SURVEY AND SUMMARY

RNA G-quadruplexes (rG4s): genomics and biological functions

Kaixin Lyu [1,†], **Eugene Yui-Ching Chow** [2,†], **Xi Mou** [1,†], **Ting-Fung Chan** [2,*] **and Chun Kit Kwok** [1,3,*]

[1]Department of Chemistry and State Key Laboratory of Marine Pollution, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China, [2]School of Life Sciences, and State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China and [3]Shenzhen Research Institute of City University of Hong Kong, Shenzhen, China

ABSTRACT

G-quadruplexes (G4s) are non-classical DNA or RNA secondary structures that have been first observed decades ago. Over the years, these four-stranded structural motifs have been demonstrated to have significant regulatory roles in diverse biological processes, but challenges remain in detecting them globally and reliably. Compared to DNA G4s (dG4s), the study of RNA G4s (rG4s) has received less attention until recently. In this review, we will summarize the innovative high-throughput methods recently developed to detect rG4s on a transcriptome-wide scale, highlight the many novel and important functions of rG4 being discovered *in vivo* across the tree of life, and discuss the key biological questions to be addressed in the near future.

INTRODUCTION

The origin of research into the G-quadruplex (G4) can be traced back more than a century to 1910, when Bang reported the formation of a gel structure from an RNA monomer, guanosine monophosphate, at high concentration (1) (Figure 1A). Around 50 years later, this peculiar structure was solved by Gellert and colleagues using X-ray diffraction, and the G-quartet motif was proposed (2) (Figure 1B). In a G-quartet, each guanine base serves as both a hydrogen bond donor and a hydrogen bond acceptor in two base pairs to form a planar structure, and two or more G-quartets can stack on top of each other to form a G4 (Figure 1B). The first demonstration of the formation and bio-

logical significance of the DNA G-quadruplex (dG4) structure, at the telomere region and immunoglobulin switch regions, was reported by Sen and Gilbert in 1988 (Figure 1A) (3). Since then, many initial structural and biochemical studies have shown that G-rich DNA or RNA sequences with G4-forming potential can fold into dG4s and RNA G-quadruplex (rG4) structures (4–9), and play vital roles in various cellular functions (10–17). The potential to form G4s can be predicted from the primary sequences. Canonical putative G-quadruplex sequences (PQSs) contain four runs of three consecutive guanines separated by three loops, with the loop length ranging from one to seven nucleotides ($GGGN_{1-7}GGGN_{1-7}GGGN_{1-7}GGG$), or G_3N_{1-7} for short (Figure 1B) (18). The formation of G4s is stabilized by cations such as K^+, which can fit into the space between two quartets (Figure 1B). The G4-stabilizing effects of monovalent ions decrease in the following order: $K^+ > Na^+ > Li^+$ (14). Similarly, small molecules that target G4 can stabilize the structure, and the first G4-targeting compound was developed to inhibit the telomerase-mediated telomere extension (19). G4s can be classified as intermolecular and intramolecular structures (Figure 1C) (20). Intermolecular G4s are formed by two or four separate strands of DNA or RNA, whereas intramolecular G4s are formed by a single DNA or RNA strand (Figure 1C). G4s can also be classified topologically as parallel, anti-parallel or hybrid (Figure 1C). rG4s prefer to form in the parallel topology due to the anti-conformation of glycosidic bonds in ribonucleosides, and except for a few special cases (21–23), most of the reported rG4s to date adopt a parallel topology. In contrast, dG4s can adopt parallel, antiparallel or hybrid topologies and switch from one to another, depending on physiochem-

A

Gel formation of guanylic acid (GMP) observed [1] — **1910**

Guanine-quartet motif elucidated [2] — **1962**

1988 — Formation and biological implication of dG4 reported [3]

Formation of intermolecular rG4 observed [7] — **1991**

1991 — Inhibition of telomerase by dG4 indicated [6]

1993 — Telomeric dG4 binding protein reported [41]

1997 — dG4 ligand and its inhibition effect demonstrated [19]

rG4 binding protein reported [10,41] — **2001**

2001 — Formation of telomeric dG4 in ciliates observed [8]

2002 — Formation and function of dG4 in the promoter region indicated [11]

Formation and function of rG4 in the 5′ UTR region indicated [16] — **2007**

2005 — Prediction of dG4s in the human genome demonstrated [18,25]

Prediction of rG4s in the human transcriptome demonstrated [27] — **2008**

Selective rG4 molecule reported [36] — **2012**

2013 — Imaging of dG4s in fixed cells using G4-specific antibody showed [43]

Imaging of rG4s in fixed cells using G4-specific antibody showed [44] — **2014**

Imaging of dG4 and rG4 in live cells using fluorescent microscopy revealed [45] — **2015**

2015 — Imaging of dG4 and rG4 in live cells using fluorescent microscopy revealed [45]

2015 — Genome-wide mapping of dG4s in the human genome reported [48]

Transcriptome-wide mapping of rG4s in the human transcriptome demonstrated [50] — **2016**

Unfolded state of rG4 regions in eukaryotic cells and depletion in bacteria reported [51]

2016 — Genome-wide mapping of dG4s in human chromatin indicated [49]

Transcriptome-wide mapping of transient rG4s in the human transcriptome demonstrated [53] — **2018**

Detection rG4 structures in plants by SHALiPE-seq reported [97] — **2020**

2020 — Live-cell single-molecule imaging of dG4 formation reported [47]

B

G-tract1 loop1 G-tract2 loop2 G-tract3 loop3 G-tract4

GGG N_{1-7} GGG N_{1-7} GGG N_{1-7} GGG

G-quartet G-quadruplex

- Guanine
- K^+ K⁺
- G-quartet

C

Antiparallel Parallel Parallel Antiparallel Hybrid

Intermolecular Intramolecular

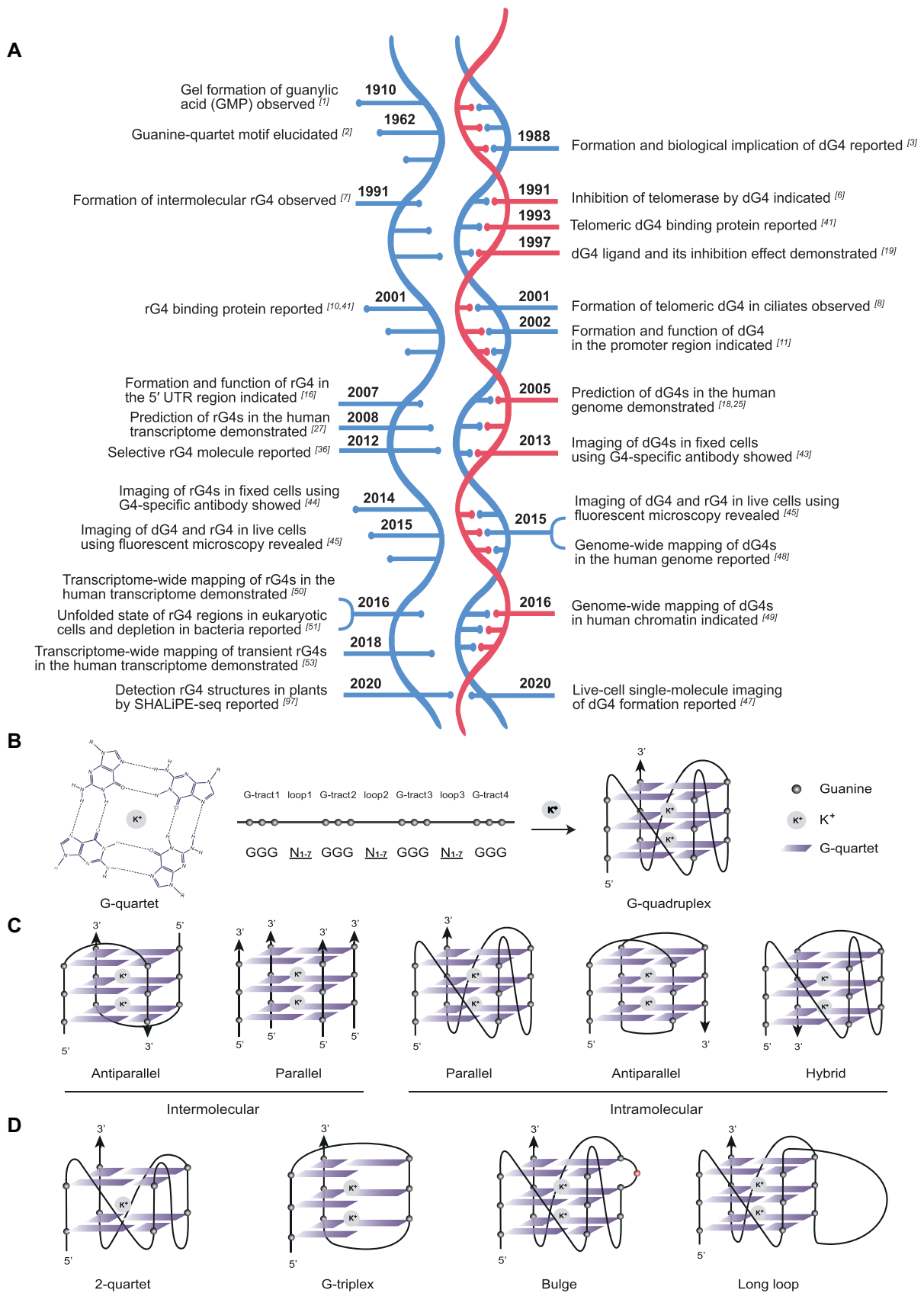**D**

2-quartet G-triplex Bulge Long loop

Figure 1. Milestones in G-quadruplex (G4) and the structural diversity of G4s. (**A**) Key discoveries and developments in rG4 (left) and dG4 (right). This is a huge topic and we apologize for any omission. (**B**) Structure of a G-quartet, and canonical G4 with G3N1-7 sequence consensus. Potassium ion (K^+) stabilizes the G4. (**C**) Representative topologies of intra- and inter-molecular G4s. (**D**) Representative type of non-canonical G4s.

ical conditions and sequence properties (24). Based on systematic computational analyses of the human genome, over 375 000 canonical PQSs were reported for the first time in 2005 (Figure 1A) (18,25). PQSs are now known to be prevalent and non-random, localizing in functionally important regions of the genome, such as in telomeres (5), promoters regions (26) and untranslated regions (UTRs) of messenger RNAs (mRNAs) (27). Moreover, many PQSs are highly conserved among mammalian species (28), which further implies their biological functionality.

Although dG4s and rG4s share the basic structural motif—the G-quartet—they are distinct. First, rG4s are more compact and thermally stable than their DNA counterparts (29–32) due to the 2′-hydroxyl group in the ribose sugar and the networks of water-mediated contacts within the grooves of RNA (33). This compactness allows more intramolecular interactions within rG4s. Second, dG4s and rG4s show distinct preferences in binding specificity with cations (34) and ligands (35,36). K^+ adequately stabilizes both types of G4, whereas Na^+ sometimes has a strong effect only on dG4s but not rG4s (37). Due to the 2′-hydroxyl groups in RNA and their effects on the groove and loop widths, some G4 ligands interact less effectively with rG4s than with dG4s (35). Third, in eukaryotes, DNA mostly exists in a classical double-stranded helix conformation in the nucleus, whereas RNA is single-stranded in nature and located in both the nucleus and cytoplasm. Therefore, RNA in general has more freedom to fold into diverse secondary structures including rG4s, and has access to a great diversity of binding partners, such as proteins, in different cellular compartments (38,39). The first dG4-binding protein, the β subunit of the Oxytricha telomere-binding protein, was identified in 1993 (40), but it was until 2001 that the first rG4-binding protein, fragile X mental retardation protein (FMRP), was reported (Figure 1A) (10,41). Over the last two decades, several of G4-binding proteins have been identified to interact with rG4 structures and play regulatory roles by folding or unfolding the G4 structures or recruiting more binding partners (38,42). In 2013–2014, G4s were found to exist in fixed human cells in studies using G4-specific antibodies, BG4 (Figure 1A) (43,44). Similarly, G4-specific turn-on probes, and more recently a single-molecule fluorescence imaging probe (SiR-PyPDS), were developed to detect G4s in live cells (45–47). Shortly after the report of genome-wide mapping of dG4s in genomic DNA *in vitro* and chromatin *in situ* (48,49), transcriptome-wide mapping of rG4s in the human transcriptome was reported in 2016 (Figure 1A) (50,51). In these works, besides canonical G4s, non-canonical G4s such as two-quartet G4s, G-triplexes, bulged G4s and long-loop G4s were found to be prevalent (Figure 1D), but the *in vivo* folding status of many of these G4s, especially rG4s, detected *in vitro* was in question (50–52). Several studies in the last few years have suggested a dynamic and transient behaviour of rG4s *in vivo* (53,54), and many new rG4s with critical functions have been revealed not only in mRNAs, but also in non-coding RNAs (ncRNAs) (55,56).

Compared with dG4s, rG4s had received less attention until recently. With the continual emergence of data on the innovative transcriptome-wide detection and novel cellular roles of rG4s, this review aims to provide a comprehensive summary and critical analysis of the latest advancements in the rG4 field, and particularly on rG4 transcriptomics and functions in different biological systems, which are currently missing in recent reviews. We summarize the high-throughput techniques developed to map rG4s on a global scale. We also highlight the crucial functions of rG4s and their mechanisms elucidated in mammalian species, viruses, plants and bacteria. In each respective section, we also discuss current challenges and future perspectives to address the open questions in the rG4 field.

rG4 genomics in its development stage—from genome- to transcriptome-wide rG4 detection

Investigation of rG4 biology from a genomic approach requires knowledge of many, if not the entirety of rG4 structures harbored in the genome of an organism-of-interest. A comprehensive and accurate list of rG4s would support the inference of associations between rG4s and other genomic elements, where such information holds promise to reveal the general functional significance of rG4s in biological systems. Unfortunately, because most of the classical experimental rG4 detection methods are low-throughput, the lack of a suitable genomic-wide rG4 profiling method has been the major obstacle in adopting a genomics approach in rG4 research (52).

To circumvent that limitation, pioneering researchers opted for an alternative approach by first computationally predicting PQSs in a genome-of-interest, then selecting the subset of sequences overlapping with transcribed gene regions as the input of rG4 genomic studies. Therefore, it is common to find existing rG4 genomic studies based predominantly on computationally predicted rG4 structures (57,58). Logically, the fidelity of these studies depended on the performance of the prediction methods in capturing the biophysical phenomena underlying rG4 structures. To offer a clearer picture of the theoretical basis of rG4 genomic studies, we first revisit the major methods and results that contributed to the bulk of our existing knowledge on rG4s, which was obtained through the iterations of

1. Low-throughput experimental validation of individual PQSs to identify rG4-forming sequences (Figure 2A)
2. Generalization of biophysical rules governing the rG4 formation propensity from experimental evidence (Figure 2B)
3. Establishment of new rG4 prediction methods built upon the generalized findings and propensity rules (Figure 2C)

In an ideal situation, rG4 prediction methods would identify rG4 structures from input transcript sequences reliably, thus alleviating the need for experimental validation. However, all existing rG4 prediction methods have limited predictive power and can only identify PQSs, whereas their rG4 formation capability remains uncertain and requires confirmation using low-throughput rG4 detection assays, e.g. circular dichroism. The requirement for low-throughput assays then bottlenecks the discovery rate of new rG4 structures, which further limits the number of rG4s available for use to generalize biophysical factors in rG4 formation,
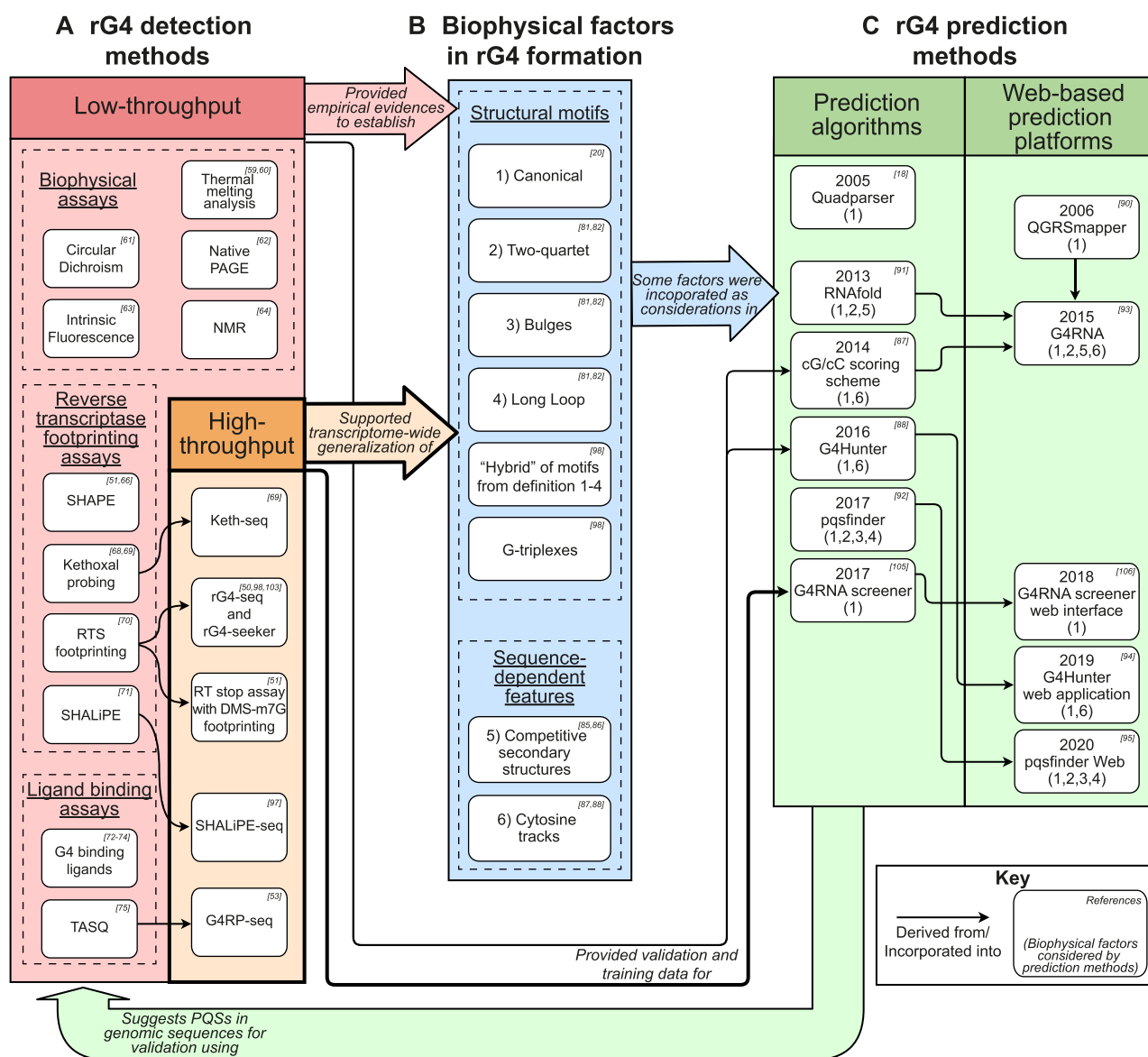
Figure 2. An overview on the current paradigm of rG4 genomic research. (**A**) rG4 detection methods identify rG4 structures in transcribed genomic sequences experimentally. (**B**) Biophysical factors in rG4 formation are generalized from the repository of experimentally confirmed rG4 structures. (**C**) rG4 prediction methods provide the genome-wide screening of PQSs, through identifying genomic sub-sequences that fulfill a subset of biophysical factors in rG4 formation.

and thus degrades the capabilities of the factors to fully describe the rG4 formation propensity. As a result, rG4 prediction methods incorporating these factors can only achieve mediocre prediction power, thus effectively leading to a cyclic problem that ultimately prevents reliable identification of all rG4s in genomes.

To overcome this problem, significant efforts have been invested to develop high-throughput experimental rG4 detection methods that profile rG4s at a transcriptome-wide scope, without requiring PQS input from rG4 prediction methods. These novel methods share a common approach of coupling classical rG4 detection chemistries with RNA sequencing (RNA-seq) technology, which has greatly enhanced their throughput from one to hundreds/thousands

of rG4s per experiment. Importantly, the structural features of rG4s detected with high-throughput methods are coherent with our understanding of their biophysics and the results from their low-throughput counterparts. In the second part of this section, we review several high-throughput rG4 detection methods available to date, summarize their capabilities and notable features and discuss their general significance in rG4 research.

Nevertheless, the emergence of high-throughput rG4 detection methods has introduced not only new opportunities in rG4 genomic research, but also new challenges. These include their high sample requirements and sensitivity to transcript abundance in RNA samples, as well as the large number of rG4 detections, which can be overwhelming for

human interpretation. In the last part of this section, we discuss our perspectives on how the challenges imposed by high-throughput methods may be overcome, and how their findings could be utilized to answer pending questions in different facets of rG4 research.

Low-throughput rG4 detection methods

The general objective of rG4 detection methods is to determine the presence/absence of rG4 structures within an RNA region-of-interest. Low-throughput assays are most applicable to genomic regions that have already been suggested to harbour PQSs, but whose rG4 formation capabilities are yet to be experimentally confirmed. These methods typically evaluate one short RNA sequence (<100 nt) at a time. The rG4 detection methods developed to date can be broadly classified into three major types based on their working principles (Figure 2A):

-**Biophysical assays** are the earliest type of assay available for detecting rG4s. These techniques either evaluate changes in biophysical metrics, such as RNA melting temperature (59,60), spectrometric absorbances (61), migration velocity in acrylamide gel (62) or intrinsic fluorescence (63), which vary between the unfolded and folded forms of rG4s; or directly determine the unique imino proton chemical shifts of G-quartet base pairing or the 3D structure of a PQS (64). Because these metrics are unique to rG4 motifs, these assays remain the gold standards for rG4 detection (65). Nonetheless, biophysical assays might not fully reflect the influence of flanking sequences and *in vivo* environments in rG4 formation, as they typically measure rG4 folding of homogeneous and short RNA oligonucleotides.

-**Reverse transcriptase (RT) footprinting assays** are a genre of techniques that exploit the intrinsic behaviour of reverse transcriptases, where cDNA synthesis can be blocked by covalently modified RNA residues (66) or thermostable RNA structures (67) on the template RNA. The footprints of blocked cDNA synthesis at some positions on a transcript, also known as a reverse transcriptase stalling (RTS) event, can then be detected by techniques such as denaturing polyacrylamide gel electrophoresis (PAGE).

The major innovation underpinning the detection of RNA structure by RT footprinting involves the development of chemicals that selectively modify unstructured nucleotide residues. Structured and unstructured regions on RNA molecules can then be identified based on the differences in chemical modification rate, which are indicated by the observed frequency of RTS events. Methods that see applications in rG4 detection include selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) (51,66) and kethoxal footprinting (68,69). Although these methods are compatible with capturing *in vivo* RNA structures, they are unable to clearly differentiate between double-stranded RNA (dsRNA) and rG4s, and could be ambiguous in situations where an rG4 competes with an alternative RNA structure, e.g. dsRNA, to form at locus-of-interest.

An alternative approach to detect rG4 specifically with RT footprinting involves incorporating an *in vitro* rG4 folding step: when supplied with an rG4-stabilizing buffer solution of physiologically relevant K^+ concentration and optionally in the presence of rG4-stabilizing ligands, rG4 structures can fold stably to block reverse transcription at their 3′ end. The reaction is then repeated in a non-rG4-stabilizing buffer, typically of Li^+ concentration of the same ionic strength, which prevents rG4s from folding. RTS events that are specific to rG4-stabilizing conditions can then be identified to infer the presence of rG4 structures. Because the formation of dsRNA structures is not known to be sensitive to monovalent cations and rG4 ligands, the method can effectively differentiate rG4s from dsRNA (70). However, the method is unable to capture the *in vivo* folding status of rG4s as native RNA structures are difficult to preserve during cell lysis and RNA extraction procedures. Moreover, since an rG4 structure typically only leaves RT footprint at its 3′ end: if 5 or more G-tracts are present in adjacent on a transcript, it may be difficult to determine the sequence span of a rG4 motif utilizing only footprinting information.

To overcome these limitations, variations of the technique have been further developed by preserving *in vivo* structural information with chemical modifications. One such method is known as selective 2′-hydroxyl acylation with lithium ion-based primer extension (SHALiPE), which compares the *in vivo* and *in vitro* SHALiPE data of PQSs to identify rG4s that can fold in both conditions from their characteristic SHALiPE footprints (71). An alternative method described by Guo *et al*. utilizes dimethylsulphate (DMS) to methylate guanine tracts of unfolded *in vivo* rG4s at N7 positions, where N7-methyl-guanine (m7G) does not support rG4 refolding under *in vitro* rG4-stabilizing conditions (51). In the respective methods, the additional SHALiPE and m7G information can both reveal the *in vivo* folding status of rG4 structures and cover the entire span of rG4 motif.

In general, because a sequence-specific reverse transcription primer can be used in PAGE, RT footprinting assays are typically compatible with native RNA extracted from biological samples. In contrast, these methods may face ambiguity when the targeted footprint is not unique to rG4s.

-**Ligand-binding assays** rely on chemical ligands that possess a binding specificity to G4s. In general, many of these ligands are approximately planar in shape and interact with G4s by approaching from the anterior/posterior of the structure by stacking (72). The ligand–G4 interaction can then be detected by ligand-induced fluorescence enhancement in spectroscopy (73), engineering of the ligand to carry fluorophores for imaging (74,75) or attachment of molecular tags such as biotins for pull-down (53). These applications allow the spatial detection of rG4s and the enrichment of rG4-containing RNAs for further analysis, which are unique advantages of ligand-binding assays. However, some G4 ligands can induce and stabilize G4s (76) that do not naturally occur under physiological conditions (77,78). Moreover, given the diversity of rG4 structural conformations, a single type of ligand may not be capable of ubiquitously binding and detecting all rG4s in the genome.

Biophysical rules governing dG4 and rG4 formation propensity

It has been established that the presence of an rG4-forming sequence of four G-tracts spaced by three connective loops is not always sufficient to infer rG4 folding and formation. Other biophysical factors, such as the lengths of G-tracts/loops and the nucleotide sequence context, have been hypothesized and later experimentally validated to affect the propensity for rG4 formation (79–82). Therefore, there have been numerous attempts to deduce a generalized understanding of how these biophysical factors influence G4 formation. Most of the generalizations were established based on human interpretation of common features (e.g. length and nucleotide compositions) among a few hundreds of rG4s identified with low-throughput detection assays, followed by experiments that tested for the boundary cases. These features can be classified as structural motifs and sequence-dependent features (Figure 2B):

-**Structural motifs** are definitions of the range of G-tracts and loop lengths that are most likely to support G4 formation. They are generally assumed to be interoperable between dG4s and rG4s.
The most primitive structural motif is referred to as the canonical motif, which refers to rG4s with G-tract lengths of ≥3 nt and loop lengths of 1–7 nt (20). Meanwhile, G4-forming sequences that deviate from the canonical motif definition have been also discovered (83,84). At the expense of reduced thermostability, these G4s tolerate additional structural imperfections in having only two layers of G-quartet instead of three, significantly longer connecting loops, or bulges in one or more G-tracts. Eventually, two studies in 2013 systematically evaluated the landscape of these imperfections using dG4s (82) or rG4s (81) as models, generalized the correlation between their sizes and negative effect on the stability of G4 structures, and offered new sets of non-canonical structural motifs now known as 'long-loop', 'bulged' and 'two-quartet' to describe these G4s (81,82).

-**Sequence-dependent features** refer to the biophysical factors that exhibit variable influences of G4 formation under different nucleotide sequence contexts. To date, competitive secondary structures and cytosine tracks are the only two factors that have been systemically characterized and are believed to apply to rG4s.
The competitive secondary structure factor expresses the likelihood of a quadruplex sequence to fold into a dsRNA structure instead of a quadruplex structure under favourable free energy conditions (85), for instance in cases where the flanking sequence and the quadruplex sequence are highly complementary. The phenomenon was first experimentally verified in 2012, where a hairpin-to-rG4 conformation transition of an RNA was achieved by modifying the concentration of rG4-stabilizing cations in buffer (86).
The cytosine track factor expresses the likelihood of cytosine residues within or adjacent to quadruplex sequences to sequester critical guanines (e.g. constituents of the G-quartet) via Watson–Crick base pairing, and thus inhibit rG4 formation. The hypothesis was first experimentally

proven and established as a scoring system in a 2014 study (87). The relative ratio between guanine tracks and cytosine tracks near a quadruplex sequence was then shown capable of forecasting the rG4 formation propensity in 2016 (88).

dG4 and rG4 prediction methods

In outline, G4 prediction methods utilize our understanding of G4 formation propensity to conduct searches for PQSs within a specific nucleotide sequence or in the entire genome, and optionally report prediction score derived from biophysical factors that influence the G4 formation propensity (52,89). Except for methods that are considered rG4-specific, G4 prediction results are generally assumed to be applicable to both dG4s and rG4s. It is worth noting that not all biophysical factors are considered by some prediction methods (Figure 2C).

Meanwhile, G4 prediction methods are commonly classified as either prediction algorithms or Web-based prediction platforms based on their user interfaces. Prediction algorithms are typically mathematical formulas or computer software that implement the PQSs screening logics, whereas web-based prediction platforms incorporating these algorithms serve as a more accessible interface for researchers (Figure 2C). At the researcher's discretion, the complete list of predicted PQSs can then be used as input for a genomic study, or a few PQSs overlapping biological regions-of-interest (e.g. a gene-of-interest, regulatory elements) can be handpicked for experimental validations and downstream structural and functional investigations.

The earliest iterations of G4 prediction methods are primitive in nature, and only search for PQSs matching specified G-tract lengths and loop lengths in an input nucleotide sequence. The canonical structural motif was first incorporated as the basis of prediction in two 2005 studies (18,25), where one of the studies consolidated their *quadparser* algorithm into a command-line program (18). *QGRSmapper*, which was published shortly after the two studies, brought improvements including a web interface and options for users to customize the G-tract lengths and loop lengths to search for (90). The simplicity of *QGRSmapper*'s design has enabled it to remain as a popular tool for rG4 research to this day. However, although *QGRSmapper* offers a 'G-score' to evaluate the likelihood for PQSs to form a stable G4, the correlation between G-score and rG4 formation propensity has not been formally certified with empirical data.

Following the establishment of several rG4-specific biophysical factors in the early 2010s, the next iteration of prediction methods newly incorporated non-canonical rG4 structural motifs for PQS searching, and/or sequence-dependent features for deriving metrics that reflect the likelihood of rG4 formation. They included *RNAfold* (91), the *cG/cC scoring scheme* that considered guanine/cytosine tracts (87), *G4Hunter* (88) and *pqsfinder* (92). Notably, the *cG/cC scoring scheme* and *G4Hunter* also leveraged the list of experimentally verified rG4s to validate their algorithms (Figure 2A, C), suggesting both rG4 detections and biophysical factors can contribute to the formulation of prediction methods. Later, the functionalities of *QGRSmapper*,

RNAfold and the *cG/cC scoring scheme* were integrated into a web-based tool named *G4RNA* (93). Meanwhile, the authors of both *G4Hunter* and *pqsfinder* offered web interfaces for their algorithms in 2019 (94) and 2020 (95), respectively. The timeline of development, dependencies and the biophysical factors that these prediction algorithms and web-based prediction platforms considered is illustrated in Figure 2C. In comparison to *QGRSmapper/quadparser*, these newer methods can capture non-canonical PQSs more comprehensively and offer more metrics/scores that illustrate the rG4 formation propensity of PQSs. They also leverage experimentally confirmed rG4s to establish and/or benchmark the prediction methods. Nevertheless, although the new metrics/scores are correlated with the rG4 formation propensity, these prediction methods do not utilize such information to differentiate between PQSs that fold into rG4 structures and those that do not.

Most importantly, all of the above-reviewed G4/rG4 prediction tools assume that users will experimentally validate the predicted PQSs for rG4 formation ability before applying them as input for rG4 biological studies, as they do not offer predictions of the rG4 folding/non-folding status of PQSs. Therefore, for rG4 genomics studies that directly utilize PQS prediction results without further validation, their findings must be interpreted with a caveat that the list of PQSs might contain a significant proportion of false positives.

In this section, we focused on discussing the subset of methods that are either foundational to rG4 research or have considered rG4-specific factors. A recent compresensive review by Lombardi *et al.* (89) provides additional background of diverse G4 prediction methods regarding their implementations and performances.

High-throughput rG4 detection methods

High-throughput RNA structural determination methods first emerged in the early 2010s following the maturation and popularity of RNA-seq technology. High-throughput assays are a development of their low-throughput counterparts that can evaluate a heterogeneous mixture of RNA sequences in parallel. They are usually applicable to a transcriptome-wide screening of rG4s without requiring PQS predictions *a priori*. Here we review a subset of these methods that are suitable for rG4 detection (Figure 3):

-Non-rG4-specific RT footprinting methods
Keth-seq is currently the only high-throughput RNA structural determination method that can claim a detection of rG4s alongside dsRNA regions (69). It was applied to map *in vivo* RNA structures including rG4s in human and mouse by employing cell models (69). The method works similarly to its low-throughput counterpart by using N^3-kethoxal to label unstructured guanine bases and induce RTS. The RT footprints are then captured via sequencing to infer the *in vivo* folding/non-folding status of rG4s based on the absence/presence of footprints at the G-tracts. Given an input of PQSs, Keth-seq is effective in determining their respective *in vivo* rG4 folding status by evaluating whether the putative G-tracts are structured and protected from kethoxal modification; however,

the possibility that the PQSs will form a dsRNA structure must also be accounted for. In contrast, Keth-seq is less suitable for identifying rG4s that fall outside the predicted list of PQSs, for instance novel non-canonical rG4s based on kethoxal footprints, as these footprints might originate from dsRNA regions instead.

-rG4-specific, RT footprinting methods
The rG4-seq method (50) and the RT stop profiling method with DMS-m^{7}G footprinting (51) were the first methods for specific rG4 detection by RT footprinting. rG4-seq was first applied to map *in vitro* rG4 in human by employing a cell model (50) and later to plants and bacteria (96,97); while the RT stop profiling method was applied to map *in vitro* and *in vivo* rG4 in yeast and *Escherichia coli* cells (51). The two methods share an *in vitro* rG4 refolding step, which allows them to determine the 3′ end of *in vitro* rG4s in a similar fashion to their PAGE-based low-throughput counterpart. The RT stop profiling method differs from the rG4-seq method by the incorporation of an additional *in vivo* DMS treatment step, which methylates guanine residues in unfolded rG4s at their N^7 position. Because m^{7}G does not support G-tetrad and G4 formation, the phenomenon can be exploited to infer the *in vivo* folding status of rG4s. Later, in 2020, the SHALiPE-seq method, which was derived from its low-throughput counterpart SHALiPE, was developed and applied to map *in vivo* rG4s in *Arabidopsis thaliana (A. thaliana)* and *Oryza sativa* (*O. sativa*, or Rice) (97). In SHALiPE-seq, *in vivo* rG4s and dsRNA are probed using SHAPE chemistry coupled with sequencing. Simultaneously, *in vitro* SHALiPE probing is conducted in both rG4-stabilizing and rG4-non-stabilizing conditions to specifically identify *in vitro* rG4s. Finally, by combining and comparing the three sets of data, rG4s that fold both *in vivo* and *in vitro* can be identified from their characteristic SHALiPE footprints.

The unique advantage of RT footprinting assays with rG4 refolding lies in their intrinsic rG4-specificity, which allows them to work without an input of known PQSs. Combined with tailored bioinformatic software such as rG4-seeker, which was engineered to work with rG4-seq, an *ab initio* rG4 identification can be achieved by detecting all RTS sites that occur only under rG4-stabilizing conditions. In contrast, this genre of assays must rely on external information, such as DMS-m^{7}G footprinting or SHALiPE, to determine the *in vivo* folding status of rG4s. Consequently, *in vivo* rG4s that do not refold under *in vitro* conditions cannot be detected and profiled using these assays.

-rG4-specific, ligand-binding methods
G4RP-seq is currently the only high-throughput rG4 detection method that takes a ligand-binding approach in cells (53). The method uses a biotinylated G4-specific ligand/probe, BioTASQ, to achieve affinity-pull down of *in vivo* rG4s and the enrichment of rG4-containing transcripts. The changes in relative abundances of transcripts are then evaluated via RNA-seq. The G4RP-seq method was applied to identify *in vivo* rG4s in human by employing a cell model (53). Although the technique represents a technological advancement in ligand-based rG4 detec-
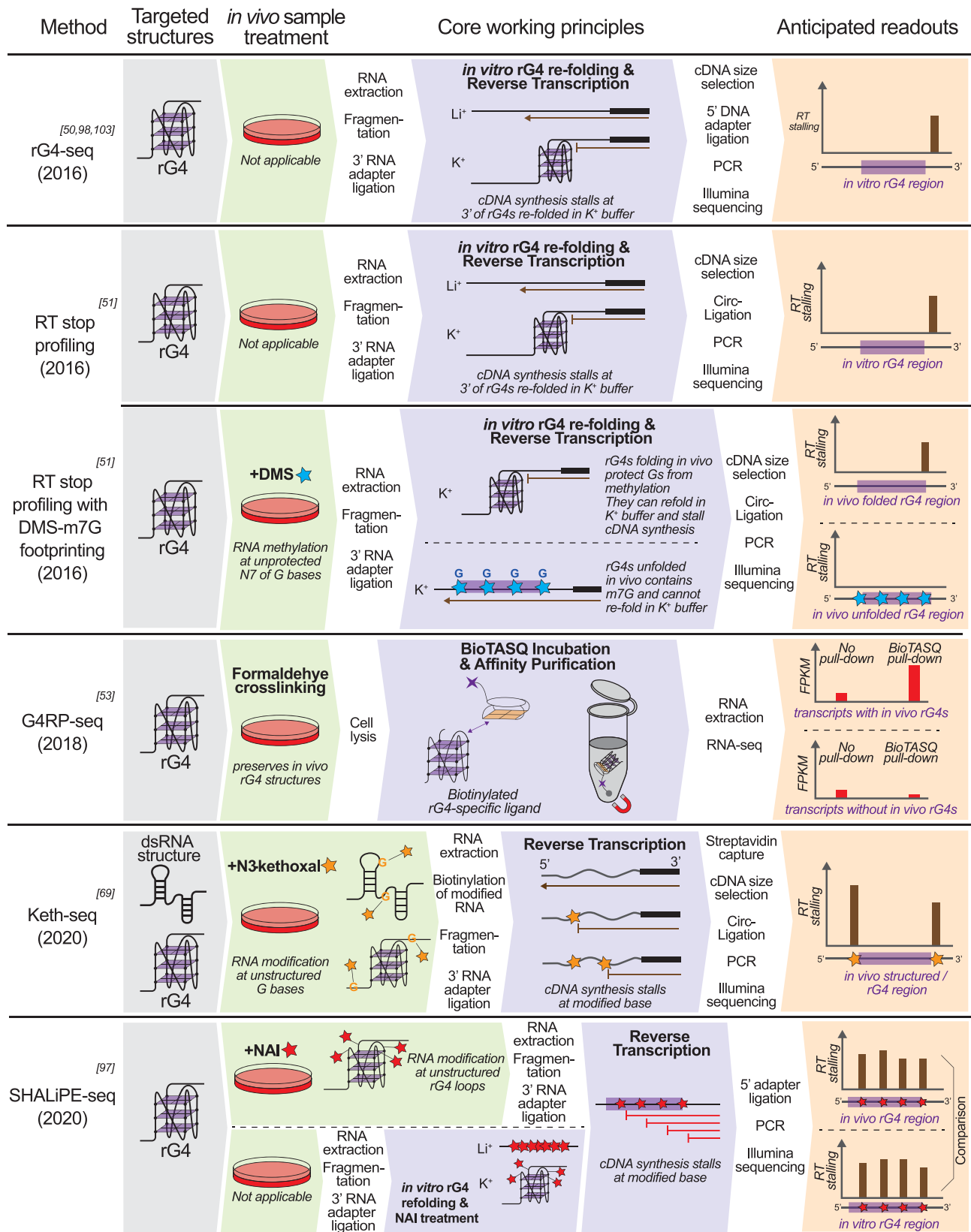
Figure 3. Summary of high-throughput rG4 detection methods.

tion, the resolution of G4RP-seq is limited to the transcript level. Because the number and the location of rG4 structures per transcript cannot be determined by G4RP-seq, its results may have limited utility in rG4 genomic studies, where the accurate sequence and position of rG4s are required.

Significance of high-throughput rG4 detection

Compared with the low-throughput and computational prediction methods, the high-throughput methods have unique strengths that are expected to bring significant improvements to the genome-wide profiling of rG4s.

First, the high-throughput methods are capable of *ab initio* detection of rG4 structures, especially for rG4s that were formed from sequences previously not considered as PQSs. A recent re-analysis of a human rG4-seq dataset published in 2016 revealed evidence for the prevalence of unusual rG4-forming sequence that fall outside of the existing rG4 structural motif definitions (98). Notable examples included two-layered rG4 structures with an extra-long loop, which could be described as a hybrid of the long-loop and two-quartet motifs, as well as RNA G-triplex motifs, which lack a fourth G-tract (98). With the expectation that more unusual rG4s in different organism models will be revealed under high-throughput methods, a better general understanding of the phenomena underlying these rG4s can be expected to emerge.

Meanwhile, when coupled with appropriate bioinformatic analysis, the high-throughput methods offer empirical evidence for rG4 detection competitive with their low-throughput counterparts. The recently proposed *rG4-seeker*, a bioinformatic program tailored for the rG4-seq method, analyses rG4-induced RTS footprints at single-nucleotide resolution and replicate-independent manner, and offers results closely resembling the outcome of a PAGE-based RTS footprinting assay (98). There are hopes that high-throughput rG4 detection methods could substitute the conventional predict-then-validate workflow for identifying novel rG4s on genes-of-interest.

Finally, other recent advancements in high-throughput methods have improved the *in vivo* detection of rG4s. The intracellular environment is known to harbour additional biophysical and biological factors that influence rG4 formation, for instance molecular crowding effects (99,100) and rG4–protein interactions (101,102) that are difficult to model *in vitro*. High-throughput methods hold promise to profile the discrepancies in the rG4 folding behaviour between *in vivo* and *in vitro* conditions, and to model the overall influence of the intracellular environment on the rG4 formation propensity.

Current challenges in high-throughput rG4 detection

One major challenge for high-throughput methods lies in their large sample input and difficulties in profiling rG4s on low-abundance transcripts. Because these methods typically require tens to hundreds of micrograms of RNA (52), their applications are mostly limited to cultured cell or whole organism models, which can be expanded to provide the necessary RNA amount. Meanwhile, because the outcomes of

rG4 detection are dependent on the number of RNA-seq reads available, the rG4 profiling results of low-abundance transcripts tend to be more ambiguous (98). Moreover, high-throughput methods frequently incorporate a poly(A) enrichment step, which compromises the success rate of rG4 discovery among the non-polyadenylated transcripts, such as non-coding RNAs, that are depleted in the process. Despite ongoing efforts to alleviate the high-input RNA requirements for these methods (103). Currently, researchers must still resort to the conventional predict-then-validate workflow when the transcript and/or model organisms-of-interest are not suitable for high-throughput detection.

Another challenge introduced by high-throughput methods lies in the subsequent analysis, which aims to reach generalized conclusions regarding biophysical features from the large numbers of rG4s detected. Conventionally, the generalizations were largely dependent on human interpretation, placing limits on both the complexity of features and the number of rG4s that can be considered in the analysis. However, rG4 formation is influenced by the complex interplay of multiple biophysical features including sequence lengths, nucleotide compositions and RNA structures in an interdependent manner. It remains unclear whether a human interpretation-driven approach can fully leverage the thousands of rG4s newly identified by high-throughput methods and offer new, better explanations of the phenomena surrounding the rG4 formation propensity.

A possible solution that may mitigate both limitations at once would be computational processing of high-throughput rG4 profiling data. This would involve applying analytic techniques from data science, such as machine learning, to capture the underlying properties and patterns shared among rG4s, including complex properties that are not intuitive to human interpretation. Based on the information obtained, improved rG4 prediction methods with better predictive power could then be implemented for rG4 screening in transcripts to which high-throughput methods are not well suited. From our review of existing G4 and rG4 prediction methods, we foresee several possible, but non-exclusive, pathways for such investigations:

1. To identify and model novel biophysical factors that influence the rG4 formation propensity. In the neighbouring dG4 research field, the results from high-throughput genome-wide profiling of dG4s are processed by machine learning, which enables identification of biophysical factors that are relevant to dG4 folding potentials, and the modelling of their corresponding influences on G4 formation (104). Given the similarities between dG4s and rG4s, the dG4 study model might be transferrable to rG4 research by using the profiling results of the latter in place of the former as input.

2. To automatically capture the defining properties of rG4s using an artificial neural network. A 2017 study suggested that the properties of rG4s can be automatically captured and modelled using machine learning techniques given input data consisting of non-rG4 sequences and rG4 sequences detected by high-throughput methods (Figure 2A and C). Based on that approach, the study proposed the *G4RNA screener* software, which predicts rG4s from arbitrary input nucleotide sequences

(105,106). Although *G4RNA screener* can only evaluate rG4s with the canonical structural motif, it is possible that similar prediction tools that evaluate both canonical and non-canonical rG4s could be developed using a similar approach but including input data from high-throughput rG4 detection methods.

3. To establish discriminative models that infer the rG4 folding outcome of PQSs. Despite the proposal of multiple schemes to calculate the rG4 folding propensity of PQSs by numerical metrics, a successful attempt at devising a decision boundary that differentiates rG4-folding and non-rG4-folding PQSs using these metrics has yet been reported. The lack of high-quality training data containing large numbers of PQSs with correctly labelled rG4 folding/non-folding status appear to have hampered the ability of the existing discriminative modelling process to devise the required decision boundaries. Interestingly, the results from some high-throughput rG4 detection methods such as *rG4-seq* offer the possibility to label PQSs as 'detected' and 'not detected'. Furthermore, a comparison of human rG4s detected/not detected in an *rG4-seq* experiment revealed differences in the distributions of two of their metrics, namely competitive secondary structure free energy and cytosine track abundance, which are well understood to associate with the rG4 folding propensity (50). This evidence suggests that the detection status of PQSs could be considered a proxy of their rG4 folding status, where high-throughput rG4 detection results may serve as suitable training data for discriminative modelling. Ultimately, the approach holds promise to establish statistical models that consider multiple rG4 structural features as input and express the rG4 folding propensity of PQSs in probabilistic terms. In such a process, high-throughput rG4 detections results would serve as both training data to construct and improve the model, and testing data for validity evaluation—similar to the approaches already used by the *cG/cC scoring scheme* (87), *G4Hunter* (88) and *G4RNA screener* (105) (Figure 2A,C).

Approaches to connect rG4 genomics with rG4 biology

The central task in rG4 genomics is to connect the presence of rG4s to their general functions within biological systems, thus revealing, for instance, the categories of pathways and mechanisms in which rG4s play a central role, as well as the classes of proteins and other biomolecules that are designated to interact with rG4s. Although high-throughput rG4 detection methods will play a pivotal role in addressing this task by revealing more thorough information on the rG4s in genomes, it is likely that not all possible forms of rG4s will be detected by these methods, nor that all detected rG4s will be biologically relevant to the central task. Therefore, selection for functionally significant rG4s would be necessary to further approach a full understanding of rG4 genomics. For this reason, we speculate that rG4 genomics studies in the near future will be characterized by the integration of different genomics data and techniques from other disciplines of biology to identify the subsets of functional rG4s and annotate them based on the genomic functional elements that they are related to.

One viable approach could be to investigate rG4 genomics from an evolutionary biology perspective, by evaluating the conservation of rG4 sequences across multiple organisms. Given the drastic differences in the genome-wide rG4 landscape between humans (50) and other forms of life (51,96,97), any conserved rG4 structures would likely be biologically significant and could provide an entry point to identify important, conserved rG4-dependent functions.

Meanwhile, some RNA-binding proteins (RBPs), especially RNA helicases, bind rG4s on specific genes with functional implications (107–109). By comparing rG4 detection data with transcriptome-wide RNA–protein interaction probing data, for instance, cross-linking immunoprecipitation-sequencing (CLIP-seq) datasets, it might be possible to reveal the families of rG4-binding proteins and annotate rG4s based on our knowledge of their RBP interaction partners.

Finally, recent studies have suggested that some naturally occurring RNA modifications such as m6As may colocalize with rG4 structures with functional implications (110,111). Although the mechanisms underlying this colocalization have not yet been identified, recent methodological advancements in RNA modification detection (112,113) have opened new possibilities to directly compare modification loci with rG4 sequences on a transcriptome-wide scale. Further progress promises to establish new associations between rG4s and other types of modifications and could reveal connections between epitranscriptomic regulation and rG4 biology.

Biological functions of mammalian rG4s in mRNAs and ncRNAs

The bioinformatic discovery of the widespread occurrence of PQSs in the human transcriptome (27), initially in mRNA, has stimulated the extensive experimental characterization of rG4 formation and function in mRNA, and over the years, numerous important roles of rG4s have been revealed (39,42,114) (Figure 4). In addition, given that the majority of the human transcriptome is composed of ncRNAs (115), recent studies have begun to report the existence and unusual role of rG4s in many classes of ncRNA such as long non-coding RNAs (lncRNAs), microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) (42,116) (Figure 5). Below we highlight and discuss the functions of mammalian rG4s reported in mRNAs and ncRNAs thus far.

5′-UTR rG4s. Earlier studies focused on the effect of 5′-UTR rG4s on cap-dependent translation. Schaeffer *et al.* first reported the formation of an rG4 in the coding region of *FMR1* transcript, and when this specific *FMR1* rG4 motif was inserted into the 5′-UTR in a reporter gene system, it was demonstrated to suppress translation in rabbit reticulocyte lysate (41). A few years later, Balasubramanian *et al.*, using the reporter gene in rabbit reticulocytes lysate, showed that an rG4 motif naturally occurring at the 5′-UTR of the proto-oncogene *NRAS* inhibited translation (16), and their study also highlighted that the density of putative rG4s at 5′-UTR was enriched as compared with the genome average. Since then, many other 5′-UTR rG4 candidates (84,117–
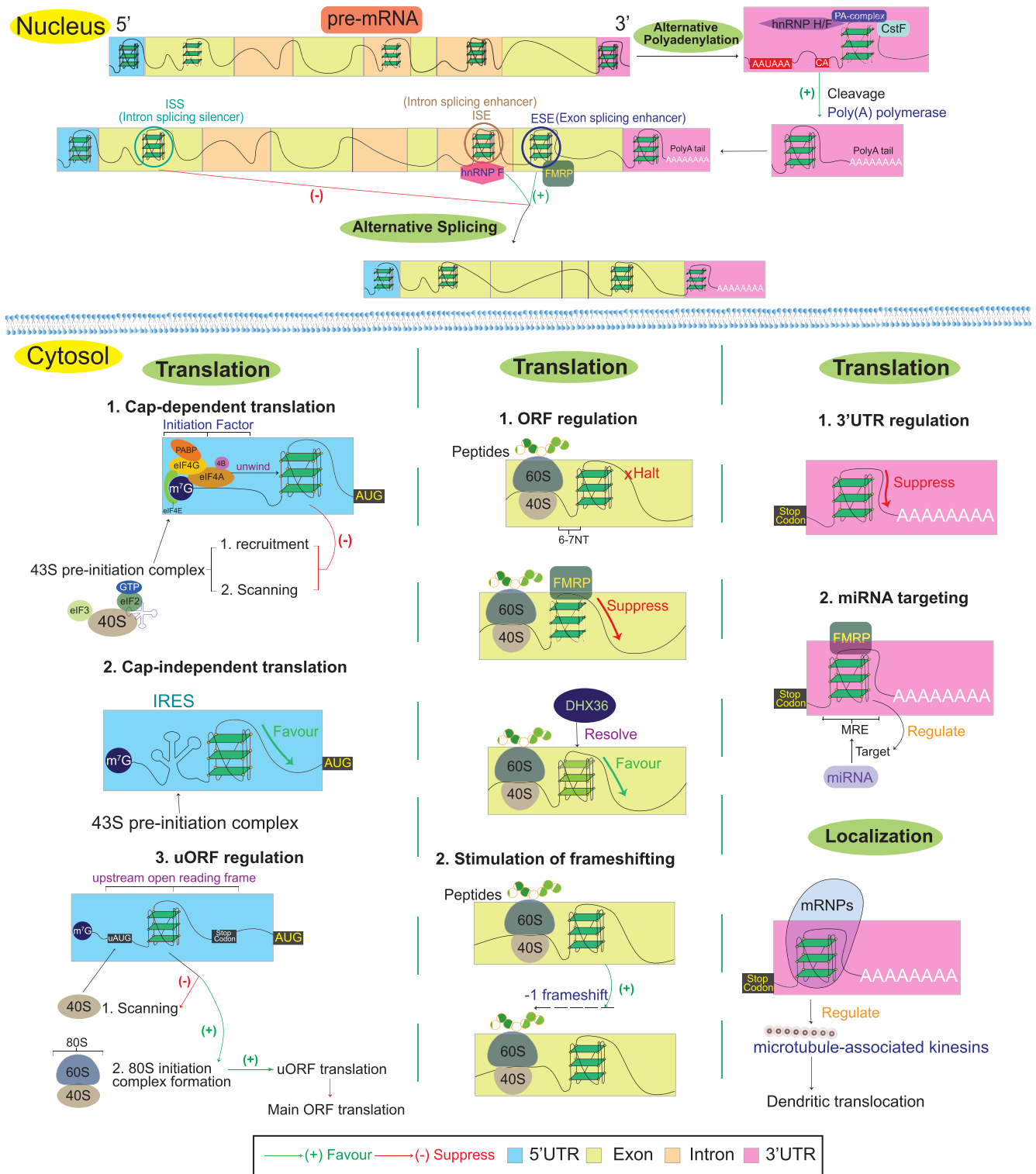
Figure 4. Representative roles of messenger rG4 on diverse biological processes in cells. Messenger RNAs rG4 on the 5′UTRs of mRNAs can primarily impair eukaryotic cap-dependent translation, while they promote the cap-independent translation. Meanwhile, rG4s in the uORF affect the downstream gene translation through the regulation of ribosome initiation complex formation and the ORF rG4s can regulate translational elongation and ribosomal frameshifting progression. The pre-mRNA intron rG4s in ORF can either enhance or silence RNA alternative splicing, while the pre-mRNA exon rG4s act as exon splicing enhancers. rG4s on the 3′UTR affect translation, alternative polyadenylation, alternative splicing, and RNA localization.
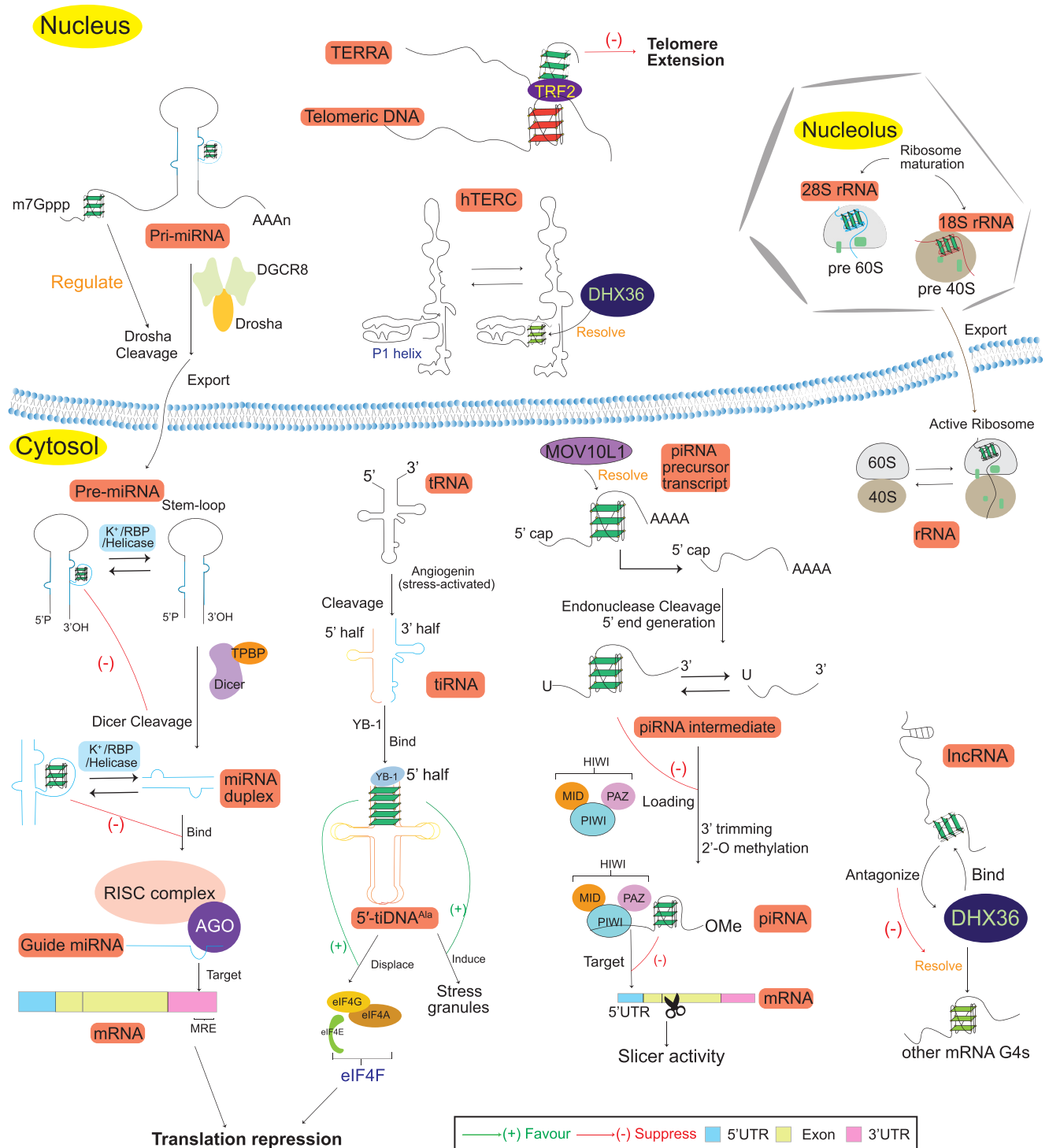
Figure 5. Representative roles of non-coding rG4 on myraid of biological processes in cells. rG4 in *TERRA* lncRNAs can regulate telomere length through interaction with telomere-binding protein and telomeric DNA. rG4 in *hTERC* lncRNA can be resolved by DHX36, which then facilitate P1 helix formation and its template boundary function. rG4 in lncRNA can bind to DHX36 helicase, and therefore, antagonize its unwinding activity to other RNA targets containing rG4. pri-mRNA rG4s regulate Drosha-mediated processing, pre-miRNAs rG4s inhibit DICER-mediated maturation, and miRNA rG4s abolish its loading onto RISC. rG4s can regulate the biogenesis of piRNA and lead to its inhibition of binding to the HIWI-PAZ domain as well as mRNA targeting. rG4 on the 5′ tiRNA can displace initiation factor and trigger translation repression as well as stress granules formation. The functions of rRNA rG4 are not clearly known, but may be involved in ribosomal protein recruitment and ribosome assembly.

125) have also been reported to suppress translation in studies using either cell-free or cell-based reporter gene systems, suggesting the general repressor role of 5′-UTR rG4s in translation (Figure 4). Some notable cases exist, such as 5′-UTR rG4 in TGFβ2, in which the insertion of an rG4 motif alone in the reporter gene inhibited translation in cells, whereas the insertion of an rG4 in the context of the entire 5′-UTR of TGFβ2 augmented translation (126). Similarly, using a cell-based reporter gene, translation was suppressed by a 5′-UTR rG4 containing $(CGG)_{99}$ repeats in *FMR1*, whereas translation was augmented by the same 5′-UTR rG4 containing only $(CGG)_{30}$ repeats (127). Moreover, it has been suggested that the relative location of rG4s to the 5′ cap of mRNA and the thermostability of 5′-UTR rG4s can control the level of translation (128), highlighting that the neighbouring sequence context, rG4 motif density, rG4 position and stability should be considered in 5′-UTR rG4-mediated translational control.

The biochemical mechanism on 5′-UTR rG4s regulating cap-dependent translation has not been fully characterized, but a few studies have provided us an initial glimpse of this complex regulatory pathway. Eukaryotic initiation factor 4A (eIF4A) is an RNA helicase that unwinds structured 5′-UTRs to help with 43S pre-initiation complex recruitment, and a subsequent mRNA 5′-UTR scanning step is to look for the translation start codon. In one study, transcriptome-wide ribosome footprinting was performed under eIF4A inhibition with silvestrol reduced the translation efficiency of mRNAs with longer 5′-UTRs, and such mRNAs were found to be enriched in the $(CGG)_4$ motif, which was later verified to form two-quartet rG4s using biophysical assays (129). These results indicated that rG4s can likely interfere with the recruitment and/or scanning of 43S pre-initiation complex, and thus with translation efficiency. This was further corroborated by a reporter gene assay under eIF4A knockdown (129), supporting an eIF4A-dependent 5′-UTR rG4-mediated translational control. More investigations are warranted to further investigate whether native transcripts containing the 5′-UTR rG4 motif function to be similarly in this system, and what other proteins, besides eIF4A, are involved in this tightly regulated process. Moreover, different 5′-UTR rG4 motifs may interplay with cis-regulatory elements such as uORF (Figure 4) (101,125), alternative competing RNA secondary structures (50,86), or recruit a distinct set of proteins (130) to modulate cap-dependent translation efficiency.

Besides cap-dependent translation, mRNA can sometimes, under stress conditions, undergo translation initiation by internal ribosome entry sites (IRESs) (131). Bonnal *et al.* conducted a deletion analysis on human fibroblast growth factor 2 (*FGF-2*) mRNA and found a 176-nt region containing two stem-loops and an rG4 motif to be a 'structure determinant' for the internal ribosome recruitment by IRES-containing cellular mRNAs (132). Another 5′-UTR rG4-containing transcripts, *hVEGF*, was also reported to promote IRES-mediated cap-independent translation (133) (Figure 4). The mechanistic role of rG4s in IRES-mediated translation is still unclear. Studies reported that the *hVEGF* 5′-UTR rG4 was functionally dispensable (134), and the *BAG-1* 5′-UTR rG4 was not a structural part of the IRES domain, unlike *FGF-2* and *hVEGF*, yet rG4 mutation on

BAG-1 can facilitate the reporter gene expression of both cap-dependent and cap-independent translation (135). Nevertheless, another study has shown that *hVEGF* 5′-UTR G4 can directly recruit the 40S ribosome to initiate the IRES-mediated translation (136). Further work is warranted to elucidate whether different cell types, physiological conditions and sequence contexts have a significant impact on the role of 5′-UTR rG4s in IRES-mediated translation. It may be possible to identify more 5′-UTR rG4 candidates and their functions in IRES-mediated translation by inhibiting cap-dependent translation using chemicals such as 4EGI-1 (137) or rapamycin (138), such that only cap-independent translation will be observed. In addition, it is worth considering the potential involvement of unidentified proteins that may control the equilibrium between cap-dependent and cap-independent translation processes under normal and stress conditions.

3′-UTR rG4s. Similar to 5′-UTR rG4s, 3′-UTR rG4s generally act as suppressors in translational regulation (Figure 4), as illustrated in *PIM1* (139) and *APP* (140). Besides their role in translational regulation, 3′-UTR rG4s regulate other processes, such as alternative polyadenylation, miRNA targeting and mRNA localization (Figure 4).

Beaudoin *et al.* reported that the density of putative rG4s at 3′-UTRs was enriched as compared with the genome-average G4 density (141). In the same study, focusing on *LRP5* and *FXR1* transcripts bearing rG4s at the 3′-UTR, they found that 3′-UTR rG4s of both transcripts stimulated the reporter gene expression when compared with an rG4 mutant. For *LRP5*, the presence of the 3′-UTR rG4 also facilitated the efficiencies of alternative polyadenylation sites in *LRP5*, suggesting that rG4 functions as a polyadenylation regulatory element that positively modulates the use of internal polyadenylation sites. For *FXR1*, the presence of a 3′-UTR rG4 promoted the use of alternative polyadenylation sites, which resulted in the predominant generation of shorter mRNA isoforms over longer isoforms. This 3′-UTR shortening in turn impaired *FXR1*'s miRNA regulation and gene expression (141). As 3′-UTR rG4s are associated with alternative polyadenylation sites, detailed study is necessary to identify and characterize more candidate 3′-UTR rG4s that interplay with alternative polyadenylation sites and determine their impact on gene regulation.

Earlier studies reported that the 3′-UTR of *PSD* mRNA contains a G-rich region that binds to FMRP, and the G-rich region overlaps with the miR-125-binding site (142,143). Stefanovic *et al.* further demonstrated that the G-rich region can fold into thermostable rG4s and interact with FMRP, or form a double-stranded RNA conformation by hybridizing with miR-125. This alternative structure can inhibit rG4 formation and disrupt rG4–FMRP binding (144). Another study by Rouleau *et al.* reported >44 000 putative 3′-UTR rG4s to overlap with predicted miRNA-binding sites (145). Using the *FADS2* mRNA/miR-331-3p pair as an example, they showed that the formation of a *FADS2* 3′-UTR rG4 prevented the targeting of *FADS2* mRNA by miR-331-3p in a cell-based reporter gene system (145). A promising direction for further experiments would be to study the cellular factors and mechanism of

3′-UTR rG4-mediated regulation in miRNA binding and post-transcriptional regulation.

Subramanian *et al.* predicted that rG4s would be enriched in the 3′-UTR dendritic mRNAs and focused on *PSD-95* and *CaMKIIa*, two postsynaptic transcripts that were shown to be involved with the localization processes (143,146–148). Using mRNA reporters in neurons, they found that the 3′-UTR rG4 was necessary and sufficient for dendritic mRNA localization, whereas the deletion of the 3′-UTR rG4 in the reporter mRNA caused a significant reduction in mRNA translocation and also the loss of neurite signals (148). This result suggested that the 3′-UTR rG4 is a cis-acting element and can function as a zip-code for mRNA localization. Future work will be required to discover trans-acting proteins that bind with 3′-UTR rG4s in *PSD-95* and *CaMKIIa* mRNA to form the mRNA–protein complex, which facilitates dendritic mRNA transport to synapses for local translation.

ORF rG4s. Studies have reported that thermostable ORF rG4s can pause ribosome processing, initiate proteolysis and induce ribosomal frameshifting (FS) (Figure 4). Using a synchronized translation assay and cell-based reporter gene system, Endoh *et al.* found that rG4s in ORF can halt the translation elongation step, reporting that the ribosome stalled 6–7 nucleotides before the rG4 (149). This rG4-mediated translational elongation stalling was later found to affect the folding and proteolysis of hERalpha protein (149–151), as well as ribosomal FS events (mostly −1 type FS) in other rG4-containing transcripts in cells (150). The effect of these events was modulated by the G-runs and loop length of the rG4 involved, as well as addition of G4 ligands to the system, which can perturb the rG4 thermostability (150,152). Later, it was demonstrated that periodic fluctuation of translation suppression occurs every three nucleotides within the ORF, but not at the 5′-UTR, providing mechanistic insights on how the ribosome translocates and unfolds rG4 structures (153).

Several ORF rG4s regulate translation via their interaction with rG4-binding proteins. Over the years, FMRP has been one of the best-studied rG4-binding proteins, and the RGG motif in FMRP has been shown to be important for binding rG4s (154–156). Westmark *et al.* reported that the ORF rG4 in *APP* mRNA interacted with FMRP to inhibit its translation, without affecting mRNA stability (157). Thandapani *et al.* showed that the ORF rG4s in *MLL1* and *MLL4* mRNA interacted with the RGG motif of Aven and increased *MLL1* and *MLL4* polysomal association and thus, translation (158). In addition, two other proteins, PRMT1 and DHX36, were also reported in that study to promote *MLL1* and *MLL4* translation through Aven arginine methylation (by PRMT1) and resolving rG4s (by DHX36) (158). Interestingly, recent transcriptome-wide studies have revealed that rG4 proteins and helicases have roles in translation (101,102,159,160). It would therefore be of interest to explore whether and how, mechanistically, these proteins interact with ORF rG4s to control translation elongation, ribosomal stalling and FS.

rG4s in exons and introns regulate pre-mRNA splicing (Figure 4). With regards to exonic rG4s, Didiot *et al.* identified two rG4s in an *FMR1* mRNA-coding sequence (exon 15) and showed that they interacted with FMRP, with no effect on translation (161). Instead, these two rG4s served as exonic splicing enhancers, and bound with FMRP to regulate the *FMR1* alternative splicing and thus the level of short and long FMRP isoforms in cells (161). Fisette *et al.* demonstrated that rG4s in *BACE1* exon 3 recruit hnRNP H to control *BACE1* alternative splicing and the level of the BACE1 isoform that can proteolytically cleave APP to produce amyloid beta peptide (Aβ), which is linked to Alzheimer's disease (162). With regards to intronic rG4s, Gomez *et al.* found an rG4 in intron 6 of the human telomerase (*hTERT*) transcript to function as an intronic splicing silencer and thereby affect the *hTERT* alternative splicing in cells, and the effect was exacerbated by titrating G4-stabilizing ligands (13). Marcel *et al.* reported that an rG4 in intron 3 of the *TP53* transcript worked as an intronic splicing enhancer to promote the splicing of intron 2, which led to differential expression of transcripts encoding distinct p53 isoforms (163). Similarly, Huang *et al.* found that an rG4 in intron 8 of *CD44* directly recruited hnRNP F and operated as an intronic splicing enhancer to regulate splicing and led to exon inclusion, which was shown to maintain the epithelial phenotype and modulate the epithelial–mesenchymal transition (EMT), an important process that drives cancer (164). Other intronic rG4 examples were also recently illustrated to regulate splicing efficiency (165–167). Future investigations may focus on whether the stability of rG4s, their proximity to RNA splicing sites, or other trans-acting factors have an impact on alternative splicing.

lncRNA rG4s. rG4s have been reported in telomere-associated lncRNAs (Figure 5). Telomerase is responsible for telomere maintenance, and was found to be upregulated in most cancer cells (168). Human telomerase consists of the telomerase protein (hTERT) and the telomerase RNA (*hTERC*) components, and studies have reported that the 5′-terminal of human *hTERC* contains an rG4 that interferes with the formation of a critical structural element, P1 helix, in defining the template boundary for reverse transcription (169–171). The ATP-dependent RNA helicase DHX36 was reported to bind and unwind *hTERC* rG4s in the presence of ATP (172), which enables the formation of the P1 helix structure necessary for telomerase function (173). Another telomere-associated lncRNA, telomeric repeat-containing RNA (*TERRA*), was also reported to form rG4s and be involved in the regulation of telomerase activity and DNA telomere length (174–176). *TERRA* RNA is localized to chromosome ends in the nucleus (175), suggesting a link between *TERRA* rG4s and telomere function. For example, *TERRA* rG4s can interact with telomeric dG4s to form intermolecular hybrid G4 structures to suppress telomerase activity (177). In addition, *TERRA* rG4s recruit TRF2 (178), a protein that regulates the association of *TERRA* and telomeric DNA (179). As the rG4s in both *hTERC* and *TERRA* recruit specific sets of proteins to play important roles in telomere homeostasis and genome stability, future investigations may profit from rG4 mutation/deletion and protein knockdown/knockout experiments to establish the relationship of rG4s with telomere length and genomic instability, and to assess their impact on ageing and cancer development.

For non-telomere-associated lncRNAs (Figure 5), Matsumura *et al*. reported the lncRNA FLJ39051, also referred to as G-quadruplex-forming sequence containing lncRNA (*GSEC*), to be localized in the cytoplasm and upregulated in colorectal cancer cells (55). *GSEC* is a 753-nt-long lncRNA, in which an rG4 structure was detected at nucleotide positions 11–26. Using multiple assays, the rG4 in *GSEC* was shown to interact directly and specifically with DHX36, and antagonized the function of DHX36 by acting as a molecular decoy to block DHX36 from interacting with its target RNAs, such as the 3′-UTR of *PITX1* (55). Mutational and cellular analysis further verified that *GSEC* promoted the motility of colon cancer cells by suppressing the function of DHX36 via the rG4 structure in *GSEC* (55). It is currently unclear how many lncRNA rG4s there are, and future studies may discover novel lncRNA rG4–protein interactions, as well as their regulatory mechanism *in vivo*.

miRNA rG4s. rG4s play roles in every step of the miRNA biogenesis and function (Figure 5). For primary miRNA (pri-miRNA), Rouleau *et al*. predicted >9% of human pri-miRNAs contains rG4s (180). They identified rG4s near the Drosha cleavage site in human tumour suppressor pri-miRNAs such as pri-miR200c, pri-miR451a and pri-miR497, and reported that their formation *in vitro* was favoured in the presence of PhenDC3 (180). To highlight the pri-miRNA rG4s' gene regulatory roles in cells, they performed mutation analysis of the effect of rG4s on mature miRNA production (180). For pri-miR497, they found that disruption of the rG4s caused a decrease in the mature miRNA level, indicating the rG4 motifs could contribute to the regulation of pri-mRNA processing. However, this effect varied depending on the rG4 location and sequence context, as pri-miR451a showed an increase in the mature miRNA level when rG4 was mutated (180). Considering the tumour-related regulatory role of the above-mentioned miRNAs, the rG4 on pri-miRNA could be further investigated as a potential target for cancer therapeutics.

Several key studies have demonstrated that rG4s form in precursor miRNAs (pre-miRNAs) (71,181,182), and are in competition with the RNA stem-loop conformation. The stabilization of rG4 or destabilization of the stem-loop conformation in these pre-miRNAs by potassium ions, G4 ligands or lock nucleic acids (LNAs) was shown to inhibit Dicer recognition and processing, which in turn affected the mature miRNA production and downstream gene regulation (71,181-183). In addition, two studies found that the expression of mature miRNAs could be restored by treatment with TmPyP4, an rG4-destabilizing ligand, to unfold rG4 (182,184). Recently, it has been reported that a single-nucleotide polymorphism (185), an RBP (186), and a G4 helicase (187) were involved in shifting the equilibrium between the rG4 and stem-loop conformations, providing plausible biochemical mechanisms for further investigation.

For mature miRNA, initial studies primarily focused on the *in vitro* biophysical characterization of rG4 formation in mature miRNAs (188–190). In 2018, Chan *et al*. predicted >100 rG4s in human miRNAs, and focused on one of the candidates, miR765, which is associated with liver, prostate and bone cancers (56). They identified and verified the formation of the rG4 in miR-765 *in vitro*, and fur-

ther performed reporter gene assay in cells to illustrate the role of the miRNA rG4 in gene regulation, showing that its function can be amplified by the addition of the rG4-stabilizing ligand NMM (56). Other examples of mature miRNA rG4s have also been reported (56). However, like for pre-miRNA rG4s, these experiments were performed using reporter genes. Therefore, their regulatory role over native mRNA targets and the effect on resultant protein products remain to be elucidated.

piRNA rG4s. rG4s also play roles in piRNA biogenesis and function (Figure 5). For piRNA precursor transcripts, Vourekas *et al*. first analysed the CLIP data of MOV10L1, an RNA helicase important in piRNA biogenesis, and identified a significant enrichment in piRNA clusters containing more Gs in the bound areas (191). Using the G4-specific antibody BG4 in testis lysate, they reported an increased level of rG4 formation within piRNA precursor transcripts in a MOV10L1 knockout as compared with wildtype mice, suggesting that MOV10L1 interacted with and resolved rG4s (191). This rG4-specific binding and unwinding activity of MOV10L1 was recently validated (192). The overall effect of MOV10L1 helicase activity on RNA secondary structures such as rG4s is to unwind them to single-stranded piRNA precursor transcripts, enabling them to be cleaved by the endonuclease. For mature piRNA, Balaratnam *et al*. predicted a 5.5% rate of rG4-forming sequences in human piRNA (193). They demonstrated, using biophysical assays and RNase T1 structure mapping, that piR-48164 can form an rG4 (193). It was then shown, using a reporter gene system, that the formation of rG4 in piRNA prevented its binding to the PAZ domain of HIWI proteins and base-paring to mRNA targets, which in turn inhibited target gene silencing in cells (193). This implies that piRNA rG4s can affect gene silencing, and it is of interest to test these effects in native transcripts and to determine the molecular basis of this piRNA rG4-mediated regulatory mechanism.

tRNA rG4s. rG4s have been reported in tRNA-derived stress-induced RNAs (tiRNAs) (Figure 5). Ivanova first reported that the 5′-terminal oligoguanine (TOG) motif in 5′-tiRNA^{Ala} and 5′-tiRNA^{Cys} folded into an intermolecular rG4 in tiRNAs, and interacted with the cold shock domain (CSD) of translational repressor YB1 protein (194). Lyons *et al.* further revealed that the tetramolecular rG4 structure consisted of five G-tetrads formed by the TOG motif of four tiRNA^{Ala} copies, and the rG4 was in equilibrium with a hairpin structure (195). The biological relevance of this rG4 was then studied by using 7-deazaguanine (7-deazaG) to replace the guanine in 5′-tiRNA^{Ala}, and comparing the resulting construct with the wild-type 5′-tiRNA^{Ala}. Biochemical and functional analyses demonstrated that the wild-type 5′-tiRNA^{Ala}, but not the 7-deazaG-substituted 5′-tiRNA^{Ala}, interacted with YB1 protein, displaced the eIF4F complexes and triggered the formation of stress granules, which then led to the inhibition of translation (195). These results demonstrated that rG4s are important for tiRNA bioactivity, and as tiRNAs are stress-induced, the next step will be to examine the rG4s' role in diverse stress conditions and diseases.

rRNA rG4s. The study of rG4s in rRNA was very limited until recently (Figure 5). Mestre-Fos *et al.* first identified G-tracts in expansion segment (ES) tentacles in rRNA large ribosomal subunits (LSUs), and found them to be conserved in chordates. They verified the thermostable rG4 formation in ES7 and ES27 *in vitro*, and further identified RNA-binding proteins such as FIP1, FUS, DDX3 and hnRNP H, some of which had been previously reported as rG4-binding proteins (196). The same group also later reported additional rG4s on the surfaces of small ribosomal subunits (SSUs) *in vitro*, such as es3 and es6, which suggested that both the LSUs and the SSUs of the human ribosome contain multiple rG4s (197), which may assist in ribosome assembly or associated protein recruitment. One recent study showed that these rRNA rG4s form in cells and can control heme bioavailability (198). It is of interest to explore whether future RNA structure mapping or 3D structure determination techniques can capture a folded rG4 state in the ribosome, or the binding of an rG4 with associated ribosomal proteins *in vivo*.

Biological functions of viral, plant and bacterial rG4s

Besides being reported in mammalian systems as described above, rG4s also form and have functions in other species, including but not limited to viruses, plants and bacteria. The existence and roles of viral G4s have been excellently reviewed elsewhere recently (199,200), and therefore we highlight some representative viral rG4 examples and then focus on SARS-CoV-2 below. The study of rG4s in plants and bacteria, however, is still in its infancy, and below we highlight key findings from recent years and compare them to the results in mammalian systems.

rG4s in viruses. The prediction of PQSs in the genome of all known DNA and RNA viruses that can infect humans has been comprehensively performed recently, and the analysis showed that the occurrence and location of PQSs in the viral genome are orderly arranged and features characteristic of each virus family and species (57). Over the years, experimental studies have been carried out to verify the viral rG4 formation, to identify viral rG4 binding proteins and to develop antiviral G4 ligands (199,200). For DNA viruses, Epstein–Barr virus (EBV)-encoded protein EBV nuclear antigen 1 (EBNA1) was reported to be bind to rG4s through its linking region 1 (LR1) and LR2 (201). Besides, the EBNA1 mRNA can form rG4 itself and take part in the cis-acting regulation of viral mRNA translation (202). Similar to EBNA1, the mRNA of latency-associated nuclear antigen (LANA) in Kaposi's sarcoma-associated herpes virus (KSHV) was found to form rG4, which inhibited the translation of LANA (203). Recently, an rG4 named PQS18-1 in the 3′UTR of *IE180* gene in Pseudorabies virus (PRV) was reported to form *in vivo* and regulates the replication of PRV by enhancing the expression of IE180 (204). For RNA viruses, conserved rG4s have been predicted in the genomes and later experimentally validated in both positively strand virus like Zika, hepatitis C virus (HCV) and the (SARS-CoV) and negatively stranded virus like Ebola virus (205–207). In addition, rG4 ligands were reported to intererfere with the gene activity of different RNA

viruses (205–208). Similarly, two highly conserved rG4s that are located in the Nipah virus G and L genes respectively were recently demonstrated to interact with TMPyP4 (209). Besides that, the nucleocapsid protein (NCp7) from human immunodeficiency virus type 1 (HIV-1) was found as the first viral protein to unwind rG4s for the reverse transcription to proceed, and the effect was counteracted by BRACO-19 G4 ligand (210).

Severe acute respiratory syndrome coronavirus (SARS-CoV) is a highly contagious human viruses which was identified in 2003 (211). Since 2019, a novel enveloped RNA betacoronavirus named SARS-CoV-2 which caused coronavirus pneumonia (COVID-19) has become a severe threat to global public health (212), causing >2.6 million human death and >121 million infected so far. Comparing to SARS-CoV, fewer rG4s were predicted in SARS-CoV-2 (213). In the RNA genome of SARS-CoV-2, about 25 PQSs were found and localized in the open reading frames of ORF1 ab, spike (S), ORF3a, membrane (M) and nucleocapsid (N) genes (214,215). Two rG4s from ORF1ab and S respectively were found to interact with viral helicase nsp13, which play a role in facilitating the replication and transcription of the SARS-CoV-2 (214). Moreover, an rG4 named RG-1, which locates in the CDS region of SARS-CoV-2 nucleocapsid phosphoprotein (N) has been demonstrated to form in live cells, and can be further stabilized by G4 ligand PDS derivative (PDP). The expression of SARS-CoV-2 N were decreased both *in vitro* and *in vivo* by PDP treatment, which indicates that rG4 in SARS-CoV-2 may be a novel target for developing antiviral drugs against COVID-19 (216). It was reported that the SARS-unique domain (SUD), which is thought to be related to its pathogenicity in the SARS-CoV showed binding preference to rG4s (217). Recently, Zhang *et al.* found that SARS-CoV-2 possesses a similar SUD domain which has eight key conserved amino acid in numerous SARS-CoV-2 samples across the world. It indicates that the SARS-CoV-2 may have similar mechanism to SARS-CoV in gene regulation, and the interaction between rG4 and rG4 binding protein domain, e.g. SUD domain of nsp3, possibly leads to dimerization and the instability of transcription or the translation efficiency (218,219). Cumulated evidences have showed that rG4s have involved in various biological processes in viral system like transcription, reverse transcription, replication and translation, highlighting that rG4s are promising targets for antiviral drug development. The next step is to develop individual rG4-specific tools for precise viral rG4 targeting.

rG4s in plants. In *A. thaliana*, rG4s were predicted to be over-represented in protein-coding genes, and under-represented in ncRNAs such as rRNAs and tRNAs (220). In addition, the same study predicted >400 putative 5′-UTR rG4s in the transcriptome of *A. thaliana* (220). Kwok *et al.* conducted structural analysis and functional characterization of an rG4 located in the 5′-UTR of *ATR* mRNA in *A. thaliana*, and it was shown using a cell-based reporter gene system that the rG4 played an inhibitory role in translation (221). Comparative sequence analysis of the *ATR* transcript indicated that the rG4 motif was conserved among 14 plant species, and biophysical assays verified that

these rG4s folded into thermostable forms (221). Notably, the translational suppression role of the *ATR* 5′-UTR rG4 in *A. thaliana* resembled that reported for most 5′-UTR rG4s in mammalian systems (discussed above), suggesting that similar molecular mechanisms may apply to control 5′-UTR rG4-mediated translation processes across plants and mammals.

Cho *et al*. performed a comparative transcriptome analysis of the phloem–cambium region of three plant species and identified an uncharacterized zinc-finger (ZnF) protein referred to as JULG1. It was shown to specifically bind and induce a 5′-UTR rG4 of SMXL4 and SMXL5, which are key regulators of phloem formation, and suppress their translation. Reduced production of JULG1 restricted phloem differentiation and strikingly increased sink strength per seed (222). Notably, the mechanism underlying post-transcriptional regulation of phloem differentiation is exclusively conserved in vascular plants (222), suggesting that this rG4 and its binding protein can influence the plant development through post-transcriptional regulation. Interestingly, ZnF proteins have also been reported to bind to G4s in humans. For example, the ZnF protein CNBP/ZNF9 specifically binds to G-rich regions in the target mRNAs and promotes their translation, potentially by resolving rG4 structures in the target mRNAs (159). Taken together, these results show that ZnF-containing proteins could directly interact with G4s and may have diverse roles in G4-associated gene regulation in plant and mammalian species.

More recently, Yang *et al*. carried out transcriptome-wide mapping of rG4s in plants and identified hundreds of rG4s *in vivo* (97). In agreement with computational prediction, *in vivo* rG4 mapping data showed that two-quartet rG4s were identified more frequently than three-quartet rG4s, and in the coding region, the rG4s were enriched relative to UTRs (97). The results were largely consistent between *A. thaliana* and *O. sativa* (97). The authors further characterized one of the 3′-UTR rG4 candidates, *HIRD11*, which encodes a KS-type dehydrin, and demonstrated that the rG4 has an inhibitory role in translation. Using a plant root growth assay, the same study also showed a significant difference in phenotype between the candidate 3′-UTR rG4 and an rG4 mutant (97). In contrast to the earlier report by Guo *et al*. of the global unfolding of rG4s in mice, humans and yeast (51), rG4s in plants show a greater tendency to be folded *in vivo*, highlighting that the physiological environment may be quite different between plants and mammalian species. It is also possible that rG4s only fold in specific cell types or environments in mammalian systems. As future subjects of study, we propose dissecting the biochemical mechanisms that underlie the above-mentioned phenomena and identifying any novel rG4-binding proteins in plants, as well as performing transcriptome-wide rG4 mapping under diverse cellular conditions across different species.

rG4s in bacteria. Genome-wide predictions of PQSs have been performed in the genomes of *Escherichia coli* (223), *Deinococcus radiodurans* (224), *Xanthomonas* sp., *Nostoc* sp. (225) and *Mycobacterium tuberculosis* (226), and experimental-based DNA G4-seq was performed in *E. coli* to identify dG4s *in vitro* (227). In 2016, Guo *et al*. reported

that rG4s were depleted in bacteria, and proposed that unlike in mammalian cells, which have evolved mechanisms to deal with rG4s (e.g. rG4-unwinding helicases and single-stranded G-rich RNA-binding proteins), rG4s may be evolutionarily selected against in bacteria (51). In addition, they demonstrated that artificial insertion of an rG4 into a bacterial reporter gene system caused a reduction in bacterial growth, as well as aberrant translation products (51).

Recently, Shao *et al*. used QUMA-1, an rG4-specific fluorescent probe, to demonstrate the presence of rG4s *in vitro* and *in vivo* in 10 diverse model bacterial species (96). Moreover, through rG4-seq, they obtained 168 and 161 *in vitro* rG4 sites in *E. coli* and *Pseudomonas Aeruginosa*, respectively, which were all distributed in the CDS regions (96). Using biophysical assays and a cell-based reporter gene system, they verified the formation of rG4 in the coding region of *hemL*, a metabolic gene, and demonstrated an rG4-dependent regulatory system in *E. coli* (96). Similarly, they revealed the function of rG4 in the coding region of *bswR*, a virulence gene, in *P. aeruginosa*, and illustrated by mutagenesis and phenotypic assays that the presence of rG4 affected the bacterial flagella and biofilm formation of *P. aeruginosa* by upregulating the expression of *bswR* (96).

Research thus far into bacterial rG4s (51,96,228,229), reveals that they display both positive and negative regulatory roles in translation. This difference may be dependent on rG4 stability, rG4 position and the flanking sequence context of the transcript-of-interest. Currently, studies of naturally occurring rG4s in bacteria are very limited. Further research is warranted to identify new bacterial rG4s using the tools developed in the mammalian and plant systems mentioned above, and explore and characterize their functions in different bacterial strains and stress conditions, which will help establish the general roles and effects of rG4s in bacterial systems, and allow a comprehensive comparison with the mammalian, viral and plant systems.

Current challenges and future perspectives of studying rG4 biological functions

Our understanding of the function of rG4s has expanded to numerous classes of RNA and diverse living organisms (Table 1). However, challenges remain with regard to establishing the relationship of rG4 motifs with other cis-regulatory elements on RNA to fine-tune gene activity, revealing the direct interactions of rG4s with the plethora of biomolecules that prevail in cellular milieu, elucidating the biological mechanisms and consequences of many rG4-mediated cellular processes, and developing rG4-specific tools to selectively target functionally and pathologically important rG4s for different applications.

Regarding the interplay between rG4s and cis-regulatory elements, several pilot studies have illustrated that rG4 formation can control the accessibility of cis-regulatory elements on a transcript-specific basis (141,144). In addition, transcriptome-wide rG4 mapping studies have shown strong positional correlations between rG4s and polyadenylation signals and miRNA target sites in mRNA (50), suggesting that rG4s and these cis-regulatory elements might coordinate to regulate RNA metabolism and gene expression on a global scale. It remains to be deciphered whether

Table 1. Representative rG4 functions and examples reported to date

Region	Representative functions	Key examples	References
5′UTR rG4	Suppress or promote translation	*FMR1, NRAS, Zic-1, ESRα, CCND3, TRF2, ADAM10, MT3, BCL-2, TGFβ2, FGF-2, hVEGF, BAG-1, ATR, SMXL4/5*	(16,41,117–121, 123,124,126,127, 132–136,221, 222)
3′UTR rG4	Suppress translation, regulate alternative polyadenylation, interfere with miRNA targeting, control mRNA localization	*PIM1, APP, HIRD11, LRP5, FXR1, FADS2, PSD-95, CaMKIIa*	(97,139–141, 145,148)
ORF rG4	Regulate translation, control alternative RNA splicing	*APP, MLL1, MLL4, hemL, bswR, hTERT, FMR1, BACE1, TP53, CD44*	(13,96,157,158, 163,164)
lncRNA rG4	Regulate telomerase activity, telomere homeostasis and genome stability, antagonize DHX36 helicase function	*hTERC, TERRA, GSEC*	(55,172,177–179)
miRNA rG4	Regulate miRNA maturation and post-transcriptional regulation, control miRNA targeting	*pri-mir200c, pri-mir451a, pri-mir497, pre-miR92b, pre-let7e, pre-miR149, miR765*	(56,71,145,181, 182)
piRNA rG4	Regulate piRNA metabolism and control piRNA targeting	*piR-48164*	(193)
tRNA rG4	Inhibit translation	*5′-tiRNA^{Ala}, 5′-tiRNA^{Cys}*	(194,195)
rRNA rG4	Mediate ribosome assembly and associated protein recruitment	*ES7, ES27, es3, es6*	(196,197)

such dynamic regulation is linked to specific cellular status and cell type. Moreover, given that there are hundreds of naturally occurring RNA epigenetic marks (230), several of which are on guanines (e.g. internal m7G), their impact on rG4 formation and biological role in cells will be worthy topics for future investigation. One initial study has shown that internal m7G in pre-miRNA precludes rG4 formation, resulting in mature miRNA regulation (231).

Regarding the rG4 interactors in cells, a dozen G4-binding proteins have been identified over the years (38,232); however, our current understanding of the protein sequence and structural requirements for rG4 binding is still limited, making it challenging to fully appreciate the complex network mediated by rG4–protein interactions. The RGG and ZnF motifs have been identified in many G4-binding proteins (233,234), and lately, a new protein motif was also reported to interact with G4 specifically (235). With accumulating information about G4-binding proteins as well as the better availability of high-resolution G4–protein complexes, it may be possible to predict new rG4-binding proteins using primarily protein sequence/structure information. Moreover, with the growing list of cell-based RNA-centric methods to detect transcriptome-wide RNA–protein interactions (236,237), we anticipate that with some degree of adaptation, these approaches can be readily applied to reveal rG4–protein interactions *in vivo*, which will greatly expand our repertoire of known rG4-binding proteins. Likewise, many existing techniques for detecting RNA–RNA and RNA–chromatin interactions (236,237) can also likely be repurposed for the detection of potential rG4–RNA and rG4–chromatin interactions. Besides those biomolecules mentioned above, rG4s have also been reported to interact with small molecules in cells, such as heme (198,238) and polyamines (239,240), which necessitates new approaches to identify rG4-interacting small molecules/metabolites *in vivo*.

Regarding the biological impact of rG4s and/or rG4-binding partners in cellular processes, reports have uncovered novel functions of rG4s that warrant further in-depth investigation. First, rG4s can facilitate RNA accumulation and phase separation in human and plant cells (241,242), and it will be interesting to see if this has any link with the assembly and function of subcellular compartments, such as stress granules and p-bodies, and under what conditions these membraneless organelles and processes can be regulated, for example by the expression level of rG4 helicases (102). Second, rG4-binding proteins and G-rich-binding proteins such as DHX36 and hnRNP F/H mediate translation by resolving rG4s and keeping them in single-stranded G-rich conformations, and this was linked to genomic instability and therapy resistance in glioblastoma (160). Elsewhere, in mice, DHX36 has been reported to unwind rG4s in pre-miRNA26a to facilitate mature miRNA26a production under healthy conditions but not in obese mice (187). This rG4-mediated regulation was shown to contribute to hepatic insulin resistance and the dysregulation of liver metabolism (187). Future studies may also focus on the biological consequence of rG4s or rG4-binding partners in other diseases and physiological changes, such as ageing, cell differentiation and cell development (243). Last, rG4s have been found in approximately 30% of dendritic transcripts, such as CaMKIIa and PSD95, which are localized in cortical neurites (148). Moreover, several key neuronal proteins are G4-specific, such as FMRP (156) and FUS (244), implying that rG4s may be linked to gene regulation and function in neurological diseases.

Regarding rG4-targeting tools, several hundred G4-selective ligands have been developed so far (245); however, given the structural similarity of dG4s and rG4s, only a few ligands have been reported to possess some degree of specificity towards rG4s, such as cPDS (36) and QUMA-1 (54). Few studies have coupled chemically functionalized or anchoring guanine with antisense oligonucleotides or peptides to achieve greater target specificity (246–248). New approaches will be needed to enable the selective targeting of rG4 classes (over dG4s and non-G4s) or even of individual rG4s for different applications. Recently, two L-RNA aptamers were developed to target rG4 structure, one of which prefers to bind generally to many rG4s (249), whereas the

other prefers to bind specifically to individual rG4s (250), suggesting that the development of universal rG4 and selective rG4 binders is feasible. For these above-mentioned approaches, the cell permeability and associated cellular cytotoxicity need to be studied in more detail across different cell lines and species in the future. In addition, their specific target binding in native intracellular transcript contexts and their ability to interfere with different biological processes need to be verified to illustrate their utility. As rG4s have been associated with diseases (42,116,251), rG4-targeting tools will help decipher and manipulate the biochemical mechanism of these rG4-associated gene regulations and functions.

CONCLUSION

The study of rG4 has progressed significantly over the past few years, and the advancement in technology has naturally led to important new methods and basic discoveries in rG4 biology. We anticipate that the next key challenge will be to further investigate where, when and how rG4s are being formed and dynamically controlled by cis- and/or trans-regulatory elements in cells, as well as the specific underlying rG4-mediated biochemical mechanism and functional consequences in different organisms (see future perspective sections above). Collectively, these fundamental insights will facilitate us to explore the origin and evolution of rG4s, and enable us to decipher the missing link between rG4s and diseases. The outcomes of these will undoubtedly be highly valuable in the design and development of tools to target rG4s and rG4-associated pathways for diverse biological, biomedical and biotechnological applications. We look forward with enthusiasm to new suite of methodologies and biological breakthroughs to be unraveled in the near future.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Bang,I. (1910) Untersuchungen über die Guanylsäure. *Biochem. Z.*, **26**, 293–311.
2. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. USA*, **48**, 2013–2018.
3. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
4. Sundquist,W.I. and Klug,A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
5. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871–880.
6. Zahler,A.M., Williamson,J.R., Cech,T.R. and Prescott,D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718–720.
7. Kim,J., Cheong,C. and Moore,P.B. (1991) Tetramerization of an RNA oligonucleotide containing a GGGG sequence. *Nature*, **351**, 331–332.
8. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Pluckthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. USA*, **98**, 8572–8577.
9. Paeschke,K., Simonsson,T., Postberg,J., Rhodes,D. and Lipps,H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.*, **12**, 847–854.
10. Darnell,J.C., Jensen,K.B., Jin,P., Brown,V., Warren,S.T. and Darnell,R.B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*, **107**, 489–499.
11. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. USA*, **99**, 11593–11598.
12. Duquette,M.L., Handa,P., Vincent,J.A., Taylor,A.F. and Maizels,N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.*, **18**, 1618–1629.
13. Gomez,D., Lemarteleur,T., Lacroix,L., Mailliet,P., Mergny,J.L. and Riou,J.F. (2004) Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.*, **32**, 371–379.
14. Neidle,S. and Balasubramanian,S. (2006) In: *Quadruplex nucleic acids*. Royal Society of Chemistry. Cambridge, UK.
15. Wieland,M. and Hartig,J.S. (2007) RNA quadruplex-based modulation of gene expression. *Chem. Biol.*, **14**, 757–763.
16. Kumari,S., Bugaut,A., Huppert,J.L. and Balasubramanian,S. (2007) An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218–221.
17. Lipps,H.J. and Rhodes,D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, **19**, 414–422.
18. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
19. Sun,D., Thompson,B., Cathers,B.E., Salazar,M., Kerwin,S.M., Trent,J.O., Jenkins,T.C., Neidle,S. and Hurley,L.H. (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.*, **40**, 2113–2116.
20. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
21. Huang,H., Suslov,N.B., Li,N.S., Shelke,S.A., Evans,M.E., Koldobskaya,Y., Rice,P.A. and Piccirilli,J.A. (2014) A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.*, **10**, 686–691.
22. Warner,K.D., Chen,M.C., Song,W., Strack,R.L., Thorn,A., Jaffrey,S.R. and Ferre-D'Amare,A.R. (2014) Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.*, **21**, 658–663.

23. Xiao,C.D., Shibata,T., Yamamoto,Y. and Xu,Y. (2018) An intramolecular antiparallel G-quadruplex formed by human telomere RNA. *Chem. Commun. (Camb.)*, **54**, 3944–3946.

24. Tang,C.F. and Shafer,R.H. (2006) Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. *J. Am. Chem. Soc.*, **128**, 5966–5973.

25. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.

26. Huppert,J.L. (2008) Hunting G-quadruplexes. *Biochimie*, **90**, 1140–1148.

27. Huppert,J.L., Bugaut,A., Kumari,S. and Balasubramanian,S. (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.*, **36**, 6260–6268.

28. Rhodes,D. and Lipps,H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.

29. Bugaut,A. and Balasubramanian,S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry*, **47**, 689–697.

30. Arora,A., Nair,D.R. and Maiti,S. (2009) Effect of flanking bases on quadruplex stability and Watson-Crick duplex competition. *FEBS J.*, **276**, 3628–3640.

31. Joachimi,A., Benz,A. and Hartig,J.S. (2009) A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg. Med. Chem.*, **17**, 6811–6815.

32. Zhang,A.Y., Bugaut,A. and Balasubramanian,S. (2011) A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry*, **50**, 7251–7258.

33. Zhang,D.H., Fujimoto,T., Saxena,S., Yu,H.Q., Miyoshi,D. and Sugimoto,N. (2010) Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry*, **49**, 4554–4563.

34. Bhattacharyya,D., Mirihana Arachchilage,G. and Basu,S. (2016) Metal cations in G-quadruplex folding and stability. *Front. Chem.*, **4**, 38.

35. Collie,G., Reszka,A.P., Haider,S.M., Gabelica,V., Parkinson,G.N. and Neidle,S. (2009) Selectivity in small molecule binding to human telomeric RNA and DNA quadruplexes. *Chem. Commun. (Camb.)*, **48**, 7482–7484.

36. Di Antonio,M., Biffi,G., Mariani,A., Raiber,E.A., Rodriguez,R. and Balasubramanian,S. (2012) Selective RNA versus DNA G-quadruplex targeting by in situ click chemistry. *Angew. Chem. Int. Ed.*, **51**, 11073–11078.

37. Guiset Miserachs,H., Donghi,D., Borner,R., Johannsen,S. and Sigel,R.K. (2016) Distinct differences in metal ion specificity of RNA and DNA G-quadruplexes. *J. Biol. Inorg. Chem.*, **21**, 975–986.

38. Mendoza,O., Bourdoncle,A., Boule,J.B., Brosh,R.M. Jr and Mergny,J.L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.*, **44**, 1989–2006.

39. Fay,M.M., Lyons,S.M. and Ivanov,P. (2017) RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.*, **429**, 2127–2147.

40. Fang,G. and Cech,T.R. (1993) The beta subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell*, **74**, 875–885.

41. Schaeffer,C., Bardoni,B., Mandel,J.L., Ehresmann,B., Ehresmann,C. and Moine,H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.*, **20**, 4803–4813.

42. Kharel,P., Balaratnam,S., Beals,N. and Basu,S. (2020) The role of RNA G-quadruplexes in human diseases and therapeutic strategies. *Wiley Interdisc. Rev. RNA*, **11**, e1568.

43. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.

44. Biffi,G., Di Antonio,M., Tannahill,D. and Balasubramanian,S. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.*, **6**, 75–80.

45. Shivalingam,A., Izquierdo,M.A., Marois,A.L., Vysniauskas,A., Suhling,K., Kuimova,M.K. and Vilar,R. (2015) The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.*, **6**, 8178.

46. Laguerre,A., Hukezalie,K., Winckler,P., Katranji,F., Chanteloup,G., Pirrotta,M., Perrier-Cornet,J.M., Wong,J.M. and Monchaud,D. (2015) Visualization of RNA-quadruplexes in live cells. *J. Am. Chem. Soc.*, **137**, 8521–8525.

47. Di Antonio,M., Ponjavic,A., Radzevicius,A., Ranasinghe,R.T., Catalano,M., Zhang,X., Shen,J., Needham,L.M., Lee,S.F., Klenerman,D. *et al.* (2020) Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.*, **12**, 832–837.

48. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.

49. Hansel-Hertsch,R., Beraldi,D., Lensing,S.V., Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M. *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.

50. Kwok,C.K., Marsico,G., Sahakyan,A.B., Chambers,V.S. and Balasubramanian,S. (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods*, **13**, 841–844.

51. Guo,J.U. and Bartel,D.P. (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, **353**, aaf5371.

52. Kwok,C.K., Marsico,G. and Balasubramanian,S. (2018) Detecting RNA G-quadruplexes (rG4s) in the transcriptome. *Cold Spring Harb. Perspect.*, **10**, a032284.

53. Yang,S.Y., Lejault,P., Chevrier,S., Boidot,R., Robertson,A.G., Wong,J.M.Y. and Monchaud,D. (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.*, **9**, 4730.

54. Chen,X.C., Chen,S.B., Dai,J., Yuan,J.H., Ou,T.M., Huang,Z.S. and Tan,J.H. (2018) Tracking the dynamic folding and unfolding of RNA G-quadruplexes in live cells. *Angew. Chem. Int. Ed.*, **57**, 4702–4706.

55. Matsumura,K., Kawasaki,Y., Miyamoto,M., Kamoshida,Y., Nakamura,J., Negishi,L., Suda,S. and Akiyama,T. (2017) The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene*, **36**, 1191–1199.

56. Chan,K.L., Peng,B., Umar,M.I., Chan,C.Y., Sahakyan,A.B., Le,M.T.N. and Kwok,C.K. (2018) Structural analysis reveals the formation and role of RNA G-quadruplex structures in human mature microRNAs. *Chem. Commun. (Camb.)*, **54**, 10878–10881.

57. Lavezzo,E., Berselli,M., Frasson,I., Perrone,R., Palu,G., Brazzale,A.R., Richter,S.N. and Toppo,S. (2018) G-quadruplex forming sequences in the genome of all known human viruses: a comprehensive guide. *PLoS Comput. Biol.*, **14**, e1006675.

58. Lee,D.S.M., Ghanem,L.R. and Barash,Y. (2020) Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. *Nat. Commun.*, **11**, 527.

59. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.

60. Mergny,J.L., Lacroix,L., Teulade-Fichou,M.P., Hounsou,C., Guittat,L., Hoarau,M., Arimondo,P.B., Vigneron,J.P., Lehn,J.M., Riou,J.F. *et al.* (2001) Telomerase inhibitors based on quadruplex ligands selected by a fluorescence assay. *Proc. Natl. Acad. Sci. USA*, **98**, 3062–3067.

61. Paramasivan,S., Rujan,I. and Bolton,P.H. (2007) Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods*, **43**, 324–331.

62. Henderson,E., Hardin,C.C., Walk,S.K., Tinoco,I. Jr and Blackburn,E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, **51**, 899–908.

63. Kwok,C.K., Sherlock,M.E. and Bevilacqua,P.C. (2013) Effect of loop sequence and loop length on the intrinsic fluorescence of G-quadruplexes. *Biochemistry*, **52**, 3019–3021.

64. Webba da Silva,M. (2007) NMR methods for studying quadruplex nucleic acids. *Methods*, **43**, 264–277.

65. Vorlickova,M., Kejnovska,I., Sagi,J., Renciuk,D., Bednarova,K., Motlova,J. and Kypr,J. (2012) Circular dichroism and guanine quadruplexes. *Methods*, **57**, 64–75.

66. Merino,E.J., Wilkinson,K.A., Coughlan,J.L. and Weeks,K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.

67. Harrison,G.P., Mayo,M.S., Hunter,E. and Lever,A.M. (1998) Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5′ and 3′ of the catalytic site. *Nucleic Acids Res.*, **26**, 3433–3442.

68. Ziehler,W.A. and Engelke,D.R. (2001) Probing RNA structure with chemical reagents and enzymes. *Curr. Protoc. Nucleic Acid Chem.*, doi:10.1002/0471142700.nc0601s00.

69. Weng,X., Gong,J., Chen,Y., Wu,T., Wang,F., Yang,S., Yuan,Y., Luo,G., Chen,K., Hu,L. *et al.* (2020) Keth-seq for transcriptome-wide RNA structure mapping. *Nat. Chem. Biol.*, **16**, 489–492.

70. Kwok,C.K. and Balasubramanian,S. (2015) Targeted detection of G-quadruplexes in cellular RNAs. *Angew. Chem. Int. Ed.*, **54**, 6751–6754.

71. Kwok,C.K., Sahakyan,A.B. and Balasubramanian,S. (2016) Structural analysis using SHALiPE to reveal RNA G-quadruplex formation in human precursor microRNA. *Angew. Chem. Int. Ed.*, **55**, 8958–8961.

72. Haudecoeur,R., Stefan,L., Denat,F. and Monchaud,D. (2013) A model of smart G-quadruplex ligand. *J. Am. Chem. Soc.*, **135**, 550–553.

73. Chilka,P., Desai,N. and Datta,B. (2019) Small molecule fluorescent probes for G-quadruplex visualization as potential cancer theranostic agents. *Molecules*, **24**, 752.

74. Renaud de la Faverie,A., Guedin,A., Bedrat,A., Yatsunyk,L.A. and Mergny,J.L. (2014) Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Res.*, **42**, e65.

75. Laguerre,A., Stefan,L., Larrouy,M., Genest,D., Novotna,J., Pirrotta,M. and Monchaud,D. (2014) A twice-as-smart synthetic G-quartet: PyroTASQ is both a smart quadruplex ligand and a smart fluorescent probe. *J. Am. Chem. Soc.*, **136**, 12406–12414.

76. Muller,S., Kumari,S., Rodriguez,R. and Balasubramanian,S. (2010) Small-molecule-mediated G-quadruplex isolation from human cells. *Nat. Chem.*, **2**, 1095–1098.

77. Rodriguez,R., Muller,S., Yeoman,J.A., Trentesaux,C., Riou,J.F. and Balasubramanian,S. (2008) A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.*, **130**, 15758–15759.

78. Rodriguez,R., Miller,K.M., Forment,J.V., Bradshaw,C.R., Nikan,M., Britton,S., Oelschlaegel,T., Xhemalce,B., Balasubramanian,S. and Jackson,S.P. (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.*, **8**, 301–310.

79. Guedin,A., Alberti,P. and Mergny,J.L. (2009) Stability of intramolecular quadruplexes: sequence effects in the central loop. *Nucleic Acids Res.*, **37**, 5559–5567.

80. Guedin,A., Gros,J., Alberti,P. and Mergny,J.L. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.*, **38**, 7858–7868.

81. Pandey,S., Agarwala,P. and Maiti,S. (2013) Effect of loops and G-quartets on the stability of RNA G-quadruplexes. *J. Phys. Chem. B*, **117**, 6896–6905.

82. Mukundan,V.T. and Phan,A.T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.*, **135**, 5017–5028.

83. Pan,B., Xiong,Y., Shi,K. and Sundaralingam,M. (2003) Crystal structure of a bulged RNA tetraplex at 1.1 a resolution: implications for a novel binding site in RNA tetraplex. *Structure*, **11**, 1423–1430.

84. Jodoin,R., Bauer,L., Garant,J.M., Mahdi Laaref,A., Phaneuf,F. and Perreault,J.P. (2014) The folding of 5′-UTR human G-quadruplexes possessing a long central loop. *RNA*, **20**, 1129–1141.

85. Lorenz,R., Bernhart,S.H., Externbrink,F., Qin,J., Höner zu Siederdissen,C., Amman,F., Hofacker,I.L. and Stadler,P.F. (2012) In: Souto,M.C.P.D. and Kann,M.G. (eds). *Advances in Bioinformatics and Computational Biology*. Springer, Berlin Heidelberg, Vol. **7409**, pp. 49–60.

86. Bugaut,A., Murat,P. and Balasubramanian,S. (2012) An RNA hairpin to G-quadruplex conformational transition. *J. Am. Chem. Soc.*, **134**, 19953–19956.

87. Beaudoin,J.D., Jodoin,R. and Perreault,J.P. (2014) New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Res.*, **42**, 1209–1223.

88. Bedrat,A., Lacroix,L. and Mergny,J.L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.*, **44**, 1746–1759.

89. Lombardi,E.P. and Londono-Vallejo,A. (2020) A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.*, **48**, 1603.

90. Kikin,O., D'Antonio,L. and Bagga,P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, W676–682.

91. Lorenz,R., Bernhart,S.H., Qin,J., Honer zu Siederdissen,C., Tanzer,A., Amman,F., Hofacker,I.L. and Stadler,P.F. (2013) 2D meets 4G: G-quadruplexes in RNA secondary structure prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **10**, 832–844.

92. Hon,J., Martinek,T., Zendulka,J. and Lexa,M. (2017) pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics*, **33**, 3373–3379.

93. Garant,J.M., Luce,M.J., Scott,M.S. and Perreault,J.P. (2015) G4RNA: an RNA G-quadruplex database. *Database (Oxford)*, **2015**, bav059.

94. Brazda,V., Kolomaznik,J., Lysek,J., Bartas,M., Fojta,M., Stastny,J. and Mergny,J.L. (2019) G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics*, **35**, 3493–3495.

95. Labudova,D., Hon,J. and Lexa,M. (2020) pqsfinder web: G-quadruplex prediction using optimized pqsfinder algorithm. *Bioinformatics*, **36**, 2584–2586.

96. Shao,X., Zhang,W., Umar,M.I., Wong,H.Y., Seng,Z., Xie,Y., Zhang,Y., Yang,L., Kwok,C.K. and Deng,X. (2020) RNA G-quadruplex structures mediate gene regulation in bacteria. *mBio*, **11**, e02926-19.

97. Yang,X., Cheema,J., Zhang,Y., Deng,H., Duncan,S., Umar,M.I., Zhao,J., Liu,Q., Cao,X., Kwok,C.K. *et al.* (2020) RNA G-quadruplex structures exist and function in vivo in plants. *Genome Biol.*, **21**, 226.

98. Chow,E.Y., Lyu,K., Kwok,C.K. and Chan,T.F. (2020) rG4-seeker enables high-confidence identification of novel and non-canonical rG4 motifs from rG4-seq experiments. *RNA Biol.*, **17**, 903–917.

99. Nakano,S., Miyoshi,D. and Sugimoto,N. (2014) Effects of molecular crowding on the structures, interactions, and functions of nucleic acids. *Chem. Rev.*, **114**, 2733–2758.

100. Matsumoto,S., Tateishi-Karimata,H., Takahashi,S., Ohyama,T. and Sugimoto,N. (2020) Effect of Molecular Crowding on the Stability of RNA G-Quadruplexes with Various Numbers of Quartets and Lengths of Loops. *Biochemistry*, **59**, 2640–2649.

101. Murat,P., Marsico,G., Herdy,B., Ghanbarian,A.T., Portella,G. and Balasubramanian,S. (2018) RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.*, **19**, 229.

102. Sauer,M., Juranek,S.A., Marks,J., De Magis,A., Kazemier,H.G., Hilbig,D., Benhalevy,D., Wang,X., Hafner,M. and Paeschke,K. (2019) DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nat. Commun.*, **10**, 2421.

103. Yeung,P.Y., Zhao,J., Chow,E.Y., Mou,X., Hong,H., Chen,L., Chan,T.F. and Kwok,C.K. (2019) Systematic evaluation and optimization of the experimental steps in RNA G-quadruplex structure sequencing. *Sci. Rep.*, **9**, 8091.

104. Sahakyan,A.B., Chambers,V.S., Marsico,G., Santner,T., Di Antonio,M. and Balasubramanian,S. (2017) Machine learning model for sequence-driven DNA G-quadruplex formation. *Sci. Rep.*, **7**, 14535.

105. Garant,J.M., Perreault,J.P. and Scott,M.S. (2017) Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics*, **33**, 3532–3537.

106. Garant,J.M., Perreault,J.P. and Scott,M.S. (2018) G4RNA screener web server: User focused interface for RNA G-quadruplex prediction. *Biochimie*, **151**, 115–118.

107. Chen,M.C., Tippana,R., Demeshkina,N.A., Murat,P., Balasubramanian,S., Myong,S. and Ferre-D'Amare,A.R. (2018) Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, **558**, 465–469.

108. Wu,G., Xing,Z., Tran,E.J. and Yang,D. (2019) DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation. *Proc. Natl. Acad. Sci. USA*, **116**, 20453–20461.

109. Tippana,R., Chen,M.C., Demeshkina,N.A., Ferre-D'Amare,A.R. and Myong,S. (2019) RNA G-quadruplex is resolved by repetitive and ATP-dependent mechanism of DHX36. *Nat. Commun.*, **10**, 1855.

110. Fleming,A.M., Nguyen,N.L.B. and Burrows,C.J. (2019) Colocalization of m(6)A and G-Quadruplex-Forming Sequences in Viral RNA (HIV, Zika, Hepatitis B, and SV40) Suggests Topological Control of Adenosine N (6)-Methylation. *ACS Cent. Sci.*, **5**, 218–228.

111. Jara-Espejo,M., Fleming,A.M. and Burrows,C.J. (2020) Potential G-Quadruplex Forming Sequences and N(6)-Methyladenosine Colocalize at Human Pre-mRNA Intron Splice Sites. *ACS Chem. Biol.*, **15**, 1292–1300.

112. Helm,M. and Motorin,Y. (2017) Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.*, **18**, 275–291.

113. Anreiter,I., Mir,Q., Simpson,J.T., Janga,S.C. and Soller,M. (2021) New Twists in Detecting mRNA Modification Dynamics. *Trends Biotechnol.*, **39**, 72–89.

114. Millevoi,S., Moine,H. and Vagner,S. (2012) G-quadruplexes in RNA biology. *Wiley Interdiscip. Rev. RNA*, **3**, 495–507.

115. Mattick,J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.*, **2**, 986–991.

116. Simone,R., Fratta,P., Neidle,S., Parkinson,G.N. and Isaacs,A.M. (2015) G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Lett.*, **589**, 1653–1668.

117. Arora,A., Dutkiewicz,M., Scaria,V., Hariharan,M., Maiti,S. and Kurreck,J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA*, **14**, 1290–1296.

118. Balkwill,G.D., Derecka,K., Garner,T.P., Hodgman,C., Flint,A.P. and Searle,M.S. (2009) Repression of translation of human estrogen receptor alpha by G-quadruplex formation. *Biochemistry*, **48**, 11487–11495.

119. Morris,M.J. and Basu,S. (2009) An unusually stable G-quadruplex within the 5′-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry*, **48**, 5313–5319.

120. Gomez,D., Guedin,A., Mergny,J.L., Salles,B., Riou,J.F., Teulade-Fichou,M.P. and Calsou,P. (2010) A G-quadruplex structure within the 5′-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res.*, **38**, 7187–7198.

121. Shahid,R., Bugaut,A. and Balasubramanian,S. (2010) The BCL-2 5′ untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry*, **49**, 8300–8306.

122. Beaudoin,J.D. and Perreault,J.P. (2010) 5′-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.*, **38**, 7022–7036.

123. Lammich,S., Kamp,F., Wagner,J., Nuscher,B., Zilow,S., Ludwig,A.K., Willem,M. and Haass,C. (2011) Translational repression of the disintegrin and metalloprotease ADAM10 by a stable G-quadruplex secondary structure in its 5′-untranslated region. *J. Biol. Chem.*, **286**, 45063–45072.

124. Weng,H.Y., Huang,H.L., Zhao,P.P., Zhou,H. and Qu,L.H. (2012) Translational repression of cyclin D3 by a stable G-quadruplex in its 5′ UTR: implications for cell cycle regulation. *RNA Biol.*, **9**, 1099–1109.

125. Jodoin,R. and Perreault,J.P. (2018) G-quadruplexes formation in the 5′UTRs of mRNAs associated with colorectal cancer pathways. *PLoS One*, **13**, e0208363.

126. Agarwala,P., Pandey,S., Mapa,K. and Maiti,S. (2013) The G-quadruplex augments translation in the 5′ untranslated region of transforming growth factor beta2. *Biochemistry*, **52**, 1528–1538.

127. Khateb,S., Weisman-Shomer,P., Hershco-Shani,I., Ludwig,A.L. and Fry,M. (2007) The tetraplex (CGG)n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res.*, **35**, 5775–5788.

128. Kumari,S., Bugaut,A. and Balasubramanian,S. (2008) Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5′ UTR of the NRAS proto-oncogene. *Biochemistry*, **47**, 12664–12669.

129. Wolfe,A.L., Singh,K., Zhong,Y., Drewe,P., Rajasekhar,V.K., Sanghvi,V.R., Mavrakis,K.J., Jiang,M., Roderick,J.E., Van der Meulen,J. *et al.* (2014) RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature*, **513**, 65–70.

130. Serikawa,T., Spanos,C., von Hacht,A., Budisa,N., Rappsilber,J. and Kurreck,J. (2018) Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5′ UTR of tumor-associated mRNAs. *Biochimie*, **144**, 169–184.

131. Hellen,C.U. and Sarnow,P. (2001) Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.*, **15**, 1593–1612.

132. Bonnal,S., Schaeffer,C., Creancier,L., Clamens,S., Moine,H., Prats,A.C. and Vagner,S. (2003) A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J. Biol. Chem.*, **278**, 39330–39336.

133. Morris,M.J., Negishi,Y., Pazsint,C., Schonhoft,J.D. and Basu,S. (2010) An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *J. Am. Chem. Soc.*, **132**, 17831–17839.

134. Cammas,A., Dubrac,A., Morel,B., Lamaa,A., Touriol,C., Teulade-Fichou,M.P., Prats,H. and Millevoi,S. (2015) Stabilization of the G-quadruplex at the VEGF IRES represses cap-independent translation. *RNA Biol.*, **12**, 320–329.

135. Jodoin,R., Carrier,J.C., Rivard,N., Bisaillon,M. and Perreault,J.P. (2019) G-quadruplex located in the 5′UTR of the BAG-1 mRNA affects both its cap-dependent and cap-independent translation through global secondary structure maintenance. *Nucleic Acids Res.*, **47**, 10247–10266.

136. Bhattacharyya,D., Diamond,P. and Basu,S. (2015) An Independently folding RNA G-quadruplex domain directly recruits the 40S ribosomal subunit. *Biochemistry*, **54**, 1879–1885.

137. Moerke,N.J., Aktas,H., Chen,H., Cantel,S., Reibarkh,M.Y., Fahmy,A., Gross,J.D., Degterev,A., Yuan,J., Chorev,M. *et al.* (2007) Small-molecule inhibition of the interaction between the translation initiation factors eIF4E and eIF4G. *Cell*, **128**, 257–267.

138. Beretta,L., Gingras,A.C., Svitkin,Y.V., Hall,M.N. and Sonenberg,N. (1996) Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation. *EMBO J.*, **15**, 658–664.

139. Arora,A. and Suess,B. (2011) An RNA G-quadruplex in the 3′ UTR of the proto-oncogene PIM1 represses translation. *RNA Biol.*, **8**, 802–805.

140. Lyu,K., Chen,S.B., Chan,C.Y., Tan,J.H. and Kwok,C.K. (2019) Structural analysis and cellular visualization of APP RNA G-quadruplex. *Chem. Sci.*, **10**, 11095–11102.

141. Beaudoin,J.D. and Perreault,J.P. (2013) Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.*, **41**, 5898–5911.

142. Muddashetty,R.S., Kelic,S., Gross,C., Xu,M. and Bassell,G.J. (2007) Dysregulated metabotropic glutamate receptor-dependent translation of AMPA receptor and postsynaptic density-95 mRNAs at synapses in a mouse model of fragile X syndrome. *J. Neurosci.*, **27**, 5338–5348.

143. Zalfa,F., Eleuteri,B., Dickson,K.S., Mercaldo,V., De Rubeis,S., di Penta,A., Tabolacci,E., Chiurazzi,P., Neri,G., Grant,S.G. *et al.* (2007) A new function for the fragile X mental retardation protein in regulation of PSD-95 mRNA stability. *Nat. Neurosci.*, **10**, 578–587.

144. Stefanovic,S., Bassell,G.J. and Mihailescu,M.R. (2015) G quadruplex RNA structures in PSD-95 mRNA: potential regulators of miR-125a seed binding site accessibility. *RNA*, **21**, 48–60.

145. Rouleau,S., Glouzon,J.S., Brumwell,A., Bisaillon,M. and Perreault,J.P. (2017) 3′ UTR G-quadruplexes regulate miRNA binding. *RNA*, **23**, 1172–1179.

146. Mori,Y., Imaizumi,K., Katayama,T., Yoneda,T. and Tohyama,M. (2000) Two cis-acting elements in the 3′ untranslated region of alpha-CaMKII regulate its dendritic targeting. *Nat. Neurosci.*, **3**, 1079–1084.

147. Dictenberg,J.B., Swanger,S.A., Antar,L.N., Singer,R.H. and Bassell,G.J. (2008) A direct role for FMRP in activity-dependent dendritic mRNA transport links filopodial-spine morphogenesis to fragile X syndrome. *Dev. Cell*, **14**, 926–939.

148. Subramanian,M., Rage,F., Tabet,R., Flatter,E., Mandel,J.L. and Moine,H. (2011) G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep.*, **12**, 697–704.
149. Endoh,T., Kawasaki,Y. and Sugimoto,N. (2013) Suppression of gene expression by G-quadruplexes in open reading frames depends on G-quadruplex stability. *Angew. Chem. Int. Ed.*, **52**, 5522–5526.
150. Endoh,T. and Sugimoto,N. (2013) Unusual -1 ribosomal frameshift caused by stable RNA G-quadruplex in open reading frame. *Anal. Chem.*, **85**, 11435–11439.
151. Endoh,T., Kawasaki,Y. and Sugimoto,N. (2013) Stability of RNA quadruplex in open reading frame determines proteolysis of human estrogen receptor alpha. *Nucleic Acids Res.*, **41**, 6222–6231.
152. Yu,C.H., Teulade-Fichou,M.P. and Olsthoorn,R.C. (2014) Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res.*, **42**, 1887–1892.
153. Endoh,T. and Sugimoto,N. (2016) Mechanical insights into ribosomal progression overcoming RNA G-quadruplex from periodical translation suppression in cells. *Sci. Rep.*, **6**, 22719.
154. Phan,A.T., Kuryavyi,V., Darnell,J.C., Serganov,A., Majumdar,A., Ilin,S., Raslin,T., Polonskaia,A., Chen,C., Clain,D. *et al.* (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.*, **18**, 796–804.
155. Vasilyev,N., Polonskaia,A., Darnell,J.C., Darnell,R.B., Patel,D.J. and Serganov,A. (2015) Crystal structure reveals specific recognition of a G-quadruplex RNA by a beta-turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci. USA*, **112**, E5391–5400.
156. Goering,R., Hudish,L.I., Guzman,B.B., Raj,N., Bassell,G.J., Russ,H.A., Dominguez,D. and Taliaferro,J.M. (2020) FMRP promotes RNA localization to neuronal projections through interactions between its RGG domain and G-quadruplex RNA sequences. *Elife*, **9**, e52621.
157. Westmark,C.J. and Malter,J.S. (2007) FMRP mediates mGluR5-dependent translation of amyloid precursor protein. *PLoS Biol.*, **5**, e52.
158. Thandapani,P., Song,J., Gandin,V., Cai,Y., Rouleau,S.G., Garant,J.M., Boisvert,F.M., Yu,Z., Perreault,J.P., Topisirovic,I. *et al.* (2015) Aven recognition of RNA G-quadruplexes regulates translation of the mixed lineage leukemia protooncogenes. *Elife*, **4**, e06234.
159. Benhalevy,D., Gupta,S.K., Danan,C.H., Ghosal,S., Sun,H.W., Kazemier,H.G., Paeschke,K., Hafner,M. and Juranek,S.A. (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.
160. Herviou,P., Le Bras,M., Dumas,L., Hieblot,C., Gilhodes,J., Cioci,G., Hugnot,J.P., Ameadan,A., Guillonneau,F., Dassi,E. *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.
161. Didiot,M.C., Tian,Z., Schaeffer,C., Subramanian,M., Mandel,J.L. and Moine,H. (2008) The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.*, **36**, 4902–4912.
162. Fisette,J.F., Montagna,D.R., Mihailescu,M.R. and Wolfe,M.S. (2012) A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. *J. Neurochem.*, **121**, 763–773.
163. Marcel,V., Tran,P.L., Sagne,C., Martel-Planche,G., Vaslin,L., Teulade-Fichou,M.P., Hall,J., Mergny,J.L., Hainaut,P. and Van Dyck,E. (2011) G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis*, **32**, 271–278.
164. Huang,H., Zhang,J., Harvey,S.E., Hu,X. and Cheng,C. (2017) RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes Dev.*, **31**, 2296–2309.
165. Ribeiro,M.M., Teixeira,G.S., Martins,L., Marques,M.R., de Souza,A.P. and Line,S.R. (2015) G-quadruplex formation enhances splicing efficiency of PAX9 intron 1. *Hum. Genet.*, **134**, 37–44.
166. Verma,S.P. and Das,P. (2018) Novel splicing in IGFN1 intron 15 and role of stable G-quadruplex in the regulation of splicing in renal cell carcinoma. *PLoS One*, **13**, e0205660.
167. Verma,S.P. and Das,P. (2018) G-quadruplex structure at intron 2 of TFE3 and its role in Xp11.2 translocation and splicing. *Biochim. Biophys. Acta Gen. Subj.*, **1862**, 630–636.
168. Kim,N.W., Piatyszek,M.A., Prowse,K.R., Harley,C.B., West,M.D., Ho,P.L., Coviello,G.M., Wright,W.E., Weinrich,S.L. and Shay,J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011–2015.
169. Chen,J.L. and Greider,C.W. (2003) Template boundary definition in mammalian telomerase. *Genes Dev.*, **17**, 2747–2752.
170. Pennarun,G., Granotier,C., Gauthier,L.R., Gomez,D., Hoffschir,F., Mandine,E., Riou,J.F., Mergny,J.L., Mailliet,P. and Boussin,F.D. (2005) Apoptosis related to telomere instability and cell cycle alterations in human glioma cells treated by new highly selective G-quadruplex ligands. *Oncogene*, **24**, 2917–2928.
171. Gros,J., Guedin,A., Mergny,J.L. and Lacroix,L. (2008) G-Quadruplex formation interferes with P1 helix formation in the RNA component of telomerase hTERC. *ChemBioChem*, **9**, 2075–2079.
172. Lattmann,S., Stadler,M.B., Vaughn,J.P., Akman,S.A. and Nagamine,Y. (2011) The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. *Nucleic Acids Res.*, **39**, 9390–9404.
173. Booy,E.P., Meier,M., Okun,N., Novakowski,S.K., Xiong,S., Stetefeld,J. and McKenna,S.A. (2012) The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acids Res.*, **40**, 4110–4124.
174. Xu,Y., Kaminaga,K. and Komiyama,M. (2008) G-quadruplex formation by human telomeric repeats-containing RNA in Na+ solution. *J. Am. Chem. Soc.*, **130**, 11179–11184.
175. Xu,Y., Suzuki,Y., Ito,K. and Komiyama,M. (2010) Telomeric repeat-containing RNA structure in living cells. *Proc. Natl. Acad. Sci. USA*, **107**, 14579–14584.
176. Martadinata,H. and Phan,A.T. (2013) Structure of human telomeric RNA (TERRA): stacking of two G-quadruplex blocks in K(+) solution. *Biochemistry*, **52**, 2176–2183.
177. Cusanelli,E. and Chartrand,P. (2015) Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. *Front Genet*, **6**, 143.
178. Biffi,G., Tannahill,D. and Balasubramanian,S. (2012) An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J. Am. Chem. Soc.*, **134**, 11974–11976.
179. Deng,Z., Norseen,J., Wiedmer,A., Riethman,H. and Lieberman,P.M. (2009) TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell*, **35**, 403–413.
180. Rouleau,S.G., Garant,J.M., Bolduc,F., Bisaillon,M. and Perreault,J.P. (2018) G-Quadruplexes influence pri-microRNA processing. *RNA Biol.*, **15**, 198–206.
181. Mirihana Arachchilage,G., Dassanayake,A.C. and Basu,S. (2015) A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem. Biol.*, **22**, 262–272.
182. Pandey,S., Agarwala,P., Jayaraj,G.G., Gargallo,R. and Maiti,S. (2015) The RNA Stem-Loop to G-Quadruplex Equilibrium Controls Mature MicroRNA Production inside the Cell. *Biochemistry*, **54**, 7067–7078.
183. Mirihana Arachchilage,G., Kharel,P., Reid,J. and Basu,S. (2018) Targeting of G-Quadruplex Harboring Pre-miRNA 92b by LNA Rescues PTEN Expression in NSCL Cancer Cells. *ACS Chem. Biol.*, **13**, 909–914.
184. Ghosh,A., Ekka,M.K., Tawani,A., Kumar,A., Chakraborty,D. and Maiti,S. (2019) Restoration of miRNA-149 Expression by TmPyP4 Induced Unfolding of Quadruplex within Its Precursor. *Biochemistry*, **58**, 514–525.
185. Imperatore,J.A., Then,M.L., McDougal,K.B. and Mihailescu,M.R. (2020) Characterization of a G-Quadruplex Structure in Pre-miRNA-1229 and in Its Alzheimer's Disease-Associated Variant rs2291418: Implications for miRNA-1229 Maturation. *Int. J. Mol. Sci.*, **21**, 767.
186. O'Day,E., Le,M.T., Imai,S., Tan,S.M., Kirchner,R., Arthanari,H., Hofmann,O., Wagner,G. and Lieberman,J. (2015) An RNA-binding Protein, Lin28, Recognizes and Remodels G-quartets in the MicroRNAs (miRNAs) and mRNAs It Regulates. *J. Biol. Chem.*, **290**, 17909–17922.

187. Liu,G., Du,W., Xu,H., Sun,Q., Tang,D., Zou,S., Zhang,Y., Ma,M., Zhang,G., Du,X. *et al.* (2020) RNA G-quadruplex regulates microRNA-26a biogenesis and function. *J. Hepatol.*, **73**, 371–382.

188. Tan,W., Zhou,J., Gu,J., Xu,M., Xu,X. and Yuan,G. (2016) Probing the G-quadruplex from hsa-miR-3620-5p and inhibition of its interaction with the target sequence. *Talanta*, **154**, 560–566.

189. Tan,W., Zhang,L.L., Zhou,J., Chen,H. and Yuan,G. (2017) Formation of a G-quadruplex structure from human mature miR-5196-5p. *Spectrosc. Lett.*, **50**, 489–493.

190. Tan,W., Yi,L., Zhu,Z., Zhang,L., Zhou,J. and Yuan,G. (2018) Hsa-miR-1587 G-quadruplex formation and dimerization induced by NH4(+), molecular crowding environment and jatrorrhizine derivatives. *Talanta*, **179**, 337–343.

191. Vourekas,A., Zheng,K., Fu,Q., Maragkakis,M., Alexiou,P., Ma,J., Pillai,R.S., Mourelatos,Z. and Wang,P.J. (2015) The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. *Genes Dev.*, **29**, 617–629.

192. Zhang,X., Yu,L., Ye,S., Xie,J., Huang,X., Zheng,K. and Sun,B. (2019) MOV10L1 binds RNA G-quadruplex in a structure-specific manner and resolves it more efficiently than MOV10. *iScience*, **17**, 36–48.

193. Balaratnam,S., Hettiarachchilage,M., West,N., Piontkivska,H. and Basu,S. (2019) A secondary structure within a human piRNA modulates its functionality. *Biochimie*, **157**, 72–80.

194. Ivanov,P., O'Day,E., Emara,M.M., Wagner,G., Lieberman,J. and Anderson,P. (2014) G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. *Proc. Natl. Acad. Sci. USA*, **111**, 18201–18206.

195. Lyons,S.M., Gudanis,D., Coyne,S.M., Gdaniec,Z. and Ivanov,P. (2017) Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. *Nat. Commun.*, **8**, 1127.

196. Mestre-Fos,S., Penev,P.I., Suttapitugsakul,S., Hu,M., Ito,C., Petrov,A.S., Wartell,R.M., Wu,R. and Williams,L.D. (2019) G-quadruplexes in human ribosomal RNA. *J. Mol. Biol.*, **431**, 1940–1955.

197. Mestre-Fos,S., Penev,P.I., Richards,J.C., Dean,W.L., Gray,R.D., Chaires,J.B. and Williams,L.D. (2019) Profusion of G-quadruplexes on both subunits of metazoan ribosomes. *PLoS One*, **14**, e0226177.

198. Mestre-Fos,S., Ito,C., Moore,C.M., Reddi,A.R. and Williams,L.D. (2020) Human ribosomal G-quadruplexes regulate heme bioavailability. *J. Biol. Chem.*, **295**, 14855–14865.

199. Ruggiero,E. and Richter,S.N. (2018) G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res.*, **46**, 3270–3283.

200. Ruggiero,E. and Richter,S.N. (2020) Viral G-quadruplexes: new frontiers in virus pathogenesis and antiviral therapy. *Annu. Rep. Med. Chem.*, **54**, 101–131.

201. Norseen,J., Johnson,F.B. and Lieberman,P.M. (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.

202. Murat,P., Zhong,J., Lekieffre,L., Cowieson,N.P., Clancy,J.L., Preiss,T., Balasubramanian,S., Khanna,R. and Tellam,J. (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.*, **10**, 358–364.

203. Dabral,P., Babu,J., Zareie,A. and Verma,S.C. (2020) LANA and hnRNP A1 regulate the translation of LANA mRNA through G-quadruplexes. *J. Virol.*, **94**, e01508-19.

204. Zhang,Y., Liu,S., Jiang,H., Deng,H., Dong,C., Shen,W., Chen,H., Gao,C., Xiao,S., Liu,Z.F. *et al.* (2020) G2-quadruplex in the 3′UTR of IE180 regulates Pseudorabies virus replication by enhancing gene expression. *RNA Biol.*, **17**, 816–827.

205. Fleming,A.M., Ding,Y., Alenko,A. and Burrows,C.J. (2016) Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. *ACS Infect Dis*, **2**, 674–681.

206. Wang,S.R., Min,Y.Q., Wang,J.Q., Liu,C.X., Fu,B.S., Wu,F., Wu,L.Y., Qiao,Z.X., Song,Y.Y., Xu,G.H. *et al.* (2016) A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Sci. Adv.*, **2**, e1501535.

207. Wang,S.R., Zhang,Q.Y., Wang,J.Q., Ge,X.Y., Song,Y.Y., Wang,Y.F., Li,X.D., Fu,B.S., Xu,G.H., Shu,B. *et al.* (2016) Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. *Cell Chem Biol*, **23**, 1113–1122.

208. Perrone,R., Butovskaya,E., Daelemans,D., Palu,G., Pannecouque,C. and Richter,S.N. (2014) Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *J. Antimicrob. Chemother.*, **69**, 3248–3258.

209. Majee,P., Kumar Mishra,S., Pandya,N., Shankar,U., Pasadi,S., Muniyappa,K., Nayak,D. and Kumar,A. (2020) Identification and characterization of two conserved G-quadruplex forming motifs in the Nipah virus genome and their interaction with G-quadruplex specific ligands. *Sci. Rep.*, **10**, 1477.

210. Butovskaya,E., Solda,P., Scalabrin,M., Nadai,M. and Richter,S.N. (2019) HIV-1 Nucleocapsid Protein Unfolds Stable RNA G-Quadruplexes in the Viral Genome and Is Inhibited by G-Quadruplex Ligands. *ACS Infect. Dis.*, **5**, 2127–2135.

211. Peiris,J.S.M., Lai,S.T., Poon,L.L.M., Guan,Y., Yam,L.Y.C., Lim,W., Nicholls,J., Yee,W.K.S., Yan,W.W., Cheung,M.T. *et al.* (2003) Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet*, **361**, 1319–1325.

212. Coronaviridae Study Group of the International Committee on Taxonomy of, V. (2020) The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.*, **5**, 536–544.

213. Cui,H. and Zhang,L. (2020) G-quadruplexes are present in human coronaviruses including SARS-CoV-2. *Front. Microbiol.*, **11**, 567317.

214. Ji,D., Juhas,M., Tsang,C.M., Kwok,C.K., Li,Y. and Zhang,Y. (2020) Discovery of G-quadruplex-forming sequences in SARS-CoV-2. *Brief. Bioinform.*, bbaa114.

215. Panera,N., Tozzi,A.E. and Alisi,A. (2020) The G-quadruplex/helicase world as a potential antiviral approach against COVID-19. *Drugs*, **80**, 941–946.

216. Zhao,C., Qin,G., Niu,J., Wang,Z., Wang,C., Ren,J. and Qu,X. (2021) Targeting RNA G-quadruplex in SARS-CoV-2: a promising therapeutic target for COVID-19? *Angew. Chem. Int. Ed.*, **60**, 432–438.

217. Tan,J., Vonrhein,C., Smart,O.S., Bricogne,G., Bollati,M., Kusov,Y., Hansen,G., Mesters,J.R., Schmidt,C.L. and Hilgenfeld,R. (2009) The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog.*, **5**, e1000428.

218. Zhang,R., Xiao,K., Gu,Y., Liu,H. and Sun,X. (2020) Whole genome identification of potential G-quadruplexes and analysis of the G-quadruplex binding domain for SARS-CoV-2. *Front. Genet.*, **11**, 587829.

219. Hognon,C., Miclot,T., Garci,A.I.C., Frances-Monerris,A., Grandemange,S., Terenzi,A., Marazzi,M., Barone,G. and Monari,A. (2020) Role of RNA guanine quadruplexes in favoring the dimerization of SARS unique domain in Coronaviruses. *J. Phys. Chem. Lett.*, **11**, 5661–5667.

220. Mullen,M.A., Olson,K.J., Dallaire,P., Major,F., Assmann,S.M. and Bevilacqua,P.C. (2010) RNA G-Quadruplexes in the model plant species Arabidopsis thaliana: prevalence and possible functional roles. *Nucleic Acids Res.*, **38**, 8149–8163.

221. Kwok,C.K., Ding,Y., Shahid,S., Assmann,S.M. and Bevilacqua,P.C. (2015) A stable RNA G-quadruplex within the 5′-UTR of Arabidopsis thaliana ATR mRNA inhibits translation. *Biochem. J.*, **467**, 91–102.

222. Cho,H., Cho,H.S., Nam,H., Jo,H., Yoon,J., Park,C., Dang,T.V.T., Kim,E., Jeong,J., Park,S. *et al.* (2018) Translational control of phloem development by RNA G-quadruplex-JULGI determines plant sink strength. *Nat. Plants*, **4**, 376–390.

223. Rawal,P., Kummarasetti,V.B., Ravindran,J., Kumar,N., Halder,K., Sharma,R., Mukerji,M., Das,S.K. and Chowdhury,S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. *Genome Res.*, **16**, 644–655.

224. Kota,S., Dhamodharan,V., Pradeepkumar,P.I. and Misra,H.S. (2015) G-quadruplex forming structural motifs in the genome of Deinococcus radiodurans and their regulatory roles in promoter functions. *Appl. Microbiol. Biotechnol.*, **99**, 9761–9769.

225. Rehm,C., Wurmthaler,L.A., Li,Y., Frickey,T. and Hartig,J.S. (2015) Investigation of a quadruplex-forming repeat sequence highly enriched in Xanthomonas and Nostoc sp. *PLoS One*, **10**, e0144275.

226. Perrone,R., Lavezzo,E., Riello,E., Manganelli,R., Palu,G., Toppo,S., Provvedi,R. and Richter,S.N. (2017) Mapping and

characterization of G-quadruplexes in Mycobacterium tuberculosis gene promoter regions. *Sci. Rep.*, **7**, 5743.

227. Marsico,G., Chambers,V.S., Sahakyan,A.B., McCauley,P., Boutell,J.M., Antonio,M.D. and Balasubramanian,S. (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.*, **47**, 3862–3874.

228. Holder,I.T. and Hartig,J.S. (2014) A matter of location: influence of G-quadruplexes on Escherichia coli gene expression. *Chem. Biol.*, **21**, 1511–1521.

229. Wieland,M. and Hartig,J.S. (2009) Investigation of mRNA quadruplex formation in Escherichia coli. *Nat. Protoc.*, **4**, 1632–1640.

230. Zhao,L.Y., Song,J., Liu,Y., Song,C.X. and Yi,C. (2020) Mapping the epigenetic modifications of DNA and RNA. *Protein Cell*, **11**, 792–808.

231. Pandolfini,L., Barbieri,I., Bannister,A.J., Hendrick,A., Andrews,B., Webster,N., Murat,P., Mach,P., Brandi,R., Robson,S.C. *et al.* (2019) METTL1 promotes let-7 MicroRNA processing via m7G methylation. *Mol. Cell*, **74**, 1278–1290.

232. Kwok,C.K. and Merrick,C.J. (2017) G-quadruplexes: prediction, characterization, and biological application. *Trends Biotechnol.*, **35**, 997–1013.

233. Huang,Z.L., Dai,J., Luo,W.H., Wang,X.G., Tan,J.H., Chen,S.B. and Huang,Z.S. (2018) Identification of G-quadruplex-binding protein from the exploration of RGG motif/G-quadruplex interactions. *J. Am. Chem. Soc.*, **140**, 17945–17955.

234. Brazda,V., Cerven,J., Bartas,M., Mikyskova,N., Coufal,J. and Pecinka,P. (2018) The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors. *Molecules*, **23**, 2341.

235. Ding,D., Wei,C., Dong,K., Liu,J., Stanton,A., Xu,C., Min,J., Hu,J. and Chen,C. (2020) LOTUS domain is a novel class of G-rich and G-quadruplex RNA binding domain. *Nucleic Acids Res.*, **48**, 9262–9272.

236. Zhao,J., Qian,X., Yeung,P.Y., Zhang,Q.C. and Kwok,C.K. (2019) Mapping in vivo RNA structures and interactions. *Trends Biochem. Sci.*, **44**, 555–556.

237. Qian,X., Zhao,J., Yeung,P.Y., Zhang,Q.C. and Kwok,C.K. (2019) Revealing lncRNA structures and interactions by sequencing-based approaches. *Trends Biochem. Sci.*, **44**, 33–52.

238. Gray,L.T., Puig Lombardi,E., Verga,D., Nicolas,A., Teulade-Fichou,M.P., Londono-Vallejo,A. and Maizels,N. (2019) G-quadruplexes sequester free heme in living cells. *Cell Chem Biol*, **26**, 1681–1691.

239. Lightfoot,H.L., Hagen,T., Clery,A., Allain,F.H. and Hall,J. (2018) Control of the polyamine biosynthesis pathway by G2-quadruplexes. *Elife*, **7**, e36352.

240. Yamaguchi,K., Asakura,K., Imamura,M., Kawai,G., Sakamoto,T., Furihata,T., Linhardt,R.J., Igarashi,K., Toida,T. and Higashi,K. (2018) Polyamines stimulate the CHSY1 synthesis through the unfolding of the RNA G-quadruplex at the 5′-untraslated region. *Biochem. J.*, **475**, 3797–3812.

241. Zhang,Y., Yang,M., Duncan,S., Yang,X., Abdelhamid,M.A.S., Huang,L., Zhang,H., Benfey,P.N., Waller,Z.A.E. and Ding,Y. (2019) G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Res.*, **47**, 11746–11754.

242. Teng,Y., Tateishi-Karimata,H. and Sugimoto,N. (2020) RNA G-quadruplexes facilitate RNA accumulation in G-rich repeat expansions. *Biochemistry*, **59**, 1972–1980.

243. Armas,P. and Calcaterra,N.B. (2018) G-quadruplex in animal development: contribution to gene expression and genomic heterogeneity. *Mech. Dev.*, **154**, 64–72.

244. Imperatore,J.A., McAninch,D.S., Valdez-Sinon,A.N., Bassell,G.J. and Mihailescu,M.R. (2020) FUS recognizes G quadruplex structures within neuronal mRNAs. *Front. Mol. Biosci.*, **7**, 6.

245. Li,Q., Xiang,J.F., Yang,Q.F., Sun,H.X., Guan,A.J. and Tang,Y.L. (2013) G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.*, **41**, D1115–D1123.

246. Chen,S.B., Hu,M.H., Liu,G.C., Wang,J., Ou,T.M., Gu,L.Q., Huang,Z.S. and Tan,J.H. (2016) Visualization of NRAS RNA G-quadruplex structures in cells with an engineered fluorogenic hybridization probe. *J. Am. Chem. Soc.*, **138**, 10382–10385.

247. Tan,D.J.Y., Das,P., Winnerdy,F.R., Lim,K.W. and Phan,A.T. (2020) Guanine anchoring: a strategy for specific targeting of a G-quadruplex using short PNA, LNA and DNA molecules. *Chem. Commun. (Camb.)*, **56**, 5897–5900.

248. He,Y.D., Zheng,K.W., Wen,C.J., Li,X.M., Gong,J.Y., Hao,Y.H., Zhao,Y. and Tan,Z. (2020) Selective targeting of guanine-vacancy-bearing G-quadruplexes by G-quartet complementation and stabilization with a guanine-peptide conjugate. *J. Am. Chem. Soc.*, **142**, 11394–11403.

249. Umar,M.I. and Kwok,C.K. (2020) Specific suppression of D-RNA G-quadruplex-protein interaction with an L-RNA aptamer. *Nucleic Acids Res.*, **48**, 10125–10141.

250. Chan,C.Y. and Kwok,C.K. (2020) Specific Binding of a d-RNA G-Quadruplex Structure with an l-RNA Aptamer. *Angew. Chem. Int. Ed.*, **59**, 5293–5297.

251. Cammas,A. and Millevoi,S. (2017) RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.*, **45**, 1584–1595.