 (2016).

1 100 Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P. A. & Burge, C. B.
2 RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding
3 specificity of RNA binding proteins. *Mol Cell* **54**, 887-900, (2014).

4 101 Riley, K. J. & Steitz, J. A. The "Observer Effect" in genome-wide surveys of protein-
5 RNA interactions. *Mol Cell* **49**, 601-604, (2013).

6 102 Steckelberg, A. L., Boehm, V., Gromadzka, A. M. & Gehring, N. H. CWC22 connects
7 pre-mRNA splicing and exon junction complex assembly. *Cell Rep* **2**, 454-461,
8 (2012).

9 103 Haberman, N., Huppertz, I., Attig, J., Konig, J., Wang, Z., Hauer, C., . . . Ule, J.
10 Insights into the design and interpretation of iCLIP experiments. *Genome Biol* **18**, 7,
11 (2017).

12 104 Jarvelin, A. I., Noerenberg, M., Davis, I. & Castello, A. The new (dis)order in RNA
13 regulation. *Cell Commun Signal* **14**, 9, (2016).

14 105 Basu, S. & Bahadur, R. P. A structural perspective of RNA recognition by intrinsically
15 disordered proteins. *Cell Mol Life Sci* **73**, 4075-4084, (2016).

16 106 Thandapani, P., O'Connor, T. R., Bailey, T. L. & Richard, S. Defining the RGG/RG
17 motif. *Mol Cell* **50**, 613-623, (2013).

18 107 Phan, A. T., Kuryavyi, V., Darnell, J. C., Serganov, A., Majumdar, A., Ilin, S., . . .
19 Patel, D. J. Structure-function studies of FMRP RGG peptide recognition of an RNA
20 duplex-quadruplex junction. *Nat Struct Mol Biol* **18**, 796-804, (2011).

21 108 Han, T. W., Kato, M., Xie, S., Wu, L. C., Mirzaei, H., Pei, J., . . . McKnight, S. L. Cell-
22 free formation of RNA granules: bound RNAs identify features and components of
23 cellular assemblies. *Cell* **149**, 768-779, (2012).

24 109 Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., . . . McKnight, S. L. Cell-free
25 formation of RNA granules: low complexity sequence domains form dynamic fibers
26 within hydrogels. *Cell* **149**, 753-767, (2012).

27 110 Vuzman, D., Azia, A. & Levy, Y. Searching DNA via a "Monkey Bar" mechanism: the
28 significance of disordered tails. *J Mol Biol* **396**, 674-684, (2010).

29 111 Casu, F., Duggan, B. M. & Hennig, M. The arginine-rich RNA-binding motif of HIV-1
30 Rev is intrinsically disordered and folds upon RRE binding. *Biophys J* **105**, 1004-
31 1017, (2013).

32 112 Gold, L., Janjic, N., Jarvis, T., Schneider, D., Walker, J. J., Wilcox, S. K. & Zichi, D.
33 Aptamers and the RNA world, past and present. *Cold Spring Harb Perspect Biol* **4**,
34 (2012).

35 113 Shi, H. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å
36 resolution: a classic structure revisited. *RNA* **6**, 1091-1105, (2000).

37 114 de Breyne, S., Yu, Y., Unbehauen, A., Pestova, T. V. & Hellen, C. U. Direct functional
38 interaction of initiation factor eIF4G with type 1 internal ribosomal entry sites. *Proc*
39 *Natl Acad Sci U S A* **106**, 9197-9202, (2009).

40 115 Sizova, D. V., Kolupaeva, V. G., Pestova, T. V., Shatsky, I. N. & Hellen, C. U.
41 Specific interaction of eukaryotic translation initiation factor 3 with the 5'
42 nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. *J*
43 *Virol* **72**, 4775-4782, (1998).

44 116 Quade, N., Boehringer, D., Leibundgut, M., van den Heuvel, J. & Ban, N. Cryo-EM
45 structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution.
46 *Nat Commun* **6**, 7646, (2015).

47 117 Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess,
48 A. & Lawrence, J. B. An architectural role for a nuclear noncoding RNA: NEAT1 RNA
49 is essential for the structure of paraspeckles. *Mol Cell* **33**, 717-726, (2009).

- 1 118 Chu, E., Koeller, D. M., Casey, J. L., Drake, J. C., Chabner, B. A., Elwood, P. C., . . .
2 Allegra, C. J. Autoregulation of human thymidylate synthase messenger RNA
3 translation by thymidylate synthase. *Proc Natl Acad Sci U S A* **88**, 8977-8981,
4 (1991).
- 5 119 Muckenthaler, M. U., Rivella, S., Hentze, M. W. & Galy, B. A Red Carpet for Iron
6 Metabolism. *Cell* **168**, 344-361, (2017).
- 7 120 Walden, W. E., Selezneva, A. I., Dupuy, J., Volbeda, A., Fontecilla-Camps, J. C.,
8 Theil, E. C. & Volz, K. Structure of dual function iron regulatory protein 1 complexed
9 with ferritin IRE-RNA. *Science* **314**, 1903-1908, (2006).
- 10 121 Tristan, C., Shahani, N., Sedlak, T. W. & Sawa, A. The diverse functions of GAPDH:
11 views from different subcellular compartments. *Cell Signal* **23**, 317-323, (2011).
- 12 122 Rodriguez-Pascual, F., Redondo-Horcajo, M., Magan-Marchal, N., Lagares, D.,
13 Martinez-Ruiz, A., Kleinert, H. & Lamas, S. Glyceraldehyde-3-phosphate
14 dehydrogenase regulates endothelin-1 expression by a novel, redox-sensitive
15 mechanism involving mRNA stability. *Mol Cell Biol* **28**, 7139-7155, (2008).
- 16 123 Xing, S. & Poirier, Y. The protein acetylome and the regulation of metabolism.
17 *Trends Plant Sci* **17**, 423-430, (2012).
- 18 124 Arif, W., Datar, G. & Kalsotra, A. Intersections of post-transcriptional gene regulatory
19 mechanisms with intermediary metabolism. *Biochim Biophys Acta* **1860**, 349-362,
20 (2017).
- 21 125 Campbell, Z. T. & Wickens, M. Probing RNA-protein networks: biochemistry meets
22 genomics. *Trends Biochem Sci* **40**, 157-164, (2015).
- 23 126 König, J., Zarnack, K., Luscombe, N. M. & Ule, J. Protein-RNA interactions: new
24 genomic technologies and perspectives. *Nat Rev Genet* **13**, 77-83, (2011).
- 25 127 Nostrand, E. L., Van, F., P., Pratt, G. A., Wang, X., Wei, X., Blue, S. M. & Yeo, G. W.
26 A Large-Scale Binding and Functional Map of Human RNA Binding Proteins. *bioRxiv*
27 **179648**, (2017).
- 28 128 Papasaikas, P., Tejedor, J. R., Vigevani, L. & Valcarcel, J. Functional splicing
29 network reveals extensive regulatory potential of the core spliceosomal machinery.
30 *Mol Cell* **57**, 7-22, (2015).
- 31 129 Tejedor, J. R., Papasaikas, P. & Valcarcel, J. Genome-wide identification of
32 Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. *Mol Cell*
33 **57**, 23-38, (2015).
- 34 130 Choudhury, N. R., Nowak, J. S., Zuo, J., Rappsilber, J., Spoel, S. H. & Michlewski,
35 G. Trim25 Is an RNA-Specific Activator of Lin28a/TuT4-Mediated Uridylation. *Cell*
36 *Rep* **9**, 1265-1272, (2014).
- 37 131 Manokaran, G., Finol, E., Wang, C., Gunaratne, J., Bahl, J., Ong, E. Z., . . . Ooi, E.
38 E. Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for
39 epidemiological fitness. *Science* **350**, 217-221, (2015).
- 40 132 Gebhardt, A., Habjan, M., Benda, C., Meiler, A., Haas, D. A., Hein, M. Y., . . .
41 Pichlmair, A. mRNA export through an additional cap-binding complex consisting of
42 NCBP1 and NCBP3. *Nat Commun* **6**, 8192, (2015).
- 43 133 Popow, J., Alleaume, A. M., Curk, T., Schwarzl, T., Sauer, S. & Hentze, M. W.
44 FASTKD2 is an RNA-binding protein required for mitochondrial RNA processing and
45 translation. *RNA* **21**, 1873-1884, (2015).
- 46 134 Gaspar, I., Sysoev, V., Komissarov, A. & Ephrussi, A. An RNA-binding atypical
47 tropomyosin recruits kinesin-1 dynamically to oskar mRNPs. *EMBO J* **36**, 319-333,
48 (2017).

- 1 135 Yang, S. Y., He, X. Y., Isaacs, C., Dobkin, C., Miller, D. & Philipp, M. Roles of
2 17beta-hydroxysteroid dehydrogenase type 10 in neurodegenerative disorders. *J*
3 *Steroid Biochem Mol Biol* **143**, 460-472, (2014).
- 4 136 Rauschenberger, K., Scholer, K., Sass, J. O., Sauer, S., Djuric, Z., Rumig, C., . . .
5 Zschocke, J. A non-enzymatic function of 17beta-hydroxysteroid dehydrogenase
6 type 10 is required for mitochondrial integrity and cell survival. *EMBO Mol Med* **2**, 51-
7 62, (2010).
- 8 137 Holzmann, J., Frank, P., Loffler, E., Bennett, K. L., Gerner, C. & Rossmann, W.
9 RNase P without RNA: identification and functional reconstitution of the human
10 mitochondrial tRNA processing enzyme. *Cell* **135**, 462-474, (2008).
- 11 138 Vilardo, E. & Rossmann, W. Molecular insights into HSD10 disease: impact of
12 SDR5C1 mutations on the human mitochondrial RNase P complex. *Nucleic Acids*
13 *Res* **43**, 6649, (2015).
- 14 139 Antonicka, H. & Shoubridge, E. A. Mitochondrial RNA Granules Are Centers for
15 Posttranscriptional RNA Processing and Ribosome Biogenesis. *Cell Rep*, (2015).
- 16 140 Ghezzi, D., Saada, A., D'Adamo, P., Fernandez-Vizarra, E., Gasparini, P., Tiranti, V.,
17 . . . Zeviani, M. FASTKD2 nonsense mutation in an infantile mitochondrial
18 encephalomyopathy associated with cytochrome c oxidase deficiency. *Am J Hum*
19 *Genet* **83**, 415-423, (2008).
- 20 141 Nagata, T., Suzuki, S., Endo, R., Shirouzu, M., Terada, T., Inoue, M., . . . Yokoyama,
21 S. The RRM domain of poly(A)-specific ribonuclease has a noncanonical binding site
22 for mRNA cap analog recognition. *Nucleic Acids Res* **36**, 4754-4767, (2008).
- 23 142 Faehle, C. R., Walleshauser, J. & Joshua-Tor, L. Mechanism of Dis3l2 substrate
24 recognition in the Lin28-let-7 pathway. *Nature* **514**, 252-256, (2014).
- 25 143 Gack, M. U., Shin, Y. C., Joo, C. H., Urano, T., Liang, C., Sun, L., . . . Jung, J. U.
26 TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral
27 activity. *Nature* **446**, 916-920, (2007).
- 28 144 Chapman, E. G., Costantino, D. A., Rabe, J. L., Moon, S. L., Wilusz, J., Nix, J. C. &
29 Kieft, J. S. The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA)
30 production. *Science* **344**, 307-310, (2014).
- 31 145 Kanakkanthara, A., Jeganathan, K. B., Limzerwala, J. F., Baker, D. J., Hamada, M.,
32 Nam, H. J., . . . van Deursen, J. M. Cyclin A2 is an RNA binding protein that controls
33 Mre11 mRNA translation. *Science* **353**, 1549-1552, (2016).
- 34 146 Pashev, I. G., Dimitrov, S. I. & Angelov, D. Crosslinking proteins to nucleic acids by
35 ultraviolet laser irradiation. *Trends Biochem Sci* **16**, 323-326, (1991).
- 36 147 Van Nostrand, E. L., Gelboin-Burkhart, C., Wang, R., Pratt, G. A., Blue, S. M. & Yeo,
37 G. W. CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously
38 tagged RNA binding proteins. *Methods* **118**, 50-59, (2016).
- 39 148 Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., . . . He, C. N(6)-
40 methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388-
41 1399, (2015).
- 42 149 Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S. R. & Qian, S. B. Dynamic m(6)A
43 mRNA methylation directs translational control of heat shock response. *Nature* **526**,
44 591-594, (2015).
- 45 150 Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., . . . He, C. N6-
46 methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**,
47 117-120, (2014).
- 48 151 Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z. & Zhao, J. C. N6-
49 methyladenosine modification destabilizes developmental regulators in embryonic
50 stem cells. *Nat Cell Biol* **16**, 191-198, (2014).

1 152 Batista, P. J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., . . . Chang, H. Y. m(6)A
2 RNA modification controls cell fate transition in mammalian embryonic stem cells.
3 *Cell Stem Cell* **15**, 707-719, (2014).

4 153 Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N.,
5 Salmon-Divon, M., . . . Hanna, J. H. Stem cells. m6A mRNA methylation facilitates
6 resolution of naive pluripotency toward differentiation. *Science* **347**, 1002-1006,
7 (2015).

8 154 Zhao, B. S., Wang, X., Beadell, A. V., Lu, Z., Shi, H., Kuuspalu, A., . . . He, C. m6A-
9 dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic
10 transition. *Nature* **542**, 475-478, (2017).

11 155 Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M. & Pan, T. N(6)-methyladenosine-
12 dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**,
13 560-564, (2015).

14 156 Lex, A., Gehlenborg, N., Strobel, H., Vuilleumot, R. & Pfister, H. UpSet: Visualization
15 of Intersecting Sets. *IEEE Trans Vis Comput Graph* **20**, 1983-1992, (2014).

16 157 Altelaar, A. F., Munoz, J. & Heck, A. J. Next-generation proteomics: towards an
17 integrative view of proteome dynamics. *Nat Rev Genet* **14**, 35-48, (2013).

18 158 Prevot, D., Darlix, J. L. & Ohlmann, T. Conducting the initiation of protein synthesis:
19 the role of eIF4G. *Biol Cell* **95**, 141-156, (2003).

20 159 Balachandran, S. & Barber, G. N. PKR in innate immunity, cancer, and viral
21 oncolysis. *Methods Mol Biol* **383**, 277-301, (2007).

22

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 RBPs¹ (a) or with SONAR-identified RBPs² (b). (c) Venn diagram showing the overlap between the *Saccharomyces cerevisiae* RNA interactome superset³⁻⁵ with the protein array-based, in vitro approach to identify RBPs^{6,7} and SONAR-identified RBPs in yeast². (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}.

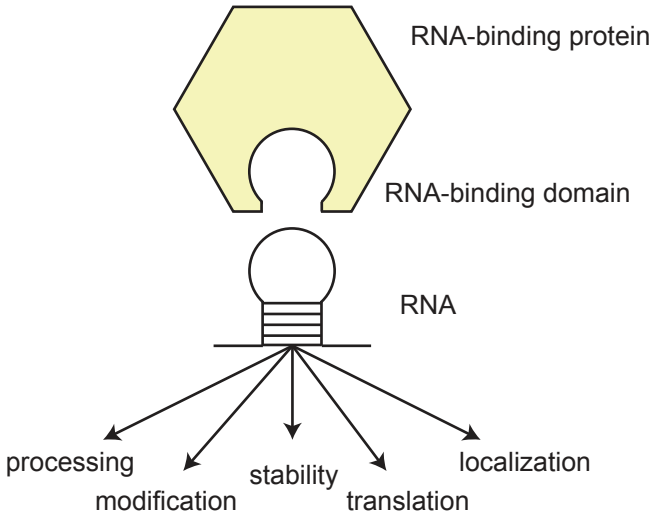
Supplementary Table 1. Curated table of RBP supersets identified in *Homo sapiens*^{4,13-17}, *Mus musculus*^{8,9,11,12}, *Saccharomyces cerevisiae*³⁻⁵, *Drosophila melanogaster*^{18,19}, *Arabidopsis thaliana*²⁰⁻²² and *Caenorhabditis elegans*⁵ using RNA interactome capture. RNA-binding proteomes generated 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., Schwanhauser, 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

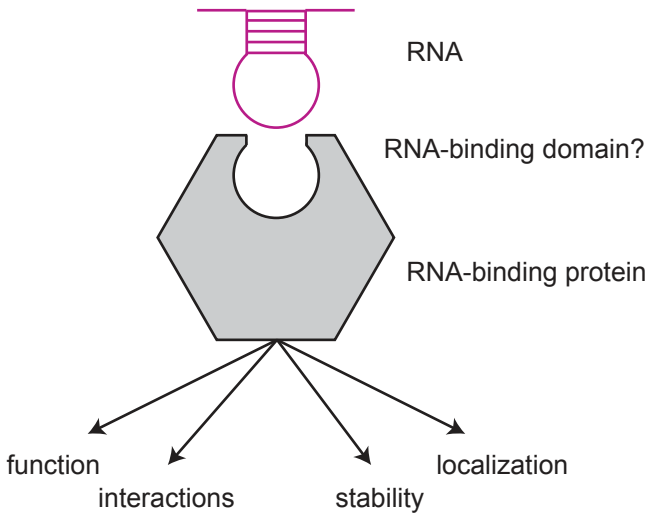
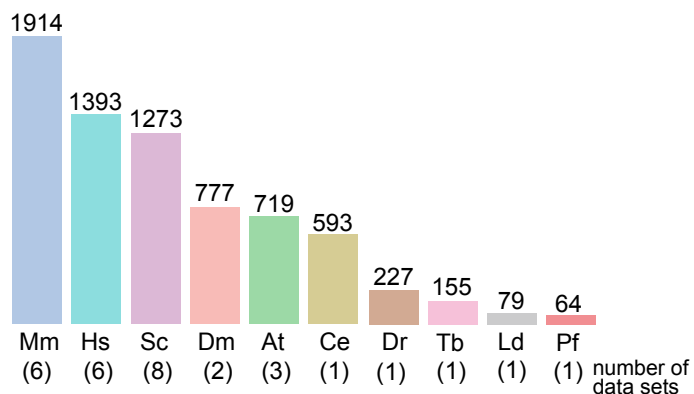
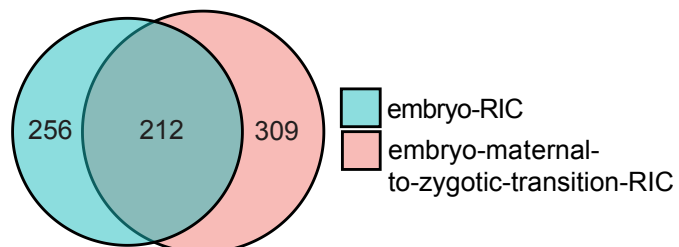


Figure 1

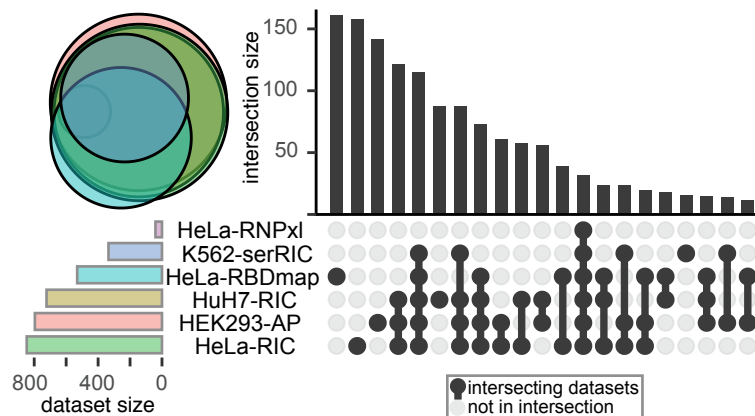
a Identified RNA-binding proteins



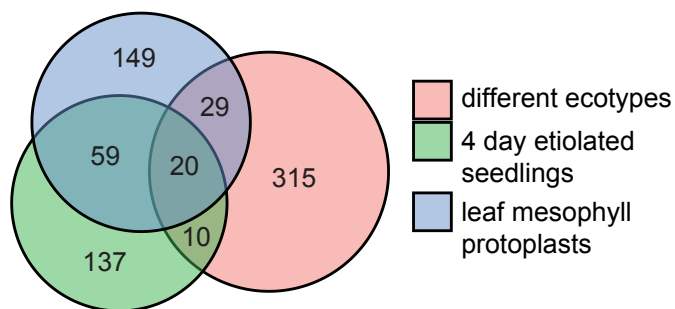
e *Drosophila melanogaster* RBPs



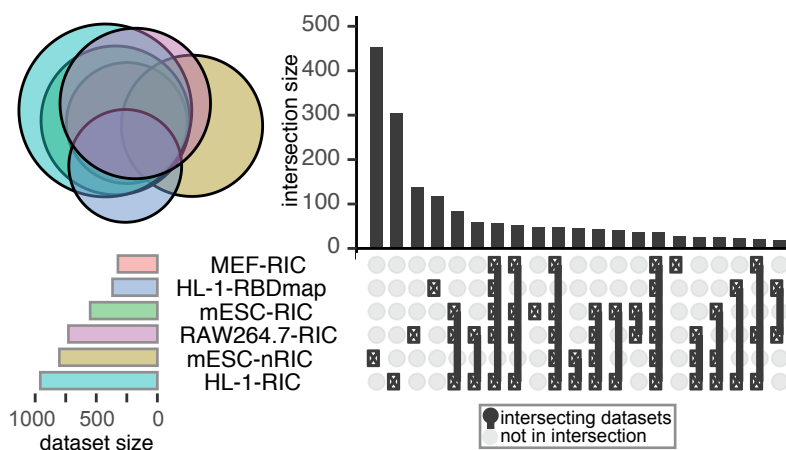
b *Homo sapiens* RBPs



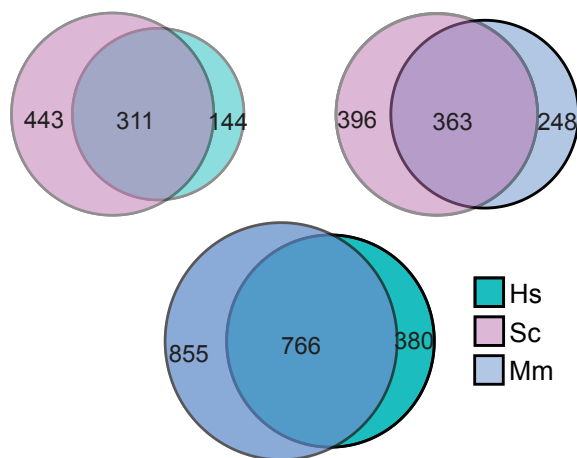
f *Arabidopsis thaliana* RBPs



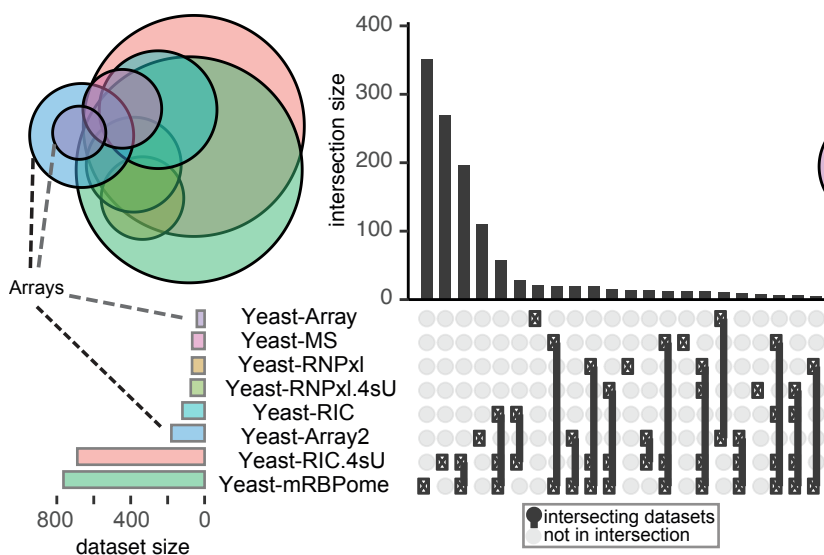
c *Mus musculus* RBPs



g InParanoid 6 cluster comparison



d *Saccharomyces cerevisiae* RBPs



h Human orthologues

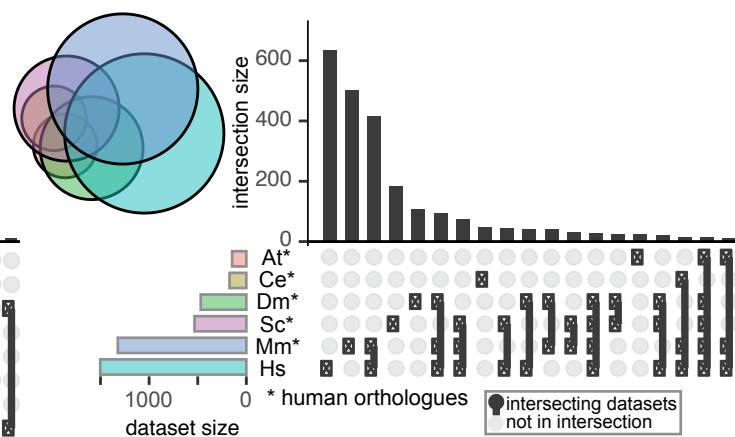


Figure 2

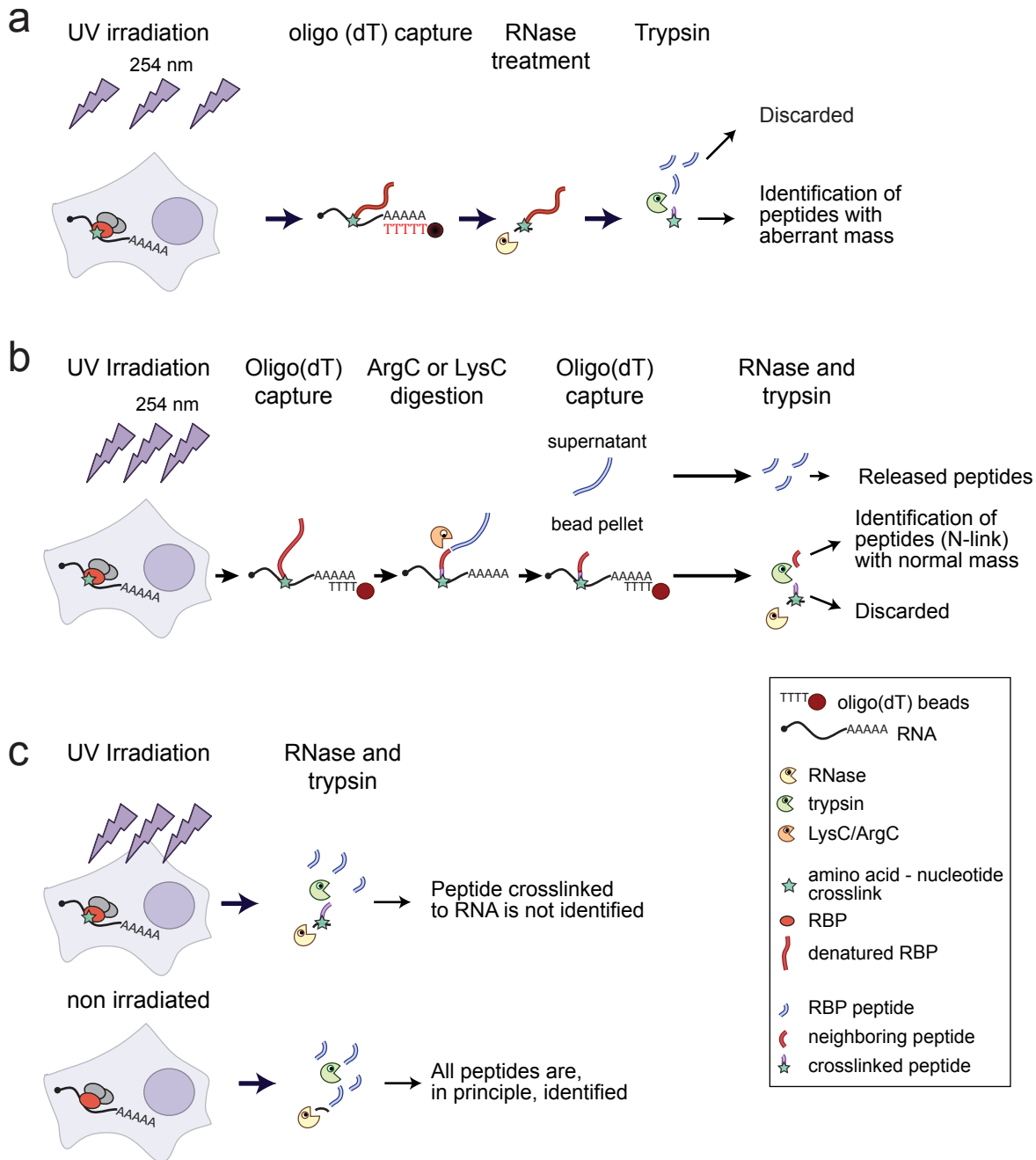


Figure 3

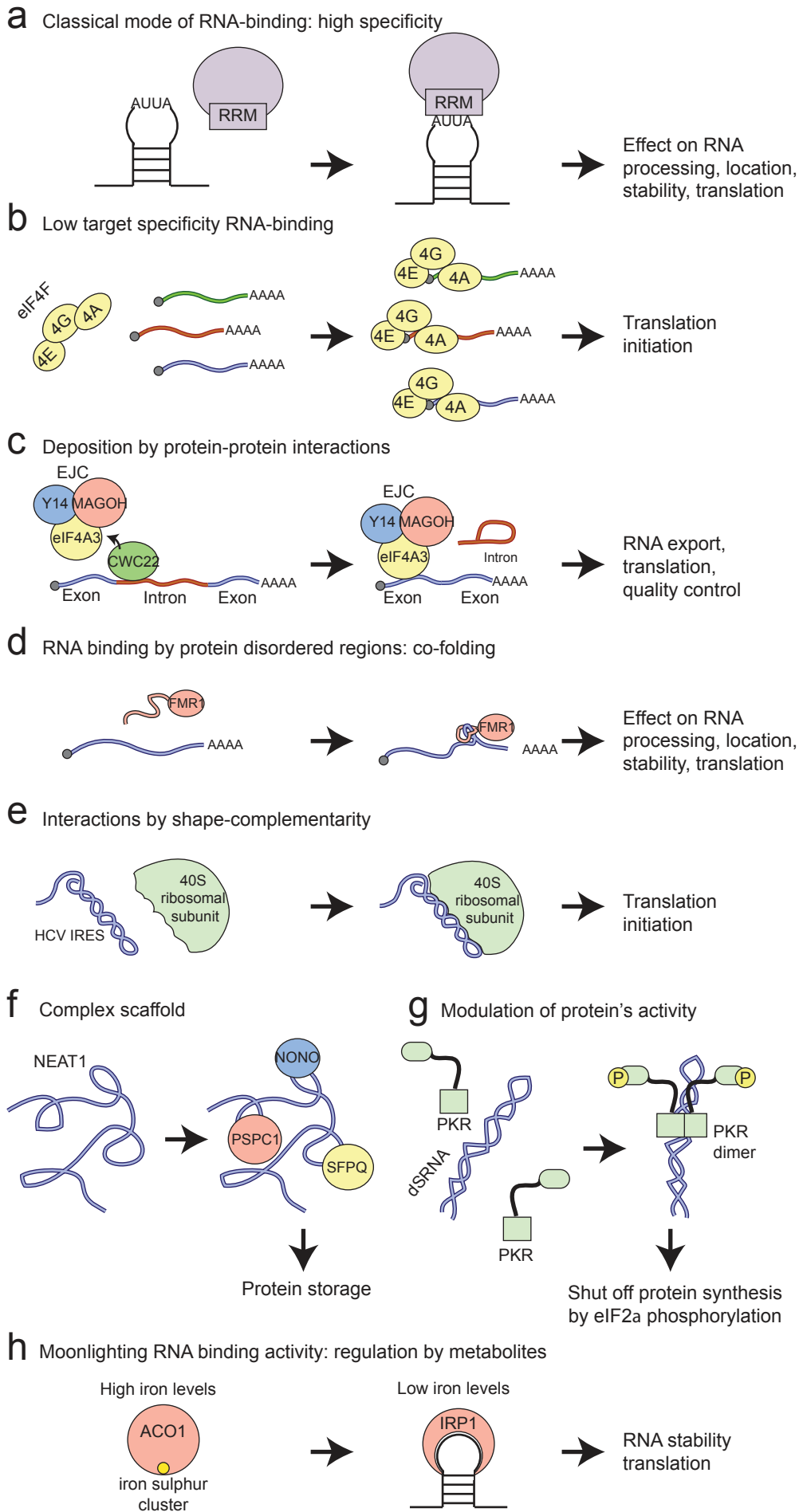


Figure 4

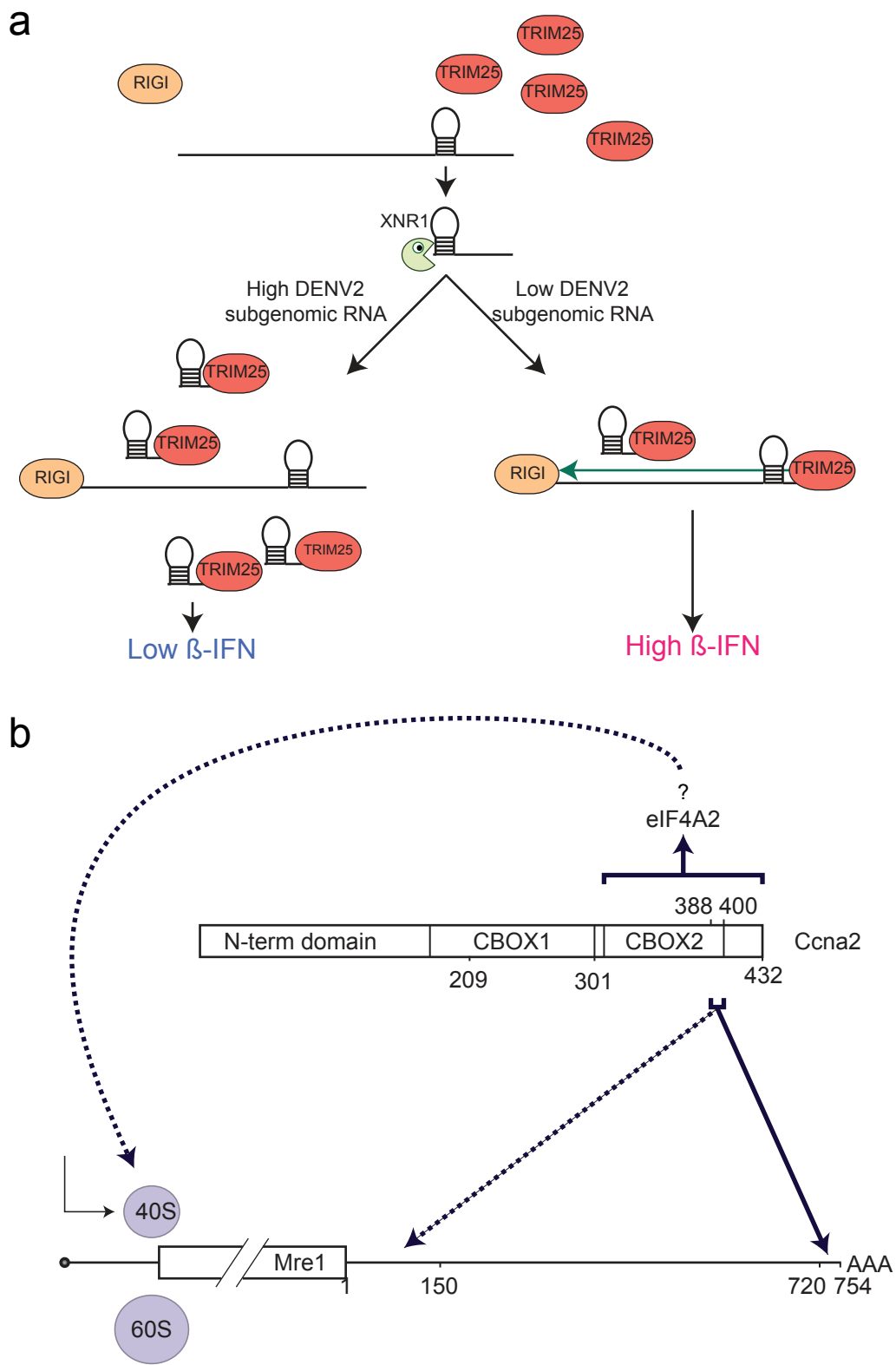


Figure 5

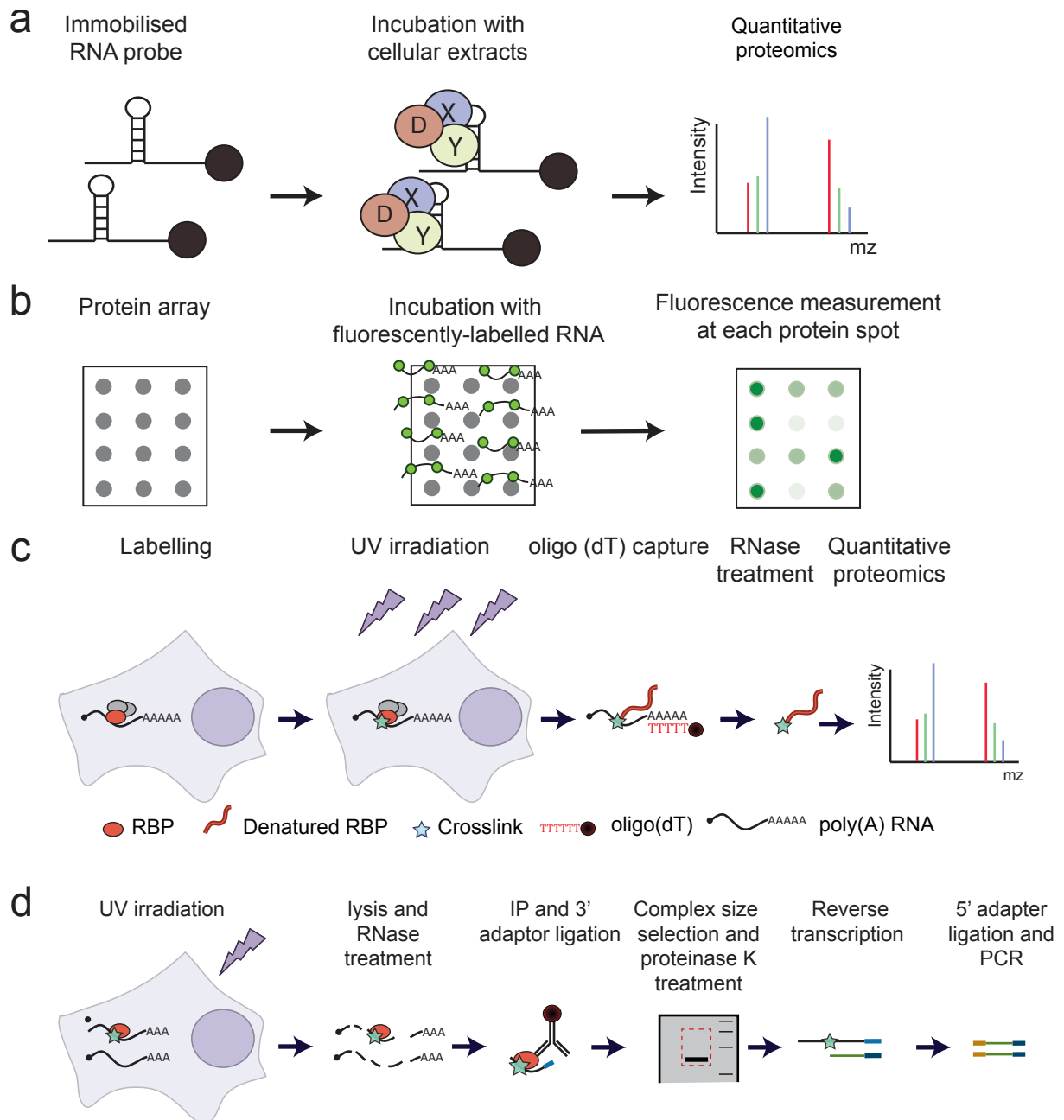
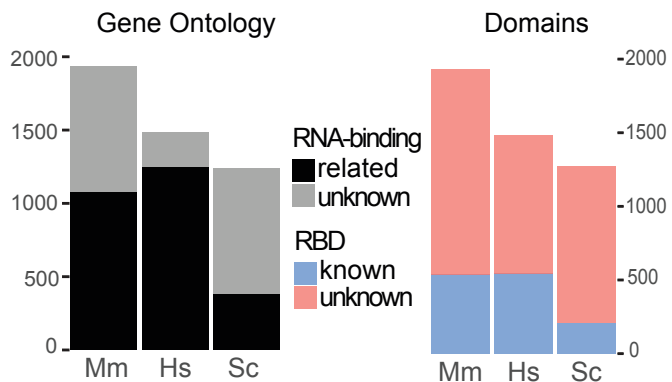
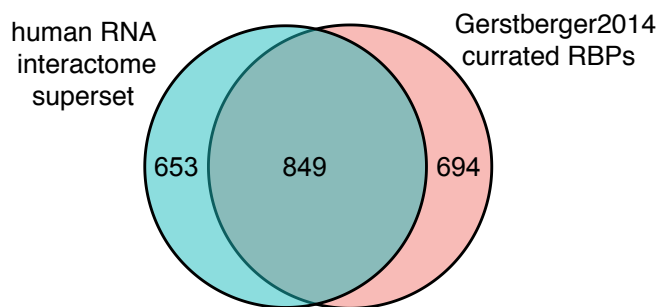


Figure for Box 1

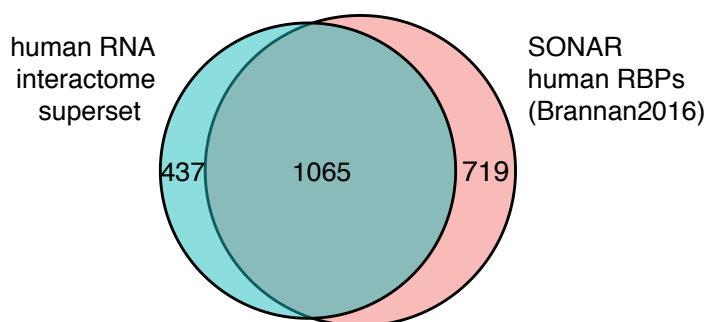
a Functional analysis of supersets



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

