

RESEARCH HIGHLIGHT

The greatest catch: big game fishing for mRNA-bound proteins

Christopher R Sibley[†], Jan Attig[†] and Jernej Ule^{*}

Abstract

Purification of proteins cross-linked to mRNAs has identified 800 mRNA-binding proteins and their characteristics.

Keywords CLIP, 4-thiouridine, RBP, RNA-binding domain, disordered proteins

From the moment an RNA is transcribed, its fate largely depends on its interactions with RNA-binding proteins (RBPs). RBPs can recognize short sequence motifs (Nova, TDP-43, U2AF65), secondary structures (Staufen-1, DGCR8), post-transcriptional modifications (eIF4E, CBP20), RNA duplexes (AGO1-4, helicases), or bind indiscriminately along transcripts (FUS, helicases). Collectively this dynamic interplay with multiple RBPs in an mRNA's life cycle helps determine the function and metabolism of an mRNA within a cell. Further, genotypic variations in many of these RBPs are responsible for diseases including fragile X syndrome, neurologic disorders and certain forms of cancer [1]. Collectively it makes studying RBP-RNA interactions essential due to the implications in both health and disease.

Here we discuss two recent studies from the laboratories of Matthias Hentze [2] and Markus Landthaler [3], which have now globally captured and defined the RBPs bound to mRNAs within cultured human cells. Both identify an unexpected wealth of RBPs, many of which were not previously known to interact with RNA, and greatly expand our understanding of the mRNA interactome.

Catching the interactome

Experimental studies on RBP-RNA interactions have typically employed top-down approaches such as RNA

immunoprecipitation, UV cross-linked immunoprecipitation (CLIP) [4] and photoactivatable ribonucleoside (PAR)-CLIP [5] to study RNA interactions of individual RBPs, using the protein as the bait in enrichment steps. More recently, others have reversed this approach and cast a tagged RNA as the bait in order to identify proteins interacting with it [6]. In an attempt to identify all proteins interacting with a pool of RNA, *in vitro* studies have incubated protein microarrays with labeled RNA, and identified a number of enzymes as unexpected RNA binders [7,8]. However, such *in vitro* studies could miss many context-dependent interactions that are present in physiological settings.

The two recent studies from the Hentze and Landthaler laboratories have taken the RNA bait concept a step further by capturing proteins by covalent cross-linking to polyadenylated RNA, mainly corresponding to mRNAs, and thereby generated a global mRNA interactome in HeLa [2] or HEK-293 [3] cell culture systems. In both studies, RNA-binding proteins are physically cross-linked to mRNAs using 254 nM UV-C light [2] or 365 nM UV in conjunction with PARs [2,3]. Next, cellular mRNA and the bound interactome are efficiently captured with oligo-dT-coated beads and purified under stringent conditions to eliminate contamination from non-cross-linked proteins, including those deriving from non-cross-linked protein-protein interactions. Finally, proteins are released by RNase digestion, gel resolved and analyzed by quantitative mass spectrometry (MS) approaches to reveal the protein interactome of the cellular mRNAs (Figure 1a).

Landthaler and colleagues used SILAC-based MS, where the control non-cross-linked sample is grown in the presence of amino acids with heavy isotopes, generating quantitative MS data of cross-linked versus non-cross-linked proteins. Thereby, they could exclude 135 proteins that appeared as contaminants based on a label-swap experiment, and they identified 797 proteins that are high-confidence RNA-binders [3]. Hentze and colleagues compiled an mRNA interactome of 860 proteins with a false discovery rate of less than 0.01 based on spectral count and peptide ion count of identified proteins [2]. Both studies validated their data set by

[†]Contributed equally

^{*}Correspondence: jule@mrc-lmb.cam.ac.uk

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

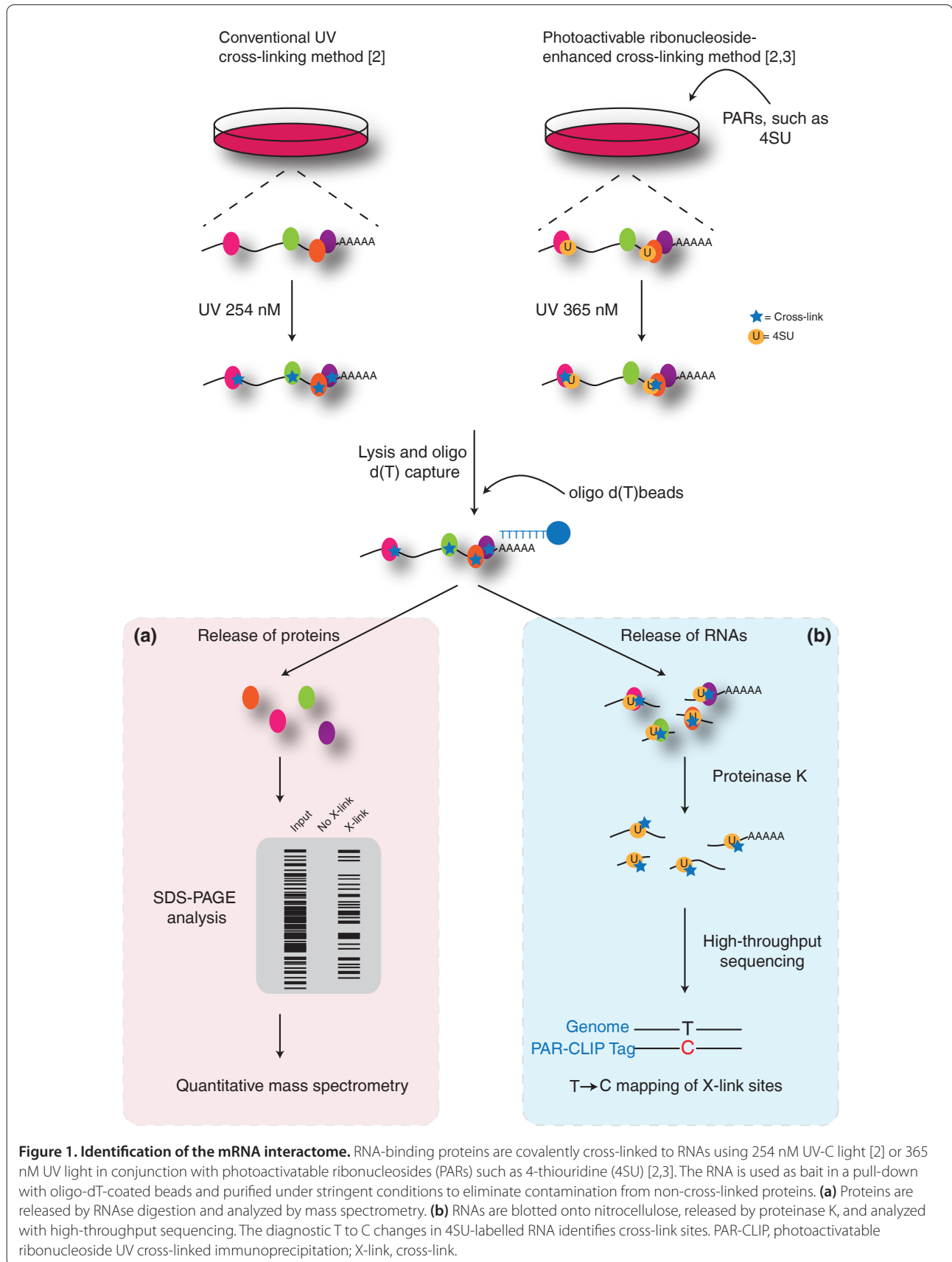


Figure 1. Identification of the mRNA interactome. RNA-binding proteins are covalently cross-linked to RNAs using 254 nM UV-C light [2] or 365 nM UV light in conjunction with photoactivatable ribonucleosides (PARs) such as 4-thiouridine (4SU) [2,3]. The RNA is used as bait in a pull-down with oligo-dT-coated beads and purified under stringent conditions to eliminate contamination from non-cross-linked proteins. **(a)** Proteins are released by RNase digestion and analyzed by mass spectrometry. **(b)** RNAs are blotted onto nitrocellulose, released by proteinase K, and analyzed with high-throughput sequencing. The diagnostic T to C changes in 4SU-labelled RNA identifies cross-link sites. PAR-CLIP, photoactivatable ribonucleoside UV cross-linked immunoprecipitation; X-link, cross-link.

testing around 20 candidates for RNA binding after immunoprecipitation, achieving a greater than 80% validation rate. However, only one study tested candidates for direct protein-RNA interaction using a gel imaging system [3], and only nine identified proteins have had their bound mRNAs experimentally probed with high-throughput sequencing [2,3]. It therefore remains unknown how many of the newly identified RBPs bind RNA with sequence specificity, and which are non-sequence-specific binders. Moreover, it remains possible that some of the identified proteins strongly interact with other RBPs and thereby elude the stringent wash steps, but do not directly cross-link to RNA.

Caught on the line

With the stringent filtering criteria, both groups found that 800 to 850 RBPs were confidently enriched in captured fractions, representing roughly 15% of cellular proteins [2]. Interestingly, a slightly larger number of RBPs were identified when using UV-C compared with PAR cross-linking [2]. The approach purified poly-A-containing transcripts, and therefore proteins binding to intronic RNA and other non-polyadenylated RNAs may not be identified. Thus, the interactome is a conservative estimate of the total number of RBPs.

The main surprise was that both datasets yielded many previously unannotated RBPs with no RNA-related Gene Ontology (GO) classifications, homology to known RBPs or experimental validations (315 [2] versus 245 [3]). Several were DNA-binding factors, others were kinases, while 17 metabolic enzymes were confidently identified as RBPs by Hentze and colleagues [2], 4 of which were confirmed in the other study [3].

The range of proteins on the RBP market

The authors analyzed the enrichment of GO annotations, known RNA-binding domains and other structural domains in the interactome, giving further insight into their catch. First, they found that all previously known RNA-binding domains were present within the mRNA interactome, constituting half [2] of all mRNA-bound proteins. Among the previously unknown RBPs, Hentze and colleagues identified proteins with the SAP and basic-amino-acid-rich WD40 domains, both of which are known to commonly bind DNA but have isolated family members that have previously been reported to bind RNA. This was complemented with other enriched motifs with no previous confirmation of RNA-binding ability such as AKAP95 and HC5HC2H subtypes of Znf domains, and the RAP and FASTK domains present in the family of *fas*-activated serine/threonine kinases. Encouragingly, overlap of some of these domains was seen in the interactome of the other study (Table 1) [3]. This greatly expands the repertoire of RNA-binding

Table 1. Overview of protein domains that are enriched among proteins not previously known to bind RNA

Putative non-classical RNA-binding domain	Identified in study led by:		Domain previously described to bind RNA
	Hentze [2]	Landthaler [3]	
AlbA-like DNA binding		✓	Yes
CBF	✓		No
DZF		✓	No
FAST and RAP	✓	✓*	No
HAT helix		✓	No
HMG box	✓*	✓	Yes
MMR HSR1	✓		No
PAP associated	✓	✓*	No
polyK-patches	✓		No
SAP	✓	✓	Yes
SWAP/SURP		✓	Yes
WD40	✓		Yes
'Winged helix' DNA binding		✓	No
YGG boxes	✓		No
ZNF motif			
Subtype AKAP95	✓	✓*	No
Subtype HC5HC2H	✓		No

*Corrected *P*-value >0.05; **P*-value >0.1.

domains and will direct new homology searches for other candidate RBPs across evolution.

Second, Landthaler and colleagues report another intriguing observation. Using their dataset to generate a protein-protein interaction network, they found a remarkable enrichment for the GO terms 'DNA damage' and 'transcription', which are almost as enriched as RNA-related processes. Seventeen proteins with the GO annotation 'response to DNA damage' are part of the interactome. Three members are annotated as RBPs and the authors found that three others have been previously recognized to bind to a specific RNA. It will be interesting to determine if these RBPs bind RNA as part of the DNA damage response, perhaps as part of transcription-coupled repair or as effectors for the small RNA double-stranded break markers, or alternatively if their RNA-binding property is used for other processes.

Third, Hentze and colleagues observed that RBPs are remarkably enriched in disordered regions. For approximately 50% of all the mRNA-interacting proteins, more than 40% of their residues are predicted to be in a disordered state [2]. Why are RNA-binding proteins enriched in disordered regions? Partially, the authors explain this occurrence by the fact that SR and RGG boxes are disordered regions that are known to bind RNA and are present in classical RNA-binding proteins.

Disordered domains are able to adopt multiple interacting structures, thereby providing flexibility in binding to multiple proteins but also to nucleic acids, particularly RNA. The dynamic nature of folded and unfolded RNA states enables induced-fit interactions between RNA and flexible protein domains [9]. Yet another explanation is offered by the recent observation that low-complexity domains (which are predicted to be disordered) form granules *in vitro*, which might resemble intracellular RNA granules [10]. In this case, the overwhelming percentage of RBPs with unstructured domains would hint at the functional importance of diverse RNA granules.

Kinks in the line

Landthaler and colleagues further examined the binding sites of the interactome on the mRNAs. To do this, the oligo-dT purified, PAR-labeled RNAs were partially RNase digested around cross-link sites, cleared of unbound RNA, and then proteinase digested in order to release cross-linked RNAs that were adapted for high-throughput sequencing. Evaluating locations of T→C transitions in sequence tags that occur at or near cross-link sites when using PARs allows identification of protein-binding sites on mRNAs with near nucleotide resolution (Figure 1b) [5]. Specifically, the analysis focused on cross-links in the 3' UTRs, where RBP-binding sites can be distinguished from those of the translating ribosomes. With 28% of all 3' UTR T nucleotides converted to C nucleotides, the approach convincingly demonstrates widespread binding of RBPs across this region of the mRNA [3]. Further, an enhanced evolutionary conservation surrounding these sites implies selective pressure to maintain many of these interactions across multiple species.

Re-casting for another catch

The major challenge following these reports will be the functional validation of all identified RBPs and putative RNA-binding domains, including the 86 interactome RBPs associated with human Mendelian diseases [1]. Further, it is expected that applying interactome capture to model organisms and specific biological paradigms will continue to be fruitful. Significant remodeling of protein-RNA interactions is expected under conditions

of hypoxia, during the cell cycle, after drug treatment or knockdown of important proteins, among others. Alternatively, interactome capture across samples from multiple species would identify stages in evolution where new RNA-binding capabilities emerged. Understanding the re-modeling of the mRNA interactome will therefore provide great insight into the dynamic interplay of RNAs and RBPs during an mRNA life cycle.

Abbreviations

CLIP, UV cross-linked immunoprecipitation; GO, Gene Ontology; MS, mass spectrometry; PAR, photoactivatable ribonucleoside; RBP, RNA-binding protein.

Competing interests

The authors declare that they have no competing interests.

Published: 17 July 2012

References

1. Cooper TA, Wan L, Dreyfuss G: **RNA and disease**. *Cell* 2009, **136**:777-793.
2. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW: **Insights into RNA biology from an atlas of mammalian mRNA-binding proteins**. *Cell* 2012, **149**:1393-1406.
3. Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M: **The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts**. *Mol Cell* 2012, **46**:674-690.
4. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB: **CLIP identifies Nova-regulated RNA networks in the brain**. *Science* 2003, **302**:1212-1215.
5. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T: **Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP**. *Cell* 2010, **141**:129-141.
6. Butter F, Scheibe M, Morl M, Mann M: **Unbiased RNA-protein interaction screen by quantitative proteomics**. *Proc Natl Acad Sci U S A* 2009, **106**:10626-10631.
7. Scherrer T, Mittal N, Janga SC, Gerber AP: **A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes**. *PLoS One* 2010, **5**:e15499.
8. Tsvetanova NG, Klass DM, Salzman J, Brown PO: **Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae***. *PLoS One* 2010, **5**:e12671.
9. Dyson HJ: **Roles of intrinsic disorder in protein-nucleic acid interactions**. *Mol Biosyst* 2012, **8**:97-104.
10. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, McKnight SL: **Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels**. *Cell* 2012, **149**:753-767.

doi:10.1186/gb-2012-13-7-163

Cite this article as: Sibley CR, et al.: The greatest catch: big game fishing for mRNA-bound proteins. *Genome Biology* 2012, **13**:163.