

## Review

G-Quadruplexes in RNA Biology:  
Recent Advances and Future DirectionsLeïla Dumas,<sup>1,2,4</sup> Pauline Herviou,<sup>1,2,4</sup> Erik Dassi,<sup>3</sup> Anne Cammas,<sup>1,2</sup> and Stefania Millevoi<sup>1,2,\*</sup>

**RNA G-quadruplexes (RG4s) are four-stranded structures known to control gene expression mechanisms, from transcription to protein synthesis, and DNA-related processes. Their potential impact on RNA biology allows these structures to shape cellular processes relevant to disease development, making their targeting for therapeutic purposes an attractive option. We review here the current knowledge on RG4s, focusing on the latest breakthroughs supporting the notion of transient structures that fluctuate dynamically *in cellulo*, their interplay with RNA modifications, their role in cell compartmentalization, and their deregulation impacting the host immune response. We emphasize RG4-binding proteins as determinants of their transient conformation and effectors of their biological functions.**

### RNA G-Quadruplexes Research Keeps Pace with the Latest Advances in RNA Biology

The central dogma ‘DNA encodes RNA, RNA encodes protein’ states that RNA is an essential molecule in the flow of genetic information in cells. Beyond serving as a bridge between DNA and protein, RNA is critical in a cell to regulate this information, both qualitatively and quantitatively, thus constituting an essential control point for cellular processes. The corollary adverse effect of the central role of RNA is its susceptibility to being vulnerable, placing it as the cause or contributor to various dysfunctions leading to human pathologies [1]. Two major features allow the RNA molecule to fulfill its key function: the first is that the RNA is versatile and can adopt secondary structures defined as canonical when using A–T and C–G base pairing (according to Watson–Crick rules) or non-canonical if disobeying Watson–Crick canons; the second is that RNA is rarely naked, with protein factors [RNA-binding proteins (RBPs)] as the partners of excellence [2].

RNA G-quadruplexes (RG4s) are non-canonical structures that have gained growing importance in RNA biology and disease. These are stable four-stranded conformations formed by stacks of guanine tetrads (G-quartets) held together by Hoogsteen hydrogen bonds (Box 1). The first reason for this growing interest is the prevalence of these structures, amounting to several thousand in mammalian transcriptomes [3,4], which together with their enrichment in certain RNA regions (specifically, untranslated regions [5]) or classes (mRNAs, non-coding RNAs (ncRNAs) [6,7]), have contributed to the notion that RG4s are pervasive and could be biologically significant. A second reason lies in the tunable nature of RG4s whose folding can be controlled by cations and **small-molecule ligands** (see Glossary; Box 1). This not only has been instrumental for their identification and characterization as key regulators of cellular physiology in the context of health and disease, but has also opened up opportunities to use them as molecular switches that can be targeted for therapeutic purposes [8,9]. The consensus from studies at the transcriptome level or on individual mRNAs is that RG4s are important regulators of all post-transcriptional steps. There, they can operate in *cis* or in *trans* to control when, where, and how much protein needs to be synthesized and modify the coding capacity of the genome [9–11] (Figure 1). Misregulation of RG4-mediated RNA biology is associated with perturbations

### Highlights

RG4s are four-stranded structures that have gained growing importance in RNA biology, RNA-associated human diseases, and RNA-based therapeutics.

Besides telomere maintenance and gene expression mechanisms, recent advances have highlighted new functions of RG4s in the regulation of RNA expression in mitochondria, in phase separation mechanisms underscoring the formation of membrane-less organelles, and in chemical modifications within transcripts resulting in dynamic shaping of post-transcriptional gene expression pathways.

RG4-binding proteins are key players in regulating the dynamic equilibrium of their formation/dissolution in the cell, controlling their biological functions and driving their deregulation associated with human diseases.

RG4s may play a role in the strategies that pathogenic organisms or cancer cells use to evade the host’s immune responses.

<sup>1</sup>Cancer Research Centre of Toulouse, INSERM UMR 1037, 31037 Toulouse, France

<sup>2</sup>Université Toulouse III – Paul Sabatier, 31330 Toulouse, France

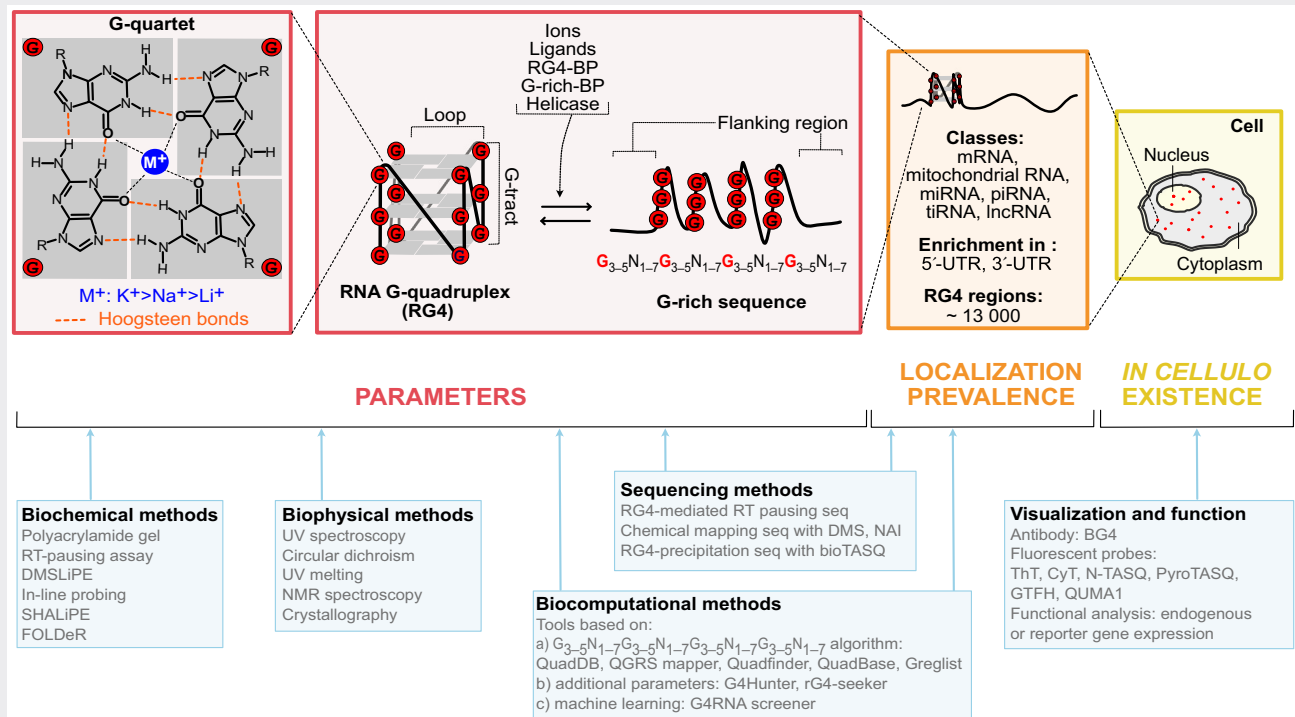
<sup>3</sup>Laboratory of RNA Regulatory Networks, Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, 38123 Trento, (TN), Italy

<sup>4</sup>These authors contributed equally to this work

\*Correspondence: stefania.millevoi@inserm.fr (S. Millevoi).

Box 1. RNA G-Quadruplexes (RG4s)

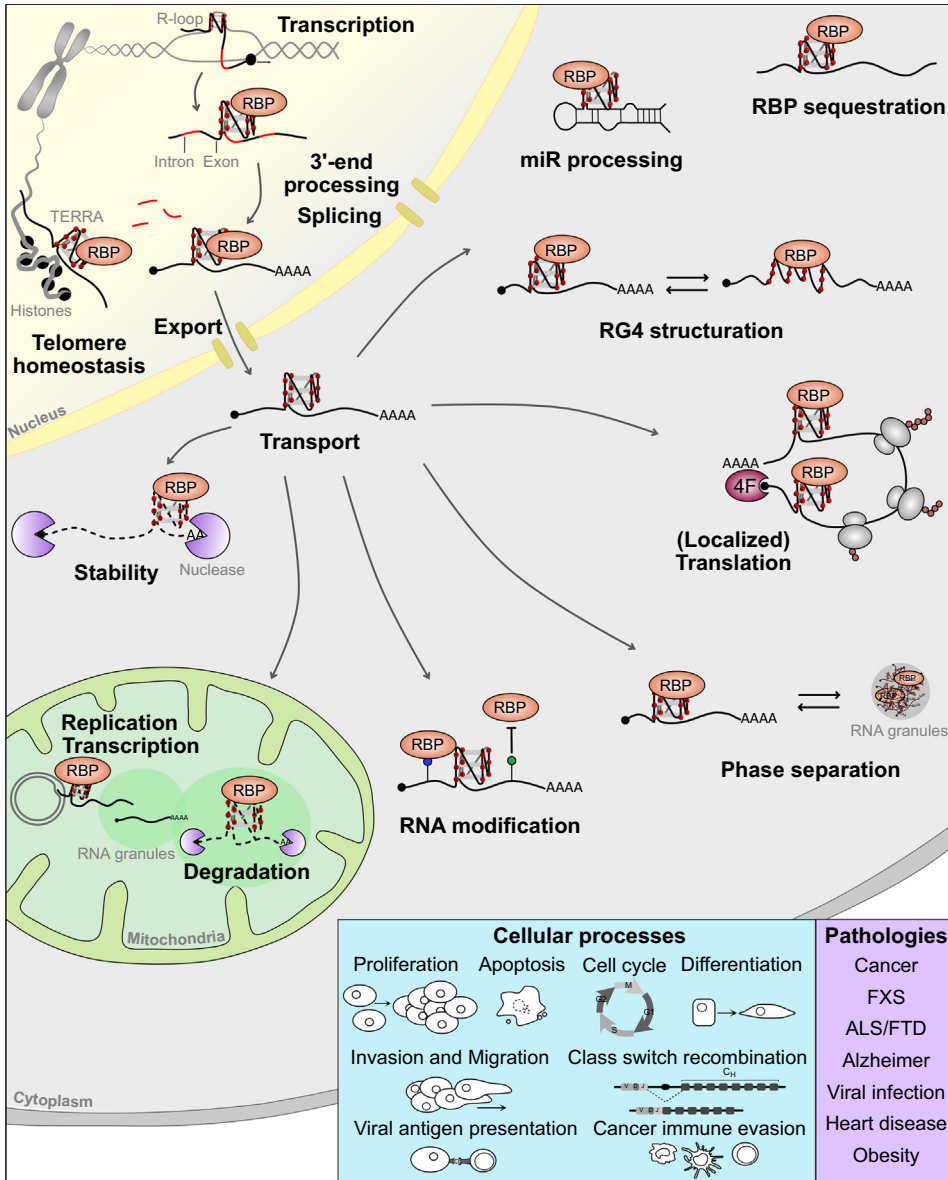
RG4s are thermodynamically stable secondary structures in which the guanines, bonded by Hoogsteen hydrogen base pairing, are organized into planar G-quartets stacked onto one another (Figure 1). The stability of the negatively charged core of the G-quartet composed of O6 atoms, the coordination between the G-quartets, and the base-stacking interactions are governed by intercalated monovalent cations [11]. Biophysical studies highlighted several parameters influencing RG4 conformations including the number of G-quartet stacks, the length/sequence composition of the loops, the occurrence of bulges, the availability/nature of the central ion, and the sequence in flanking regions. The prevalence of RG4s in the human transcriptome has been revealed by computational analysis based on the early algorithm searching for the G<sub>3-5</sub>N<sub>1-7</sub>G<sub>3-5</sub>N<sub>1-7</sub>G<sub>3-5</sub>N<sub>1-7</sub>G<sub>3-5</sub>N<sub>1-7</sub> consensus sequence [99,100], or on more recent tools that take into account structural variants (longer loops, bulges, mismatches, surrounding sequences), or on machine learning approaches [10,101]. Complementing these predictions, RG4 profiling methods based on RG4-mediated reverse transcriptase pausing or RG4-chemical mapping coupled with high-throughput sequencing have identified approximately 13 000 RG4 forming regions in 3 000 human mRNAs [3,4]. Interestingly, both computational and sequencing studies underscored not only the enrichment of RG4 structures in untranslated regions [4,5,102], but also their presence in ncRNAs [8,9,103,104]. This was key to drive subsequent functional analyses (e.g., [19,36,37]) that revealed the biological relevance of RG4s in the post-transcriptional control of gene expression impacting cellular processes. Recent findings suggesting that almost all RG4s exist in unfolded conformations *in cellulo* [3] created a debate regarding their existence in living cells. However, their visualization using antibody-based [26] or small molecule-based [28,34,35] characterization dampened the skepticism and favored a model of a dynamic RG4 folding controlled by a protein machinery whose characterization constitutes an emerging challenge in the field [18–25]. The RG4 folding equilibrium is viewed as an on-off switch tuned toward a folded state by stabilizing cations, RG4-binding proteins, or RG4-stabilizing ligands (e.g., Phen-DC3, BRACO-19, or cPDS) [9]. By contrast, destabilizing ions, G-rich-binding proteins, or RNA helicases favor the unfolded conformation of RG4s [19,54]. Dysregulation of this tightly controlled equilibrium in pathological situations arises from several mechanisms (e.g., RG4-binding protein sequestration, aberrant RG4-binding protein expression and localization, disabled RG4-protein interactions [10]), impacts on the expression of disease-associated mRNAs [10,105], and opens the opportunity to explore the potential for therapeutic targeting of RG4s [8,9].



Trends in Biochemical Sciences

Figure 1. RNA G-Quadruplex Characteristics. DMSLIPE, dimethyl sulfate with lithium ion-based primer extension; FOLDeR, footprinting of long 7-deaza-guanine-substituted RNAs; G-rich-BP, G-rich-binding protein; lncRNA, long non-coding RNA; piRNA, Piwi-interacting RNA; RG4-BP, RG4-binding protein; RT, reverse transcription; SHALIPE, selective 2'-OH acylation with lithium ion-based primer extension; tiRNA, tRNA-derived stress-induced RNA; UTR, untranslated region. References for the different methods could be found in [9,10,103,106].

of cellular processes underlying several diseases [10,12] (Figure 1). A third reason, central to this review, is that RG4 research has kept pace with major advances in RNA biology in recent years. These structures have been propelled to the forefront of emerging issues, often at the intersection of multiple disciplines, that could have a major impact on RNA research over the next decade. The renewed interest in identifying the RNA protein partners has synchronized RG4 research with



Glossary

**Epitranscriptomics:** study of all functionally relevant chemical modifications of the transcriptome (epitranscriptome).

**Gel-like condensates:** nondynamic condensates that stimulate RNA granule formation by increasing the local concentration of liquid phases.

**G-skewness:** degree of G/C asymmetry between the complementary strands.

**Immune checkpoint inhibitors:** drugs that block proteins called checkpoints, key regulators of the immune system that when stimulated can dampen the immune response to an immunologic stimulus.

**Immunological surveillance:** monitoring process of the immune system that detects and destroys virally infected and neoplastically transformed cells.

**miRNA:** short single-stranded RNA molecule that targets the RNA interference silencing complex (RISC) to specific mRNAs, resulting in decreased mRNA translation and/or in mRNA degradation.

**Mitochondrial degradosome:** functionally conserved complex composed of the ATP-dependent RNA and DNA helicase SUV3 and the PNPase ribonuclease.

**Mitochondrial RNA granules:** distinct RNP structures allowing spatiotemporal control of mitochondrial RNA processing and the biogenesis of mitochondrial ribosomes.

**N<sup>6</sup>-methyladenosine (m<sup>6</sup>A):** most prevalent internal RNA modification in mRNAs consisting in methylation of the adenosine base at the nitrogen-6 position.

**Paraspeckles:** RNA-protein nuclear bodies that regulate gene expression.

**P-bodies:** processing bodies that consist primarily of mRNA decay factors and translationally repressed mRNAs.

**Poly-ADP-ribosylation:** fully reversible post-translational modification with key roles in cellular physiology.

**qRRM:** quasi-RNA-recognition motif (RRM) that binds RNAs using loop residues rather than  $\beta$ -sheet residues (as the canonical RRM).

**RAN translation:** translation mechanism that occurs at pathological repeat expansions, in the absence of an initiating codon and in all possible reading frames, generating

**Figure 1. Role of RNA G-Quadruplexes (RG4s) in RNA Biology, Cellular Processes, and Human Diseases.** RG4s can regulate any gene expression step, from transcription to protein synthesis, including pre-mRNA maturation [including 3'-end processing which results in poly-A tail addition and removal of introns (in red) through splicing], export of mature transcripts into the cytoplasm, mRNA transport, localization stability, and translation. These regulatory mechanisms may rely on RG4s found in mRNAs but also in non-coding RNAs such as miRNAs (miR). Besides regulating transcription (both nuclear and mitochondrial), RG4s in telomeric repeat-containing RNA (TERRA) or at DNA/RNA hybrid structures (R-loops) can modulate other DNA-related processes such as telomere elongation, DNA replication, and recombination. RNA-binding proteins (RBPs) are the main RG4 partners assisting them in all their cellular functions as well as regulating the equilibrium between their structured and unstructured form. In turn, the activity of RBPs can be modulated by RG4s through sequestration mechanisms. Recent advances propose an important role for RG4s in the biology of mitochondrial RNA, in the formation of condensates mediated by phase separation mechanisms, and in RNA modifications key to gene expression pathways. RG4-mediated regulation, possibly in synergy or competition with RBPs, can affect several cellular processes (blue box) and thus impinge on pathological situations associated with human diseases (purple box). Abbreviations: ALS, amyotrophic lateral sclerosis; FXS, fragile X syndrome; FTD, frontotemporal dementia.

other new avenues in the field of RNA biology. The impetus driving these factors into the spotlight has surely been the development of new technologies that capture RBPs at the system-wide level [13–17] (Box 2).

In this review, we discuss the latest breakthroughs on RG4s, focusing on the RBPs regulating their structure and function in physiological and pathological processes, and emphasizing the most recent advances in our understanding of RNA regulation. In particular, we discuss the transient nature of RG4s and outline the mechanistic link between RG4s and RNA modifications, mitochondria, phase transitions, and immune evasion. This will be implemented through an analysis that intersects proteomics data identifying RBPs [13–17] with the most recent data sets on RG4-binding factors [18–25].

### RG4 Folding Dynamics

A substantial body of literature supports the view that RG4 formation plays various functions in both RNA and DNA biology. Their dysregulation due to mutations in the RG4-forming sequences or via the alteration of the association with *trans* factors contributes to human pathologies such as neurodegenerative diseases, cancer, or microbial infections (Figure 1). The pervasive nature of RG4s has been demonstrated by surveying the extent of their structuration both *in vitro* using transcriptome-wide reverse transcriptase stalling assay [3,4] and *in cellulo* using imaging studies with antibodies [26] or molecular probes [27–31] (Box 1). By contrast, but consistent with previous *in cellulo* RNA folding studies [32], RG4s in mRNAs appeared to be predominantly unfolded in steady-state RG4-seq experiments in human cells [3]. Overall, these studies suggest that RNA structures, whether they are RG4s or others, form secondary structures if given the opportunity.

#### Box 2. RNA-Binding Proteins (RBPs)

RBPs are a key class of factors in post-transcriptional regulation of gene expression. Consisting of more than 1300 human genes, they are involved in a host of processes ranging from alternative splicing and polyadenylation to the control of mRNA localization, stability, and translation [107]. Able to compete, cooperate, and autoregulate their own expression to control a wide set of targets, RBPs give rise to a complex network of interactions allowing the fine-tuning of gene expression [108]. The role played by RBPs in modulating such fundamental cellular processes reveals their importance as regulatory nexus to control not only specific gene expression but also intracellular organization [109]. This role has profound implications for the development of several pathologies. Alterations to expression of RBPs, binding ability, and interactions are indeed increasingly associated with the onset and progression of cancer [110] and neurological diseases [111]. From a translational point of view, it underlines the importance of these factors in the clinic, offering potential opportunities for developing therapeutic strategies [112].

#### A Modular Lifestyle

The structure of many RBPs is modular, building on a limited number of RNA-binding domains (RBDs) that can be further complemented by auxiliary domains. The most frequent RBD in higher vertebrates is the RNA-recognition motif (RRM). The RRM is a small domain of around 80 amino acids that can occur in one or multiple copies whose arrangement can lead to different RNA-binding specificities. Combinations of different RBDs are also observed, allowing to further increase the diversity and complexity of targets recognized by RBPs. More recently, a number of unknown RBDs have been discovered, suggesting multifunctional domains combining RNA binding with enzymatic activity or protein–protein interaction capabilities to be frequent, thus potentially further expanding the functions of RBPs [113].

#### New RBPs and New Functions

The known repertoire of RBPs may still be partial, particularly if considering those that could function through yet unknown RBDs. Therefore, several approaches have recently been developed to probe for proteins interacting with different RNA species at the transcriptome-wide level and *in cellulo* [13–17]. These methods allowed to identify several new classes of RBPs, including intrinsically disordered proteins and many metabolic enzymes whose role in RNA regulation is still to be understood. These unorthodox RBPs may thus ‘moonlight’ from their primary role, thus contributing to increase the complexity of this layer of gene expression regulation.

homopolymeric proteins with glutamine, serine, or alanine repeats.

**Repeat expansion RNAs:** repeat-containing transcripts whose accumulation leads to the formation of RNA foci, a common pathological feature of neurological and neuromuscular disorders.

**R-loops:** three-stranded DNA–RNA hybrids in which the nascent RNA strand binds to the DNA.

**RNA foci:** RNA–RBP complexes or aggregates formed by expanding RNA repeats that can become toxic to the cell, leading to neurodegenerative disorders.

**Small-molecule ligands:** natural or purely synthetic heteropolyaromatic planar chromophores which bind RG4 via  $\pi$ – $\pi$  stacking to a terminal G-quartet and is composed of one or more flexible substituents with a cationic charge that bind quadruplex grooves and loops. Based on the relative arrangement of the aromatic rings, three major families have been identified: (i) fused aromatic polycyclic systems (e.g., berberine, quarfloxin, Phen-DC3, RHPS4, BRACO-19), (ii) macrocycles (telomestatin, pyridostatin, carboxypyridostatin), and (iii) nonfused aromatic systems with flexible structural motifs.

**Stress granules:** cytoplasmic assemblies of untranslating messenger RNPs formed in stress conditions.



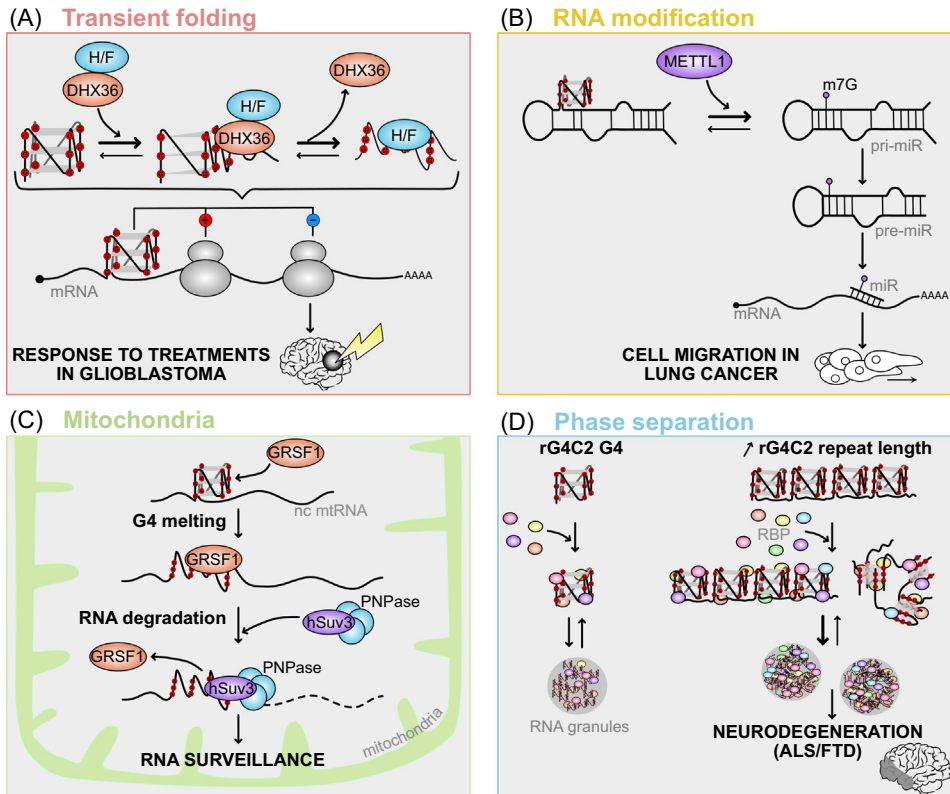
However, *in vivo*, since mRNAs must be unfolded to fulfill their messenger function, they are challenged by cellular components that convert them to either a linear form or an alternative Watson–Crick base pair structure. In view of their abundance and their ability to regulate all post-transcriptional gene expression steps, RBPs emerged as leading candidates to regulate RG4 folding *in cellulo*, paving the way for several studies that characterized the RG4-binding protein machinery using unbiased RNA affinity proteomic-wide approaches (termed RNA purification coupled with mass spectrometry, RP-MS) [18–25] and/or *in silico* analysis of the association between RG4-forming sequences and RBP-binding sites [5,33]. These studies, together with recent *in cellulo* RG4-capturing approaches [34] and live-cell imaging of RG4 folding and unfolding [35], provided evidence of transient RG4 folding while reinforcing the view that some RBPs play a role in shifting RG4s toward an unfolded state. These have been recently identified using the RP-MS approach with RNAs in which 7-deaza-guanines prevented RG4 structuration. This study highlighted a sequential mechanism in which the RNA helicase DHX36 first unwinds the RG4 followed by the binding of hnRNP H/F (heterogeneous nuclear ribonucleoprotein H/F) which maintain the RG4 unstructured, thereby regulating the translational efficiency of mRNAs playing a role in aggressive forms of brain tumors (glioblastoma; Figure 2A). This ‘bind–unfold–lock’ mechanism was also proposed for CNBP, resulting in increased protein synthesis [33]. RNA helicases proved to be important players in the RG4 dynamics linked to mRNA translation [36,37], where RBPs could also participate in recruiting them on specific RG4-containing mRNAs [38].

The transient nature of RG4s can also depend on the need to fold RG4s in particular cell types or stages. For instance, in B cells, the RNA helicase DDX1 unfolds RG4s in ncRNAs generated by splicing at immunoglobulin switch regions, resulting in the formation of **R-loops** that promote IgH class-switch recombination [39]. Another example is the antagonizing interplay between the SMaRT ncRNA and DHX36 that controls the translation of an RG4-containing mRNA in early phases of muscle differentiation [40]. Finally, the mode of action of helicases could also explain the transient and dynamic nature of RG4s, as demonstrated by the structural characterization of DHX36 bound to an RG4 showing repetitive cycles of ATP-independent unfolding and ATP-dependent RG4 refolding [41].

### RG4s and RNA Modifications

Emerging evidence supports a link between RG4s and **epitranscriptomics**. The epitranscriptome extends to all cellular RNA classes, encompassing more than 170 modifications, including internal modifications, and is dynamically regulated by the activity of enzymes. Chemical modifications within transcripts affect both the structure and the ability to interact with RBPs, resulting in dynamic shaping of post-transcriptional gene expression. While the fundamental question of the mechanisms that control selectivity in the deposition and erasing of modifications remains unanswered, the most likely clues concern the ability of effector proteins to recognize sequence or structural elements [42]. Given their conformational plasticity and their ability to selectively attract RBPs, RG4s could provide the modification enzymes with a means of specific recognition and regulatory selectivity. A possible RG4–epitranscriptome interplay is suggested by RP-MS studies, indicating that RG4-forming sequences may recruit factors that write, read, and erase these RNA marks (Figure 3 and Supplemental Table S1). These include several methyltransferases (such as TRMT112, RMNT, BUD23, or NSUN5) and RBPs that can be recruited or repelled by **N<sup>6</sup>-methyladenosine** (m<sup>6</sup>A) [43,44]. Importantly, some of these enzymes may display RG4 conformational-dependent affinity (as for the demethylase ALKBH5 or the methyltransferase METTL16 [19]).

While the role of RG4s in regulating the RNA accessibility of epitranscriptomic factors remains to be demonstrated, their importance in RG4-dependent gene expression modulation linked

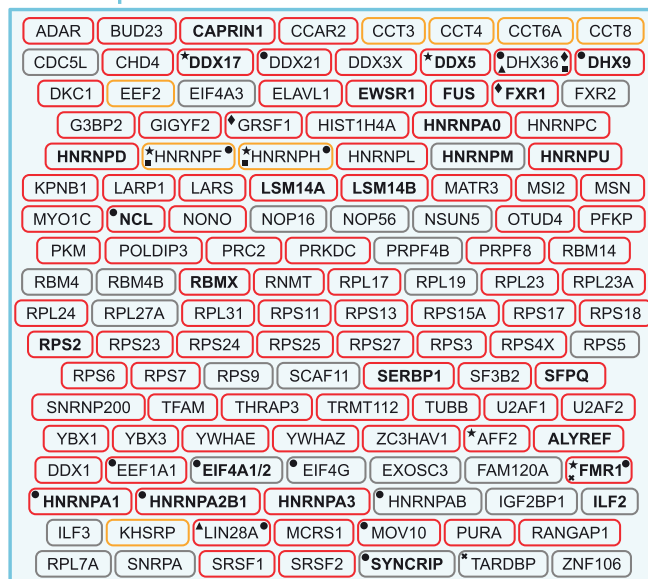


Trends in Biochemical Sciences

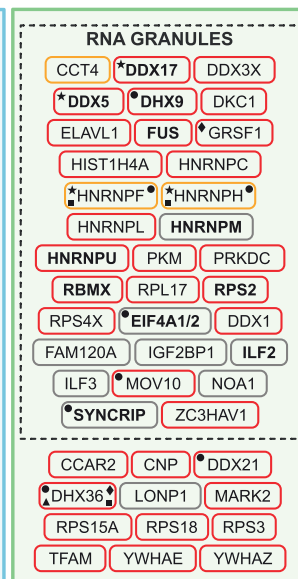
**Figure 2. Focus on Emerging Mechanisms Regulating RNA G-Quadruplex (RG4) Structuration and Function.** (A) RG4s are highly regulated *in cellulo* by proteins that specifically recognize them in their structured or unstructured form. An example of this dynamic equilibrium is that of hnRNP H/F (H/F) whose recruitment is mediated by the unwinding activity of the helicase DHX36, thus allowing its association with linearized RG4s and resulting in translational regulation of the mRNAs coding for factors involved in the response to glioblastoma treatment [19]. (B) RG4s can affect modifications of RNA bases that are essential to regulate the structure, expression, and function of mRNAs. This emerging mechanism has been proposed for METTL1, which catalyzes the methylation of guanosine on position N7 (m<sup>7</sup>G) within RG4s, thus promoting the formation of the double-stranded structures that are necessary for processing of pre-miRNAs (pre-miR) to miRs involved in the regulation of mRNA expression encoding cell migration protein relevant to lung cancer [45]. (C) In addition to transcription termination (Figure 1), RG4s may play a role in mitochondrial surveillance mechanisms. This involves GRSF1 RBP, which keeps unfolded the RG4s present in abundant non-coding RNAs antisense to functional mRNAs, thus promoting their degradation by the degradosome (including hSuv3 and PNPase) and enabling mitochondrial mRNA expression [54]. (D) RG4s together with RG4-binding proteins may promote the formation of RNA condensates through phase separation mechanisms resulting in membrane-less granules, as those formed by the RG4-forming rG4C2 repeats involved in neurodegenerative disorders, such as amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) [73,74]. mtRNA, mitochondrial RNA.

to RNA modifications has recently been demonstrated for m<sup>7</sup>G methylation by METTL1 within **miRNAs** [45]. The proposed model involved METTL1 binding to RG4-forming sequences, resulting in internal m<sup>7</sup>G deposition. Since this modification interferes with Hoogsteen but not Watson–Crick base pairing, METTL1-dependent modification impedes RG4 folding while promoting canonical base pairing required for proper processing of tumor suppressor miRNAs (Figure 2B). As METTL1 has been shown to be involved in lung cancer cell migration [45], RG4-forming sequences would confer METTL1 selectivity toward a specific pool of miRNAs regulating cellular processes linked to cancer development.

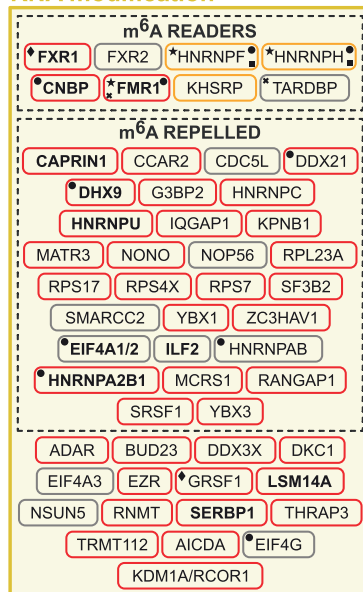
Phase separation



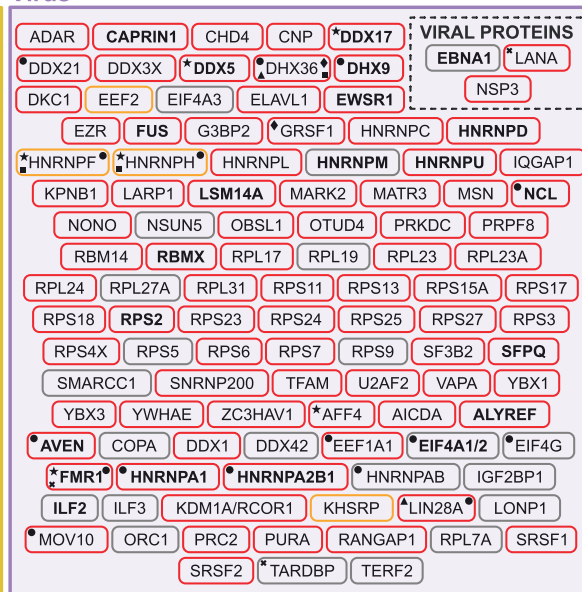
Mitochondria



RNA modification



Virus



LEGEND:

Conformation preference:

- Red box: Folded RG4
- Orange box: G-rich unfolded sequence
- White box: ND

RG4-dependent role in RNA biology:

- Star: Splicing
- Circle: Translation
- Triangle: ncRNA processing
- Square: 3'-end processing
- Diamond: mRNA stability
- Cross: mRNA localization

Protein domains:

- Bold: RGG domain

Trends in Biochemical Sciences

Figure 3. RNA G-Quadruplex (RG4) Protein Partners with Functions in Phase Separation, Mitochondria, RNA Modification, or Viral Pathogenesis. Protein factors found to be interacting with RG4s using RNA affinity chromatography followed by mass spectrometry (RP-MS), grouped according to their function in phase separation, mitochondria, RNA modification, or viral pathogenesis (Supplemental Table S1). RNA-binding proteins (RBPs) associated with these different functions were further clustered to highlight their identification in RNA granules (e.g., in mitochondria), their binding preference [i.e., whether they recognize (reader) or are repelled by the N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) RNA modification], or their cellular or viral origin. The color of each RG4-binding protein rectangle indicates whether it binds structured or linear RG4s. The symbols indicate the function of these factors in RNA biology. In bold are highlighted selected RG4-binding proteins containing an RGG domain.

These findings support the view that, in addition to hairpins, four-stranded conformations could cooperate with RNA modifications to guide RNA–protein interactions [46], thereby influencing post-transcriptional gene expression and controlling many cellular functions. In line with this, recent bioinformatics analyses of RG4s colocalizing with m<sup>6</sup>A suggest that RNA modifications and RG4s could cooperate to regulate pre-mRNA alternative splicing [47] or viral post-transcriptional gene expression [48]. Other indirect but promising pieces of evidence support that RG4s might be part of the structural ‘switch’ induced by the pseudouridylation of tRNA-derived fragments important for translation initiation impacting stem cell commitment during key developmental processes [49].

### RG4s and Mitochondria

In addition to the guanine number in G-tracts, G4 formation also depends on the **G-skewness**, which is a feature of the mitochondrial genome. Consistent with the *in silico* analysis predicting almost three times more G4s in mitochondrial DNA than in nuclear DNA [50], live-cell imaging microscopy revealed that the G4 small-molecule ligand RHPS4 localizes primarily into mitochondria and is retained by binding nucleic acids [51]. Despite this enrichment, knowledge on the role of RG4s in mitochondrial metabolism is still in an early phase. There is evidence that, in human mitochondria, transcription termination at an RG4-forming sequence near the replication origin drives mitochondrial DNA replication by generating an RNA primer. The underlying molecular mechanism involved an antitermination complex, in which the transcription factor TEFM prevents the formation of the RG4 required for transcription termination, thus promoting the synthesis of the replication primer [52].

Recently, TEFM was also found to regulate RNA processing and interact with RNA processing factors, including known RG4-binding proteins, such as GRSF1 [53]. While the role of TEFM in connecting transcription and RNA processing, possibly via RG4s, requires further molecular investigation, it appears that the G-rich binding activity of GRSF1 is key to control mitochondrial RNA levels. Specifically, GRSF1 melts the RG4s in abundant ncRNAs that are antisense to functional mRNAs, promoting their degradation by the **mitochondrial degradosome** [54] (Figure 2C). Similar to hnRNP H/F [55], GRSF1 uses **quasi-RNA-recognition motif (qRRMs)** to sequester G-tracts and maintain them in a single-stranded conformation [54]. The role of additional factors in this surveillance mechanism remains to be determined (e.g., LRPPRC or the transcription factor TFAM, both interacting with TEFM [53] and RG4s [19]). The growing inventory of RG4-binding proteins (Figure 3 and Supplemental Table S1) is expected to contribute to future investigations, further supporting the role of RG4s in mitochondrial gene expression regulation and coordination within **mitochondrial RNA granules** [56] or the synchronization of cytoplasmic-mitochondrial translation [57].

### RG4s and Phase Transitions

RG4s have been proposed to contribute to liquid–liquid phase separation (LLPS), in which metastable demixing of proteins and RNAs triggers the temporal and spatial organization of biochemical reactions [58]. LLPS underlines the formation of membrane-less organelles in the nucleus and the cytosol, such as cytoplasmic **P-bodies**, **stress granules**, **paraspeckles**, and **RNA foci** formed from **repeat expansion RNAs**. These ribonucleoprotein (RNP) granules, formed by transient multivalent protein–protein, RNA–RNA, and protein–RNA interactions, are involved in multiple aspects of RNA metabolism [59] and are linked to diseases, including viral infection, neurodevelopmental disorders, and cancer [60]. Specificity for the protein-driven LLPS can be achieved by proteins having repetitive modular domains and intrinsically disordered regions with weakly adhesive motifs. Post-translational modifications of RBPs (including **poly-ADP-ribosylation**, e.g., [61]), phosphorylation (e.g., [62]), and arginine methylation (e.g., [63])



or their protein partners (e.g., [64]) emerged as important regulators of LLPS. Intermolecular interactions between RNAs, promoted by high local RNA concentrations and driven by specific sequences and structures (canonical or non-canonical), are believed to trigger RNA condensation necessary for RNP formation. Other RNA properties potentially contributing to LLPS include the length and level of translation [65], the RNA-to-protein ratio [66], and the presence of modified bases (specifically m<sup>6</sup>A [67]).

Several RG4 features qualify them as candidate contributors to LLPS. First, at high concentrations, poly-guanosine can form gel-like structures in aqueous solutions [11]. These nondynamic **gel-like condensates** might stimulate LLPS by increasing the local concentration of liquid phases [68]. Second, RG4s formed *in cis* or *in trans* may promote the recruitment of protein factors leading to protein condensates potentially involved in LLPS-induced RNA granules. As an example, RG4s derived from tRNAs have been recently proposed to assemble in tetramolecular RG4 structures that could act as molecular scaffolds to promote high-local concentrations of RG4-binding proteins, thus triggering phase transitions that nucleate stress granule formations [69]. Similarly, the interaction between NONO and several RG4s identified within the ncRNA NEAT1 could contribute to seed paraspeckles formation [24]. Third, RG4-binding proteins often contain arginine-/glycine-rich regions (or RGG domain) [70] that are intrinsically disordered, thus presenting conformational flexibility mediating degenerate specificity in RNA binding [71]. This degeneracy may allow RBP oligomerization along the transcripts or multivalent interactions with multiple RNAs at the same time, both important to create RNP assemblies promoting LLPS. Moreover, RGG domains mediate protein–protein interactions, and tandem or triplet RGG gives rise to robust liquid–liquid demixing even in the absence of RNA [72].

While several features link the dynamics of RG4s to LLPS, understanding of this connection is limited. The requirement of RG4s for LLPS-mediated formation of RNA granules was demonstrated for the transcripts generated by the GGGGCC hexanucleotide repeats (rG4C2) in C9orf72 gene, the most common mutation associated with amyotrophic lateral sclerosis and frontotemporal dementia (C9-ALS/FTD) [73]. Similar to the relationship between the repeat length of C9orf72 and the pathogenesis of C9-ALS/FTD, rG4C2 repeat length correlates with the degree of protein condensation [73,74], suggesting a link between RNA phase separation and the number of G-quartets (Figure 2D). Although rG4C2 foci formation may lead to phase separation without the requirement of *trans*-acting factors, the rG4C2 repeat number required for RNA condensation is higher *in cellulo* than *in vitro* [74], supporting the notion of a protein machinery unfolding RG4s in living cells [3]. hnRNP H is a candidate for modulating RNA foci formation since it binds rG4C2 repeats [75] and can maintain RG4s in an unfolded conformation [19,75]. RG4s may also indirectly influence C9-ALS/FTD-linked LLPS by modulating **repeat-associated non-AUG (RAN) translation** occurring at C9orf72 repeats, which generates toxic arginine-rich dipeptides that in turn can promote LLPS [76]. In the case of the RG4 formed at the SHORT ROOT (SHR) RNA, which is important in plant development, phase separation is enhanced by a higher number of G-quartets and longer loops, strengthening the notion that both RG4 stability and their multivalent interactions are critical for RG4-triggered phase separation [77].

Although RG4s appear to confer structural specificity when initiating phase separation [73], the role of *trans*-acting factors remains unclear. Several RG4-binding proteins have been found in the LLPS protein collection [78] (Figure 3 and Supplemental Table S1), paving the way for future studies. As several of these factors contain an RGG domain (Figure 3) and RGG methylation is involved in both RG4-dependent regulation [38] and LLPS [63,76], it would be interesting to investigate whether and how arginine methylation is involved in RG4-dependent LLPS

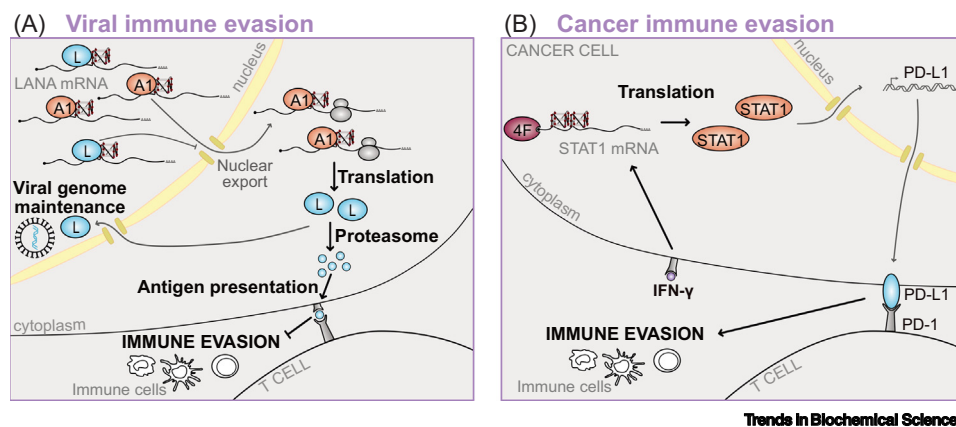
mechanisms of RNA granule formation. Other challenging questions regarding the RG4–LLPS interplay raised by recent findings in both RG4 field (this review) and RNA-mediated regulation of LLPS [79] are whether (i) this interplay is modified by RNA modifications or is involved in translational regulation [80], (ii) RG4 structuration is similar in the condensed phase compared with the dilute one, as suggested in [81], and (iii) given the role of RG4s in RNA transport [82], whether this interplay impacts RG4-containing mRNAs or ncRNAs localization or cell compartmentalization.

### RG4 Misregulation in Disease: Focus on Immune Evasion

The importance of RG4s in disease was initially suggested by the observation that RG4 folding modulates the expression of transcripts having a central role in human pathologies, such as those encoded by the oncogene NRAS, the tumor suppressor TP53, and the Epstein-Barr virus (EBV) protein EBNA1 [10]. Several models explaining how RG4s affects diseases have been proposed. This includes *cis*-mechanisms where RG4s resulting from repeat expansion mutations in the untranslated regions modify mRNA translation (by blocking it or inducing alternative translation initiation), or *trans*-mechanisms that either use RG4s to sequester gene expression regulatory protein factors or result from the mutation and altered expression of RG4-binding proteins [10]. Among the cellular processes hijacked by altered RG4 regulation, there are proliferation and survival, cell cycle, apoptosis, differentiation, invasion and migration, antigen presentation, and immunoglobulin class switch recombination (Figure 1). In addition to cancer, microbial pathogenesis, and neurodegeneration, recent studies point to a possible implication of RG4 dysregulation in congenital heart disease [83] and obesity [84].

Recent evidence suggests that RG4s may play a role in the strategies that pathogenic organisms or cancer cells use to evade the host immune response, termed immune evasion. For viruses, RG4s repress the expression of viral proteins, some of which play immunomodulatory roles by restricting antigen presentation to cytotoxic T cells, allowing the virus to persist in infected cells without being recognized by the host immune system. Two different mechanisms have been reported for the viral proteins EBNA1 and LANA, which are functional homologs in EBV and Kaposi's sarcoma-associated herpesvirus, respectively. The first exploits the ability of RG4s in both EBNA1 and LANA mRNAs coding sequences to inhibit the translation of these proteins. The structure–immune function relationship was demonstrated by modulating RG4s folding, which then resulted in the altered expression of the viral proteins and, in turn, in modified antigen presentation [85–87]. Although the molecular mechanism awaits further investigation, host cell protein factors (hnRNP A1 and nucleolin for LANA and EBNA1 mRNAs, respectively) take part in this regulation by interacting with RG4 structures [86,87]. This is consistent with the observation that terms associated with viral infection are over-represented in cellular RG4–protein interaction data sets [5,18,19,87]. Additional work supports the notion of nucleolin as a host factor for antiviral immunity and suggests that nucleolin expression is regulated by viral infection [88]. The proposed model involves the induction of nucleolin by hepatitis C virus infection, which binds viral core RG4s, resulting in the suppression of hepatitis C virus replication.

The second mechanism implies the direct binding of viral proteins to RG4s. This was anticipated for EBNA1 [89] but has only recently been demonstrated in detail for LANA, which uses this RG4-binding activity to bind its mRNA. LANA thus self-regulates its expression by mRNA sequestration in the nucleus and competing with hnRNP A1 for association with RG4s at the LANA mRNA [87] (Figure 4A). This mechanism was also reported for the nsp3 of SARS coronavirus—severe acute respiratory syndrome coronavirus (SARS-CoV) [90] and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [91]. The interaction between RG4s and viral proteins might also contribute to LLPS-dependent viral RNA packaging and host proteins co-opting [92–94]. Considering that SARS-CoV-2 coronavirus is the key culprit responsible for the coronavirus



**Figure 4. Focus on RNA G-Quadruplex (RG4) Functions and Immune Response.** (A) The viral protein LANA (L) can self-regulate its protein expression by binding to an RG4 within its own mRNA, thereby inhibiting its export to the cytoplasm. The RBP hnRNP A1 (A1) competing with LANA (L) for RG4 binding promotes LANA mRNA export and translation. This results in an increase in LANA-derived antigens and stimulation of the antiviral cellular response [87]. (B) The eIF4F (4F) initiation complex regulates the RG4-dependent translation of the mRNA encoding STAT1, a transcription factor known to regulate the interferon- $\gamma$  (IFN- $\gamma$ )-induced expression of PD-L1. PD-L1 binding to PD-1 triggers the inhibition of cytotoxic T cells proliferation and activity, resulting in a dampened immune response against tumor cells [97].

disease 2019 (COVID-19) pandemic and that the expression of host genes in SARS-CoV-2-infected cells is extremely inhibited [95], investigating whether, when, and how SARS-RG4 interactions impair the immune response of host cells is of the outmost importance.

The potential for RG4s to act as immunomodulators has also been proposed for cancer cells via the regulation of signaling transducers modulating the expression of **immune checkpoint inhibitors**. Specifically, both G4-stabilizing ligands and natural compounds inhibiting eIF4A, an initiation factor with potential RG4-resolving activity [96], modulate mRNA translation of STAT1, an upstream transcriptional activator of the negative immune checkpoint PD-L1 [97] (Figure 4B). It remains to be fully determined whether the STAT1–PD-L1 axis, impacting tumor immune escape, relies on RG4-dependent regulation.

### Concluding Remarks

The renewed interest in RBP identification and the latest advances in RNA biology regulation, fostered by cutting-edge approaches at the transcriptome and proteome scales, have shaped the future directions of research in the field of RG4s over the coming decades. There is recent evidence that these structures, in synergy with RBPs, can be dynamically regulated *in cellulo* and play a key role in regulating chemical RNA composition or cell compartmentalization (both membrane-bound and membrane-less). Nevertheless, many questions in fundamental and translational biology remain unanswered, and new questions have arisen (see Outstanding Questions). Future research is expected to continue focusing on these research lines to improve our understanding of the underlying molecular mechanisms and uncover the extent of these regulations, their biological consequences, and their connections to diseases. The molecular partnership between heme and RG4s also points to a new research direction, aimed at exploring the link between RG4s and metabolism [98]. Another major area of interest for RG4s is their potential impact in **immunological surveillance**, which is the target of cutting-edge therapeutic strategies against cancer and viral infections. The identification of several cellular and viral factors (including those of SARS coronavirus) that bind RG4s and play a role in viral infections (Figure 3) opens up new perspectives to further investigate this connection and define whether targeting RG4s or their protein partners may offer an opportunity to counteract viral replication.

### Outstanding Questions

Interplay between RG4s and RBPs: what are the upstream regulatory signaling pathways? How do post-translational modifications of RBPs and/or RG4 epitranscriptomic modifications impact on RG4-associated biology and disease? Do specific interactions coordinate the expression of functionally correlated genes?

RG4s and cell compartmentalization: in addition to RNA granules, do RG4s contribute to the formation of other cell compartments? What are the molecular determinants governing the formation of specific condensates?

RG4 binders: What else besides cations, RBPs, and synthetic small molecules? Are RG4s able to sequester molecules from metabolic pathways as shown for heme? What are the consequences for the cell?

RG4s and disease: How does the RG4 structural equilibrium shift and what are the underlying molecular mechanisms? Aside from translational regulation, which other processes are involved in modulating the immune response? Does RG4 targeting represent a real therapeutic opportunity?

## Acknowledgments

L.D. was supported by MENRT and P.H. by ANR. Research on RNA G-quadruplexes was funded by institutional grants (from INSERM, Université Toulouse III - Paul Sabatier, CNRS) and by funding from LNCC (Ligue Nationale Contre le Cancer), ARC (Association pour la Recherche contre le Cancer), Emergence Cancéropole GSO, and ANR (ANR-17-CE12-0017-01). We apologize to all colleagues whose important findings could not be cited owing to space limitations.

## Supplemental Information

Supplemental information associated with this article can be found online at <https://doi.org/10.1016/j.tibs.2020.11.001>

## References

- Rajesh, P. (2020) *RNA-Based Regulation in Human Health and Disease*, Academic Press
- Singh, G. *et al.* (2015) The clothes make the mRNA: past and present trends in mRNP fashion. *Annu. Rev. Biochem.* 84, 325–354
- Guo, J.U. and Bartel, D.P. (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* 353, aaf5371
- Kwok, C.K. *et al.* (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* 13, 841–844
- Lee, D.S.M. *et al.* (2020) Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. *Nat. Commun.* 11, 527
- Kwok, C.K. *et al.* (2016) Structural analysis using SHALiPE to reveal RNA G-quadruplex formation in human precursor microRNA. *Angew Chem Int Ed Engl* 55, 8958–8961
- Chan, K.L. *et al.* (2018) Structural analysis reveals the formation and role of RNA G-quadruplex structures in human mature microRNAs. *Chem. Commun. (Camb.)* 54, 10878–10881
- Kharel, P. *et al.* (2020) The role of RNA G-quadruplexes in human diseases and therapeutic strategies. *Wiley Interdiscip. Rev. RNA* 11, e1568
- Varshney, D. *et al.* (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* 21, 459–474
- Cammás, A. and Millevoi, S. (2017) RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* 45, 1584–1595
- Fay, M.M. *et al.* (2017) RNA G-quadruplexes in biology: principles and molecular mechanisms. *J. Mol. Biol.* 429, 2127–2147
- Saranathan, N. and Vivekanandan, P. (2019) G-quadruplexes: more than just a kink in microbial genomes. *Trends Microbiol.* 27, 148–163
- Trendel, J. *et al.* (2019) The human RNA-binding proteome and its dynamics during translational arrest. *Cell* 176, 391–403 e19
- Queiroz, R.M.L. *et al.* (2019) Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat. Biotechnol.* 37, 169–178
- Bao, X. *et al.* (2018) Capturing the interactome of newly transcribed RNA. *Nat. Methods* 15, 213–220
- Urdaneta, E.C. *et al.* (2019) Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. *Nat. Commun.* 10, 990
- Perez-Perri, J.I. *et al.* (2018) Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat. Commun.* 9, 4408
- Herdý, B. *et al.* (2018) Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res.* 46, 11592–11604
- Herviu, P. *et al.* (2020) hnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma. *Nat. Commun.* 11, 2661
- Haeuſtler, A.R. *et al.* (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507, 195–200
- McRae, E.K.S. *et al.* (2017) Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Res.* 45, 6656–6668
- Serikawa, T. *et al.* (2018) Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. *Biochimie* 144, 169–184
- von Hacht, A. *et al.* (2014) Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res.* 42, 6630–6644
- Simko, E.A.J. *et al.* (2020) G-quadruplexes offer a conserved structural motif for NONO recruitment to NEAT1 architectural lncRNA. *Nucleic Acids Res.* 48, 7421–7438
- Bolduc, F. *et al.* (2020) The small nuclear ribonucleoprotein polypeptide A (SNRPA) binds to the G-quadruplex of the BAG-1 5'UTR. *Biochimie* 176, 122–127
- Biffi, G. *et al.* (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* 6, 75–80
- Chen, S.B. *et al.* (2016) Visualization of NRAS RNA G-quadruplex structures in cells with an engineered fluorogenic hybridization probe. *J. Am. Chem. Soc.* 138, 10382–10385
- Di Antonio, M. *et al.* (2020) Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.* 12, 832–837
- Laguerre, A. *et al.* (2015) Visualization of RNA-quadruplexes in live cells. *J. Am. Chem. Soc.* 137, 8521–8525
- Lyu, K. *et al.* (2019) Structural analysis and cellular visualization of APP RNA G-quadruplex. *Chem. Sci.* 10, 11095–11102
- Xu, S. *et al.* (2015) Directly lighting up RNA G-quadruplexes from test tubes to living human cells. *Nucleic Acids Res.* 43, 9575–9586
- Ramos, S.B. and Laederach, A. (2014) Molecular biology: a second layer of information in RNA. *Nature* 505, 621–622
- Benhalevy, D. *et al.* (2017) The human CCHC-type zinc finger nucleic acid-binding protein binds G-rich elements in target mRNA coding sequences and promotes translation. *Cell Rep.* 18, 2979–2990
- Yang, S.Y. *et al.* (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* 9, 4730
- Chen, X.C. *et al.* (2018) Tracking the dynamic folding and unfolding of RNA G-quadruplexes in live cells. *Angew Chem. Int. Ed. Engl.* 57, 4702–4706
- Murat, P. *et al.* (2018) RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.* 19, 229
- Sauer, M. *et al.* (2019) DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nat. Commun.* 10, 2421
- Thandapani, P. *et al.* (2015) Aven recognition of RNA G-quadruplexes regulates translation of the mixed lineage leukemia protooncogenes. *Elife* 4, e06234
- Ribeiro de Almeida, C. *et al.* (2018) RNA helicase DDX1 converts RNA G-quadruplex structures into R-loops to promote IgH class switch recombination. *Mol. Cell* 70, 650–662.e8
- Martone, J. *et al.* (2020) SMArT lncRNA controls translation of a G-quadruplex-containing mRNA antagonizing the DHX36 helicase. *EMBO Rep.* e49942
- Tippiana, R. *et al.* (2019) RNA G-quadruplex is resolved by repetitive and ATP-dependent mechanism of DHX36. *Nat. Commun.* 10, 1855
- Roundtree, I.A. *et al.* (2017) Dynamic RNA modifications in gene expression regulation. *Cell* 169, 1187–1200
- Edupuganti, R.R. *et al.* (2017) N(6)-methyladenosine (m<sup>6</sup>A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* 24, 870–878
- Arguello, A.E. *et al.* (2017) RNA chemical proteomics reveals the N<sup>6</sup>-methyladenosine (m<sup>6</sup>A)-regulated protein-RNA interactome. *J. Am. Chem. Soc.* 139, 17249–17252

45. Pandolfini, L. *et al.* (2019) METTL1 promotes let-7 microRNA processing via m7G methylation. *Mol. Cell* 74, 1278–1290.e9
46. Lewis, C.J. *et al.* (2017) RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* 18, 202–210
47. Jara-Espejo, M. *et al.* (2020) Potential G-quadruplex forming sequences and N<sup>6</sup>-methyladenosine colocalize at human pre-mRNA intron splice sites. *ACS Chem. Biol.* 15, 1292–1300
48. Fleming, A.M. *et al.* (2019) Colocalization of m<sup>6</sup>A and G-quadruplex-forming sequences in viral RNA (HIV, Zika, hepatitis B, and SV40) suggests topological control of adenosine N<sup>6</sup>-methylation. *ACS Cent. Sci.* 5, 218–228
49. Guzzi, N. *et al.* (2018) Pseudouridylation of tRNA-derived fragments steers translational control in stem cells. *Cell* 173, 1204–1216.e26
50. Bedrat, A. *et al.* (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* 44, 1746–1759
51. Falabella, M. *et al.* (2019) G-quadruplex dynamics contribute to regulation of mitochondrial gene expression. *Sci. Rep.* 9, 5605
52. Hillen, H.S. *et al.* (2017) Mechanism of transcription anti-termination in human mitochondria. *Cell* 171, 1082–1093.e13
53. Jiang, S. *et al.* (2019) TEFM regulates both transcription elongation and RNA processing in mitochondria. *EMBO Rep.* 20, e48101
54. Pietras, Z. *et al.* (2018) Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria. *Nat. Commun.* 9, 2558
55. Dominguez, C. *et al.* (2010) Structural basis of G-tract recognition and engaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.* 17, 853–861
56. Jourdain, A.A. *et al.* (2013) GRSF1 regulates RNA processing in mitochondrial RNA granules. *Cell Metab.* 17, 399–410
57. Pearce, S.F. *et al.* (2017) Regulation of mammalian mitochondrial gene expression: recent advances. *Trends Biochem. Sci.* 42, 625–639
58. Boeynaems, S. *et al.* (2018) Protein phase separation: a new phase in cell biology. *Trends Cell Biol.* 28, 420–435
59. Tauber, D. *et al.* (2020) Mechanisms and regulation of RNA condensation in RNP granule formation. *Trends Biochem. Sci.* 45, 764–778
60. Alberti, S. and Dormann, D. (2019) Liquid-liquid phase separation in disease. *Annu. Rev. Genet.* 53, 171–194
61. Leung, A.K.L. (2020) Poly(ADP-ribose): a dynamic trigger for biomolecular condensate formation. *Trends Cell Biol.* 30, 370–383
62. Kim, T.H. *et al.* (2019) Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. *Science* 365, 825–829
63. Hofweber, M. *et al.* (2018) Phase separation of FUS is suppressed by its nuclear import receptor and arginine methylation. *Cell* 173, 706–719.e13
64. Liu, Z. *et al.* (2020) Hsp27 chaperones FUS phase separation under the modulation of stress-induced phosphorylation. *Nat. Struct. Mol. Biol.* 27, 363–372
65. Khong, A. *et al.* (2017) The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Mol. Cell* 68, 808–820.e5
66. Maharana, S. *et al.* (2018) RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* 360, 918–921
67. Ries, R.J. *et al.* (2019) m<sup>6</sup>A enhances the phase separation potential of mRNA. *Nature* 571, 424–428
68. Putnam, A. *et al.* (2019) A gel phase promotes condensation of liquid P granules in *Caenorhabditis elegans* embryos. *Nat. Struct. Mol. Biol.* 26, 220–226
69. Lyons, S.M. *et al.* (2017) Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. *Nat. Commun.* 8, 1127
70. Huang, Z.L. *et al.* (2018) Identification of G-quadruplex-binding protein from the exploration of RGG motif/G-quadruplex interactions. *J. Am. Chem. Soc.* 140, 17945–17955
71. Ozdilek, B.A. *et al.* (2017) Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res.* 45, 7984–7996
72. Schuster, B.S. *et al.* (2018) Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* 9, 2985
73. Fay, M.M. *et al.* (2017) ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions *in vitro* and in cells. *Cell Rep.* 21, 3573–3584
74. Jain, A. and Vale, R.D. (2017) RNA phase transitions in repeat expansion disorders. *Nature* 546, 243–247
75. Conlon, E.G. *et al.* (2016) The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS patient brains. *Elife* 5, e17820
76. Boeynaems, S. *et al.* (2017) Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Mol. Cell* 65, 1044–1055.e5
77. Zhang, Y. *et al.* (2019) G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Res.* 47, 11746–11754
78. You, K. *et al.* (2020) PhaSepDB: a database of liquid-liquid phase separation related proteins. *Nucleic Acids Res.* 48, D354–D359
79. Roden, C. and Gladfelter, A.S. (2020) RNA contributions to the form and function of biomolecular condensates. *Nat. Rev. Mol. Cell Biol.* Published online July 6, 2020. <https://doi.org/10.1038/s41580-020-0264-6>
80. Tsang, B. *et al.* (2019) Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. *Proc. Natl. Acad. Sci. U. S. A.* 116, 4218–4227
81. Langdon, E.M. *et al.* (2018) mRNA structure determines specificity of a polyQ-driven phase separation. *Science* 360, 922–927
82. Goering, R. *et al.* (2020) FMRP promotes RNA localization to neuronal projections through interactions between its RGG domain and G-quadruplex RNA sequences. *Elife* 9, e52621
83. Nie, J. *et al.* (2015) Post-transcriptional regulation of Nkx2-5 by RHAU in heart development. *Cell Rep.* 13, 723–732
84. Liu, G. *et al.* (2020) RNA G-quadruplex regulates microRNA-26a biogenesis and function. *J. Hepatol.* 73, 371–382
85. Murat, P. *et al.* (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.* 10, 358–364
86. Lista, M.J. *et al.* (2017) Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA. *Nat. Commun.* 8, 16043
87. Dabral, P. *et al.* (2020) LANA and hnRNP A1 regulate the translation of LANA mRNA through G-quadruplexes. *J. Virol.* 94, e01508-19
88. Bian, W.X. *et al.* (2019) Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication. *Nucleic Acids Res.* 47, 56–68
89. Norseen, J. *et al.* (2009) Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J. Virol.* 83, 10336–10346
90. Tan, J. *et al.* (2009) The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog.* 5, e1000428
91. Zhang, R. *et al.* (2020) Whole genome identification of potential G-quadruplexes and analysis of the G-quadruplex binding domain for SARS-CoV-2. *bioRxiv* Posted online June 5, 2020. <https://doi.org/10.1101/2020.06.05.135749>
92. Cubuk, J. *et al.* (2020) The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *bioRxiv* Posted online June 18, 2020. <https://doi.org/10.1101/2020.06.17.158121>
93. Lu, S. *et al.* (2020) The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *bioRxiv* Posted online July 31, 2020. <https://doi.org/10.1101/2020.07.30.228023>
94. Perdikari, T.M. *et al.* (2020) SARS-CoV-2 nucleocapsid protein undergoes liquid-liquid phase separation stimulated by RNA and partitions into phases of human ribonucleoproteins. *bioRxiv* June 10, 2020. <https://doi.org/10.1101/2020.06.09.141101>
95. Kim, D. *et al.* (2020) The architecture of SARS-CoV-2 transcriptome. *Cell* 181, 914–921.e10



96. Wolfe, A.L. *et al.* (2014) RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* 513, 65–70
97. Cerezo, M. *et al.* (2018) Translational control of tumor immune escape via the eIF4F-STAT1-PD-L1 axis in melanoma. *Nat. Med.* 24, 1877–1886
98. Gray, L.T. *et al.* (2019) G-quadruplexes sequester free heme in living cells. *Cell Chem. Biol.* 26, 1681–1691.e5
99. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* 33, 2908–2916
100. Todd, A.K. *et al.* (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* 33, 2901–2907
101. Garant, J.M. *et al.* (2017) Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics* 33, 3532–3537
102. Huppert, J.L. *et al.* (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.* 36, 6260–6268
103. Kwok, C.K. *et al.* (2018) Detecting RNA G-quadruplexes (rG4s) in the transcriptome. *Cold Spring Harb. Perspect. Biol.* 10, a032284
104. Perreault, J.-P. *et al.* (2020) Where are G-quadruplexes located in the human transcriptome? *NAR Genomics Bioinforma.* 2, lqaa035
105. Maizels, N. (2015) G4-associated human diseases. *EMBO Rep.* 16, 910–922
106. Kwok, C.K. and Merrick, C.J. (2017) G-quadruplexes: prediction, characterization, and biological application. *Trends Biotechnol.* 35, 997–1013
107. Gerstberger, S. *et al.* (2014) A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845
108. Quattrone, A. and Dassi, E. (2019) The architecture of the human RNA-binding protein regulatory network. *iScience* 21, 706–719
109. Sanders, D.W. *et al.* (2020) Competing protein-RNA interaction networks control multiphase intracellular organization. *Cell* 181, 306–324.e28
110. Pereira, B. *et al.* (2017) RNA-binding proteins in cancer: old players and new actors. *Trends Cancer* 3, 506–528
111. Nussbacher, J.K. *et al.* (2019) Disruption of RNA metabolism in neurological diseases and emerging therapeutic interventions. *Neuron* 102, 294–320
112. (2019) Delivering the promise of RNA therapeutics. *Nat. Med.* 25, 1321.
113. Castello, A. *et al.* (2016) Comprehensive identification of RNA-binding domains in human cells. *Mol. Cell* 63, 696–710