# RNA Regulatory Networks as a Control of Stochasticity in Biological Systems

Marylène Vandevenne*, Michael Delmarcelle and Moreno Galleni

InBioS – Center for Protein Engineering, University of Liège, Liège, Belgium

The discovery that the non-protein coding part of human genome, dismissed as "junk DNA," is actively transcribed and carries out crucial functions is probably one of the most important discoveries of the past decades. These transcripts are becoming the rising stars of modern biology. In this review, we have casted a new light on RNAs. We have placed these molecules in the context of life origins, evolution with a big emphasize on the "RNA networks" concept. We discuss how this view can help us to understand the global role of RNA networks in modern cells, and can change our perception of the cell biology and therapy. Finally, although high-throughput methods as well as traditional case-to-case studies have laid the groundwork for our current knowledge of transcriptomes, we would like to discuss new strategies that are better suited to uncover and tackle these integrated and complex RNA networks.

Keywords: regulatory RNA networks, non-coding RNAs, cell stochasticity/determinism, RNA world theories, origins of life

## INTRODUCTION: NON-CODING RNA IN THE SPOTLIGHTS

Proteins have been for a long time considered as major effectors of most cellular processes involved in cell metabolism, homeostasis, and genetic regulation. If DNA had the second role of genetic information storage, RNA was reduced to a simple genetic intermediate step between DNA and proteins. It's only recently that the scientific community has gain increasing interests in RNAs for several observations. One of the most intriguing one is the contradiction between the number of protein-coding genes and the complexity of organisms. For example, the size of the human genome is currently estimated to 19 000 protein-coding genes (Ezkurdia et al., 2014) which is far below the 100 000 genes that were initially predicted. These 19 000 coding genes are unexpectedly slightly fewer compared to those of nematodes (*Caenorhabditis elegans* (Hillier et al., 2005) with ≈ 20 000 protein-coding genes or flies (*Drosophila melanogaster*) with 14 000 protein-coding genes and only five fold more abundant compared to bacteria (*Escherichia coli*) with ≈ 4 500 protein-coding genes. The source of organism complexity and diversity might therefore rather rely on how these genes are used and regulated (King and Wilson, 1975; Franchini and Pollard, 2017). A second important observation was highlighted in 2012, when the international ENCODE project has established for the first time that 75% of the human genome is transcribed into RNAs while only 2% of these transcripts are translated into proteins (ENCODE Project Consortium, 2012). This indicates that 98% of the transcripts are not translated into proteins. Therefore RNAs are much more abundantly represented in the cell compared to proteins.

The last past years have witnessed a strong interest in these, so-called, non-coding RNAs (ncRNAs), since many of them have emerged as key players in the cell biology with important regulatory roles and, therefore, were associated to numerous diseases ranging from cancers to neurological disorders (Taft et al., 2010; Fu, 2014; Barta and Jantsch, 2017). Thus, ncRNAs represents a gold mine of potential new biomarkers and drug targets. However, we have only started to scratch the surface of regulatory ncRNAs. Their structures, precise molecular mechanisms, biological functions and overall role of this huge amount of RNAs in the cell remains poorly understood.

Importantly, if various classes of regulatory RNAs (e.g., miRNAs, siRNAs, ...) are single-stranded and act by base pairing with other nucleic acids (RNAs or DNAs), it is very likely that a vast majority of non-coding transcripts adopts complex 3D structure(s) to achieve their biological functions. These "structured" RNAs act using very diverse mechanisms including RNA-RNA, RNA-ligand, RNA-protein, RNA-DNA, and RNA-substrate interactions (Wang and Chang, 2011). However, currently, less than 1% of all the structures reported in the Protein Data Bank (PDB) are RNA structures. Like proteins, RNA structures have different organization levels: the first one consists in the nucleotide sequence that folds on itself via Watson–Crick base-pairing to form secondary structure elements (e.g., hairpins, bulges...) and unpaired regions. Finally, these elements are precisely organized in space to form the tertiary structure of the RNA that is, in most cases, stabilized by divalent ions, e.g., $Mg^{2+}$ (Westhof, 2000). Finally, it is worth to mention that RNA structures are highly dynamic and modulated by binding to partners, which add another degree of complexity to these structures.
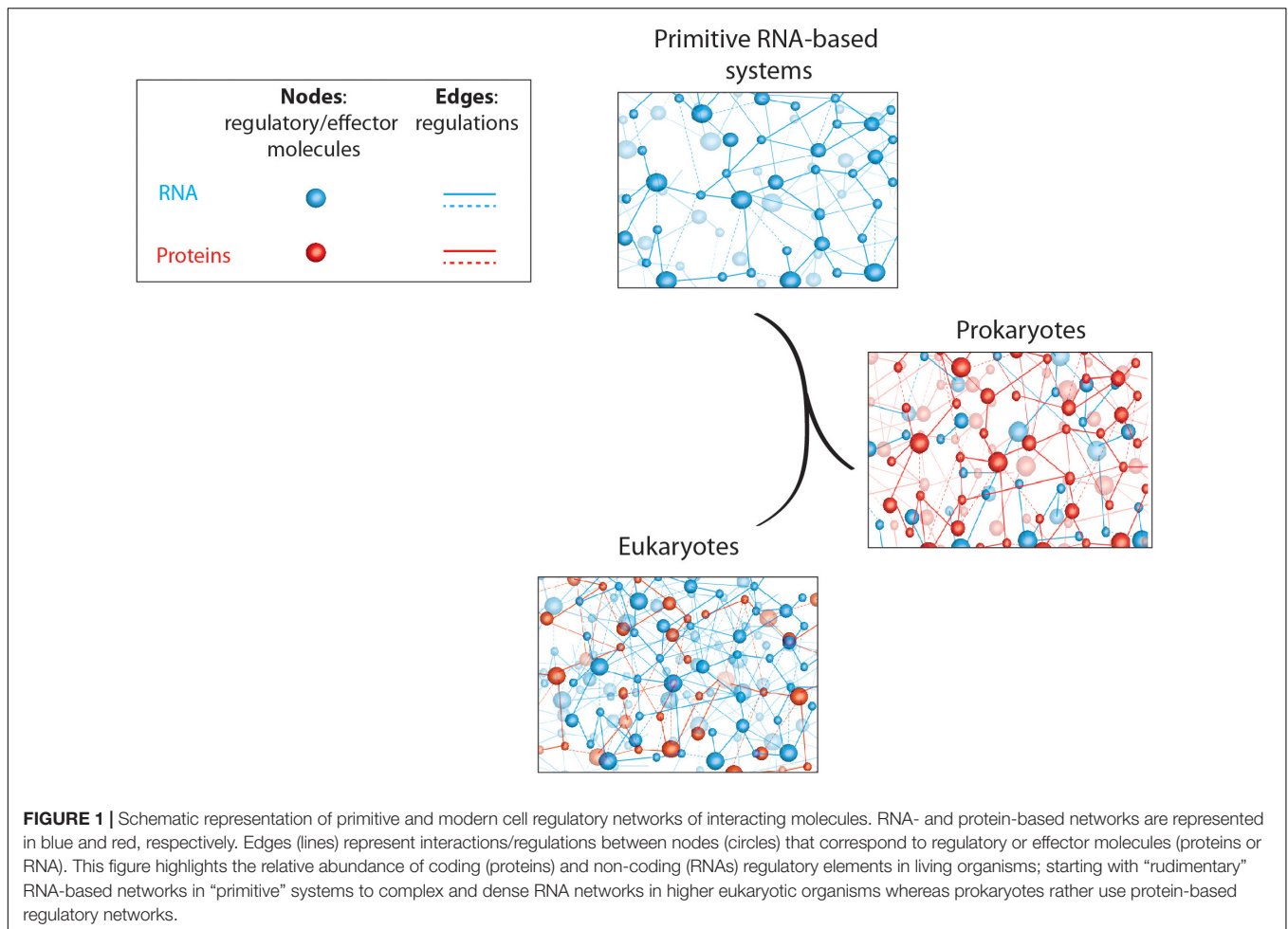
## THE "RNA WORLD" AND RNA NETWORKS THEORIES OF LIFE ORIGINS

The molecular mechanisms and actors that have led to the origin of life billions of years ago remain among the most fundamental unsolved enigmas of modern science. In this context, RNAs were suggested to be the first biological molecules on Earth, mostly because of their ability to do both: store genetic information and catalyze various biochemical reactions (Yanagawa, 1994; Higgs and Lehman, 2015). In addition, RNA has different properties that make it the ideal candidate as the predecessor of proteins and DNA: (i) it can exists in a single-stranded form, in duplexes or adopt more complex structures; (ii) RNA subunits (e.g., ATP) constitute a source of energy; (iii) and finally RNA has the ability to evolve under selective pressure, as demonstrated in SELEX experiments (Systematic Evolution of Ligands by Exponential enrichment) (Alberts et al., 2002; Harris, 2010; Yarus, 2010). Finally, self-replicating RNAs have been developed *in vitro* (Ekland and Bartel, 1996; Johnston et al., 2001; Shechner and Bartel, 2011; Robertson and Joyce, 2014) using engineered ribozymes (catalytic RNAs) with RNA-template RNA ligase activities that join oligonucleotide substrates to form complementary RNA products (James and Ellington, 1999;

Levy and Ellington, 2001; Robertson and Joyce, 2014). This multitask property and high plasticity in terms of structures and activities of RNAs strongly support the hypothesis of the, so-called, "RNA world" theory. According to this hypothesis, it's only later in evolutionary time that DNA arose and took over the storage of genetic information whereas proteins supported the catalysis tasks in the cells. Therefore, it is reasonable to state that life has started with non-specialized molecules (RNAs) able to accomplish different tasks but with limited efficiencies. Evolution has led to the selection of more specialized molecules (DNA, proteins) able to take over restricted functions in the cells but with much higher efficiencies. Indeed, catalytic RNAs increase reaction rates by up to $10^{11}$-fold with reaction efficiencies ($k_{cat}/K_m$) up to $10^8$ $M^{-1}$ $min^{-1}$, which is $10^3$-fold less than what is observed for proteins catalyzing equivalent reactions (Cech, 1993; Narlikar and Herschlag, 1997; Tanner, 1999). Thus, compared to proteins and DNA, RNAs can be seen as the most fundamental elements in the cell which explains why, nowadays, RNAs are found in all fundamental processes in the cell (tRNA, rRNA, and mRNA...).

Besides the simplified and "individualistic" view of a unique auto-replicative RNA molecule at the origins of life, another theory, rather "communistic," postulated that life started with ensembles of RNA molecules (Yeates and Nehman, 2016). This theory is seducing, because the definition of "life" consists in an ensemble of physical entities that carry out biological processes, and form a system that is self-sustaining and capable of Darwinian evolution. Therefore, this communistic view of life origins where a self-sustaining system arose from different populations of RNAs that interact with each other and have complementary tasks to manage different processes (ex: catalysis, support for genetic information, and substrates/products of chemical reactions) is an hypothesis that has been well admitted by the scientific community (Eigen and Schuster, 1977; Ganti, 2003; Vaidya et al., 2012; Vasas et al., 2012; Hordijk and Steel, 2013; Higgs and Lehman, 2015; Yeates and Nehman, 2016). In this "*RNA network*" hypothesis, each individual RNA harbors one or several function(s) that complement(s), or partially overlap(s) with the functions carried out by other RNAs. This concept of prebiotic networks constituted of interacting RNA species that evolve and act co-operatively has been demonstrated experimentally (Vaidya et al., 2012) and mathematically modeled (Hordijk and Steel, 2013). Indeed, co-operating molecules with complementary activities make the biological system more robust to external and internal changes.

Notably, the "individualistic" and "communistic" theories are not contradictory. We can imagine that the first theory evolves toward the second one, and similarly, the second one evolves toward the first one to give auto-replicative entities. However, the communistic theory is surely more admitted by the scientific community. First, co-operative replication within these ensembles of RNAs is easier than self-replication of a single RNA (Kauffman, 1993). Secondly, cooperative molecular networks have demonstrated fitness benefits and selection preferences compared to selfish entities (Eigen and Schuster, 1977; Ganti, 2003; Vaidya et al., 2012; Vasas et al., 2012; Hordijk and Steel, 2013; Higgs and Lehman, 2015).

**FIGURE 1 |** Schematic representation of primitive and modern cell regulatory networks of interacting molecules. RNA- and protein-based networks are represented in blue and red, respectively. Edges (lines) represent interactions/regulations between nodes (circles) that correspond to regulatory or effector molecules (proteins or RNA). This figure highlights the relative abundance of coding (proteins) and non-coding (RNAs) regulatory elements in living organisms; starting with "rudimentary" RNA-based networks in "primitive" systems to complex and dense RNA networks in higher eukaryotic organisms whereas prokaryotes rather use protein-based regulatory networks.
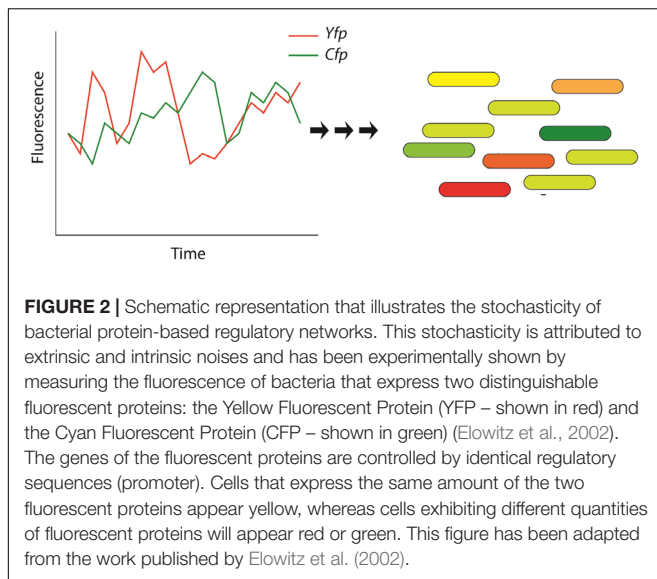
## UNDERSTANDING THE GLOBAL ROLES OF RNAs IN THE CELL BIOLOGY

It is interesting to note that, in response to specific nutritional or environmental conditions, all the cell signaling pathways starting from stimuli perception, activation of appropriate genes and then modification of the cell behavior have been described with proteins as main actors of cell decisions and homeostasis. If this is a valid simplification in prokaryotic cells, where only 12% of prokaryotic genomes are non-protein coding DNA (Ahnert et al., 2008). This cannot be true for eukaryotic cells where the percentage of non-protein coding DNA increases quadratically with organism complexity (Ahnert et al., 2008) to reach 98% in human (Mattick, 2001). This huge energetic cost associated with massive transcription of the genome cannot be due to random or residual RNA polymerases activities: it has to have an important purpose for the cell biology that proteins and DNA are not able to carry out. An increasing number of evidences show that besides housekeeping functions, numerous RNAs carry out important regulatory roles in both eukaryotic and also in prokaryotic cells using highly diverse mechanisms [for reviews on eukaryotic and prokaryotic mechanisms of RNAs (see Waters and Storz, 2009; Marchese et al., 2017)].

This high diversity in RNA mechanisms is directly associated with their high plasticity in terms of structures, partners of interaction and therefore functions (Ancel and Fontana, 2000). However, it is interesting to note that, so far, we don't really have yet a comprehensive and overall understanding of the global roles of RNA networks in the cell and how RNA and protein networks are integrated to regulate gene expression and cell fate.

Based on the observations described above, it seems that prokaryotes encode a large majority of their regulatory overheads in proteins, whereas eukaryotes rather recruit regulatory RNAs for this purpose (Pheasant and Mattick, 2007; Taft et al., 2007; Ahnert et al., 2008). This observation is somehow surprising if we consider the RNA world theories discussed in this review. Indeed, modern cells evolved from the most primitive life form, which presumably consisted in organized RNA networks, and gave rise to prokaryotes (archaebacteria and eubacteria) and eukaryotes. If it is a common thought to consider prokaryotes more "primitive" than eukaryotes, it is therefore surprising to observe that prokaryotes use more recent and evolved regulatory molecules (proteins) whereas complex eukaryotes rather use the good old, and in a way more "primitive," RNA networks for regulatory purposes (**Figure 1**).

**FIGURE 2 |** Schematic representation that illustrates the stochasticity of bacterial protein-based regulatory networks. This stochasticity is attributed to extrinsic and intrinsic noises and has been experimentally shown by measuring the fluorescence of bacteria that express two distinguishable fluorescent proteins: the Yellow Fluorescent Protein (YFP – shown in red) and the Cyan Fluorescent Protein (CFP – shown in green) (Elowitz et al., 2002). The genes of the fluorescent proteins are controlled by identical regulatory sequences (promoter). Cells that express the same amount of the two fluorescent proteins appear yellow, whereas cells exhibiting different quantities of fluorescent proteins will appear red or green. This figure has been adapted from the work published by Elowitz et al. (2002).

## Regulatory Networks Are Stochastic

A particularly important property of genetic regulatory networks is their intrinsic "*stochasticity*" (**Figure 2**; Elowitz et al., 2002; Kaern et al., 2005; Losick and Desplan, 2008; Silva-Rocha and de Lorenzo, 2010; Locke et al., 2011; Ben-Jacob et al., 2014). This stochasticity in gene regulation has important impacts on the cell since the amount of many cellular components (DNA, regulatory molecules) is very low (Elowitz et al., 2002) and explains, for a large amount, the heterogeneity or cell-to-cell variations often observed in clonal populations of cells that are submitted to identical environmental/stress conditions. This observation has been well documented in various bacterial strains (e.g., *Bacillus subtilis*, *E. coli*...) (Ben-Jacob et al., 2014; Davis and Isberg, 2016) as well as complex eukaryotic organisms (Kar et al., 2009; Salari et al., 2012; Dacheux et al., 2017). One of the most famous and relevant single-cell experiments that explored stochastic gene expression is illustrated in **Figure 2**. In this study, Elowitz et al. (2002) analyzed the variability of expression of a specific promoter of *E. coli* (*E. coli*). The authors inserted two copies of this promoter in the genome of *E. coli*: one copy driving the expression of the Cyan Fluorescent Protein (CFP) and the second driving the expression the Yellow Fluorescent Protein (YFP). The authors reported high variability on the fluorescence type that was emitted by individual bacteria. The source of this variability relies on the stochasticity of genetic regulation that is explained by two different factors or noises (Elowitz et al., 2002; Silva-Rocha and de Lorenzo, 2010). First, "extrinsic" noises that arise because the expression of each gene/protein is controlled by the concentrations, fluctuations in the amounts, activities and locations of metabolites and regulatory molecules (e.g., polymerases, ribosomes...). Secondly, "*intrinsic*" noises imply that, even if the concentrations and states of every cellular component would be identical in every cell, the rate of expression of particular genes would also vary from cell to cell due to stochastic microscopic events that are intrinsic to

transcription/translation events and influence gene regulation (e.g., collision rate) (Elowitz et al., 2002).

In prokaryotes and "simple" unicellular organisms, stochastic gene regulatory networks can be seen as advantageous since it allows sub-populations of bacteria to be prepared/conditioned to face and adapt very quickly to drastic environment changes (Schultz et al., 2009, 2013) and, therefore, can be considered as beneficial in various contexts: metabolism (ex: lactose utilization in *E. coli* (Ozbudak et al., 2004; Mettetal et al., 2006), stress response (ex: competence and sporulation in *B. subtilis* (Maamar et al., 2007), and pathogenesis (ex: antibiotic resistance in *Mycobacterium tuberculosis* (Stewart et al., 2003). On the other hand, it substantially limits the precision of gene regulation; which can be harmful. This is particularly true for complex eukaryotic organisms, in which the ultimate manifestations of this stochasticity can lead to aging (ex: murine cardiac myocytes (Bahar et al., 2006), cancers (Davies and Agus, 2015), neurodegenerative (ex: Alzheimer disease (Hadjichrysanthou et al., 2018), and autoimmune (Fatehi et al., 2018) disorders. Therefore, it seems that life, in particular in complex organisms, relies on a good compromise between randomness and determinism/finality. From randomness (stochasticity and trend to chaos of complex dynamic living systems) to the precise coordination of development, higher organisms have found a way to balance these two apparently opposite aspects of their internal way of working (Raj and van Oudenaarden, 2008).

## Complex Eukaryotes Require a Large Quantity of Non-coding RNAs for Their Regulation

As mentioned above, one of the main differences between prokaryotic and eukaryotic genomes is the proportion of non-protein coding DNA. Furthermore, this ncDNA portion increases with organism complexity: the ncDNA size follows a quadratic equation that is function of the total length of exonic DNA (Ahnert et al., 2008). Why do eukaryotes produce so much non-coding transcripts and how can this be correlated to their complexity? First, we have to consider that one of the main differences of complex eukaryotic organisms compared to prokaryotes is the spatio-temporal differentiation: with the specialization of cells into tissues, with all the different tissues that act together to coordinate organism homeostasis. Secondly, the generation time of complex eukaryotes can be much longer compared to prokaryotes or unicellular eukaryotes (examples: 20 min for *E. coli* (bacteria), 80 min for *Saccharomyces cerevisiae* (yeast) to several decades for human neurons). Since random events can be quantified by a frequency, organisms with longer lifetime are more prone to stochasticity and chaotic drift. Therefore, it is more obvious that eukaryotic cells need to undergo a much tighter regulation compared to prokaryotes since it is more crucial for eukaryotes to compensate and regulate randomness and better control cell fate.

How do eukaryotes satisfy this required tighter regulation? The answer most likely relies on the complexity of their regulatory networks. Indeed, recent studies have shown that more complex networks are better at coping with both: intrinsic

and extrinsic noises that are the sources of stochasticity. Intrinsic noise tends to decrease with network complexity, and extrinsic noise tends to have less impact in complex regulatory networks (Cardelli et al., 2016). Nonetheless, the problem with protein-based regulatory networks is that their complexity is limited because of their restricted capacity to make interactions/regulations (Mattick and Gagen, 2005; Ahnert et al., 2008). Indeed, an interesting study conducted in yeast has revealed that the number of protein interfaces available for regulation or receiving regulation of other molecules is limited to 14 (Kim et al., 2006). With this restriction, various mathematical and theoretical studies have revealed that, given the genome sizes and the number of coding genes (proteins) found in complex eukaryotes, global regulation of all the genome components cannot possibly be achieved by proteins, since the number of regulations that proteins are able to do is way too low (Mattick, 2004; Taft et al., 2007; Ahnert et al., 2008). This is why it is very likely that the huge non-coding portions of eukaryotic genomes account for a large majority to the control of most genome components and regulate cell fate and homeostasis (Mattick, 2004; Pheasant and Mattick, 2007; Taft et al., 2007). In conclusion, ncRNAs are abundantly required in eukaryotes because they scale up the number of regulatory connections that is required for a fully integrated regulatory network (Ahnert et al., 2008).

Importantly, if the number of possibilities to receive/give regulations has not been established yet for RNAs, it is very likely that this number would be significantly higher compared to proteins. Indeed, proteins are more specialized molecules, whereas RNAs exhibit much higher plasticity in terms of structures, binding partners and therefore functions. For example, another interesting mechanism that RNAs could potentially use to increase the dynamic and plasticity of their high interaction potential is their possibility to interact with each other (trans interactions) and generate specific structural motifs that could modulate their interaction to binding partners (Doyle and Tenenbaum, 2014). Furthermore, in analogy to epigenetic for DNA, RNAs are also known to undergo diverse biochemical modifications. To date, well over hundreds of modifications were reported for RNAs (Grosjean, 2005). These modifications (with methylation predominating) can be added post-transcriptionally to every positions of either purine or pyrimidine rings (Cantara et al., 2011). Like epigenome, the, so-called, epitranscriptome is highly dynamic and include "writers" and "erasers" that modify coding or non-coding RNAs and "readers" that can translate these modifications into functional changes (Yang et al., 2018). Even if the functional relevance and molecular mechanisms of epitranscriptomic remain largely unexplored, it is very likely that these modifications shape RNA structure, stability and therefore adds many additional possibilities for regulating RNA interactions.

Interestingly, prokaryotes do not seem to require such a complex regulatory RNA network with only 12% of genomic ncDNA. Given their very short generation time and poor differentiation, it seems that harmful effects of stochasticity are limited. Instead, stochasticity is rather beneficial for them in order to adapt, respond and evolve much faster compared to complex eukaryotic organisms.

## Global Role of Regulatory RNA Networks

How can we integrate RNA- and protein-based regulatory networks from a global point of view? How are these two types of regulatory molecules linked to each other to control cell fate and decision? First, we have to consider that one of the main characteristics of RNAs, in contrast to proteins, is their high plasticity as mentioned above. Based on this property and what we discussed in this review, we postulate that one plausible global role of regulatory RNA ensembles/networks would be that using their high interaction/connection potential, they can "buffer" the stochasticity of genetic regulatory networks in order to guide cells toward an appropriate response observed upon specific stimuli and maintain homeostasis. According to this hypothesis, we might rather see regulatory RNAs as major "moderators" that supervise cellular pathways and guide cell decisions in order to prevent cells from chaotic drifts and death. This concept is interesting if we considered that primitive RNA networks led to the first life forms and remained conserved in modern cells. Consequently, RNA networks could be seen as the balance between randomness and determinism; it is therefore probably not surprising that life might have emerged from these RNA networks.

A recent study, conducted by Du et al. (2016), illustrates very well this hypothesis of RNA networks as regulators of stochasticity. In this study, the authors showed that an RNA regulatory network (composed of long-non-coding RNAs-lncRNAs) affects the expression of numerous protein-coding prostate cancer driver genes by acting as "sponges" by binding to miRNAs and preventing them to destabilize protein-coding transcripts. They demonstrated that this RNA network regulation is multiple: many protein-coding genes were regulated by only one lncRNA. Finally, they showed that restoring this lncRNA network is sufficient to suppress tumor activities in prostate cell lines. This study represents a good example where uncontrolled stochasticity (cancer) can be muted using an RNA network that acts as master regulator of several regulatory proteins/RNAs.

## CONCLUDING REMARKS AND FUTURE PERSPECTIVES

This view of RNA networks as master regulators that control the balance between randomness and determinism to control cell fate is important because it can change our perception of cell biology and provide new opportunities to design better therapeutic interventions. Indeed, we could imagine more efficient strategies to prevent or restore control on cancers, neurological disorders or senescence by focusing on these RNA regulatory networks rather than developing protein-based therapeutics (ex: antibodies, enzyme inhibitors...).

The emerging question is: how can we tackle the huge complexity of regulatory RNA networks given that we have only scratched the surface of protein-based networks? Nowadays, the current trend to fill in the gap of knowledge in the biology, functions and structures of regulatory RNAs is undoubtedly the use of high-throughput methods (Weidmann et al., 2016). Next-generation sequencing (NGS) is used for genome-wide

measurements of inter- and intra-molecular RNA duplexes in living cells [PARIS, LIGR-Seq, and SPLASH (Aw et al., 2016; Lu et al., 2016; Sharma et al., 2016)] as well as for the identification of protein-binding partners [e.g., HITS-CLIP, PAR-CLIP (Zhang and Darnell, 2011; Li et al., 2014; Spitzer et al., 2014)], and several variants iCLIP, iCLAP (Huppertz et al., 2014; Li et al., 2014). If these methods have undoubtedly brought crucial information (big data) in the field of RNA, they didn't really bring, so far, the expected breakthrough. This is probably because these data need to be used and transposed to detailed mechanisms of action in order to delineate general rules for regulatory RNA networks. In a similar manner to what we did with proteins, we need to use old school/traditional approaches (e.g., mutations to see effects on RNA structure and function) and study specific cases of regulatory RNAs with a particularly big emphasize on the structural data that currently cruelly lack.

## Fundamental Differences Between Proteins and RNAs and the Need of New and Adapted Approaches to Tackle the Complexity of RNA Networks

Whereas protein-based networks are highly specialized with delimited functions and tasks, RNA-based networks are much more complex systems to study. Indeed, RNAs are less specialized molecules that act as part of bigger networks where all individual RNAs can interact with many different other proteins and nucleic acids. RNAs also exhibit much more dynamic and plasticity in their structures, biochemical modifications (epitranscriptomic) and therefore binding partners and functions. In this regard, the recent progresses made in cryo-electronic microscopy (cryoEM) will certainly facilitate RNA structure determination and will also allow assessing the dynamic and plasticity of these structures.

In addition, this high functional plasticity of RNAs might explain why inactivation of a regulatory RNA can be trickier to analyze and will most likely generate more subtle changes with only partial destabilization of the downstream cell pathway(s). This is why we need to adapt our experimental approaches to the complexity of RNA networks. Therefore, the idea is to detect differences in the dispersion but not necessary in the average of the affected pathway(s). This is why single-cell analysis constitutes an attractive approach to highlight subtle changes in the cell (sub)-populations. In this approach, single cell are sorted and isolated using several well-established methods [e.g., micromanipulation, laser-capture microdissection, and fluorescence-activated cell sorting (Navin and Hicks, 2011; Hodzic, 2016)] as well as more recent techniques such as microfluidic (Yilmaz and Singh, 2012; Saliba et al., 2014). Then the use of high-throughput sequencing (whole-transcriptome analysis) on the isolated cells will allow amplification of small differences and will permit expression profiling and sequencing of coding and non-coding RNAs present in a single cell. Combined with strong statistical analysis, this offers the possibility to assess the transcriptomic heterogeneity and subtle changes occurring upon inactivation of a specific regulatory RNA. In addition,

these high-throughput techniques can help us to identify the different RNA members of a network. Based on this identification, mutations of several members of the network can be envisaged to generate a detectable phenotype on the affected cell pathway.

## The Use of the Information Theories to Model the Complexity of RNA Networks

Finally, we need to better understand how the different regulatory RNAs work as networks and how they interact and are connected to each other as well as with DNA and regulatory proteins to better understand their roles in the cell biology.

How can we address this network complexity without considering the information theories? Indeed, a good analogy of complex biological regulatory network is the information network (Battail, 2013). In this analogy, all biochemical processes present in the cell can be seen as transfer of information. For example: (i) an hormone that binds to its receptor located on the cell surface to activate specific genes; and (ii) a population of bacteria that starts to sporulate under limiting nutrient conditions. These information transfers through the regulatory networks need to be robust and protected against the intrinsic and extrinsic noises, namely stochasticity, of the cell. It is surprising to note that information networks work exactly the same way since information transfer consists in sending "signals" through a network and the main purpose of this network is to protect these signals against noises until delivery point. In information theories, this protection of signals against noises is achieved using a network of, so-called, "correction codes" [ex: low density parity code (Wang et al., 2018), turbo codes (Valenti and Sun, 2002)]. These correction codes protect the original signals from noises in order to preserve the original content of these signals.

From this point of view, the analogy between RNA and information networks is striking. Since, as discussed in this review, we hypothesized that the major purpose of complex RNA networks such as the ones found in complex eukaryotic organisms is to "buffer" the stochasticity of biological systems. In other words, RNAs can protect living cells/organisms against chaotic drifts. With this view, RNA networks are analogous to "correction codes" found in the information theory. Therefore this involves that we could imagine to model RNA regulatory networks by extending information theories and thereby reach a wider and global comprehension of these molecules.

In conclusion, tighter regulation of complex organisms doesn't rely on their genome size (number of nodes/genes), but rather relies on the possibility of each molecule (nodes) to receive or exert more regulations (edges). In higher and complex organisms, spatio-temporal regulation is crucial and needs to be more tightly regulated with much more noise control. This noise control network is probably strongly intricate with the catalyzers and lower level regulators from the biotic era. Life has adopted two strategies: either it diminishes the impact of noises by shortening the lifetime using fast replication (thanks to the efficiency of proteins) or it increases the noise control

in complex organisms. As demonstrated by Ahnert et al. (2008) only RNAs can make a satisfactory number of possible edges compared to proteins and therefore create a sufficiently dense network for this purpose.

## AUTHOR CONTRIBUTIONS

MV wrote most of the manuscript. MD participated to the manuscript writing. MV and MD contributed the most to the ideas presented in the manuscript. MD and MG gave regular intellectual inputs and proofread the manuscript.

## REFERENCES

Ahnert, S. E., Fink, T. M., and Zinovyev, A. (2008). How much non-coding DNA do eukaryotes require? *J. Theor. Biol.* 252, 587–592. doi: 10.1016/j.jtbi.2008.02.005

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *The RNA World and the Origins of Life, in Molecular BIology of the Cell*, 4th Edn. New York, NY: Garland Science.

Ancel, L. W., and Fontana, W. (2000). Plasticity, evolvability, and modularity in RNA. *J. Exp. Zool.* 288, 242–283. doi: 10.1002/1097-010x(20001015)288%3A3%3C242%3A%3Aaid-jez5%3E3.0.co%3B2-o

Aw, J. G., Shen, Y., Wilm, A., Sun, M., Lim, X. N., Boon, K. L., et al. (2016). In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell.* 62, 603–617. doi: 10.1016/j.molcel.2016.04.028

Bahar, R., Hartmann, C. H., Rodriguez, K. A., Denny, A. D., Busuttil, R. A., Dolle, M. E., et al. (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* 441, 1011–1014. doi: 10.1038/nature04844

Barta, A., and Jantsch, M. F. (2017). RNA in disease and development. *RNA Biol.* 14, 457–459. doi: 10.1080/15476286.2017.1316929

Battail, G. (2013). Biology needs information theory. *Biosemiotics* 6, 77–103. doi: 10.1007/s12304-012-9152-6

Ben-Jacob, E., Lu, M., Schultz, D., and Onuchic, J. N. (2014). The physics of bacterial decision making. *Front. Cell Infect. Microbiol.* 4:154. doi: 10.3389/fcimb.2014.00154

Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., and Zhang, X. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* 39, D195–D201. doi: 10.1093/nar/gkq1028

Cardelli, L., Csikasz-Nagy, A., Dalchau, N., Tribastone, M., and Tschaikowski, M. (2016). Noise reduction in complex biological switches. *Sci. Rep.* 6:20214. doi: 10.1038/srep20214

Cech, T. R. (1993). *Structure and Mechanism of the Large Catalytic RNAs: Group I and Group II Introns and Ribonuclease P. in The RNA World*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 239–269.

Dacheux, E., Malys, N., Meng, X., Ramachandran, V., Mendes, P., and McCarthy, J. E. G. (2017). Translation initiation events on structured eukaryotic mRNAs generate gene expression noise. *Nucleic Acids Res.* 45, 6981–6992. doi: 10.1093/nar/gkx430

Davies, P. C., and Agus, D. B. (2015). Stochasticity and determinism in cancer creation and progression. *Converg. Sci. Phys. Oncol.* 1:026003. doi: 10.1088/2057-1739/1/2/026003

Davis, K. M., and Isberg, R. R. (2016). Defining heterogeneity within bacterial populations via single cell approaches. *Bioessays* 38, 782–790. doi: 10.1002/bies.201500121

Doyle, F., and Tenenbaum, S. A. (2014). Trans-regulation of RNA-binding protein motifs by microRNA. *Front. Genet.* 5:79. doi: 10.3389/fgene.2014.00079

Du, Z., Sun, T., Hacisuleyman, E., Fei, T., Wang, X., Brown, M., et al. (2016). Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nat. Commun.* 7:10982.

Eigen, M., and Schuster, P. (1977). The hypercycle. A principle of natural self-organization. Part A: emergence of the hypercycle. *Naturwissenschaften* 64, 541–565. doi: 10.1007/bf00450633

Ekland, E. H., and Bartel, D. P. (1996). RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* 382, 373–376. doi: 10.1038/382373a0

Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183–1186. doi: 10.1126/science.1070919

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247

Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., and Harrow, J. (2014). Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum. Mol. Genet.* 23, 5866–5878. doi: 10.1093/hmg/ddu309

Fatehi, F., Kyrychko, S. N., Ross, A., Kyrychko, Y. N., and Blyuss, K. B. (2018). Stochastic effects in autoimmune dynamics. *Front. Physiol.* 9:45.

Franchini, L. F., and Pollard, K. S. (2017). Human evolution: the non-coding revolution. *BMC Biol.* 15:89. doi: 10.1186/s12915-017-0428-9

Fu, X. D. (2014). Non-coding RNA: a new frontier in regulatory biology. *Natl. Sci. Rev.* 1, 190–204. doi: 10.1093/nsr/nwu008

Ganti, T. (2003). *The Principles of Life*. Oxford: Oxford University Press.

Grosjean, H. (2005). *Fine-Tuning of RNA Functions by Modification and Editing in Topics on Current Genetics*. Berlin: Springer.

Hadjichrysanthou, C., Ower, A. K., de Wolf, F., and Anderson, R. M. (2018). The development of a stochastic mathematical model of Alzheimer's disease to help improve the design of clinical trials of potential treatments. *PLoS One* 13:e0190615. doi: 10.1371/journal.pone.0190615

Harris, T. (2010). Evidence for RNA origins. *Nature* 464:494. doi: 10.1038/464494a

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

Hillier, L. W., Coulson, A., Murray, J. I., Bao, Z., Sulston, J. E., and Waterston, R. H. (2005). Genomics in C. elegans: so many genes, such a little worm. *Genome Res.* 15, 1651–1660. doi: 10.1101/gr.3729105

Hodzic, E. (2016). Single-cell analysis: advances and future perspectives. *Bosn. J. Basic Med. Sci.* 16, 313–314. doi: 10.17305/bjbms.2016.1371

Hordijk, W., and Steel, M. (2013). A formal model of autocatalytic sets emerging in a RNA replicator system. *J. Syst. Chem.* 4:3.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L. E., Sibley, C. R., Sugimoto, Y., et al. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* 65, 274–287. doi: 10.1016/j.ymeth.2013.10.011

James, K. D., and Ellington, A. D. (1999). The fidelity of template-directed oligonucleotide ligation and the inevitability of polymerase function. *Orig. Life Evol. Biosph.* 29, 375–390.

Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* 292, 1319–1325. doi: 10.1126/science.1060760

Kaern, M., Elston, T. C., Blake, W. J., and Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6, 451–464. doi: 10.1038/nrg1615

Kar, S., Baumann, W. T., Paul, M. R., and Tyson, J. J. (2009). Exploring the roles of noise in the eukaryotic cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6471–6476. doi: 10.1073/pnas.0810034106

Kauffman, S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford: Oxford University Press.

Kim, P. M., Lu, L. J., Xia, Y., and Gerstein, M. B. (2006). Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* 314, 1938–1941. doi: 10.1126/science.1136174

King, M. C., and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116. doi: 10.1126/science.1090005

Levy, M., and Ellington, A. D. (2001). The descent of polymerization. *Nat. Struct. Biol.* 8, 580–582.

Li, X., Song, J., and Yi, C. (2014). Genome-wide mapping of cellular protein-RNA interactions enabled by chemical crosslinking. *Genom. Proteom. Bioinform.* 12, 72–78. doi: 10.1016/j.gpb.2014.03.001

Locke, J. C., Young, J. W., Fontes, M., Hernandez, M., Jimenez, J., and Elowitz, M. B. (2011). Stochastic pulse regulation in bacterial stress response. *Science* 334, 366–369. doi: 10.1126/science.1208144

Losick, R., and Desplan, C. (2008). Stochasticity and cell fate. *Science* 320, 65–68. doi: 10.1126/science.1147888

Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., and Robinson, J. T. (2016). RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165, 1267–1279. doi: 10.1016/j.cell.2016.04.028

Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in gene expression determines cell fate in Bacillus subtilis. *Science* 317, 526–529. doi: 10.1126/science.1140818

Marchese, F. P., Raimondi, I., and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* 18:206. doi: 10.1186/s13059-017-1348-2

Mattick, J. S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* 2, 986–991. doi: 10.1093/embo-reports/kve230

Mattick, J. S. (2004). RNA regulation: a new genetics? *Nat. Rev. Genet.* 5, 316–323. doi: 10.1038/nrg1321

Mattick, J. S., and Gagen, M. J. (2005). Mathematics/computation. *Accelerating Netw. Sci.* 307, 856–858.

Mettetal, J. T., Muzzey, D., Pedraza, J. M., Ozbudak, E. M., and van Oudenaarden, A. (2006). Predicting stochastic gene expression dynamics in single cells. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7304–7309. doi: 10.1073/pnas.0509874103

Narlikar, G. J., and Herschlag, D. (1997). Mechanistic aspects of enzymatic catalysis: lessons from comparison of RNA and protein enzymes. *Annu. Rev. Biochem.* 66, 19–59. doi: 10.1146/annurev.biochem.66.1.19

Navin, N., and Hicks, J. (2011). Future medical applications of single-cell sequencing in cancer. *Genome Med.* 3:31. doi: 10.1186/gm247

Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I., and Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 427, 737–740. doi: 10.1038/nature02298

Pheasant, M., and Mattick, J. S. (2007). Raising the estimate of functional human sequences. *Genome Res.* 17, 1245–1253. doi: 10.1101/gr.6406307

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216–226. doi: 10.1016/j.cell.2008.09.050

Robertson, M. P., and Joyce, G. F. (2014). Highly efficient self-replicating RNA enzymes. *Chem. Biol.* 21, 238–245. doi: 10.1016/j.chembiol.2013.12.004

Salari, R., Wojtowicz, D., Zheng, J., Levens, D., Pilpel, Y., and Przytycka, T. M. (2012). Teasing apart translational and transcriptional components of stochastic variations in eukaryotic gene expression. *PLoS Comput. Biol.* 8:e1002644. doi: 10.1371/journal.pcbi.1002644

Saliba, A. E., Westermann, A. J., Gorski, S. A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* 42, 8845–8860. doi: 10.1093/nar/gku555

Schultz, D., Lu, M., Stavropoulos, T., Onuchic, J., and Ben-Jacob, E. (2013). Turning oscillations into opportunities: lessons from a bacterial decision gate. *Sci. Rep.* 3:1668. doi: 10.1038/srep01668

Schultz, D., Wolynes, P. G., Ben Jacob, E., and Onuchic, J. N. (2009). Deciding fate in adverse times: sporulation and competence in Bacillus subtilis. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21027–21034. doi: 10.1073/pnas.0912185106

Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B. J. (2016). Global Mapping of Human RNA-RNA Interactions. *Mol. Cell* 62, 618–626. doi: 10.1016/j.molcel.2016.04.030

Shechner, D. M., and Bartel, D. P. (2011). The structural basis of RNA-catalyzed RNA polymerization. *Nat. Struct. Mol. Biol.* 18, 1036–1042. doi: 10.1038/nsmb.2107

Silva-Rocha, R., and de Lorenzo, V. (2010). Noise and robustness in prokaryotic regulatory networks. *Annu. Rev. Microbiol.* 64, 257–275. doi: 10.1146/annurev.micro.091208.073229

Spitzer, J., Hafner, M., Landthaler, M., Ascano, M., Farazi, T., and Wardle, G. (2014). PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods Enzymol.* 539, 113–161.

Stewart, G. R., Robertson, B. D., and Young, D. B. (2003). Tuberculosis: a problem with persistence. *Nat. Rev. Microbiol.* 1, 97–105. doi: 10.1038/nrmicro749

Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M., and Mattick, J. S. (2010). Non-coding RNAs: regulators of disease. *J. Pathol.* 220, 126–139.

Taft, R. J., Pheasant, M., and Mattick, J. S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 29, 288–299. doi: 10.1002/bies.20544

Tanner, N. K. (1999). Ribozymes: the characteristics and properties of catalytic RNAs. *FEMS Microbiol. Rev.* 23, 257–275. doi: 10.1016/s0168-6445(99)00007-8

Vaidya, N., Manapat, M. L., Chen, I. A., Xulvi-Brunet, R., Hayden, E. J., and Lehman, N. (2012). Spontaneous network formation among cooperative RNA replicators. *Nature* 491, 72–77. doi: 10.1038/nature11549

Valenti, M. C., and Sun, J. (2002). The UMTS turbo code and an efficient decoder: implementation suitable for software-defined radios. *Int. J. Wireless Inform. Netw.* 8, 202–215.

Vasas, V., Fernando, C., Santos, M., Kauffman, S., and Szathmary, E. (2012). Evolution before genes. *Biol. Direct* 7:1. doi: 10.1186/1745-6150-7-1

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. doi: 10.1016/j.molcel.2011.08.018

Wang, X., Zhang, Y., Yu, S., and Guo, H. (2018). High speed error correction for continuous-variable quantum key distribution with multi-edge type LDPC code. *Sci. Rep.* 8:10543. doi: 10.1038/s41598-018-28703-4

Waters, L. S., and Storz, G. (2009). Regulatory RNAs in bacteria. *Cell* 136, 615–628. doi: 10.1016/j.cell.2009.01.043

Weidmann, C. A., Mustoe, A. M., and Weeks, K. M. (2016). Direct duplex detection: an emerging tool in the RNA structure analysis toolbox. *Trends Biochem. Sci.* 41, 734–736. doi: 10.1016/j.tibs.2016.07.001

Westhof, E. A. (2000). "RNA tertiary structure," in *Encyclopedia of Analytical Chemistry*, ed. R. A. Meyers (Chichester: John Wiley & Sons Ltd.,) 5222–5232.

Yanagawa, H. (1994). [Origins of life and RNA world]. *Tanpakushitsu Kakusan Koso* 39, 2465–2479.

Yang, Y., Hsu, P. J., Chen, Y. S., and Yang, Y. G. (2018). Dynamic transcriptomic m(6)A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Res.* 28, 616–624. doi: 10.1038/s41422-018-0040-8

Yarus, M. (2010). *Life From an RNA World: The Ancestor Within*. Cambridge, MA: Harvard University Press, 208.

Yeates, J. A. M., and Nehman, L. (2016). RNA networks at the origins of life. *Biochem. Soc.* 38, 1–5.

Yilmaz, S., and Singh, A. K. (2012). Single cell genome sequencing. *Curr. Opin. Biotechnol.* 23, 437–443.

Zhang, C., and Darnell, R. B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* 29, 607–614. doi: 10.1038/nbt.1873