The Evolution of G-quadruplex Structure in mRNA Untranslated Region

Ting Qi1, Yuming Xu1, Tong Zhou2 and Wanjun Gu1,3,4

1State Key Laboratory of Bioelectronics, School of Biological Sciences and Medical Engineering, Southeast University, Nanjing, Jiangsu, China. 2Department of Physiology and Cell Biology, University of Nevada, Reno School of Medicine, Reno, NV, USA. 3Collaborative Innovation Center of Jiangsu Province of Cancer Prevention and Treatment of Chinese Medicine, Nanjing, Jiangsu, China. 4School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China.

ABSTRACT: The RNA G-quadruplex (rG4) is a kind of non-canonical high-order secondary structure with important biological functions and is enriched in untranslated regions (UTRs) of protein-coding genes. However, how G4 structures evolve is largely unknown. Here, we systematically investigated the evolution of RNA sequences around UTR G4 structures in 5 eukaryotic organisms. We found universal selection on UTR sequences, which facilitated rG4 formation in all the organisms that we analyzed. While G-rich sequences were preferred in the G4 structural region, C-rich sequences were selectively not preferred. The selective pressure acting on rG4 structures in the UTRs of genes with higher G content was significantly smaller. Furthermore, we found that rG4 structures experienced smaller evolutionary selection near the translation initiation region in the 5′ UTR, near the polyadenylation signals in the 3′ UTR, and in regions flanking the mRNA targets in the 3′ UTR. These results suggest universal selection for rG4 formation in the UTRs of eukaryotic genomes and the selection may be related to the biological functions of rG4s.

KEYWORDS: RNA G-quadruplex, untranslated region, evolutionary selection

Introduction

Gene expression is regulated by numerous mechanisms at the DNA and RNA levels. Post-transcriptional regulation is involved in a wide range of biological processes, including in the control of mRNA processing, stability, turnover, cellular and subcellular localization, translation.

It has been demonstrated that the genetic information for post-transcriptional regulation lies mainly in the untranslated regions (UTRs) of mRNAs and may be related to both the primary sequence and secondary structure of non-protein coding elements. The 5′ UTR contains several regulatory elements, such as binding sites for RNA binding proteins (RBPs), upstream open reading frames, and upstream start codons, that have a great impact on translation regulation.

The abundant miRNA target sites in the 3′ UTR of mRNAs are also exemplary of translation control via primary sequence recognition. In addition to the primary sequences, secondary structures formed in 5′ UTRs and 3′ UTRs can play important roles in post-transcriptional regulation by acting as target sites for RNA-binding factors or by interacting directly with the translation machinery.

In addition to canonical secondary structures such as hairpins, loops, bulges, and pseudoknots, a higher-order, non-canonical structure, namely, the G-quadruplex (G4), could also be formed by guanine-rich sequences of RNAs. G4s are folded by the stacking of G-quartets (or G-tetrads), in which 4 guanines are associated in a square planar arrangement by Hoogsteen hydrogen bonding.

The stability of a G4 depends on the kind and concentration of cation positioned within or between the planes of G-tracts. The G4 topological or conformational properties are influenced by several factors, including the cation type, the length, and sequence of loops present between the G-tracts, and the number of stacking G-quad
tets. Although first discovered in conserved DNA sequences of telomeres, G4s have been confirmed to exist in RNA sequences.

RNA G4s (rG4) are more stable than their DNA counterparts due to their chemical composition and properties. Thus, the interest in investigating the G4 structures formed in RNAs has been growing rapidly. Several computational and high-throughput experimental methods have demonstrated that G4s are frequently present in mRNAs, both in vitro and in vivo, and are especially enriched in the 5′ and 3′ UTRs. An abundance of experimental evidence has shown that G4 structures perform diverse and vital functions in a wide range of biological processes, such as pre-mRNA splicing, alternative polyadenylation, sub-cellular mRNA localization, microRNA biogenesis and targeting, and translational regulation. However, researchers have often focused on specific individual G4s or several G4s in specific regions, and few studies have tried to reveal some universal rules or mechanisms that G4s adopt to perform their functions.

In this study, we examined the evolution of rG4 structures in UTRs, which may have some implications for the functional roles played by these high-order structures. We identified the
Results

**RG4 structures were selected in untranslated sequences across 5 eukaryotic species**

We determined the rG4-specific position by using the G4HunterSeeker algorithm,\(^43,44\) and then, we calculated the rG4 formation propensity score, \(G4H_{\text{score}}\), along the mRNA sequences using a sliding window scheme. For each rG4 structure, using the location of rG4 as the origin window, we slid 13 windows upstream and downstream with a step of 30 nucleotides (nts) and calculated the \(G4H_{\text{score}}\) of each window. To estimate the background distribution of the formation propensity of rG4 structures, we randomized the mRNA sequences by shuffling nucleotides of the UTR sequences 1000 times and calculated the \(G4H_{\text{score}}\) in the corresponding sliding windows. We compared the \(G4H_{\text{score}}\) of the real mRNA sequence in a sliding window with that of 1000 corresponding sliding windows in the shuffled sequences and calculated the \(Z_{G4H_{\text{score}}}\) to assess the deviation from random expectation of the rG4 formation in the observed mRNA sequence (see Methods for details). A positive \(Z_{G4H_{\text{score}}}\) value indicated that the nucleotide sequences were selected to facilitate the formation of rG4 structures, while a negative \(Z_{G4H_{\text{score}}}\) value indicated a selective pressure that prevented the formation of rG4 structures.

We analyzed 5 eukaryotic species, namely, *Homo sapiens* (H. sapiens), *Mus musculus* (M. musculus), *Gallus gallus* (G. gallus), *Danio rerio* (D. rerio), and *Drosophila melanogaster* (D. melanogaster). We observed a significantly positive \(Z_{G4H_{\text{score}}}\) value in the window of pG4 structures in all 5 species (Figure 1). This showed that the formation of the rG4 structure was selective for specific nucleotides in these 5 organisms. When the sliding windows were moved upstream or downstream of the pG4 structures, the \(Z_{G4H_{\text{score}}}\) values dropped quickly in the flanking region of the pG4 structures and oscillated around zero for sliding windows far away from the pG4 structures (Figure 1). This suggested that no selective pressures act on nucleotides to facilitate or prevent the formation of rG4 structures when they are away from the pG4 structures in their flank region.

In addition, we found that the selection pressure acting on the 3′ UTR was higher than that on the 5′ UTR among the 5 species. When rG4 structures that were experimentally identified in vitro in human HEK293T cells and mouse mESCs were used in the analysis, we found a similar pattern of \(Z_{G4H_{\text{score}}}\) changes along the sliding windows (Supplemental Figure S1).

**More G or fewer C nucleotides were selected to facilitate rG4 formation in the UTR**

To investigate how nucleotides are selected for rG4 formation, we calculated the \(Z_g\) and \(Z_c\) of a 30-nts window for each pG4 structure in the UTR. Compared to the \(Z_{G4H_{\text{score}}}\) value of the same pG4 structure, we found a significant positive correlation between \(Z_g\) and \(Z_{G4H_{\text{score}}}\) for pG4 structures in the 5′ UTR in all 5 species (Figure 2A). In comparison, a weaker but significant negative correlation between \(Z_c\) and \(Z_{G4H_{\text{score}}}\) was also observed in all 5 species (Figure 2B). These results suggest that sequences with more Gs and/or less Cs are generally selected to facilitate the formation of rG4 structures in the UTR, while the biased usage of G-rich sequences is more relevant to rG4 formation than that of C-poor sequences (Figure 2). This pattern is consistent for rG4 structures in the 3′ UTR across all 5 species (Supplemental Figures S3 and S4). In addition, the GC content of the entire transcript is also related to the selection pressure. PG4 structures located in the 5′ UTR of genes with the top 5% G or C content tend to have smaller \(Z_{G4H_{\text{score}}}\) values than those in genes with the bottom 5% G or C content.
C content (Figure 3A and Supplemental Figure S3C). Similar results were observed for pG4 structures in 3′ UTRs (Figure 3B and Supplemental Figure S3D). The above results suggest that the GC content of a gene affects the selective pressure on the formation of rG4 structures.

Selection pressure is different for rG4 structures in different functional regions

Previous studies have found that rG4 structures have important biological functions in cells, including in alternative polyadenylation, miRNA targeting, and translational regulation.45 Therefore, we investigated the selection pressure (ZG4Hcore) trend of pG4 structures in different functional regions. First, we grouped pG4s into those in the translation initiation region (70 nts in the 5′ UTR upstream of the start codon) and those outside the translation initiation region. We discovered that the selection pressures acting on pG4 structures in the translation initiation region were significantly lower than those acting on pG4 structures outside the translation initiation region in humans and mice (Figure 3C). Next, we compared the ZG4Hcore values of pG4 structures in the flank region of the polyadenylation signal (PAS) and those far away from the PAS. Significant differences were found between pG4 structures in the flanking region of the PAS and those away from the PAS (Figure 3D). The ZG4Hcore values of pG4s near the PAS were obviously lower than those of pG4s away from the PAS. Finally, we compared the ZG4Hcore values of pG4 structures near miRNA target regions and those of pG4 structures far from miRNA target regions. We found that the ZG4Hcore values of pG4 structures near miRNA target regions were slightly smaller than those of pG4 structures far from miRNA targets in humans and mice (Figure 3E). These results suggest that the different selective pressures of rG4 structures in different genomic regions may be involved in their functional roles.

Discussion

In this study, we systematically exploited evolutionary selection on rG4 formation for all pG4 structures in UTR regions across 5 eukaryotic species. We found obvious selection signals for the formation of rG4 structures in both 5′ and 3′ UTRs across all 5 species (Figure 1). While the selection of rG4s in the 5′ UTR was slightly weaker than that of rG4s in the 3′ UTR among
$G4Hunter$-predicted rG4 sequences, the situation was quite the same for the experimentally determined rG4 sequences in both humans and mice (Figure 1, Supplemental Figure S1). Although the exact reason is not known, the selective difference between rG4s in the 5′ UTR and 3′ UTR may be explained by the different functional implications in these UTRs. We observed that $G$-rich and $C$-poor sequences are preferred in the rG4 region, facilitating the formation of rG4 structures (Figure 2A and B). Moreover, we observed that the GC content of the whole UTR may also exert considerable influence on the GC content at the genomic level.46

In addition to the sequence of UTRs, we also considered specific functional locations of rG4s and parsed rG4 structures in several specific gene regions to investigate whether the nucleotide sequences in these rG4 structures experience different selective pressures, including in translational initiation regions (Figure 3C), PAS sites (Figure 3D), and miRNA target regions (Figure 3E). It has been demonstrated that thermodynamically stable RNA canonical secondary structures (such as hairpins) in the 5′ UTR can impair eukaryotic cap-dependent translation, possibly by crippling the assembly of the translation initiation machinery at the 5′ cap of the mRNA or by impeding the scanning process of the initiation complex toward an AUG translation start codon.47-49 However, considering the variety of RNA structures, one should be equally cognizant of RNA structure forms other than stem-loop formation. In 2007, Kumari et al reported that an rG4 structure in the 5′ UTR of the $NRAS$ proto-oncogene modulates translation and that the repression of $NRAS$ translation is primarily determined by the position and stability of rG4 in the 5′ UTR.50 Notably, they found that rG4 structures proximal to the 5′ cap could strongly inhibit translation. This translation inhibition may be caused by repression of assembly of the translation preinitiation complex at the 5′ end of the mRNA.48 Since then, the G4s formed in the 5′ UTR of genes have been thoroughly studied by many researchers, and these rG4 structures are mainly correlated with translational repression of host genes, including $MT3-MMP$,51 $ERSI$,52 $BCL-2$,53 $TRF2$,54 $ADAM10$,55,56 and TGBβ2.57 As previous studies demonstrated that structures near the translation start codon could also disturb the translation initiation scanning process by blocking the AUG codon,58-60 we focused our investigation on the rG4 structures that occur in the vicinity of the AUG start codon in the 3′ end of the 5′ UTR. We observed a significantly lower mean $Z_{G4Hscore}$ of pG4s in these rG4 structures near the translation start codon compared to that of pG4s in other regions of the 5′ UTR (Figure 3C), indicating a negative trend of selection that prevents the formation of rG4 structures near the start codon. The decreased selective pressure for rG4s may imply that rG4s immediately upstream of the start codon are not favorable for efficient translation initiation.

In addition to translational regulation, Margarita et al conducted a detailed analysis of rG4 structures near the PAS in a
dataset of 244 human pre-mRNA PASs.\textsuperscript{63} They found that rG4s occurred approximately 5 times more often in the downstream +1/+70 regions (11% of pre-mRNAs in the dataset) of the PAS sites than in the upstream -1/-70 region (2%), while in the +71/+200 and -71/-200 regions from the PAS sites, G4 occurred at approximately equal frequencies (8% and 7%, respectively). The frequency of rG4 structures located far away from the PAS sites was roughly close to the frequency of rG4 located immediately downstream of the PAS sites, yet the frequency of G4 located just upstream of the PAS sites was significantly lower. This difference in the frequency of occurrences may indicate that rG4 structures immediately upstream of the PAS sites are unfavorable. Additionally, in fine mapping of pG4 structures in the 3' UTR of human mRNAs, G-PQS was extremely rare in the region stretching from 20 bases within the 3' end of the 3'UTR to 10 bases downstream from the transcription end site.\textsuperscript{63} However, just downstream of the above mentioned region, there was a very sharp peak in G-PQS density. Experimental evidence from Beaudoin et al led to the hypothesis that rG4 structures downstream of PAS sites may function as auxiliary downstream elements (AUX DSEs).\textsuperscript{33} As Chen and Wilusz\textsuperscript{63} suggested, some PAS auxiliary DSEs may promote mRNA 3'-end processing efficiency by maintaining the elements of the core poly-A signal in an opened (unstructured) form, enabling general polyadenylation factors to assemble efficiently on the pre-mRNA molecule. In contrast, G4 structures located upstream (instead of downstream) of the PAS sites may occlude the polyadenylation signal and impair efficient polyadenylation. Our result was consistent with this hypothesis since the $Z_{\text{G4 Hunter}}$ values of pG4s immediately upstream of PAS sites were moderately yet significantly lower than those of pG4s far away from the PAS sites (Figure 3D), implying that rG4 structures were not preferred in the upstream region of PAS sites to avoid disruption of efficient polyadenylation and transcription termination.

Finally, we observed a weak but significant difference between selection on rG4s near the mRNA target sites and that on rG4s far from the target region (Figure 3E). This observation may be explained by their positive or negative effects on miRNA binding. Stefanovic et al\textsuperscript{37} showed that pG4 structures in the 3' UTR of PSD-95 gene can embed mRNA target “seed” sites, thus perturb targeting regulation. However, Woo et al\textsuperscript{44} and Booy et al\textsuperscript{45} demonstrated that rG4s present in the 3' UTR of the CSF-1 or PTTX1 gene may be bound by proteins or helicases to enhance miRNA-mediated mRNA decay, respectively. Additionally, the rG4 adjacent to, but not overlapping miRNA binding sites, could also have either a positive or negative effect on miRNA binding by changing the global structure of mRNA.\textsuperscript{45}

In conclusion, our results suggested that sequences in UTRs were selected to facilitate rG4 structure formation, and the selection may be related to the specific functional roles of rG4s in different genomic regions.

**Materials and Methods**

**Data**

We downloaded the nucleotide sequences of the UTR for all protein-coding genes in 5 eukaryotic species, namely, *H. sapiens* (GRCh38.p13), *M. musculus* (GRCm38. p6), *D. rerio* (GRCz11), and *D. melanogaster* (BDGP6.28), from Ensembl BioMart\textsuperscript{66} (release 98, accessed in December 2019).\textsuperscript{66} We only considered protein-coding genes with a coding sequence more than 150 nts in length. In addition, we investigated miRNA target sites in 3' UTR of human and mouse genomes from the TargetScan (http://targetsan.org/vert_72) database.\textsuperscript{67} Moreover, the position information of PASs in humans and mice was downloaded from GENCODE (http://gencodegenes.org),\textsuperscript{68} showing the specific position of the PAS of each transcript.

**RG4 structure in untranslated regions**

To locate rG4 structures in the UTRs of mRNAs, we used the G4Hunter algorithm\textsuperscript{27} to systematically search for potential rG4 formation sites in UTR sequences in all 5 species. We ran G4Hunter App\textsuperscript{27} with a window size of 25 nts and a cut-off value of 1.2 (https://github.com/LacroixLaurent/G4HunterMulti FastaSeeker) to identify all potential rG4 structures (pG4s) in all UTRs.\textsuperscript{49} We chose the G4Hunter algorithm with these 2 parameters since a comprehensive evaluation of computational methods for rG4 prediction has suggested that G4Hunter has the best performance in predicting G4 structures.\textsuperscript{44} We identified 79560, 44600, 12812, 6097, and 3158 rG4 structures in 3' UTR sequences and 27173, 17432, 8679, 2113, and 1397 rG4 structures in 5' UTR sequences for *Homo sapiens*, *Mus musculus*, *Gallus gallus*, *Danio rerio*, and *Drosophila melanogaster*, respectively (Supplemental Table S1). To validate the observations from pG4 structures, we also downloaded experimentally derived rG4 structures in human HEK293T cells and mouse mESCs by high-throughput RT-stop techniques from the supplemental materials of Guo and Bartel.\textsuperscript{31}

**mRNA randomization**

If the nucleotide content of sequences influences the formation of rG4 structures in UTRs, the G4Hscore of mRNA sequences in the naturally occurring pG4 region should be significantly different from that of the randomized sequences. Thus, we randomly shuffled the nucleotides in the untranslated sequence, keeping the dinucleotide frequency and GC composition unchanged. For each UTR sequence, the shuffling process was repeated 1000 times to obtain a set of randomized artificial sequences. The G4Hscore of each 30-nts window in the native UTR sequence and each permuted sequence were calculated using the G4Hunter algorithm.\textsuperscript{27} The difference in G4 formation potential between the native sequence and the randomized sequences was determined by calculating the Z-score of G4Hscore for each sliding window as follows:
\[
Z_{G4Hscore} = \frac{G4Hscore_N - G4Hscore_P}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (G4Hscore_{pi} - G4Hscore_{pi})^2}}
\]

Here, \(G4Hscore_{pi}\) is the \(G4Hscore\) for the native sequence in the window, \(G4Hscore_P\) is the \(G4Hscore\) of the corresponding window of the \(i\)th randomized sequence, and \(G4Hscore_P\) is the mean of \(G4Hscore_{pi}\) over all randomized sequences. The variable \(n\) (which was 1000 here) represents the total number of randomized sequences.

Similarly, the difference between the \(G\) (or \(C\)) compositions of the native sequence and the randomized sequences was evaluated. The \(Z\)-score of the \(G\) (or \(C\)) content (\(Z_G\) or \(Z_C\)) for each sliding window can be calculated as follows:

\[
Z_G = \frac{G_N - G_P}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (G_{pi} - G_{pi})^2}} \quad \text{and} \quad Z_C = \frac{C_N - C_P}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (C_{pi} - C_{pi})^2}}
\]

\(G_N(C_N), G_P(C_P)\) and \(G_P(C_P)\) are analogous to \(G4Hscore\) but refer to the \(G\) or \(C\) content instead of the \(G\) formation potential.

Author Contributions
WG and TZ conceived this study. TQ and YX designed and performed the data analyses. TQ, YX, and WG interpreted the results. TQ, YX, and WG wrote the manuscript. All authors read and approved the final manuscript.

ORCID iD
Wanjun Gu https://orcid.org/0000-0003-4501-0539

Supplemental Material
Supplemental material for this article is available online.

REFERENCES
1. Corbett AH. Post-transcriptional regulation of gene expression and human disease. Caro Oren Cell Biol. 2018;32:96-104.
2. Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. Structural and functional features of eukaryotic mRNA untranslated regions. Gene. 2001; 276:73-81.
3. Arai K, Fukuoka T, Kato T, et al. Before it gets started: regulating translation at the 5′-UTR. Comp Funct Genomics. 2012;2012:475311.
4. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136:215-233.
5. Mandal M, Breake RR. Gene regulation by riboswitches. Nat Rev Mol Cell Biol. 2004;5:451-463.
6. Wilkie GS, Dickson KS, Gray NK. Regulation of mRNA translation by 5′- and 3′-UTR-binding factors. Trends Biochem Sci 2003;28:182-188.
7. Babendure JR, Babendure JL, Ding JH, Tseit NY. Control of mammalian translation by mRNA structure near caps. RNA. 2006;12:851-861.
8. Kozak M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem. 1991;266:19687-19687.
9. Sonenberg N. mRNA translation: influence of the 5′ and 3′ untranslated regions. Curr Opin Genet Dev. 1994;4:310-315.
10. Bochman ML, Tscheike K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. Nat Rev Genet. 2012;13:779-780.
11. Mendoza O, Boudanceau A, Boule J-B, Brosk RM Jr, Mergny J-L. G-quadruplexes and helicases. Nucleic Acids Res. 2016;44:1989-2006.
12. Rhodes D, Lippis HJ. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 2015;43:8627-8637.
13. Moriarty SP, Balaubramanian S. G-quadruplexes and helicases: secondary and structural features of G-quadruplex structures in DNA. Curr Opin Genet Dev. 2014;25:22-29.
14. Bugaut A, Balaubramanian S. G-UTR RNA G-quadruplexes: regulation translation and targeting. Nucleic Acids Res. 2012;40:4727-4741.
15. Milletti S, Moine H, Vagner S. G-quadruplexes in RNA biology. WIREs RNA. 2013;4:559-567.
16. Lane AN, Chaires JB, Gray RD, Trent JO. Stability and kinetics of G-quadruplex structures. Nucleic Acids Res. 2008;36:5482-5515.
17. Wong A, Wu G. Selective binding of monovalent cations to the stacking G-quartet structure formed by guanosine 5′-monophosphate: a solid-state NMR study. J Am Chem Soc. 2003;125:11895-11905.
18. Fay MM, Lyons SM, Ivanov P. RNA G-quadruplexes in biology: principles and molecular mechanisms. J Mol Biol. 2017;429:2137-2147.
19. Kamura T, Katsuda Y, Kitamura Y, Ihara T. G-quadruplexes in mRNA: a key structure for biological function. Biopolymers. 2020;104:2639-2646.
20. Sundquist WI, Heapsy S. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus-1 genomic RNA. Proc Natl Acad Sci U.S.A. 1993;90:3393-3397.
21. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. 2006;34:5402-5415.
22. Singh V, Azarkh M, Exner TE, Hartig JS, Drescher M. Human telomeric quadruplex conformations studied by pulsed EPR. Angew Chem Int Ed. 2009;48:9728-9730.
23. Lippis HJ, Gruissem W, Prescott DM. Higher-order DNA-structure in macro-nuclear chromatin of the hypotrichous ciliate oxytricha nova. Proc Natl Acad Sci U.S.A. 1982;79:2495-2499.
24. Sundquist WI, Klug A. Telomeric DNA dimersize by formation of guanine tetrads between hairpin loops. Nature. 1989;342:825-829.
25. Kharief P, Balatratnam S, Beali N, Bao S. The role of RNA G-quadruplexes in human diseases and therapeutic strategies. WIREs RNA. 2020;11:e1568.
26. Zhang DH, Fujimoto T, Saxena S, Yu HQ, Miyoshi D, Sugimoto N. Non-morphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. Biochemistry. 2010;49:4554-4563.
27. Bedrat A, Lacroix L, Mereny J-L. Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. 2016;44:1746-1759.
28. Kwik CK, Marsico G, Balaubramanian S. Detecting RNA G-quadruplexes (rG4s) in the transcriptome. Cell Growth Differ. 2017;28:3322-3326.
29. Kwik CK, Marsico G, Sukaluyan AR, Chambers VS, Balaubramanian S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods. 2016;13:841-844.
30. Yang X, Cheema J, Zhang Y, et al. RNA G-quadruplex structures exist and function in vivo in plants. Genome Biol. 2020;21:226.
31. Guo JU, Bartel DP. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science. 2016;353:aaf5371.
32. Huang H, Zhang J, Harvey SE, Hu X, Cheng C. RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hNRNP. Genes Dev. 2017;31:2296-2309.
33. Beaudoin JD, Perreault JP. Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. Nucleic Acids Res. 2013;41:5898-5911.
34. Balaubramanian M, Rafe T, Tabet R, Flatter E, Mandel J-L, Moine H. G-quadruplex RNA structure as a signal for mRNA targeting. EMBO Rep. 2011;12:697-704.
35. Kanai Y, Doimae N, Hirokawa N. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. Neuro. 2004;43:513-525.
36. Pandey S, Agarawal P, Jayaraj GG, Gargallo R, Maini S. The RNA stem-loop to G-quadruplex equilibrium controls mature microRNA production inside the cell. Biochemistry. 2015;54:7067-7077.
37. Stefano S, Bassell GJ, Mihaiulescu MR. G quadruplex RNA structures in PSD-95 mRNA: potential regulators of miR-125a seed binding site accessibility. RNA. 2015;21:48-60.
38. Beaudoin J-D, Perreault J-P. G-5′-UTR G-quadruplex structures acting as translational repressors. Nucleic Acids Res. 2010;38:7022-7036.
39. Kumari S, Bugaut A, Huppert JL, Balaubramanian S. An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. Nat Chem. 2007;3:218-221.
40. Moruz P, Marsico G, Herdy B, Ghanbarian AT, Portella G, Balasubramanian S. RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.* 2018;19:229.

41. Simone R, Fratta P, Neidle S, Parkinson GN, Isaacs AM. G-quadruplexes: emerging roles in neurodegenerative diseases and the non-coding transcriptome. *PFEBS Lett.* 2015;589:1653-1668.

42. Lee DSM, Ghanem LR, Barash Y. Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. *Nat. Commun.* 2020;11:527.

43. Lacroix L. *G4HunterApps.*

44. Brazda V, Kolomaznik J, Lysek J, et al. *G4Hunter web application: a web server for G-quadruplex prediction.* Bioinformatics. 2019;35:3493-3495.

45. Kharel P, Tsvetkov V. Properties and biological impact of RNA G-quadruplexes: from order to turmoil and back. *Nucleic Acids Res.* 2020;48:12534-12555.

46. Prabhu V. Symmetry observed in long nucleotide sequences. *Nucleic Acids Res.* 1993;21:5797-5800.

47. Kozak M. Inferences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl Acad Sci U.S.A.* 1986;83:2850-2854.

48. Kozak M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol.* 1989;9:5134-5142.

49. Koromilas AE, Lazaris-Karatza A, Sonenberg N. mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells over-expressing initiation factor eIF-4E. *EMBO J.* 1992;11:4153-4158.

50. Kumari S, Bugaut A, Balasubramanian S. 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.* 2008;47:12664-12669.

51. Morris MJ, Basu S. An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry.* 2015;54:38-5135-5139.

52. Balkwill GD, Derecka K, Garner TP, Hodgman C, Flint APF, Seazle MS. Repression of translation of human estrogen receptor alpha by G-quadruplex formation. *Biochemistry.* 2009;48:11487-11495.

53. Shahid R, Bugaut A, Balasubramanian S. The BCL-2 5'-untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry.* 2010;49:8300-8306.

54. Gomez D, Guerin A, Menguy JL, et al. A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res.* 2010;38:7187-7198.

55. Lammich S, Kamp-F, Wagner J, et al. Translational repression of the disintegrin and metalloprotease ADAM10 by a stable G-quadruplex secondary structure in its 5'-untranslated region. *J Biol Chem.* 2011;286:45063-45072.

56. Dai J, Liu ZQ, Wang XQ, et al. Discovery of small molecules for up-regulating the translation of anti-amyloidogenic secretase, a disintegrin and metalloprotease ADAM10, by binding to the G-quadruplex-forming sequence in the 5'-untranslated region (UTR) of its mRNA. *J Med Chem.* 2015;58:3875-3891.

57. Agarwala P, Pandey S, Mapa K, Maiti S. The G-quadruplex augments translation in the 5'-untranslated region of transforming growth factor beta 2. *Biochemistry.* 2013;52:1528-1538.

58. Kudla G, Murray AW, Tollervey D, Plotkin JB. Capping-sequence determinants of gene expression in *Escherichia coli.* *Science.* 2009;324:255.

59. Wu G, Zhou T, Wilke CO. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol.* 2010;6:e1000664.

60. Keller TE, Mis SD, Jia KE, Wilke CO. Reduced mRNA secondary-structure stability near the start codon indicates functional genes in prokaryotes. *Genome Biol Evol.* 2012;4:80-88.

61. Zaradnuya MI, Kolomiets IM, Portyakho AL, Hovorun DM. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res.* 2003;31:1375-1386.

62. Huppert JL, Bugaut A, Kumari S, Balasubramanian S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.* 2008;36:6260-6268.

63. Chen F, Wilusz J. Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. *Nucleic Acids Res.* 1998;26:2891-2898.

64. Woo H-H, Baker T, Laszlo C, Chambers S. Nucleolin mediates microRNA-directed CSF-1 mRNA deadenylation but increases translation of CSF-1 mRNA. *Mol Cell Proteomics.* 2013;12:1665-1677.

65. Boosy EP, Howard R, Maruschak O, et al. The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1. *Nucleic Acids Res.* 2014;42:3346-3351.

66. Kinsella RJ, Kähäri A, Haider S, et al. Ensemble BioMarts: a hub for data retrieval across taxonomic space. *Database.* 2011;2011:bau030.

67. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 2020;48:D127-D131.

68. Frankish A, Diekhans M, Jungreis I, et al. GENCODE 2021. *Nucleic Acids Res.* 2021;49:D916-D923.