



# The Grand Challenge of Characterizing Ribonucleoprotein Networks

Gian Gaetano Tartaglia<sup>1,2,3\*</sup>

<sup>1</sup> Bioinformatics and Genomics, Gene Function and Evolution, Bioinformatics and Genomics Programme, Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona, Spain, <sup>2</sup> Universitat Pompeu Fabra, Barcelona, Spain, <sup>3</sup> Institutio Catalana de Recerca i Estudis Avançats, Barcelona, Spain

**Keywords:** ribonucleoprotein networks, protein–RNA granules, non-coding RNA, high-throughput screening assays, molecular evolution

Protein–RNA interactions are at the heart of cell regulation. From transcription, processing, storage, and translation, all the stages in the life cycle of an RNA depend on interactions with proteins. Although technologies are making remarkable progress in unraveling the landscape of protein–RNA interactions, many key issues are unclear. We still have to identify how many proteins have RNA-binding ability, what are their targets and functional pathways. Moreover, while we know the number of protein-coding genes in the human genome, functional non-coding RNAs are still poorly defined. What is the function of the non-coding part of the eukaryotic transcriptome? A clear understanding of the biological functions of coding and non-coding transcripts would provide novel insights in molecular biology. What are the protein components binding to an RNA while it is being produced? Our lack of understanding of how ribonucleoprotein complexes assemble is a major rate-limiting factor to future progress in the field. We need to generate an in-depth characterization of protein–RNA complexes that form in cells during development and in response to external *stimuli*.

## OPEN ACCESS

### Edited by:

Annalisa Pastore,  
King's College London, UK

### Reviewed by:

Andres Ramos,  
University College London, UK

### \*Correspondence:

Gian G. Tartaglia  
gian.tartaglia@icrea.cat

### Specialty section:

This article was submitted to  
Ribonucleoprotein Networks,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 04 April 2016

**Accepted:** 25 May 2016

**Published:** 09 June 2016

### Citation:

Tartaglia GG (2016) The Grand  
Challenge of Characterizing  
Ribonucleoprotein Networks.  
*Front. Mol. Biosci.* 3:24.  
doi: 10.3389/fmolb.2016.00024

## HOW DID LIFE BEGIN AND WHAT ROLE DID RNA AND PROTEIN MOLECULES PLAY?

Life on earth might descend from an RNA world (Higgs and Lehman, 2015) although RNA and proteins could have emerged together. As protein-based molecules are essential to make nucleic acid polymers and nucleotide-based molecules are needed to synthesize proteins, protein, and RNA might have co-evolved from the very beginning of life (Chao et al., 2008): RNA would contain the instructions for life while peptides accelerate key chemical reactions to carry out the instructions. In support of this hypothesis, it has been reported that network of reactions beginning with hydrogen cyanide and hydrogen sulfide in streams of water irradiated by UV light could produce the chemical components of proteins and lipids, alongside those of RNA (Patel et al., 2015).

## DIFFERENT TYPES OF PROTEIN-RNA ASSEMBLIES

From birth to degradation, most cellular RNAs are never naked but they form complexes with partner proteins in ribonucleoprotein (RNP) particles. The assembly of functional RNPs and delivery to their destinations often involves progression through a series of intermediate complexes and subcellular compartments. For instance, *Cajal* bodies are sites of non-coding RNA maturation, concentrating assembly factors to accelerate complicated biochemical reactions (Gall, 2000). Similarly, messenger RNA bind to protein complexes that undergo constant remodeling as they travel from the site of transcription to the cytoplasm (Matera and Wang, 2014). What are the protein components binding to coding and non-coding RNAs while they move in the cell?

Many protein–RNA assemblies are observed in the nucleoplasm and cytoplasm of all cells and play fundamental roles in growth, development, and homeostasis. For instance, using an electron microscope, George Emil Palade observed dense particles or granules in the endoplasmic reticulum, which were later called ribosomes (Palade, 1955). Before their export to the cytoplasm, ribosome components are subject to a number of reactions in the *nucleoli*. Processing of the ribosomal RNAs initiates in the dense fibrillar part of the nucleolus and continues in the granular component, where the RNAs self-assemble with ribosomal proteins to form nearly completed subunits. Other examples of protein–RNA assemblies found in the nucleoplasm include speckles and promyelocytic leukemia bodies, among others (Spector and Lamond, 2011). Changes in the protein or RNA composition of these structures are associated with diseases such as Huntington’s and spinal muscular dystrophy (Irimia et al., 2014).

Studies on the composition, structure, and behavior of speckles provide precious information to understand how protein and RNA components assemble and re-organize in RNP particles such as the spliceosome (Papasaikas et al., 2015). Although highly dynamic, ribosomal, and spliceosomal components form distinct RNPs that involve hundreds of proteins and RNAs (**Figure 1**): How variable are protein–RNA assemblies in composition? How large, dynamic and structurally heterogeneous is the *spectrum* or ribonucleoprotein assemblies?

## WHAT IS THE COMPOSITION OF PROTEIN-RNA GRANULES INSIDE THE CELL?

As shown by Tony Hyman and coworkers, RNP self-assemble from soluble proteins and RNAs to form structures that grow, collapse, and fuse continuously (Brangwynne et al., 2009). These and other findings have challenged the concept of cell organization in compartments and how we study protein accumulations (Li et al., 2012). For instance, it has been shown that membrane-less granules form upon environmental insults to prevent translation and when the stress is resolved, the assembly dissolve and protein production is resumed (Buchan and Parker, 2009). Intriguingly, if the stress persists, RNA within the granules can be transferred to other RNP assemblies, called P-bodies, to be degraded (Parker and Sheth, 2007). Recent evidence suggests that ribonucleoprotein granules are also associated with the onset of neurodegenerative diseases (Wolozin, 2012).

We need to generate an in-depth characterization of protein–RNA granules that form in cells during development and in response external *stimuli*. These RNA structures assemblies play key roles in numerous aspects of cell biology and a better understanding of how and why they form will provide significant innovation (**Figure 1**). We do not have a full biochemical composition for many of these particles: What is the full range and relevance of RNA structures forming granules? Are non-coding RNA participating in RNP assemblies?

## WHAT IS THE FUNCTION OF THE LONG NON-CODING PART OF THE TRANSCRIPTOME?

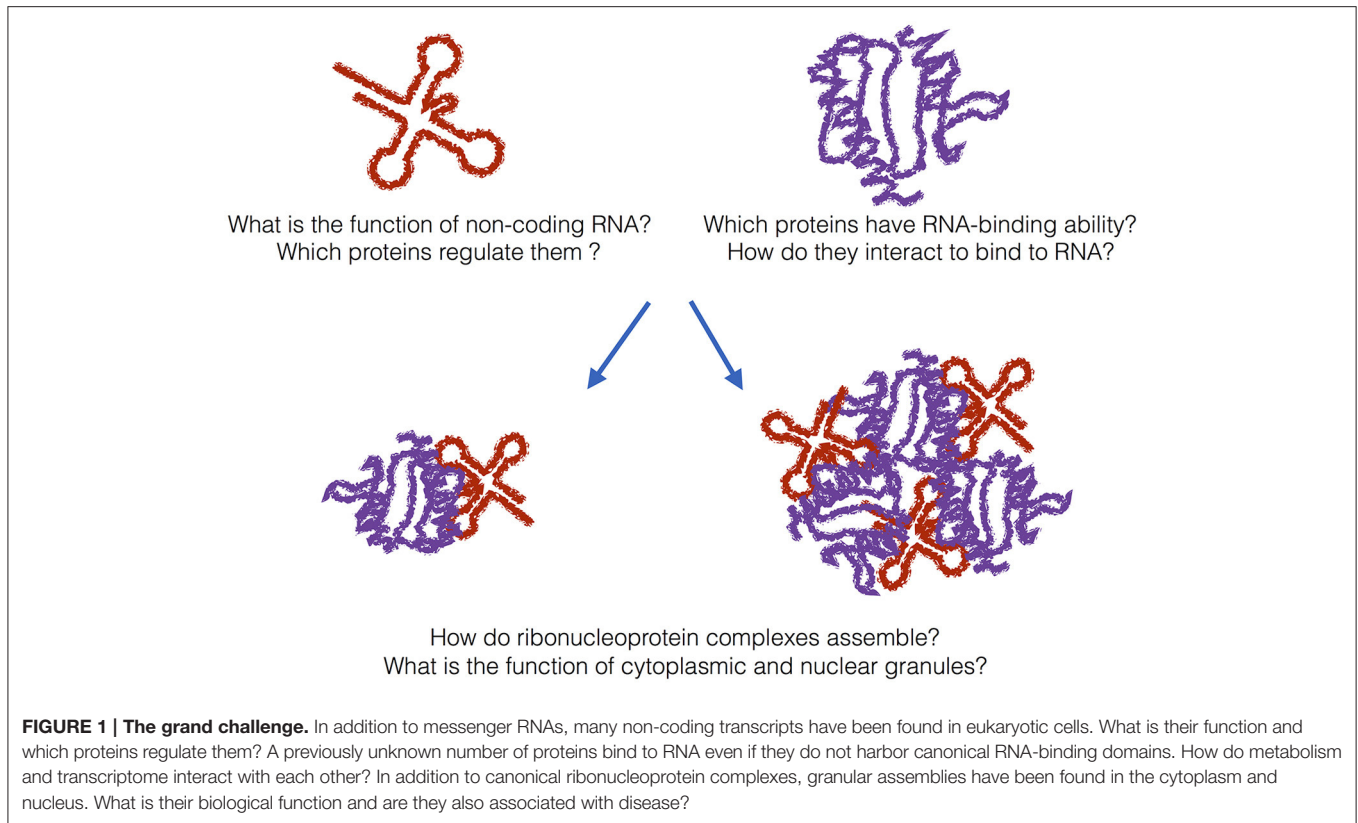
The human genome project was completed in the early 2000s (Lander et al., 2001). Recently, impacts of next-generation sequencing technologies have been massive. Plans to sequence 1000 different human genomes have been pursued and completed. Thousands of different bacteria species have been sequenced, and now over a hundred different plant species have been sequenced. Thanks to Roderic Guigo’ and other contributors of the ENCODE project, we now know that a large portion of the genome is transcribed into RNA, but not translated into proteins, resulting in more than half a million of non-coding RNA in eukaryotes (Djebali et al., 2012). Are these non-coding RNA species some kind of transcriptional noise or do they have a biological function? On the one hand, since no gene promoter can be considered silent at all times, ultra-deep sequencing could just reveal single copy RNA transcripts with little functional significance. On the other hand, non-coding RNA could be regulating a number of coding-genes contributing to the complexity of higher eukaryotes. A clear understanding of the biological functions of non-coding transcripts will provide novel insights into our understanding of the molecular biology (**Figure 1**). What protein networks assist the long non-coding RNAs and what is their specificity?

## HOW MANY PROTEINS HAVE RNA-BINDING ABILITY?

Deep-sequencing approaches based on pull-downs of proteins (e.g., PAR-CLIP and iCLIP) and RNAs (e.g., CHART and CHIRP) as well as *in vitro* evolution methods (e.g., SELEX) have started to unveil the targets of a number of RNA-binding proteins in the cell at defined conditions (König et al., 2010; Bechara et al., 2013; Chu et al., 2015; Darmostuk et al., 2015). Despite the growing amount of data collected, many questions remain to be answered. As shown by the groups of Mattias Hentze and Markus Landthaler, a previously unknown number of proteins have RNA-binding ability, although they do not harbor canonical RNA-binding domains (Baltz et al., 2012; Castello et al., 2012). Intriguingly, a large fraction of these proteins, such as for instance iron responsive protein IRP-1 (Philpott et al., 1994, 1; Cirillo et al., 2013), have a parallel, or moonlighting, activity as metabolic enzymes (Beckmann et al., 2015) (**Figure 1**). How do transcriptomic and metabolomic pathways interact with each other in the cell? What are the targets of the newly discovered RNA-binding proteins?

## LET THE GRAND CHALLENGE BEGIN!

The field of protein–RNA interactions is moving fast and a number of fascinating hypotheses have been recently formulated on ribonucleoprotein complexes. A PubMed search using the terms “protein” and “RNA” shows that the number of indexed publications increased progressively in the last decades: 736



(period 1985–1995), 1377 (1995–2005), and 1768 (2005–2016) manuscripts. The overall increase of about 250% indicates the strong interest in the field and that in the future we might witness a revolution in the study of protein–RNA networks!

Indeed, we do not have yet a complete understanding of how protein–RNA binding specificity is achieved and how the regulatory function of individual proteins is influenced by synergy and competition with other molecules. Novel approaches based on biochemical and functional studies, such as for instance SHAPE (Wan et al., 2011) and CRISPR (Hsu et al., 2014) paired with bioinformatics will lead to a better understanding of the principles underlying protein–RNA networks. In particular, advances based on high-resolution (STochastic Optical Reconstruction Microscopy, STORM) and biophysical (Nuclear Magnetic Resonance NMR) characterization will be key to derive mechanistic models for the interactions. Computational models will be an important source of information to identify new trends, understand the principles of molecular recognition and design experiments. Improvements in the theoretical models and validation of their predictions will be crucial to achieve a better description of the role of coding and non-coding RNAs in protein networks (Cirillo et al., 2014).

## REFERENCES

Baltz, A. G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., et al. (2012). The mRNA-bound proteome and its global

## AUTHOR CONTRIBUTIONS

GT wrote the “Grand Challenge” to introduce the most recent advances in the field of protein–RNA as well as the challenges that experimental and computational approaches will have to face in future studies.

## FUNDING

My research received funding from the European Union Seventh Framework Programme (FP7/2007–2013), through the European Research Council, under grant agreement RIBOMYLOME\_309545 (GT), and from the Spanish Ministry of Economy and Competitiveness (BFU2014-55054-P). I also acknowledge support from AGAUR (2014 SGR 00685), the Spanish Ministry of Economy and Competitiveness, “Centro de Excelencia Severo Ochoa 2013–2017” (SEV-2012-0208).

## ACKNOWLEDGMENTS

I would like to thank all members of the lab, Miriam Proietti and Gianni de Felice for stimulating discussions.

occupancy profile on protein-coding transcripts. *Mol. Cell* 46, 674–690. doi: 10.1016/j.molcel.2012.05.021

Bechara, E. G., Sebestyén, E., Bernardis, I., Eyra, E., and Valcárcel, J. (2013). RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control

- cancer cell proliferation. *Mol. Cell* 52, 720–733. doi: 10.1016/j.molcel.2013.11.010
- Beckmann, B. M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.-M., et al. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigms. *Nat. Commun.* 6:10127. doi: 10.1038/ncomms10127
- Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoeye, C., Gharakhani, J., et al. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324, 1729–1732. doi: 10.1126/science.1172046
- Buchan, J. R., and Parker, R. (2009). Eukaryotic stress granules: the ins and out of translation. *Mol. Cell* 36:932. doi: 10.1016/j.molcel.2009.11.020
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393–1406. doi: 10.1016/j.cell.2012.04.031
- Chao, J. A., Patskovsky, Y., Almo, S. C., and Singer, R. H. (2008). Structural basis for the coevolution of a viral RNA-protein complex. *Nat. Struct. Mol. Biol.* 15, 103–105. doi: 10.1038/nsmb1327
- Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., et al. (2015). Systematic discovery of Xist RNA binding proteins. *Cell* 161, 404–416. doi: 10.1016/j.cell.2015.03.025
- Cirillo, D., Agostini, F., Klus, P., Marchese, D., Rodriguez, S., Bolognesi, B., et al. (2013). Neurodegenerative diseases: quantitative predictions of protein-RNA interactions. *RNA* 19, 129–140. doi: 10.1261/rna.034777.112
- Cirillo, D., Marchese, D., Agostini, F., Livi, C. M., Botta-Orfila, T., and Tartaglia, G. (2014). Constitutive patterns of gene expression regulated by RNA-binding proteins. *Genome Biol.* 15:R13. doi: 10.1186/gb-2014-15-1-r13
- Darmostuk, M., Rimpelova, S., Gbelcova, H., and Ruml, T. (2015). Current approaches in SELEX: an update to aptamer selection technology. *Biotechnol. Adv.* 33, 1141–1161. doi: 10.1016/j.biotechadv.2015.02.008
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108. doi: 10.1038/nature11233
- Gall, J. G. (2000). Cajal bodies: the first 100 years. *Annu. Rev. Cell Dev. Biol.* 16, 273–300. doi: 10.1146/annurev.cellbio.16.1.273
- Higgs, P. G., and Lehman, N. (2015). The RNA World: molecular cooperation at the origins of life. *Nat. Rev. Genet.* 16, 7–17. doi: 10.1038/nrg3841
- Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. doi: 10.1016/j.cell.2014.05.010
- Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523. doi: 10.1016/j.cell.2014.11.035
- König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., et al. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* 17, 909–915. doi: 10.1038/nsmb.1838
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483, 336–340. doi: 10.1038/nature10879
- Matera, A. G., and Wang, Z. (2014). A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* 15, 108–121. doi: 10.1038/nrm3742
- Palade, G. E. (1955). Studies on the endoplasmic reticulum. *J. Biophys. Biochem. Cytol.* 1, 567–582.
- Papasaïkas, P., Tejedor, J. R., Vigevani, L., and Valcárcel, J. (2015). Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol. Cell* 57, 7–22. doi: 10.1016/j.molcel.2014.10.030
- Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. *Mol. Cell* 25, 635–646. doi: 10.1016/j.molcel.2007.02.011
- Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D., and Sutherland, J. D. (2015). Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* 7, 301–307. doi: 10.1038/nchem.2202
- Philpott, C. C., Klausner, R. D., and Rouault, T. A. (1994). The bifunctional iron-responsive element binding protein/cytosolic aconitase: the role of active-site residues in ligand binding and regulation. *Proc. Natl. Acad. Sci. U.S.A.* 91, 7321–7325.
- Spector, D. L. and Lamond, A. I. (2011). Nuclear Speckles. *Cold Spring Harb Perspect Biol.* 3:a000646. doi: 10.1101/cshperspect.a000646
- Wan, Y., Kertesz, M., Spitale, R. C., Segal, E., and Chang, H. Y. (2011). Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12, 641–655. doi: 10.1038/nrg3049
- Wolozin, B. (2012). Regulated protein aggregation: stress granules and neurodegeneration. *Mol. Neurodegener.* 7:56. doi: 10.1186/1750-1326-7-56

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Tartaglia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.