

Profiling the dress codes of RNA-binding proteins

Methods choice is broad and labs are pushing protocols to address expansive biological questions.

Vivien Marx

Clothes can be enabling. RNAs in a eukaryotic cell can be clothed, as some scientists phrase it, or unclothed¹.

Some RNAs in a cell, including noncoding RNAs, wear proteins that cloak certain RNA regions and expose others. RNA-binding proteins (RBPs) are powerful, versatile regulatory units. They play roles in cellular housekeeping, development, differentiation, metabolism, health and disease, a functional diversity that labs are beginning to characterize. Part of this research is profiling RBPs, for example, to learn which RNAs and proteins are connected^{2,3}.

“The proteins are what carries out the function, but the RNA is what coordinates the activity of all those proteins to put them together into a coherent unit to carry out a unique function,” says California Institute of Technology researcher Mitchell Guttman. “Independent of the RNA, these proteins wouldn’t be able to do that role.” That, in his view, makes the RNA the “epicenter” for organizing complex functions related to RNAs such as translation, splicing, localization. From another angle, proteins sit at the ‘epicenter’ because they mediate cellular events such as enzymatic reactions and localization, he says. RBPs and RBP complexes regulate transcription and translation, gene expression, processing of RNAs.

Proteins can be the crucial players that bend, fold or squeeze the RNA, leading to the RNA acting catalytically as in the case of the eukaryotic spliceosome, which is a kind of 3D RNA reaction center, says Henning Urlaub, who runs the bioanalytical mass spectrometry group at the Max Planck Institute for Biophysical Chemistry in Göttingen. The fact that RNAs tend to be complexed with proteins opens up numerous possibilities for thousands of protein structures and domains to be combined with presumably more than a thousand different RNA structures.

Some RBP-related methods emphasize protein capture and purification, while others highlight RNA capture, says Guttman. Mass spectrometry helps to analyze proteins attached to the RNA. In the more protein-centric approaches, UV light cross-links the RNAs and associated proteins, which are then immunoprecipitated with antibodies.



In a eukaryotic cell RNAs can be naked or clothed with proteins. Proteins bound to RNAs become RNA-binding proteins (RBPs), which are powerful gene regulators. Credit: [yogysic/DigitalVision Vectors/ Getty](#)

The ‘pulldown’ material is denatured, and the RNA is sequenced to identify the RNA’s protein-binding regions. RNA-centric methods also apply UV-based cross-linking, and the RNA is captured from lysed cells with various types of baits.

Some labs seek to identify the protein region interacting with the RNA, says Urlaub, by using mass spec after cross-linking to identify the cross-linked region and amino acids, down to several amino acids in that protein. In their experimental design, he observes that some researchers can be biased about protein–RNA interaction: they sometimes assume that proteins lacking known RNA-binding motifs don’t bind RNAs, but they do, he says. Also, some protein regions that bind strongly to RNA do not cross-link well.

The wealth of RBP-profiling methods calls scientists to make thoughtful choices, says Miguel Esteban, a researcher at Guangzhou Institutes of Biomedicine and Health. Studying RBPs with many-stepped protocols takes much methods-related troubleshooting, he says. Enrichment strategies deliver labs a certain slant to their RBP data, he says. Pulling down RNAs and finding the proteins that interact with them does not give a view of all proteins near RBPs.

Each method has experimental limitations but all methods share that they are isolation methods that enrich a target of interest, says Columbia University MD/PhD student Stefanie Gerstberger, who tallied human RBPs as part of her PhD research completed in 2016 in Tom Tuschl’s lab at Rockefeller University. With untargeted approaches “you get a very broad unspecific picture,” she says. That picture can, for example, be used to assess the general state of cells, also transcriptome-wide.

Many RBP resources exist for a data-hunting researcher, says Gerstberger, with information from a variety of cell types and with varying data depths. As is typical in biochemistry, users will see variability between experimental results, “so it’s often easier and better to do the experiment from scratch to control for these factors,” she says.



Plenty of discovery awaits labs on the subject of regulation and physiological function related to RBPs, says Stefanie Gerstberger.

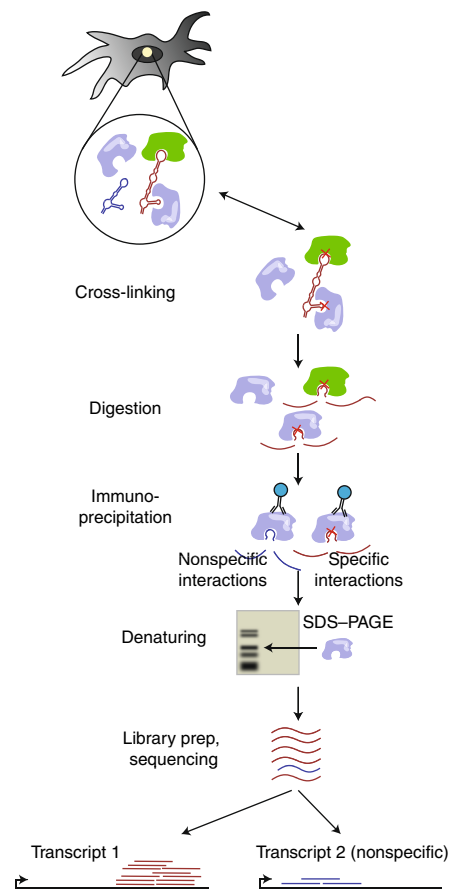
(Photo: J. Lee)

Although RBP methods can be binned as protein-centric or RNA-centric, it takes both bins to learn about the biology of RBPs, says Guttman. “I think it’s a split that comes from people’s backgrounds,” says Matthias Hentze, who directs the European Molecular Biology Laboratory in Heidelberg and who develops and uses RBP-profiling methods. Some labs focus more on nucleic acids, while others study a few proteins or take a wider proteomics approach. “Biology doesn’t care about the background of the investigator,” he says. “You need to consider both equally.”

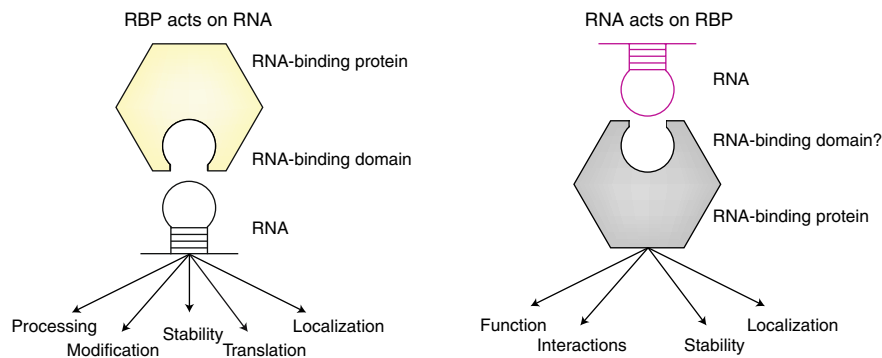
Census-taking

Around 5–10% of the human proteome can bind RNA. To date no census of all RNA-binding proteins has taken place; the number of known RBPs remains an estimate. Studies lead scientists to believe there are between 1,000 and 2,000 RBPs in mammalian cells.

As she manually curated RBPs in human cells, based on data of mRNA isoforms from the vertebrate genome browser Ensembl



Some methods focus on characterizing the protein in RNA-binding proteins. Credit: Guttman lab, Caltech. Adapted with permission from ref. ², Springer Nature



It’s two-way cross-talk: RNA-binding proteins (RBPs) can regulate RNAs, and RBPs can be regulated by RNA. Credit: Hentze lab, EMBL. Adapted with permission from ref. ³, Springer Nature

and mass spec data from RNA–protein cross-linking experiments, Gerstberger’s tally reached 1,542 RBPs in human cells. A few more might emerge, she says, such as proteins that bind RNA in a specific or nonspecific manner. They may belong to an uncharacterized family or they might be singular members of a protein class. For an average mammalian cell under standard culture conditions, around 2,000 RBPs might be close to the true number of proteins that bind RNA with a meaningful biological purpose, says Hentze. He recommends that when labs look across species, they watch for species-specific differences, such as particular RBPs an organism might have evolved.

Even as RBP tallies vary, “I think it’s more important that we understand the function and coordination of gene expression of the current ones,” says Gerstberger. Knowing about a potential ability to bind is not telling enough about RBPs, she says. Much discovery awaits related to regulation and physiological function.

A tally of RBPs, such as one completed in an organized project, would be a valuable resource, says Esteban. Even without it, data about RBPs will amass through the efforts of many labs. Discerning the different relevance of every interaction between an RNA and a protein matters, he says. For example, just because a metabolic enzyme regularly interacts with RNA does not automatically mean the RBP regulates metabolism; the RBP could play other, yet unknown, roles.

A catalog, whether compiled as a concerted effort or by the collective effort of many labs, is unlikely to come together through the use of just one method, says Guttman. As labs learn more about microRNAs, long noncoding RNAs (lncRNAs), circular RNAs, and other poorly understood RNA classes, a catalog can be a framework for evaluating function and mechanism and for exploring the RBP universe. “It no longer sort of becomes the

Wild, Wild West of every possible protein in the cell is a culprit,” he says.

Cross-linking

When his lab members are cross-linking RNAs to proteins, such as with UV light, Esteban reminds them to include a non-cross-linked sample and to set up other experimental checkpoints. This can add a day or two to the start of an experiment, but running a gel sooner can spare pain later. He also recommends never losing sight of how the more dominant proteins can overshadow others. For example, RNA species in the cell such as some ncRNAs are important for cell function, “yet they represent a very small percentage of the total RNA output,” he says.

It will be hard to pick up RNAs transcribed at low levels when labs work with whole cells, says RIKEN researcher Piero Carninci, given that “there are four to five orders of magnitude of difference in gene expression.” Scientists can try more targeted approaches to identify RNAs bound to specific compartments, he says. He and his team have been working on capturing lncRNAs attached to chromatin, for example. They capture these complexes by sequencing RNA–DNA pairs that associate to specific genomic locations. Then they statistically analyze the frequency of interaction to explore significance. “In this case, sequencing introduces a measurable way to count and assess frequency,” he says.

As Guttman explains, cross-linking data are less noisy when UV light rather than formaldehyde is used. UV light is a zero-distance cross-linker, meaning cross-links occur only when protein and RNA are touching. “Even if it is slightly removed, it will not cross-link,” he says. The solid results UV-based cross-linking delivers make it the gold standard when labs ask, “Is there interaction forming in vivo?” Cross-linking’s high degree of specificity can, however, be a downside, such as when the RNA is not quite



Many RBPs do not have recognizable RNA-binding domains, says Matthias Hentze. “Clearly there is a whole universe out there.”

(Photo: Private collection)

in the right conformation when cross-linking occurs. Another important aspect, he says, is that cross-linking tends to involve 1–5% of all RNA–protein interactions. Labs can miss plenty of interactions, and to get at these low-abundance interactions, researchers need to do experiments with many cells.

Hands on RNAs and proteins

As Gerstberger points out, RBP exploration dates back to the 1950s. Over time, high-throughput methods emerged such as RBP immunoprecipitation, then cDNA array hybridization of RNA or approaches to select RNA ligands on the basis of SELEX, systematic evolution of ligands by exponential enrichment. RNA-sequencing and mass spec have enabled profiling of RBPs at higher throughput and transcriptome-wide. Some techniques combine sequencing, immunoprecipitation and sequencing leading to so-called CLIP-seq approaches.

It has turned out to be a major methodological advance to be able to pull down RNAs via polyadenylated tails. There are two methods to do so: one developed in the Landthaler lab at Max Delbrück Center for Molecular Medicine, and the other in the Hentze lab^{4,5}.

The approaches were developed in parallel. The two labs knew of one another’s efforts, methods and results, and they worked independently and coordinated publication, says Hentze. “They’re virtually identical methods; the buffers are slightly different,” he says. In both, there is UV-light-based cross-linking of proteins with RNA in live cells, then RNA pulldown via the poly(A) tail, and the polyadenylated RNA is captured on oligo(dT) beads. Next follows mass spectrometry to identify and quantify proteins.

The use of oligo(dT)-coated beads for capturing RNAs has helped labs to discover RBPs and to appreciate the scope of proteins that bind to RNAs, says Hentze. Labs have tweaked the methods to, for example, incorporate nucleotide analogs that are subjected to biotinylation followed by the pulldown step, he says.

As Markus Landthaler of the Max Delbrück Center for Molecular Medicine

and colleagues point out, over 30 years ago, labs tried to isolate the poly(A) RNA-bound proteome using oligo(dT) Sepharose chromatography³. Since then, methods developers have reached a more comprehensive approach for identifying RBPs with UV-based cross-linking and sequencing to determine RNA-binding sites. These methods are now being used to, for example, compare healthy and diseased cells under different conditions.

Landthaler and his team used the method for studying pathways involved in sensing and repairing DNA damage⁶. In breast cancer cells, they identified over 260 RBPs that show more interaction with mRNAs in response to ionizing radiation. The team used a 4-thiouridine (4sU)-based mRNA capture approach and an isotope-based labeling technique, SILAC, to label cells and help with quantification of protein differences in samples. Integration of 4sU into RNAs can vary by cell type, so labs should test nucleoside concentration and incubation time prior to their experiments.

One approach Landthaler codeveloped with others is PAR-CLIP, for photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation. Photoreactive thionucleosides (4sU) are taken up by live cells, and the nucleosides enhance the cross-linking between RNA and protein. By scoring thymidine-to-cytidine transitions, the method reveals RBP binding sites. The researchers note that the method enables transcriptome-wide identification of RNA-binding sites of targeted RBPs.

To enable RNA capture beyond RNAs with polyadenylated tails, Esteban developed RICK, or capture of the newly transcribed RNA interactome using click chemistry. The method includes RNA labeling with 5-ethyluridine followed by a biotinylation step using click chemistry. Next is bead-based capture, after which the RNAs are sequenced and the proteins analyzed by mass spec. He and his team are applying RICK to capture newly generated RNAs in stem cells to see how the type and patterns of RBPs change under different conditions or in cells engineered to have certain mutations. Capturing newly transcribed RNAs offers hints about a facet of RBP function, he says.

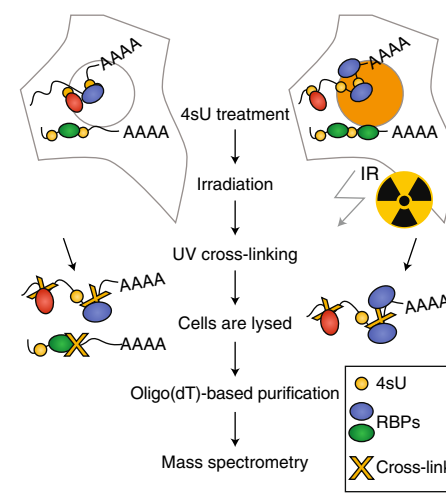
Combating noise

Signal-to-noise ratios become especially important when cells in different conditions are being compared, says Hentze. Background can hurt results and analysis. Smaller but relevant changes risk being overlooked. “I think it’s critical to develop the existing techniques in such a way that differences can be scored well,” he says.

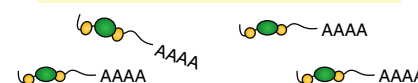
The Hentze lab and other labs are exploring the effect of replacing oligo(dT)s with locked nucleic acids (LNAs) for pulling down polyadenylated RNA. The oligos are, for example, instead of 20 deoxy(T)s, a string of a mix of deoxy(dT)s and LNAs. LNAs melt at higher temperatures than deoxy(dT)s, which lets experimenters be more stringent with washing steps for purification. The more stringent the wash, the more unspecific hangers-on can be washed away. In both his method and the one from the Landthaler lab, a “significant signal of ribosomal RNA” clutters pulldown results, he says, which the use of LNAs will lower. The polyadenylated RNA stays on the beads and the rRNA and contaminating DNA, too, can be washed away with greater stringency.

Labs have long explored how to purify RNA-binding proteins with higher stringency, says Guttman, in order to wash, keep RNA and protein and be rid of as many nonspecific hybridizations as possible. He points to a method published in 2006 that applies peptide nucleic acids to achieve stringent washing⁷. Guttman sees experimental advantages to LNAs: they are short and have a high melting temperature, which helps with higher-stringency washing. But LNAs cannot be made in the lab; they must be bought. “LNAs are expensive,” he says.

RAP-MS, a method developed in the Guttman lab, applies biotinylated probes and an RNA antisense purification strategy.



In breast cancer cells exposed to radiation, 266 proteins increased binding to mRNA



An approach to study the DNA damage response using breast cancer cells. Credit: Landthaler lab, MDC Berlin. Adapted with permission from ref. ⁶, CSH Press

The team uses single-stranded, biotinylated 90-mer oligos, which also have high melting temperatures and can be made in the lab. The hybrid is stable, says Guttman. It can weather denaturing and washing in chaotropic agents such as 6 M urea to bring back what has covalently cross-linked to the RNA. He and his team study lncRNAs that can be around 17,000 nucleotides long. It's risky to use only one probe such as an LNA, he says, given how fragile RNA can be. When it breaks, a large fraction of the total protein bound to RNA is lost. In his lab, the team tiles the lncRNA with many 90-mer probes so data can be captured even if the RNA breaks.

Addressing purification, the Hentze lab developed 2C, a method leveraging an old observation that proteins do not bind to resins in silica columns, whereas RNA does³. "I think that is a really beautiful observation because it makes life so much easier," says Guttman. The 2C method involves using silica columns to isolate cross-linked RBPs.

Conventionally, labs need to validate RBPs with immunoprecipitation, washing, labeling of the cross-linked RNA, elution and autoradiography. The 2C method avoids immunoprecipitation and radiolabeling. If a protein is cross-linked to RNA, the RNA sticks to the resin, delivering the proteins bound to it. What sticks to the resin can be stringently washed, enabling better RBP purification, he says. "I am very excited about that as a new foundation for thinking about purification of RNA-binding proteins," says Guttman.

Shape matters

Proteins are three-dimensional molecules, and a 'good' RNA-binding domain might be exposed only under certain conditions, says Esteban. "Those domains are very difficult to identify with bioinformatics," he says. RNA forms 2D structures, says Guttman, because of the thermodynamic properties of the molecule and other aspects, and the structure will also determine which proteins bind where. An RNA also changes conformation depending on its bound proteins. Structure and function go together. Bacterial riboswitches are a type of RNA that sense, for example, cellular nutrients. In reaction to nutrients, a riboswitch changes its structure, which can shape a function such as gene expression. Mammalian cells might turn out to have RNAs fulfilling these roles, he says.

Studying RNA structure remains a challenge but some methods exist to help characterize RBPs, says Guttman. With SHAPE, or selective 2-hydroxylacylation analyzed by primer extension, developed in the Weeks lab at the University of North Carolina, a small molecule marks



Some interactions between RNA and proteins might be more subtle than others, says Miguel Esteban.

(Photo: B. Huang)

accessible, dynamic RNA regions in ways that can be detected by sequencing. Another way to probe RNA structure with modifications is with DMS-MaPseq, for dimethyl sulfate mutational profiling with sequencing, a method from the Weissman lab at the University of California at San Francisco, MIT's Silvi Rouskin and colleagues. To label specific RNA motifs in RBPs, Paul Khavari at Stanford University School of Medicine and colleagues developed RNA-protein interaction detection, or RaPID. It applies biotinylation, another way to permit stringent washing during RBP purification.

Many RBPs do not have recognizable RNA-binding domains, says Hentze. "Clearly there is a whole universe out there," he says. He and his team developed RBDmap to find binding domains. It can see not-yet-identified regions that are enriched for RNA binding, such as highly unstructured protein regions and low-complexity domains, he says. In RBDmap, a second round of oligo(dT) captures the cross-linked RNA and protein along with a neighboring peptide, which is used to identify RNA-binding sites.

Speaking more generally about RBP techniques, Urlaub says that "we are still not specific and quantitative enough on the cell-wide level." Changes to interactions between RNAs and proteins need to be defined in terms of cell cycle and states of various cells, also quantitatively. Labs want to know, for example, how many copies of a protein bind to one RNA molecule and whether different protein domains act differently on different RNAs. It remains hard to study low-complexity protein regions with repetitive amino acid sequences that bind to RNA, because they are nearly impossible to study with mass spec, he says.

Perhaps, says Esteban, some RNAs bind to proteins in several different ways. "I think we need to change our minds a little bit," he says. Some binding interactions between RNA and proteins might be "more subtle, delicate" than others, and these add to the regulatory landscape. In addition, RNA and proteins can both be modified in various

ways, another aspect that can influence regulation.

RNAs can evolve the ability to bind to a particular protein surface, which means "in principle any protein surface could be an RNA-binding surface," says Hentze. A new awareness of how RNA and proteins influence one another is emerging. "RNA-binding proteins are not only proteins that bind to RNA to regulate RNA, but RNA-binding proteins are proteins that are bound by RNA to be regulated by RNA," he says. He and his team have found that a protein involved in autophagy is regulated by a type of ncRNA called a vaultRNA, in this case vRNA1-1⁹. Autophagy is the process through which cells rid themselves of debris.

The RBP in question is an enigmRBP identified in his lab. EnigmRBPs are enigmatic in their RNA-binding, meaning it's hard to identify where they are binding, says Hentze. The team shows that the small ncRNA binds the protein and regulates its function through the RNA-protein interaction in autophagy, he says. "That RNA-protein interaction is doing what we normally see protein-protein interactions doing," which is to regulate protein function. Here, RNAs act as riboregulators to regulate autophagic flux. Such findings indicate how RBPs give cellular genomes "the chance to regulate biology and biological processes in a far more direct way than we previously knew," he says.

Observations about the versatility of RBPs add an aspect to the RNA World Hypothesis, according to which life on Earth began with RNA, followed by the evolution of proteins and, later, DNA. Work in the Hentze lab, including this latest riboregulator and the regulatory role of RNAs in RBPs more generally, all fit well into the RNA World model, he says—"I could see in it, perhaps, an evolutionarily early form of biological regulation." □

Vivien Marx

Technology editor for *Nature Methods*.

e-mail: v.marx@us.nature.com

Published online: 31 August 2018

<https://doi.org/10.1038/s41592-018-0117-9>

References

1. Singh, G., Pratt, G., Yeo, G. W. & Moore, M. J. *Annu. Rev. Biochem.* **84**, 325–354 (2015).
2. McHugh, C. A., Russell, P. & Guttman, M. *Genome Biol.* **15**, 203 (2014).
3. Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. *Nat. Rev. Mol. Cell Biol.* **19**, 327–341 (2018).
4. Baltz, A. G. et al. *Mol. Cell* **46**, 674–690 (2012).
5. Castello, A. et al. *Cell* **149**, 1393–1406 (2012).
6. Milek, M. et al. *Genome Res.* **27**, 1344–1359 (2017).
7. Zielinski, J. et al. *Proc. Natl. Acad. Sci. USA* **103**, 1557–1562 (2006).
8. Asencio, C., Chatterjee, A. & Hentze, M. W. *Life Sci. Alliance* **1**, e201800088 (2018).
9. Horos, R. et al. *bioRxiv* Preprint at <https://doi.org/10.1101/177949> (2017).