

Potential G-Quadruplex Forming Sequences and N⁶-Methyladenosine Colocalize at Human Pre-mRNA Intron Splice Sites

Manuel Jara-Espejo, Aaron M. Fleming, and Cynthia J. Burrows*



Cite This: *ACS Chem. Biol.* 2020, 15, 1292–1300



Read Online

ACCESS |



Metrics & More

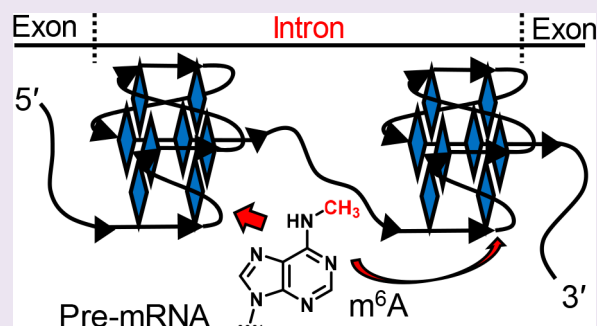


Article Recommendations



Supporting Information

ABSTRACT: Maturation of mRNA in humans involves modifying the 5' and 3' ends, splicing introns, and installing epitranscriptomic modifications that are essential for mRNA biogenesis. With respect to epitranscriptomic modifications, they are usually installed in specific consensus motifs, although not all sequences are modified suggesting a secondary structural component to site selection. Using bioinformatic analysis of published data, we identify in human mature-mRNA that potential RNA G-quadruplex (rG4) sequences colocalize with the epitranscriptomic modifications N⁶-methyladenosine (m⁶A), pseudouridine (Ψ), and inosine (I). Using the only available pre-mRNA data sets from the literature, we demonstrate colocalization of potential rG4s and m⁶A was greatest overall and occurred in introns near 5' and 3' splice sites. The loop lengths and sequence context of the m⁶A-bearing potential rG4s exhibited short loops most commonly comprised of single A nucleotides. This observation is consistent with a literature report of intronic m⁶A found in SAG (S = C or G) consensus motifs that are also recognized by splicing factors. The localization of m⁶A and potential rG4s in pre-mRNA at intron splice junctions suggests that these features could function together in alternative splicing. A similar analysis for potential rG4s around sites of Ψ installation or A-to-I editing in mRNA also found a colocalization; however, the frequency was less than that observed with m⁶A. These bioinformatic analyses guide a discussion of future experiments to understand how noncanonical rG4 structures may collaborate with epitranscriptomic modifications in the human cellular context to impact cellular phenotype.



INTRODUCTION

The process of pre-mRNA maturation is of considerable interest because each step can impact the stability, coding potential, or localization of the mature mRNA.^{1–5} Maturation of mRNA involves installation of a 5' cap, addition of a 3'-polyadenosine tail, writing of epitranscriptomic modifications, and intron excision.⁶ Processing of the 5' and 3' ends of mRNA has been studied, and recent work suggests there is more to learn.^{7,8} Understanding of writing epitranscriptomic modifications and alternative mRNA splicing are rapidly advancing as a result of next-generation sequencing (NGS) and expansion of bioinformatic tools.^{1–4} In human cells, NGS-based studies in tandem with knockdown or knockout of established epitranscriptomic protein readers, writers, or erasers have identified that these modifications are involved in all aspects of mRNA biogenesis including splicing, nuclear export, translation efficiency, and cellular half-life. The collaboration of epitranscriptomic modifications and alternative mRNA splicing has recently been described,⁹ but the structural features of mRNA driving the process remain mysterious. A better understanding of the epitranscriptome and alternative mRNA splicing will enable future researchers to modulate these pathways judiciously in cells and provide new

therapeutic opportunities to target these pathways for disease treatment.

N⁶-METHYLADENOSINE

The best studied epitranscriptomic modification in mRNA is N⁶-methyladenosine (m⁶A; Figure 1A). Writing of the methyl group on A nucleotides in mRNA occurs in specific sequence motifs by the METTL3/14 SAM-dependent methyltransferase complex at a frequency of 0.1–1% of A nucleotides; however, not all such target sequences are modified.^{1–5} The DRACH (D = A, G, or U; R = G or A; H = A, C, or U; A = m⁶A) consensus sequence appears to be a dominant site of m⁶A installation in mature mRNA; in contrast, the splicing sites within intron sequences of pre-mRNA have m⁶A installed in SAG (S = C or G) motifs.^{9,10} To date, the accessory protein WTAP appears to

Received: April 4, 2020

Accepted: May 12, 2020

Published: May 12, 2020



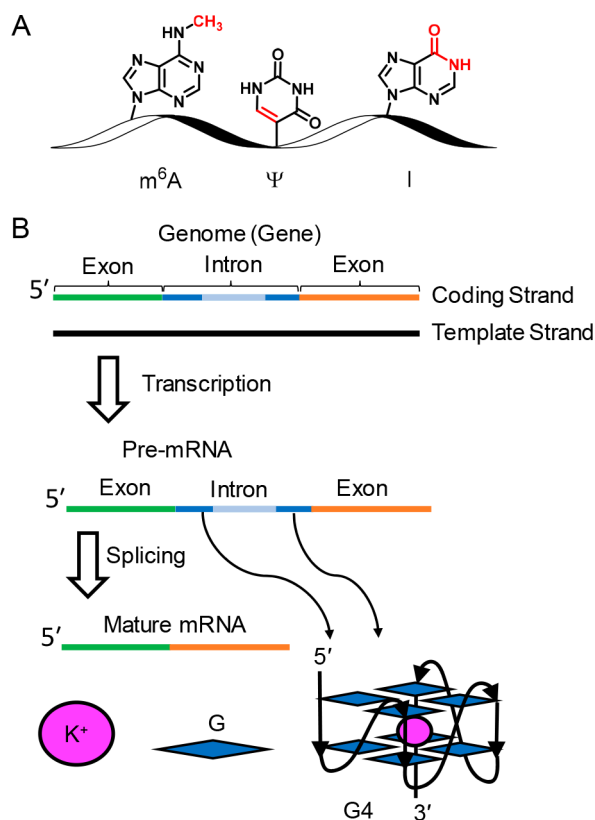


Figure 1. (A) Modifications m⁶A, Ψ, and I are epitranscriptomic. (B) Maturation of mRNA involves intron splicing; potential G-quadruplex sequences are found in introns near splice sites.

be the most essential for guiding the methylation process, while VIRMA, ZC3H13, and RBM15, as examples, also have been implicated in writing of m⁶A on mRNA.^{3,4,11} Once written on mRNA, m⁶A is read by the cytosolic YTHDF1–3 proteins, and the m⁶A modifications can be removed by the demethylases FTO or ALKBH5 imparting dynamics to this epitranscriptomic system.^{1–5} Current studies have led to the suggestion that RNA secondary structure plays a key role in selection of m⁶A modified sites.^{12–14} Future work to clarify the structural component in selection or mRNA sites epitranscriptomically modified is needed.

■ PSEUDOURIDINE AND A-TO-I EDITING

A second modification is the isomerization of uridine to pseudouridine (Ψ; Figure 1A) in mRNA that occurs at a frequency of 0.2–0.6% of U nucleotides.¹ Out of the many human pseudouridine synthases, PUS1 and PUS7 are established to isomerize U to Ψ in mRNA.^{2,15} The same questions regarding site selection of Ψ installation have been asked, with a recent report finding that PUS1 targets hairpin-type structures.² Lastly, two human adenosine deaminase RNA specific (ADAR) proteins catalyze A-to-I editing that yields inosine (I; Figure 1) as another epitranscriptomic modification.¹⁶ Extensive studies on RNA editing have been conducted and cataloged in databases for researchers to interrogate.¹⁷ Editing of RNA by ADARs is catalyzed in a central region of long dsRNA with some minor sequence and local structural context requirements noted.¹⁸ More analyses are needed to seek out and clarify the RNA structural and sequence requirements that dictate sites of m⁶A and Ψ epitranscriptomic

modification, as well as to explore whether other sequence motifs can be sites of A-to-I RNA editing.^{12,18,19}

■ PRE-mRNA AND G-QUADRUPLEXES (G4s)

Alternative splicing of nascent or pre-mRNA to yield mature mRNA is a highly regulated eukaryotic process resulting in a single gene having the potential to code for multiple proteins.²⁰ This diversification occurs by inclusion or exclusion of particular exons in the final processed mRNA. In humans, ~95% of multiexonic genes are alternatively spliced enabling the ~20 000 protein-coding genes to direct synthesis of a much greater diversity of proteins. Alternative splicing of mRNA is cell-type specific and changes with oxidative stress or disease.^{20,21} The mRNA features and epitranscriptomic components that drive when and to what extent alternative splicing occurs in particular cell types have been topics of recent studies,^{20,21} and future work is needed to better understand the RNA structural details guiding the process.

Genomes code for all the information found in RNA (Figure 1B), and in the human genome there exists enrichment of potential G-quadruplex forming sequences (PQSs) in promoters, 5′-untranslated regions, and introns near splice sites.²² G-Quadruplexes (G4s) are noncanonical folds in nucleic acids that have sequences comprised of four or more runs of G in close proximity providing the opportunity for the sequence to fold around intracellular K⁺ ions (Figure 1B).²³ In DNA G4s, stable folds each have G-runs of three or more Gs, while in RNA G4s (rG4s), two Gs per run can adopt stable folds.^{23–25} Those PQSs in the nontemplate strand (i.e., coding strand) of a gene will also be present in the RNA transcript (Figure 1B). We have noted that potential rG4s in the Zika and HIV viral RNAs colocalize with sites of m⁶A installation,¹³ however, whether a similar colocalization exists in the human transcriptome had not yet been studied.

■ BIOINFORMATIC ANALYSIS OF G4s AND EPITRANSCRIPTOMIC MODIFICATIONS

In the present set of bioinformatic studies, mapping data for m⁶A and Ψ in human mature mRNA, as well as I in the human transcriptome, were inspected for colocalization around potential rG4s. We found enrichment of each epitranscriptomic modification around potential rG4s in the data inspected. A second analysis looked at the only available modification maps in pre-mRNA looking for m⁶A, in which we found a greater colocalization of this modification with potential rG4s than observed in mature mRNA for all three modifications analyzed. This finding suggests a possible synergy in the deposition of m⁶A on pre-mRNA splice sites around potential rG4s. Biochemical studies suggest that m⁶A is written on nascent mRNA cotranscriptionally; therefore, chromatin features such as genomic sequences and structures or histone modifications that interact with the mRNA synthesis machinery can impact where and to what extent epitranscriptomic modifications are installed.^{3,9,26,27} Here, we identify that colocalized sites of m⁶A in pre-mRNA and rG4s track with the genomic G4s found in introns, suggesting a possible long-range interplay of chromatin structure and the epitranscriptome. In the final analysis of the m⁶A sites in the potential rG4s of intronic pre-mRNA, a preference for rG4 loop sequences with a modifiable A nucleotide was identified. The SAG sequence previously found for the consensus motif of m⁶A

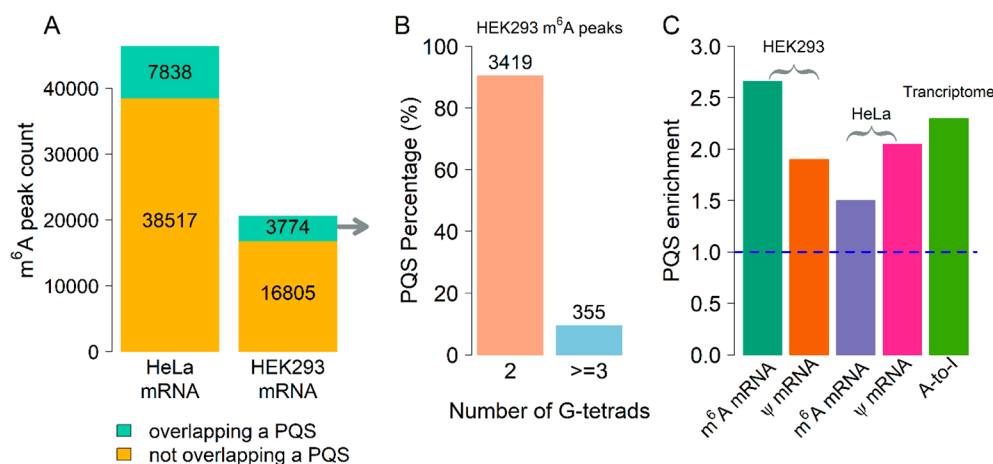


Figure 2. Inspection of human mature mRNA sequenced for m⁶A, Ψ, and A-to-I editing sites to determine whether a PQS resides in the same location. (A) Bar plot showing the number of m⁶A enriched sites in HeLa and HEK2993 mature mRNA that also have a PQS overlapping with the site of modification. (B) Classification of the PQSs found in the HEK2993 mature mRNA on the basis of the number of G-tetrads that can form from the sequence. (C) Plot of PQS enrichment (observed/expected) found in the m⁶A or Ψ epitranscriptomic modification sites in mature mRNA from HEK2993 or HeLa cells, and A-to-I editing sites in the Inosinome Atlas.^{17,27–29}

deposition in intronic sequences was found to be part of a larger rG4 structural context.

The publicly available RNA modification maps in human mRNA used in the present study are summarized in Table S1.^{9,12,17,27–29} The peak summits for each epitranscriptomic modification were identified, and then we selected a window of sequence space ± 30 nucleotides flanking the summit to inspect for PQSs computationally. The quadruplex-forming G-rich sequence (QGRS) mapper algorithm was used to look for the sequence pattern 5'-G_xL_{<7}G_xL_{<7}G_xL_{<7}G_x-3' where $x \geq 2$ nucleotides and L represents the loops.³⁰ The first data set inspected had sequenced m⁶A in mature mRNA collected from HeLa cells using the m⁶A-CLIP sequencing protocol.²⁷ From these data in mature HeLa mRNA, 17% (7838 out of 46355) of the m⁶A enriched regions were found to also contain a PQS (Figure 2A). Inspection of m⁶A mapped via miCLIP in mature mRNA from HEK2993 cells²⁹ found 18% (3774 out of 20579) of the m⁶A enriched regions colocalized with a PQS (Figure 2A). These data in two different cell lines indicate that nearly one-fifth of the m⁶A peaks colocalize with a PQS in human mature mRNA.

In the next step of the analysis, the PQSs found in sites of m⁶A enrichment were classified on the basis of the number of G-tetrads that would occur in the rG4 fold. In RNA, stable rG4 folds have been found with only two G-tetrads, which is in contrast to DNA that generally require at least three G-tetrads to adopt a stable fold.^{13,23} The greater stability in rG4s results from the 2'-OH providing an additional hydrogen bond that is not possible in DNA.²³ Prior studies have shown that two-tetrad rG4s provide a quasi-stable fold that can be harnessed as a switch to impact the fate of RNA in cells.^{25,31} In mature mRNA from HEK2993 cells, 81% could adopt a two G-tetrad rG4 and 19% could adopt a three or more G-tetrad rG4 (Figure 2B); a similar excess of two G-tetrad compared to three G-tetrad potential rG4s was observed for the mature mRNA from HeLa cells (Figure S1). The overabundance of two G-tetrad rG4s may have biological significance because this lowers the stability of the fold and possibly allows it to function as an on-off switch with dependency on the modification status. Studies of m⁶A in rG4s have yet to be conducted, although a recent report of N⁶-methyl-2'-

deoxyadenosine in the loop of a DNA G4 has found this modification to destabilize the structure.³² A similar impact on stability may exist in rG4s. In this situation, m⁶A would function to destabilize rG4 folds and possibly favor other secondary structures as the structural switch.

The statistical significance of PQSs colocalizing with m⁶A-enriched sites in human mRNA was calculated by comparison of the identified count to one obtained from randomized and shuffled sequences (Figure 2C). The presence of PQSs in the m⁶A sites in the HEK2993 and HeLa cells was found to be significant with 2.6-fold ($P < 2.2 \times 10^{-16}$) and 2.1-fold ($P < 2.2 \times 10^{-16}$; Table S2) enrichment, respectively, relative to the randomized samples on the basis of the Fisher's exact test. To summarize the m⁶A analysis in mature mRNA, the modification sites were found to be favorably enriched (>2-fold) around potential rG4s that predominantly have two G-tetrads (Figure 2).

Next, mature mRNA from HeLa or HEK2993 cells were inspected for colocalization of Ψ sites with PQSs that were chosen for study to be consistent with the m⁶A data sets analyzed (Table S1).^{12,17} In the Ψ data set from HEK2993 cells, 13% of the modified sites (308 out of 2058) also contained a PQS (Figure S1). This number of PQSs represents a significant 1.7-fold enrichment of these G-rich sequences ($P < 8.8 \times 10^{-8}$; Figure 2C and Table S2). In the HeLa cell data set, 20% of the modified sites (24 out of 115) also contained a PQS in the mature mRNA (Table S1). This represents a significant 2.0-fold enrichment of these G-rich sequences ($P < 6.4 \times 10^{-3}$; Figure 2C and Table S2). In the Ψ sites from the different cell lines, 95% of the HEK2993 and 89% of the HeLa sites had a two G-tetrad rG4s with the remainder having three or more G-tetrads (Figure S1). This identified a small but favorable enrichment of Ψ sites around potential rG4s, and the rG4s predominantly adopt two G-tetrads.

The A-to-I editing data were obtained from the Inosinome Atlas that provides a comprehensive listing of all established editing sites.¹⁷ A key difference in this analysis is that the data were obtained from the entire transcriptome and not restricted to mature mRNA. In the A-to-I analysis, 22% (1 004 026 out of 4 668 508) colocalized with a PQS (Figure 2C), which was a significant 2.3-fold enrichment ($P < 2.2 \times 10^{-16}$; Table S2). The

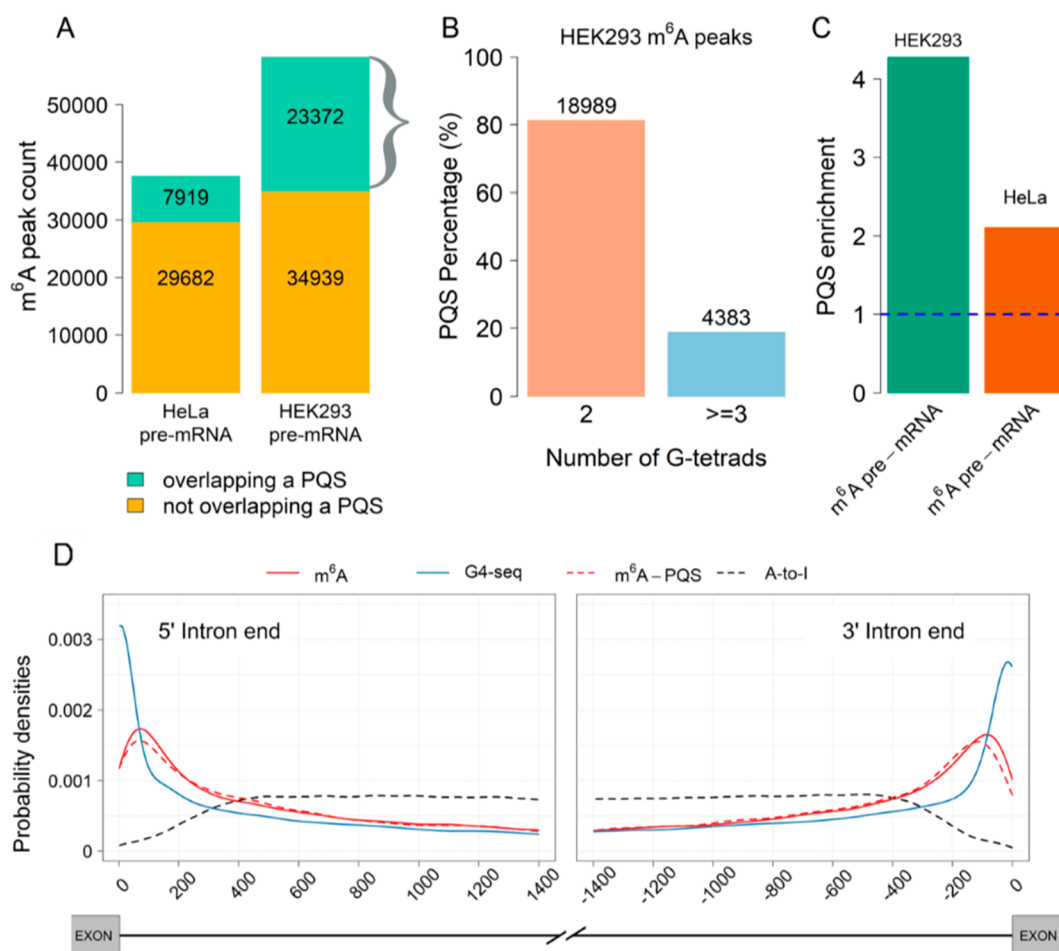


Figure 3. Analysis of enriched sites of m⁶A in human pre-mRNA for potential rG4s. (A) Counts of m⁶A sites with and without a colocalized potential rG4. (B) Break down of potential rG4s found in the HEK2993 data set for the number of G-tetrads. (C) Fold enrichment of potential rG4s in the experimental sample relative to the randomized sample. (D) Intron map illustrating the m⁶A enriched sites found in HEK2993 pre-mRNA sequenced by TNT-seq,⁹ the position of the PQSs found in the sequencing data, and comparison to the position of G4s found in the human genome via G4-seq.³³

G-tetrad count for PQSs colocalized with A-to-I editing sites was ~80% two G-tetrad G4s and ~20% three or more G-tetrad rG4s (Figure S1); these values are very similar to those found in the m⁶A and Ψ PQS colocalization data described above.

Guided by the knowledge that PQSs are enriched in human introns,³³ inspection of the chromatin-associated RNA (i.e., pre-mRNA) was then conducted. Maps for m⁶A in HeLa and HEK2993 cellular pre-mRNA are the only ones available;^{9,27} thus, we were not able to conduct a similar analysis for Ψ or I in pre-mRNA. The m⁶A enriched sites in pre-mRNA from HEK2993 cells were sequenced using transient N⁶-methyladenosine transcriptome sequencing (TNT-seq),⁹ while the HeLa pre-mRNA were sequenced using the m⁶A-CLIP protocol;²⁷ thus, there exists a difference in how these two maps were obtained and they may have different sequence and structural biases. In the HeLa pre-mRNA, 21% (7919 out of 37606) of the m⁶A sites colocalized with a potential rG4 (Figure 3A). On the other hand, in HEK2993 pre-mRNA 40% (23372 out of 58311) of the enriched m⁶A sites colocalized with a potential rG4 (Figure 3A). In the population of potential rG4s colocalized with m⁶A in HEK2993 cells, ~80% were two G-tetrad rG4s and the rest had three or more G-tetrads (Figure 3B). In the HeLa pre-mRNA, potential rG4s were found to be ~90% two G-tetrad rG4s and the remainder were three or

more G-tetrad rG4s (Figure S1). This analysis suggests a high incidence of m⁶A enriched sites occurring in potential rG4s with two G-tetrads in pre-mRNA particularly in the pre-mRNA from HEK2993 cells.

The statistical significance for enrichment of PQSs in the regions of m⁶A installation in human pre-mRNA was compared to randomized, shuffled sequences (Figure 2C). Comparison of the 23372 PQSs identified in the HEK2993 pre-mRNA to the 5461 PQSs expected by randomized shuffling found a 4.3-fold enrichment in PQSs; this finding is significant on the basis of Fisher's exact test ($P < 2.2 \times 10^{-16}$; Figure 3C and Table S2). In the HeLa mRNA analyzed for m⁶A and PQS colocalization, there was a significant 2.1-fold enrichment found in the pre-mRNA ($P < 2.2 \times 10^{-16}$; Figure 2C and Table S2). Taken together, these results support a favorable colocalization of m⁶A sites and PQSs in human mRNA. The finding of m⁶A and PQS colocalization in human mRNA, particularly introns of pre-mRNA, suggests that the rG4 secondary structure and epitranscriptomic m⁶A may be synergistic; this observation is consistent with our previous report on the colocalization of m⁶A and PQSs in viral genomic RNA.¹³

In the HEK2993 pre-mRNA m⁶A data reported by Louloup et al.,⁹ a high incidence of m⁶A residing in intronic regions was

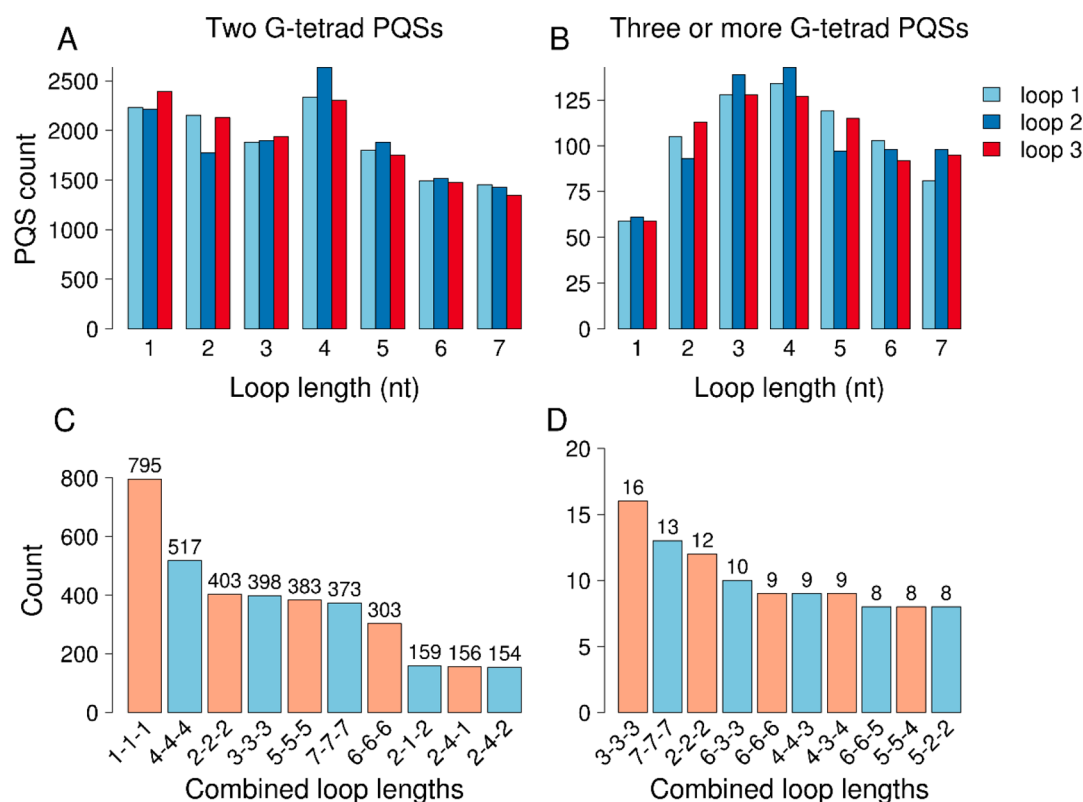


Figure 4. Loop length analysis of the PQSs that colocalized with m^6A enriched regions in the HEK293 pre-mRNA.⁹ Analysis of individual loop lengths for (A) two G-tetrad PQSs and (B) three or more G-tetrad PQSs. Analysis of loop length combinations for (C) two G-tetrad PQSs and (D) three or more G-tetrad PQSs that identify the most prevalent loop length combinations.

observed. Next, a focused inspection of PQSs in the intron regions was conducted. The m^6A data (Figure 3D, solid red) and PQS (Figure 3D, dashed red) sites tracked with each other and were favorably enriched on the intronic side of both the 5' and 3' splice sites. With the knowledge that epitranscriptomic modifications are installed cotranscriptionally and that the chromatin architecture on the genome and histones impacts the methylation process, we asked whether DNA G4s near the region coding for the RNA m^6A sites are favorably colocalized. The Balasubramanian laboratory developed G4-seq to find all sequences that could adopt G4s in the human genome.³³ There is one noteworthy point regarding this comparison, which is that stable DNA G4s are generally at least three G-tetrads,²⁴ while the present RNA analysis found a preference for two G-tetrad rG4s (Figures 2B, 3B, and S1). The comparison found the G4-seq data on the coding strand of human introns did indeed show enrichment at both the 5' and 3' splice sites (Figure 3D, blue line) that also tracked with the mRNA m^6A and PQS profiles.

How could a genomic G4 impact writing of m^6A on pre-mRNA? Events that stall the RNA pol II complex during transcription show increased deposition of m^6A on the transcript.^{3,9,26,27} It is known that template strand G4s stall polymerase bypass,³⁴ however, G4s in the template strand do not code for the G-rich sequence in the mRNA. In contrast, genomic G4s in coding strands can stall the RNA pol II complex by increasing the persistence and length of R-loops^{35,36} and at the same time code for an rG4 in the nascent mRNA. It is possible that the genomic G-rich sequence has two effects that are (1) to slow mRNA synthesis and (2) to cause greater writing of m^6A on potential rG4s in pre-mRNA.

Future experimental studies are needed to address this hypothesis derived from bioinformatic inspection of m^6A mapping data in HEK293 pre-mRNA.

■ A ROLE FOR G4s AND m^6A IN SPLICING?

A closer examination of the data around intron splice sites provided a few additional observations. Within 200 nt of each splice junction, 8040 m^6A -enriched sites on the intronic 5' splice site representing 3187 genes and 7236 m^6A -enriched sites on the intronic 3' splice site representing 2681 genes were found; these numbers represent 47% of total m^6A intronic sites in the HEK295 pre-mRNA data set. A similar distribution of m^6A and PQS colocalization sites was observed indicating a possible close association of these two RNA features near mRNA splicing sites. These observations suggest an opportunity for future studies that address whether rG4s and m^6A function synergistically to guide mRNA splicing.

The A-to-I editing sites in introns of HEK293 mRNA were plotted alongside the m^6A , PQS, and G4-seq data to find that RNA editing was not observed around splice sites and did not track with PQSs in this region (Figure 3D, black dashed line). The A-to-I editing sites appear to be depleted around intron splicing sites. The Ψ data set from HEK293 cells was conducted on mature mRNA, in which introns are not present, and therefore, no further analysis of the data was conducted.

■ G-QUADRUPLEX LOOP ANALYSIS

In mature mRNA, m^6A mapping studies have suggested a broad consensus motif for A methylation in the sequence context DRACH (D = A, G, or U; R = G or A; and H = A, C,

or U). In the work by Louloui et al., intronic m⁶A was favorably deposited in SAG (S = C or G) sequence motifs.⁹ Because the HEK293 pre-mRNA data exhibited the highest colocalization of m⁶A and PQSs, the sequence population was further interrogated to identify favorable rG4 loop profiles with respect to length and sequence. During the PQS inspection, the three loops could have 7 or fewer nucleotides. The loop analysis was conducted on 14076 two G-tetrad PQSs and 729 three or more G-tetrad PQSs. In the loop length analysis of two G-tetrad PQSs, there was a slight preference for shorter loop lengths, but there were many longer loop PQSs observed at a high relative frequency (Figure 4A). Additionally, a breakdown of the first, second, or third loops found they all had a similar length profile (Figure 4A). The minimal dependency of two G-tetrad rG4 loop lengths in the population colocalizing with m⁶A suggests a broad structural scope for writing this modification on these noncanonical folds. Additionally, the data suggest symmetry in loop length may be a feature of rG4s that are at sites of m⁶A installation.

In the PQS population with three or more G-tetrads, one nucleotide loop lengths were less common, while three and four nucleotide loop lengths were most common (Figure 4B). Inspection of the combination of all three loop lengths together in the two G-tetrad PQSs found the 1–1–1 loop length combination to be most common followed by the 4–4–4 loop length combination. In general, as the loop length combination increased or became asymmetric, the number of PQSs observed decreased (Figure 4C). For the three-loop length combination analysis of the three G-tetrad or more PQSs, the most common combination found had 3–3–3 nucleotide loop lengths; this was followed by the longer 7–7–7 and shorter 2–2–2 nucleotide loop lengths (Figure 4D). In the three or more G-tetrad PQS data, the least common were those with asymmetric loop lengths, with the exception to the least common pool being 6–6–6 nucleotide long loops (Figure 4D). This information suggest longer loops in three or more G-tetrad rG4s with which m⁶A colocalize are preferred. The reason for this difference relative to the two G-tetrad cohort is not known; however, longer loops in G-quadruplexes usually destabilize the structure,³⁷ which may be important in rG4s used as structural switches responding to the presence of m⁶A in cells.

Inspection of the loop sequences of the two G-tetrad PQSs in m⁶A-enriched regions of intronic pre-mRNA in HEK293 cells identified a nucleotide preference. In Table 1, a rank ordering of the top five most common loop sequences found in each of the three possible rG4 loops is provided. Single-

Table 1. PQS Loop Nucleotide Composition for Those Loops in m⁶A-Enriched Regions^a

	order most to least common				
	1	2	3	4	5
loop 1	A	U	GA	G	AA
count	1044	518	502	474	237
loop 2	A	U	G	GA	C
count	936	657	384	250	239
loop 3	A	U	G	CA	GA
count	999	691	459	379	332

^aData corresponds to HEK293 pre-mRNA m⁶A profile reported by Louloui et al.⁹ See Table S3 for additional data.

nucleotide loops comprised of an A nucleotide were the most common in all three loops. Because A nucleotides are potential sites of methylation, the observation of single A nucleotides in the two G-tetrad PQSs nicely fits our hypothesis that rG4 folds provide a structural motif to guide sites of m⁶A introduction in human mRNA. Furthermore, the high incidence of single A nucleotides is consistent with the work by Louloui et al.,⁹ in which the consensus motif SAG occurred with greater frequency.

The second most common loop sequence observed in the two G-tetrad PQSs was single U nucleotides in any of the three loops, although many of the top ten most prevalent loop sequences contained A nucleotides within dinucleotide motifs such as 5'-GA, 5'-AA, 5'-CA, and 5'-AG. Interestingly, the 5'-AC dinucleotide that would indicate the DRACH consensus motif within a PQS was not among the top 10 most prevalent loop sequences (Table S3). Further inspection of the 795 PQSs in which single-nucleotide loops were identified also found that homogeneous loops of single A nucleotides dominated the distribution with 142 occurrences. Inspection for homogeneous single U or C nucleotide loops found 18 and 34 occurrences, respectively, and no PQSs were observed with all-G loops. In summary, two G-tetrad PQSs in m⁶A enriched regions are biased in their nucleotide composition to have A > U > G > C and in their propensity to have the same lengths for all three loops (Figure 4 and Tables 1 and S3). This indicates that homogeneous short loops favor A nucleotides as evidenced by the loop composition analysis, although the sequences with all C or U nucleotide loops will not provide an RNA substrate for writing m⁶A. These sequences may be false positives or the methylated A may reside just beyond the G4; additional analysis of the tail sequences for the small sample of PQSs without an A was not conducted. The sequence composition for the three or more G-tetrad PQSs found colocalized with m⁶A was also analyzed and found to be rich in A nucleotides (Table S4). Additionally, the 5'-AC dinucleotide common to the DRACH consensus motif was not found in the top ten most common loop sequences, as indicated by the Louloui et al. work.⁹ Short rG4 folds such as these found in the sequencing data can adopt stable folds as was recently reported.³⁸

CONCLUSIONS AND OUTLOOK

The observations of the bioinformatic analysis of m⁶A colocalization with rG4s, especially those near intronic splice sites, can guide the design of future experiments. (1) This bioinformatic study suggests a synergy between rG4 folds with sites of m⁶A epitranscriptomic modification. Do rG4 folds function as a structural motif for methylation of the RNA by the METTL3/14 methyltransferase complex? The loop length and sequence identity of the rG4s found are essential knowledge for the design of in vitro studies to test the hypothesis that METTL3/14 favors writing of m⁶A on rG4 scaffolds. One feature of the rG4s the present data cannot address is whether the preferred folds exist in sequence contexts that are dynamic between rG4 structures or with other RNA structures; this type of information will likely need to be addressed on a sequence-by-sequence basis or via inspection of high-resolution RNA structural maps in vivo obtained by chemical probing.^{39,40} (2) The SAG consensus sequence is also recognized by some splicing proteins such as SRSF3.⁹ Do these splicing proteins bind rG4s and is their binding modulated by the presence of m⁶A in a rG4 loop? (3)

Splicing factors that bind SAG sequence motifs were found to be involved in alternative mRNA splicing.⁴¹ The role of PQSs in introns and their folding to rG4s to impact alternative splicing of specific mRNA, such as the p53 mRNA, has been noted.^{42–44} Is there synergy between rG4s and m⁶A to guide alternative mRNA splicing? At present, studies have addressed rG4s^{42,43} and m⁶A individually for guiding alternative splicing.⁹ The analysis presented here suggests that rG4s and m⁶A may collaborate in alternative splicing. (4) Whether the rG4 fold is the signal for writing or erasing m⁶A in the mRNA or the presence of m⁶A impacts the rG4 fold is not known, and is a question we previously asked.¹³ Further, rG4 folds are known to be dynamic and adopt many different structures; because each sequence is unique, this would have to be studied on a case-by-case basis. (5) The preference for A-rich two G-tetrad potential rG4s identified in the intronic m⁶A enriched sites is not understood, and further studies are needed. It is known that A-rich G4 loops destabilize the fold that may be important for stability of the fold to be altered by methylation under physiological conditions.⁴⁵

Herein, we explored the structural pattern of human RNA sites harboring m⁶A, Ψ, or A-to-I editing modifications, focusing on the presence of PQSs at these sites. The study revealed that all three modifications favorably colocalized with potential rG4s when a comparison to a randomized data set was conducted (Figures 2C and 3C). The greatest colocalization was observed between m⁶A and potential rG4s near the splice sites in introns of HEK293 cells (Figure 3A–D). This observation suggests there may be an interplay between m⁶A, rG4s, and mRNA splicing that could be a component of alternative splicing; future experimental work is needed to address this possibility. Our prior interest in the colocalization of m⁶A and PQSs focused on viral RNA genomes that showed a preference for DRACH motif methylation.¹³ The present work on human pre-mRNA is consistent with the viral RNA analysis; however, the sequence context found for the colocalization in intronic pre-mRNA occurs largely within SAG motifs (Table 1). This difference observed may reflect the fact that writing of m⁶A on pre-mRNA occurs in the nucleus, while m⁶A in viral RNA occurs in the cytosol.^{9,13} The present bioinformatic study identified colocalization of potential rG4s in mRNA and epitranscriptomic modifications that suggests many additional experimental questions to be asked. Structurally, rG4s will present to epitranscriptomic writing enzymes differently than duplex, hairpin, or single-stranded regions of RNA, which may help explain why not all consensus motifs for a given modification are modified and why they generally are not quantitatively modified.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscchembio.0c00260>.

Complete methods, tables for data sources, complete statistical values, and loop sequence composition data (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Cynthia J. Burrows – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0850, United States;

orcid.org/0000-0001-7253-8529; Email: burrows@chem.utah.edu

Authors

Manuel Jara-Espejo – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0850, United States; Department of Morphology, Piracicaba Dental School, University of Campinas-UNICAMP, Piracicaba 13414-018, Sao Paulo, Brazil; orcid.org/0000-0002-5454-7787

Aaron M. Fleming – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0850, United States; orcid.org/0000-0002-2000-0310

Complete contact information is available at: <https://pubs.acs.org/10.1021/acscchembio.0c00260>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by a grant from the U.S. National Institute of General Medical Sciences (R01 GM093099). M.J.-E. thanks S. Line (UNICAMP) and the Brazilian Coordination for the Improvement of Higher Education Personnel-CAPES-PRINT 88887.364735/2019-00 for supporting his stay at the University of Utah.

■ KEYWORDS

G-Quadruplex: A nucleic acid secondary structure formed by sequences with four or more runs of two or more G nucleotides

Epitranscriptomics: The study of RNA base modifications that alter the structure or function of mRNA.

N⁶-Methyladenosine: Methylation product of adenosine nucleotides in RNA that has epitranscriptomic properties in cells.

Pseudouridine: Isomerization product of uridine nucleotides in RNA that has epitranscriptomic properties in cells.

Inosine: Deamination product of adenosine nucleotides in RNA that has epitranscriptomic properties in cells.

Pre-mRNA: The newly synthesized mRNA from the genome prior to intron processing to yield mature mRNA.

RNA Splicing: Stage in pre-mRNA maturation that involves removal of introns and joining of exons to form the mature transcript.

Consensus Sequence: Sequence motif that is recognized by epitranscriptomic writer proteins such as those that install m⁶A in SAG motifs in introns or DRACH motifs in other human mRNA regions.

■ REFERENCES

- (1) Jones, J. D., Monroe, J., and Koutmou, K. S. (2020) A molecular-level perspective on the frequency, distribution, and consequences of messenger RNA modifications. *Wiley Interdiscip. Rev. RNA*, e1586.
- (2) Carlile, T. M., Martinez, N. M., Schaening, C., Su, A., Bell, T. A., Zinshteyn, B., and Gilbert, W. V. (2019) mRNA structure determines modification by pseudouridine synthase 1. *Nat. Chem. Biol.* 15, 966–974.
- (3) Zaccara, S., Ries, R. J., and Jaffrey, S. R. (2019) Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.* 20, 608–624.
- (4) Shi, H., Wei, J., and He, C. (2019) Where, when, and how: Context-dependent functions of RNA methylation writers, readers, and erasers. *Mol. Cell* 74, 640–650.

- (5) Hartstock, K., and Rentmeister, A. (2019) Mapping m6A in RNA: Established methods, remaining challenges and emerging approaches. *Chem. - Eur. J.* 25, 3455.
- (6) Hocine, S., Singer, R. H., and Grünwald, D. (2010) RNA processing and export. *Cold Spring Harbor Perspect. Biol.* 2, a000752–a000752.
- (7) Wang, J., Alvin Chew, B. L., Lai, Y., Dong, H., Xu, L., Balamkundu, S., Cai, W. M., Cui, L., Liu, C. F., Fu, X. Y., Lin, Z., Shi, P. Y., Lu, T. K., Luo, D., Jaffrey, S. R., and Dedon, P. C. (2019) Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. *Nucleic Acids Res.* 47, e130.
- (8) Kumar, A., Clerici, M., Muckenfuss, L. M., Passmore, L. A., and Jinek, M. (2019) Mechanistic insights into mRNA 3'-end processing. *Curr. Opin. Struct. Biol.* 59, 143–150.
- (9) Louloui, A., Ntini, E., Conrad, T., and Ørom, U. A. V. (2018) Transient N⁶-methyladenosine transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency. *Cell Rep.* 23, 3429–3437.
- (10) Lorenz, D. A., Sathe, S., Einstein, J. M., and Yeo, G. W. (2020) Direct RNA sequencing enables m6A detection in endogenous transcript isoforms at base-specific resolution. *RNA* 26, 19–28.
- (11) Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., Dai, Q., Chen, W., and He, C. (2014) A METTL3-METTL14 complex mediates mammalian nuclear RNA N⁶-adenosine methylation. *Nat. Chem. Biol.* 10, 93–95.
- (12) Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., and Yi, C. (2015) Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* 11, 592–597.
- (13) Fleming, A. M., Nguyen, N. L. B., and Burrows, C. J. (2019) Colocalization of m⁶A and G-quadruplex-forming sequences in viral RNA (HIV, Zika, hepatitis B, and SV40) suggests topological control of adenosine N⁶-methylation. *ACS Cent. Sci.* 5, 218–228.
- (14) Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015) N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 518, 560–564.
- (15) Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., Leon-Ricardo, B. X., Engreitz, J. M., Guttman, M., Satija, R., Lander, E. S., Fink, G., and Regev, A. (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159, 148–162.
- (16) Bass, B. L. (1997) RNA editing and hypermutation by adenosine deamination. *Trends Biochem. Sci.* 22, 157–162.
- (17) Picardi, E., Manzari, C., Mastropasqua, F., Aiello, I., D'Erchia, A. M., and Pesole, G. (2015) Profiling RNA editing in human tissues: towards the inosinome atlas. *Sci. Rep.* 5, 14941.
- (18) Wang, Y., Park, S., and Beal, P. A. (2018) Selective recognition of RNA substrates by ADAR deaminase domains. *Biochemistry* 57, 1640–1651.
- (19) Reich, D. P., and Bass, B. L. (2019) Mapping the dsRNA world. *Cold Spring Harbor Perspect. Biol.* 11, a035352.
- (20) Lee, Y., and Rio, D. C. (2015) Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* 84, 291–323.
- (21) Jordan, P., Goncalves, V., Fernandes, S., Marques, T., Pereira, M., and Gama-Carvalho, M. (2019) Networks of mRNA processing and alternative splicing regulation in health and disease. *Adv. Exp. Med. Biol.* 1157, 1–27.
- (22) Maizels, N., and Gray, L. T. (2013) The G4 genome. *PLoS Genet.* 9, e1003468.
- (23) 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.
- (24) Mergny, J. L., and Sen, D. (2019) DNA quadruple helices in nanotechnology. *Chem. Rev.* 119, 6290–6325.
- (25) 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.
- (26) Slobodin, B., Han, R., Calderone, V., Vrieling, J. A. F. O., Loayza-Puch, F., Elkon, R., and Agami, R. (2017) Transcription impacts the efficiency of mRNA translation via co-transcriptional N⁶-adenosine methylation. *Cell* 169, 326–337.
- (27) Ke, S., Pandya-Jones, A., Saito, Y., Fak, J. J., Vagbo, C. B., Geula, S., Hanna, J. H., Black, D. L., Darnell, J. E., Jr., and Darnell, R. B. (2017) m(6)A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev.* 31, 990–1006.
- (28) Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., and Gilbert, W. V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515, 143–146.
- (29) Linder, B., Grozhik, A. V., Orlarier-George, A. O., Meydan, C., Mason, C. E., and Jaffrey, S. R. (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* 12, 767–772.
- (30) 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–W682.
- (31) Mullen, M. A., Assmann, S. M., and Bevilacqua, P. C. (2012) Toward a digital gene response: RNA G-quadruplexes with fewer quartets fold with higher cooperativity. *J. Am. Chem. Soc.* 134, 812–815.
- (32) Laddachote, S., Nagata, M., and Yoshida, W. (2020) Destabilisation of the c-kit1 G-quadruplex structure by N(6)-methyladenosine modification. *Biochem. Biophys. Res. Commun.* 524, 472–476.
- (33) 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.
- (34) Sun, D., and Hurley, L. H. (2010) Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods Mol. Biol.* 608, 65–79.
- (35) Belotserkovskii, B. P., Tornaletti, S., D'Souza, A. D., and Hanawalt, P. C. (2018) R-loop generation during transcription: Formation, processing and cellular outcomes. *DNA Repair* 71, 69–81.
- (36) Malig, M., Hartono, S. R., Giafaglione, J. M., Sanz, L. A., and Chedin, F. (2020) Ultra-deep coverage single-molecule R-loop footprinting reveals principles of R-loop formation. *J. Mol. Biol.* 432, 2271.
- (37) 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.
- (38) Binas, O., Bessi, I., and Schwalbe, H. (2020) Structure validation of G-rich RNAs in noncoding regions of the human genome. *ChemBioChem*, DOI: 10.1002/cbic.201900696.
- (39) Smola, M. J., and Weeks, K. M. (2018) In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.* 13, 1181–1195.
- (40) Weng, X., Gong, J., Chen, Y., Wu, T., Wang, F., Yang, S., Yuan, Y., Luo, G., Chen, K., Hu, L., Ma, H., Wang, P., Zhang, Q. C., Zhou, X., and He, C. (2020) Keth-seq for transcriptome-wide RNA structure mapping. *Nat. Chem. Biol.* 16, 489.
- (41) Ajiro, M., Jia, R., Yang, Y., Zhu, J., and Zheng, Z.-M. (2016) A genome landscape of SRSF3-regulated splicing events and gene expression in human osteosarcoma U2OS cells. *Nucleic Acids Res.* 44, 1854–1870.
- (42) 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.
- (43) 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.
- (44) Zhang, J., Harvey, S. E., and Cheng, C. (2019) A high-throughput screen identifies small molecule modulators of alternative splicing by targeting RNA G-quadruplexes. *Nucleic Acids Res.* 47, 3667–3679.

(45) Puig Lombardi, E., Holmes, A., Verga, D., Teulade-Fichou, M.-P., Nicolas, A., and Londoño-Vallejo, A. (2019) Thermodynamically stable and genetically unstable G-quadruplexes are depleted in genomes across species. *Nucleic Acids Res.* 47, 6098–6113.